Chinese Academy of Sciences
Cyberspace Administration of China
Ministry of Education of the PRC
Ministry of Science and Technology of the PRC
China Association for Science and Technology
Chinese Academy of Social Sciences
National Natural Science Foundation of China
Chinese Academy of Agricultural Sciences
Editors

China's e-Science Blue Book 2020

中国工信出版集团
電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn
Springer

China's e-Science Blue Book 2020

Chinese Academy of Sciences ·
Cyberspace Administration of China ·
Ministry of Education of the PRC ·
Ministry of Science and Technology of the PRC ·
China Association for Science and Technology ·
Chinese Academy of Social Sciences ·
National Natural Science Foundation of China ·
Chinese Academy of Agricultural Sciences
Editors

China's e-Science Blue Book 2020

中国工信出版集团 電子工業出版社 PUBLISHING HOUSE OF ELECTRONICS INDUSTRY http://www.phei.com.cn

Springer

Editors
Chinese Academy of Sciences
Beijing, China

Cyberspace Administration of China
Beijing, China

Ministry of Education of the PRC
Beijing, China

Ministry of Science and Technology of the PRC
Beijing, China

China Association for Science and Technology
Beijing, China

Chinese Academy of Social Sciences
Beijing, China

National Natural Science Foundation of China
Beijing, China

Chinese Academy of Agricultural Sciences
Beijing, China

ISBN 978-981-15-8341-4 ISBN 978-981-15-8342-1 (eBook)
https://doi.org/10.1007/978-981-15-8342-1

Jointly published with Publishing House of Electronics Industry
The print edition is not for sale in China (Mainland). Customers from China (Mainland) please order the print book from: Publishing House of Electronics Industry.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Contents

Introduction

Overview of e-Science Research in China

Yang Wang, Yan Ban, and Xuehai Hong

Abstract During "the 13th Five-Year Plan" period, the Chinese experts and scholars make full use of advanced information technology to carry out scientific research work, and have achieved a series of scientific and technological achievements. It reflects the level of e-Science application in China. This article systematically summarized the e-Science research from three aspects, the application in the frontier research of science and technology, the progress of e-Science in major projects and the achievements of informatization in interdisciplinary. During the past two years, so as to provide a reference for the scholars in this field for the further e-Science research.

Keywords E-Science · Informatization · Frontier research of science and technology · Achievements of informatization

1 Introduction

The essence of e-Science is the informatization of scientific and technological innovation activities, which is one of the indispensable input elements of modern scientific research [1]. The e-Science is a key of improving innovation capacity and an important means of enhancing national scientific and technological competitiveness. It is also an effective and powerful instrument of promoting transformation of scientific research mode and improving efficiency and output of scientific research.

During "the 13th Five-Year Plan" period, the State Council of the PRC have paid high attention to e-Science and issued the "Outline of National Informatization Development Strategy" [2] which proposed to accelerate the e-Science development. It is the first time for the e-Science work was included in the national strategy. China

Y. Wang (✉) · Y. Ban · X. Hong
Computer Network Information Center, Chinese Academy of Sciences, Beijing, China
e-mail: wangyang@cnic.cn

X. Hong
Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_1

now is developing in a period of important strategic opportunities in which informatization leads overall innovation and creates new advantages of national competitiveness, and also in a key window period in which China develops strong capability in informatization. In April 2018, the secretary general Xi Jinping addressed in the National Symposium on Cyber Security and Informatization [3] that IT application brought opportunities for China. We should flexibly seize the historical opportunity of Informatization development.

For the past few years, China's scientific researchers have conscientiously implemented the deployments of "*Three Directions of Innovation in Science and Technology*" [4]. They accelerated scientific innovation in various fields, positively advanced the work related to e-Science, and achieved a series of new results, effects and breakthroughs. E-Science provides powerful supports for the development of scientific and technological innovation. In order to better introduce and summarize the development trends and results of China's e-Science in the past two years, Chinese Academy of Sciences has continued to coordinate with national related sectors and published *China's e-Science Blue Book 2020*. This book was firstly released in Chinese and English version around the world in order to better display and share major achievements, successful experiences and typical cases of China's e-Science development.

2 Overview

China's e-Science Blue Book 2020 has collected 28 research reports about China's e-Science application in the past two years to introduce development trends, major scientific achievements and progress of China's e-Science, from three aspects of the application in the frontier research of science and technology, the progress of e-Science in major projects and the achievements of informatization in interdisciplinary. In order to provide overall understanding and reference for the readers, this paper introduces the background and extracts the key points of several articles in detail.

2.1 Application in the Frontier Research of Science and Technology

The basic task or even the primary task of scientific and technological innovation is to face the frontiers science and technology over the world. China has become a scientific and technologically country with great influences, and is constantly developing into a great powerful country in science and technology. Besides, the level of science and technology is also transforming from tracking and learning to parallel and leading. In the past two years, China has achieved a great number of world-class scientific and technological achievements, in which e-Science has also played an important supporting role.

In the "First Chapter – Application in the frontier research of science and technology", there are a total of 5 articles included, covering biomedicine, dark matter, high-performance computing, water science, and artificial intelligence. These articles have described many world-class research achievements in science and technology, conducted development trends of the above domains on the future, and elaborated on the important roles played by e-science in those scientific research.

We have extracted 3 articles to introduce in detail.

1. The big data of biomedicine advances COVID-19 research.

From the beginning of 2020, an unexpected epidemic broke our peaceful life. In order to win this "battle", the whole country and even the people all over the world have made great efforts. Domestic and foreign scientists and experts have also rapidly organized scientific strengths to tackle such difficult problem, and positively played the role of think tank, thus making important contributions to fight the epidemic. The paper "*Biomedicine Big Data – Trends and Prospect*" written by Guoping Zhao, an academician of Chinese Academy of Sciences and the chief scientist of Shanghai Institute of Nutrition and Health of CAS, has traced research and development of biomedicine big data (BMBD) and development progress of transformation and application, explored the implications of BMDB in sectors such as life science research, medical and health institutions, and biotechnology and biomedicine industries in connection with the challenges and opportunities faced by social and economic development. The recent COVID-19 outbreak is used as an illustrative case study. This paper has summarized an analysis of a decade of BMBD practice, both domestically and abroad. Further depending on China's national situation, the paper has finally proposed corresponding policies, recommendations and solutions, providing useful reference for China's biomedical big data administration sectors, research and application sectors.

2. "Sunway TaihuLight" guides high performance computing stepping into ExaScale supercomputer era.

The development of high-performance computing comprehensively reflects the scientific and technological strength of a country, and is also an important part of the national innovation system. The paper "*the Progress of China's High Performance Computing in China and the Development Trends of International High Performance Computing*" written by Zuoning Chen, an academician of Chinese Academy of Engineering, described the landscape of the world TOP500 high performance computers in the past two years, analyzed the latest progress and main shortcomings of China's HPC research, explained the technological breakthrough and application of "Sunway TaihuLight" supercomputer, and finally discussed the R&D upsurge of Exascale computers in the world. In addition, this paper predicted the future development trends of high performance computing technologies.

3. Dark Matter Particle Explorer (DAMPE) has made breakthrough in electrons and high energy cosmic rays observation.

As the first satellite-based observatory of China targeting on astronomical objects, Dark Matter Particle Explorer (DAMPE, also called as "WuKong") now has operated for over three years on-orbit. It collected and analyzed over 6 billion high-energy cosmic rays(CRs) and obtained hundreds of TB scientific data. Such rich observations have largely benefited the research of electrons (positrons), protons and gammas within the high energy CRs. Therefore, such massive astronomical observation data has presented larger challenges for data storage and processing technology. Chang Jin, an academician of Chinese Academy of Sciences and the president of Purple Hills Observatory of CAS, taken DAMPE for an example in his paper "*IT Application of In-orbit Data Processing for Dark Matter Particle Explorer*", took a deep insight into DAMPE, introduced the structure and features of its data processing software DAMPESW, explained constructions of DAMPE infrastructures that are built on the demand of high data processing performance. This paper summarized the IT application of such scientific satellite in data processing, providing reference for subsequent similar satellite projects in the future.

2.2 Progress of E-Science in Major Projects

At present, national demands for strategic and scientific supports are more urgent than before. In the course of developing into a world-class in science and technology, we need to focus on national strategy, strengthen the major scientific and technological breakthroughs related to national economy and the people's livelihood, and make more scientific and innovative achievements related to production and livelihood. In the past few years, new-generation information technologies with representatives of big data, cloud computing and artificial intelligence have flourished. Multiple information technologies have been comprehensively applied in many major projects and achieved excellent results.

In the "Second Chapter – Progress of e-Science in major projects", there are 12 articles, focusing on development trends of IT application and application of core technologies in scientific research domains. These articles covered scientific investigation in Tibetan Plateau, scientific research over nuclear fusion, lunar exploration projects, cloud computer oriented to scientific research and other aspects of e-Science application.

We have extracted 5 articles to introduce in detail.

1. Big data promotes scientific research over Tibetan Plateau and Pan-Third Pole earth system.

The Pan-Third Pole region mainly includes the Tibetan Plateau and the northern intracontinental arid region of Asia, extending to the Caucasus Mountains in the west and the western Loess Plateau in the east. This region covers 20 million square kilometers and affects the environment inhabited by three billion people. The big data era has brought new opportunities and challenges for us to properly understand and solve environmental problems in Tibetan Plateau and the Pan-Third Pole region.

The paper "*Big data promotes the Tibetan Plateau and Pan-Third Pole Earth System Science*" written by Xin Li, a researcher from Institute of Tibetan Plateau Research of CAS, has introduced the system architecture, data resource integration and big data analysis methods of the Pan-Third Pole Big Data System in detail, advanced big data processing capacity in fields of study, explored a new mode of big data driven geoscientific research, and promoted scientific research over Tibetan Plateau and the Pan-Third Pole earth system.

2. The man made sun "EAST" has realized 1GHW plasma discharge for the first time.

So far, EAST is the only experimental device in the world with similar conditions as ITER and have the best ability to achieve long pulse and high performance operation on the particle balance time scale, which attracts extensive international cooperation and generates a huge amount of experimental data. However, generated massive experimental data has brought new demands and challenges for building IT application environment. The paper "*Information construction and prospect of EAST collaborative experimental platform*" written by Feng Wang, a senior engineer from Institute of Plasma Physics of HIPS of CAS, has introduced the whole process of IT application construction for EAST collaborative experimental platform and outlook for future planning in detail. Besides, The platform not only promoted the accumulation of research resources and improves the work efficiency, but also provided an open and shared way of academic exchange and promotes domestic and foreign cooperative and research.

3. Chinese VLBI network and e-VLBI technology advance lunar and deep space exploration.

The lunar is the first outer planet where human starts exploration in space. Since the beginning of this century, China started its own lunar exploration project, proposed a three-step plan of "circling, landing, and returning", also known as the Chang'E Project, completed trilogy of lunar exploration projects in 2020. In the Chinese Lunar Exploration Project, Chinese VLBI Network has successfully adopted real-time e-VLBI technology serves the TT&C system in the Chang'E series of lunar exploration missions, providing fast and accurate determination of orbit positioning services during the phases of flying to the moon, orbiting the moon, descending to the moon, and returning to earth. The paper "*Chinese VLBI network and e-VLBI technology applications in Chinee Lunar Exploration Project*" written by Zhong Chen, a senior engineer from Shanghai Astronomical Observatory of CAS, has introduced Chinese VLBI network and e-VLBI technology application in the Chinese Lunar Exploration Project, and details of system development, operation and performance of the VLBI tracking system. Finally, it gives a prospect of Chinese VLBI network for future deep space exploration missions.

4. The new-generation information communication technology 5G promotes to build smart scientific research network.

The Chinese national “the 13th Five-Year Plan” clearly puts forward the requirements and terms of developing and promoting the key technologies of the 5th generation mobile communications and ultra-wideband, and launching 5G commercial applications. As 5G network technology and cloud computing technology constantly develop, combined with current “date-intensive science and research”, the demands for massive data processing constantly increase. Compared with the network resources construction of traditional scientific research institutions, 5G scientific research cloud relies on its excellent network performance, security guarantee and efficient flexible computing resource allocation ability and simple hardware requirements. It can realize the intelligent and flexible construction of computing resources in the face of different needs, localized specialized services, cloud network collaborative optimization and traffic payment, and fully satisfies the stable and high-speed network usage requirements of researchers and students. The paper “*Advanced Scientific Research Environment Evolution and Cloud Service Architecture Design Integrating 5G Technology*” written by Xu Zhou, a researcher of Computer Network Information Center of CAS and the Director of Advanced Network and Technology Development, has introduced the construction requirements of 5G networks and their key technologies and scientific research clouds, and proposed a 5G scientific research cloud architecture. Through the analysis of typical scenarios such as research institutes, large scientific installations, field stations, and university campuses, this paper expounds the practicability and necessity of the intelligent research network constructed by 5G scientific research cloud, and finally summarizes and forecasts it.

5. The national scientific and technical literature’s information guarantee system supports scientific and technological innovation and research.

As an important national strategic resource, scientific and technical literature platform is essential for scientific and technological work and an important supporting system for innovation and development. China’s guarantee level for scientific and technological resources and service capacity will directly influence national innovation and sustainable development in science and technology. The paper “*Development and Services of Resource Discovery System of National Science and Technology Library*” written by Yiqi Peng, the director of National Science and Technology Library, introduced the development of the national scientific and technical literature’s information guarantee system, analyzed current situations and challenges of literature’s information guarantee work, described the proposal of building the national scientific and technical literature’s resources discovery platform, and propagandized the method of finding literature resources.

2.3 Achievements of Informatization in Interdisciplinary

The scientific research shall not only follow the indomittable spirit, pursue knowledge and truth, but also “benefit the people” and serve economic social development and the people. The scientific and technological level has been one of the main variables influencing world economic cycle, and also a major factor determining increase of economic aggregate. Since the 21st century, global scientific and technological innovation has been in the unprecedented development period, while new scientific and technological revolution and industrial reform are re-building the global innovation territory and re-shaping global economic structure. The science and technology have never deeply influenced future and destiny of a nation and the people's livelihood like today [5].

In the "Third Chapter – Achievements of informatization in interdisciplinary", there are 11 articles which have described that scientific innovation advances national economic development, industry, agriculture, medicine, education, resources and environment in detail, covering the IT application of China-Russia-Mongolia economic corridor for the Belt and Road, stem cell scientific research, public travel, urban governance, Digital orchard and other domains.

We have extracted 4 articles to introduce in detail.

1. Desertification assessment for China-Russia-Mongolia economic corridor provides IT application supports and decision-making support for the Belt and Road Initiative.

In 2013, the General Secretary Xi Jinping proposed the initiative of co-building the "Silk Road" economic belt and 21st Century Maritime Silk Road [6]. Then, the basic framework for co-building the Belt and Road has been formed. For over six years, the Belt and Road Initiative has promoted fruitful achievements in trade cooperation and significantly advanced economic growth of countries and regions along the "Road" and even the world. The paper "*Implementation of the informatization application scenario for prevention and control of desertification in the China-Mongolia-Russia economic corridor of the Belt and Road Initiative*" written by Juanle Wang, the deputy director of the earth data science and sharing Lab at the Institute of Geographic Sciences and Natural Resources Research of CAS, has analyzed the complex natural geography, fragile ecological environment, and serious desertification of the China-Russia-Mongolia economic corridor in the Belt and Road Initiative; introduced a desertification remote sensing inversion algorithm, a big data application platform, and multi-source data fusion and integration, and established application scenarios for desertification risk control, based on IT application and GIS technology; used modes of big data batch processing and real-time processing, the desertification information along the corridor was extracted, analyzed, and dynamically monitored. Combined with historical data, the diagnosis and testing of desertification patterns and changes within 200 km of both sides of the China-Mongolia Railway (Mongolian section) from 1990 to 2015 have been completed. Besides, the researches have provided IT and decision-making supports for preventing and controlling desertification risks in key regions of the Belt and Road.

2. "One-time Face Recognition System" Drives reform of smart security check mode for civil aviation.

With the surge of civil aviation airport passenger traffic in recent years, the airport security and screening mode is under great pressure, which has become an obstacle for achieving the target to become a leading civil aviation power. The paper "*'One-time Face Recognition System' Drives Changes in Civil Aviation Smart Security Screening Mode*" written by Yu Shi, the director of Intelligent Security Technology Research Center of Chongqing Institute of Green and Intelligent Technology of CAS, has introduced the innovative achievement—"One-time Face Recognition System" which has integrated with multiple advanced technologies and concepts. The system has been applied in Hohhot Baita International Airport and achieved a great success. Besides, the system and the operating procedures have been approved by the authorities, authorizing the use of manually assisted computerized verification to replace the original manual verification, which has started a reform of the civil aviation smart security screening mode in China.

3. The Urban Resources, Environment, and Ecology (UREE) big data platform provides decision-making supports for realizing sustainable urban development.

By the end of 2018, the urbanization rate of China had reached 59.58%. The rapid urbanization has brought a series of social and environmental dilemmas, including urban heat islands, traffic congestion, solid waste, air pollution, lack of essential services and facilities. Therefore, the urban sustainability has become the most significant urban development issue in the world. The paper "*Urban Resources, Environment, and Ecology (UREE) Big Data Platform: Construction and Application*" written by Weiqiang Chen, a professor of Sustainability Science in the Institute of Urban Environment of CAS, has introduced a seven-layer architecture UREE based on open and in-depth data and state-of-the-art technologies, including data acquisition and standardization, data query and visualization, data unified interface and fusion, and urban metabolism simulation. The platform has not only provided powerful technical supports for monitoring and studying dynamic changes of urban resources, environment, ecology and driving mechanism, but also provided decision-making support for solving the Urban Problem in the process of urbanization and how to build the sustainable cities. The UREE is an innovative platform to promote the development and research of urban environment and ecology with big data technology.

4. Digital orchard technology advances intelligent agricultural development.

The Chinese "the 13th five-year plan" explicitly calls for strengthening the integration of agriculture and information technology. In addition, in the "New Generation of Artificial Intelligence Development Plan" released by the state council in 2017, important arrangements were made for the major task of upgrading the intelligent agriculture industry, and pointing out that a number of agricultural integration application demonstrations, including intelligent orchards, should be carried out. In China, fruit industry is the third largest planting industry that after grain and vegetable and it plays an important role in the development of rural economy. The paper "*Present Situation and Development Prospect of the Digital Orchard Technology*" written by Guomin Zhou, the deputy director of the Department of Science and Technology Management of CAAS, has introduced the concept and connotation of the Digital

Orchard, summarizes the research and application status of the Digital Orchard technology, and looks forward to the future trend and key development direction of the Digital Orchard. This would provide reference for advancing development of domestic smart orchard in the future.

3 Conclusions and outlook

The year of 2020 is the end year of building a moderately prosperous society in all aspects and completing the "the 13th Five-Year Plan". During the "the 13th Five-Year Plan", China constantly follows the development direction of "three orientations" in scientific and technological innovation, which not only strives to realize leaping development in important domains of science and technology, keeps pace with and even guides world development direction in science and technology, supports national strategic demands, but also advances deep integration between science and economic and social development. In the past few years, China has been rapid development in scientific strength, achieved remarkable results, and realized multiple major breakthroughs, including aerospace technology, deep water exploration, manufacturing technology, biotechnology, new energy, new materials, and so on. Meanwhile, the new-generation IT such as big data, cloud computing, mobile Internet, and AI advances rapidly, and e-Science has provided powerful supports and guides for innovation-driven development of scientific research and achievements.

Looking into the world, the future and "the 14th Five-Year Plan", we have been deeply aware that China lacks innovative capacity, overall scientific development and supporting for economic and social development by science. The new generation of reform in science and technology results in more fierce competition. Therefore, we need to grasp the development trends of the world in science and technology. In order to advance new-generation cross development of scientific technology, we shall focus on building scientific power, further greatly develop information technologies such as big data, cloud computing, AI, and Internet of Things, drive application of disruptive technologies such as quantum computing, carbon-based ICT and block chain in scientific research, deeply integrate with scientific innovation, and play the important role of IT application in scientific and technological development. That will make more contributions to national strategic innovation in science and technology.

References

1. Fangyu Liao, Xuehai Hong et al (2019) Data and computing platform are important infrastructure driving development of modern scientific research. Front Data Comput 1(1):2–10
2. Xinhua News Agency (2016) Outline for National IT Application Development Strategy issued by the General Office of CPC Central Committee and the State Council [EB/OL]. http://www.gov.cn/xinwen/2016-07/27/content_5095297.htm. 27 July 2016

3. xinhuanet.com (2018) Xi Jinping: Self-dependent innovation advances construction of network power [EB/OL]. http://www.xinhuanet.com/politics/2018-04/21/c_1122719810.htm. 21 April 2018
4. people.com (2016) Important address Three Directions of Innovation in Science and Technology made by the General Secretary Xi Jinping in National Scientific and Technological Innovation Conference. In: 18th general meeting of academicians of CAS, 13th general meeting of academicians of CAE, 9th National Party Congress of CAST [EB/OL]. http://politics.people.com.cn/n1/2016/0602/c1001-28406379.html 30 May 2016
5. Xinhua News Agency (2018) Important address made by the general secretary Xi Jinping in 19th general meeting of academicians of CAS, and ceremony of 14th general meeting of academicians of CAE [EB/OL].http://www.cas.cn/zt/hyzt/ysdh19th/yw/201805/t20180528_4647518.shtml. 28 May 2018
6. xinhuanet.com (2015) Authorized release: vision and action of advancing the silk road economic belt and 21st century maritime silk road [EB/OL]. http://www.xinhuanet.com/world/2015-03/28/c_1114793986.htm. 28 March 2015

Yang Wang Ph.D., Senior Engineer, Head of the Information Development Strategy and Evaluation Center, Computer Network Information Center of the Chinese Academy of Sciences. He mainly focuses on the development strategy of information technology and e-Science. He has published more than 10 academic papers in SCI/EI and Chinese journals. He participated in "the 13th Five-Year Plan for Informatization Development of the Chinese Academy of Sciences" and "Data and Computing Platform Planning". He also participated in the research work on "The promotion of scientific research informatization", "National Informatization Development Strategy Research", "Information Development Strategy in Interdisciplinary Research".

Application in the Frontier Research of Science and Technology

Biomedicine Big Data—Trends and Prospect

Guoping Zhao, Yixue Li, Daming Chen, and Yan Xiong

Abstract This forward-looking review focuses on the development and applications for Biomedicine Big Data (BMBD), and its role in the engineering system for data management, scientific and technological research and development, as well as in social and economic transformation. The review starts with an elaboration on the complex connotations of BMDB from the inter-disciplinary point of view. It then explores the implications of BMDB in sectors such as life science research, medical and health institutions, and biotechnology and bio-medicine industries in connection with the challenges and opportunities faced by social and economic development. The recent COVID-19 outbreak is used as an illustrative case study. The review ends with an analysis of a decade of BMBD practice, both domestically and abroad, with suggestions for policy-making and solutions to tackle major challenges from China's perspective. It is hoped that any BMBD-related institutions, including administrative, academic, industrial, financial and social organizations,practitioners and users will benefit from this insightful summary drawn from the past decades of BMBD practice. Any critical comments and constructive suggestions are sincerely welcomed by the authors.

Keywords Biomedicine big data (BMBD) · Knowledge connotation · Service platform · Management system · Transformation and application · Interdisciplinary talents

The genomics revolution of the last decade of the twentieth century has not only made "data" as an important foundation for life sciences, but also has established the focus on "humans", particularly in biomedical researches. "Biomedicine Big Data" (BMBD), generated by systematic biomedical research, translational

G. Zhao (✉) · Y. Li
Big Data Center for BioMedicine, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, Shanghai, China
e-mail: gpzhao@sibs.ac.cn

D. Chen · Y. Xiong
Shanghai Information Center for Life Sciences, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, Shanghai, China

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_2

medicine research and precision medicine practice, has the features characterized by “4V” (volume, velocity, variety and veracity) [1] and “3H” (high dimension, high complexity and high uncertainty) [2].

From the beginning, BMBD has been interdisciplinarily developed by biology (life science and biotechnology), medicine (including pharmacy) and data science (information and computer sciences). It can be roughly classified into two categories: (1) Biological data, covering that generated from research and applications in basic life sciences, omics and systems biology, physiology and psychology, cognitive behavior, clinical medicine and public health; and (2) Environmental data, covering that from domains such as social demography and environmental exposure. The core of BMBD, therefore, is anchored in the data generated from systematic research of biomedical and translational medicine for populations, and precision medicine for individuals.

BMBD’s current fast growing phase is gradually and effectively driving the paradigm shift in biomedical research from hypothesis-driven to data-intensive discovery. Because of its interwoven data origins of both natural and social sciences, and the information gap between original data and applicable practice, it has become an urgent and critical requirement for interactive data sharing, cross-disciplinary collaboration and administrative coordination, in order to tackle unique new biomedical challenges.“BMBD Basic Technology Service Platform” is an authorized, collaborative, integrative and intelligent information ecosystem, a long-term solution to provide stable public services in data distribution, dissemination and standardization, as well as for knowledge mining and translational applications.

1 Development and Connotations of BMBD

Mathematics has been an intrinsic tool used from the very beginning in physics and astronomy. Up to the twentieth century, the accumulated data from these disciplines have exceeded the Exabyte (EB, 10^{18}) level, making them the first to enter the era of “big data” after their previous “small-data” discovery modes of experimental verification, theoretical analysis and computational simulation. The role of mathematics in chemistry was not obvious initially. It was Mendeleev’s Periodic Table of Elements, the breakthrough in the discovery of periodic rules and relationships between atomic numbers and chemical properties of chemical elements, that made math and computation into chemistry, and transformed it from a pure experimental science to a computational and theoretical science. As a result, Chemical Engineering was quickly developed as a new discipline, and has so far achieved large-scale applications in rational transformation of natural materials into artificial materials serving the economy and wellbeing of mankind.

Biology has had a long stepwise history of its relationship with mathematics and computation. In the early days of the seventeenth-eighteenth centuries, biology was initially a collection of specimens of living objects with their descriptive records (Taxidermy) followed by their classifications (Taxonomy). With the establishment of cytology, biochemistry and genetics during the late 19th and the early twentieth centuries, biology gradually developed into a scientific discipline—Life Science—that explored the common structure and function of living organisms by experimental verification and limited computations. Mathematics and computation did not play a significant role in biology until the mid-twentieth century when molecular biology was established on the basis of the double-helix DNA model of chromosome and the central dogma that links genetic code with its expression regulation, and functional molecules.Because the targeted research subjects at the time were limited to a limited number of biomolecules, the size of accumulated bio-data was also limited. Quantitative Biology [3] and Computational Biology [4] born and shaped during this period were largely providing auxiliary tools for biological research, not yet becoming mainstream biological disciplines.

1.1 Biological Connotations of BMBD

The "Human Genome Project (HGP)" launched in the 1990s, with the goal to sequence the entire genome of five representative human individuals, marked a major milestone in the holistic survey of human genetic background [5]. An initial challenge of the "Big Life Science Project" of genome sequencing was to determine the one-dimensional (1-D) sequence of thousands to billions of chemical bases, consisting of only 4 distinct types of ACGT(adenine, cytosine, guanine and thymine, respectively), of an organism, and then to annotate their genetic structure (genes). In order to tackle this unprecedented engineering challenge in biological scientific research,a "four-map" (genetic, physical, transcription and DNA sequence) strategy was designed, and new sequencing technologies were developed. Briefly, shot-gun sequencing fragments were assembled to obtain the whole genome by using mathematical algorithms (de novo or mapping) and ***h***igh-***t***hrough***p***ut (HTP)/large-scale parallel computational platforms. The genomic information, such as gene structure and functions, was then annotated to the assembled genome from known knowledge and mathematical and computational inferences, making genomes the source of meaningful codes of life. Bioinformatics was thus shaped and matured as an independent science discipline during this epic "genomic revolution". Rapid accumulated genomic sequences and associated annotations form the basis for an indispensable data foundation of modern life science, from which new approaches were created to guide experimental studies of biological systems. Anew research paradigm: Systems Biology that integrates both "wet" experimentation and "dry" computation/theoretical analysis, was born.

Technical breakthroughs in the omics platforms, such as ***n***ext-***g***eneration ***s***equencing (NGS), mass spectrometry and biochips, expedited the rapid development of various "life omics" (transcriptomics, epigenomics, proteomics, metabolomics and phenomics etc.). These platforms nurtured the inception of large-scale scientific programs in life science and medicine. Biological data, with the dramatic and rapid increase "in quantity" and multi-dimensional changes "in quality" [6], has reached the exabyte scale, making Life Science a real big data discipline as astronomy and physics. **Thus the biological connotations of BMBD include (a) systematic and standardized collection, quality control, annotation, analysis, integration and application of biomedical data, (b) modelling and simulation of biological systems using the data and, (c) quantitative description and prediction of the function, phenotype and behavior of an organism.**

1.2 Medical Connotations of BMBD

Medicine is the practice of diagnosis, treatment and prevention of various diseases by means of scientific technology, and psychological and humanistic care. It is also an academic discipline of applied science continuously evolving from clinical practice. By integrating with biological evidence and life science experiments, modern medicine has established itself as a scientific research system including the branches of basic medicine, clinical medicine and preventive medicine. A considerable amount of medical research data has naturally accumulated as a result.

In the mid-twentieth century, a series of cutting edge theoretical and technological innovations in life science, including the development of modern pharmacy, medical imaging, molecular and cellular immunology and molecular diagnosis and treatment, propelled the emergence of modern "Biomedicine" as a scientific discipline and research field. Differing from biology that concerns the common structure, function, and metabolic/growing dynamics of living organisms in general, biomedicine focuses on human health and disease as its primary research subject. In other words, biomedicine differentiates itself from biology and life science with its human-centric context in both research and application.

"Systems Biomedicine", the integration of the genomics-based holistic studies of biology and medicine, is the core connotation of modern biomedicine that drives its revolutionary development. It is these two research paradigms that have contributed to the majority of the core BMBD. Other sources of BMBD come from diverse research fields such as basic life science, omics-/systems biology, physiology, psychology, cognitive behavior, clinical medicine, public health and drug discovery, as well as those covering social demography and environmental exposure.

"Translational Medicine" takes the approach of "bedside to bench to bedside" with the aim to train a new generation of "research physicians". It is the key constituent of systems biomedicine research. A rationale for the shift "towards precision medicine" was laid out in a report from the National Research Council of the United States National Academies in 2011. Although the concept of "precision medicine" was developed on the basis of the "4P" model of translational medicine (i.e. preventive, predictive, personalized and participatory), its primary purpose, however, was to design the customized therapeutic recipes based on individual's genome, transcriptome, proteome, metabolome and other internal environment, so as to maximize therapeutic effect and minimize its side effects. It is a new medical approach for disease prevention and treatment. By collecting multi-omics experimental data and the corresponding clinical information from individual "small samples" ($n = 1$) to the aggregated individuals “big samples” (Σn), it generates a new set of “big data” as an important resource for BMBD.

In summary, **the medical and health science connotations of BMBD are (a) a large volume of population-based complex “biomedicine data” generated from the research platforms of system biomedine and translational medicine emerged from the integration of modern medicine/pharmacology and biology; (b) the “real-world data” collected from the rich dimensions of “precision medicine” (individual-based healthcare and medical services) by the application of extacted and mined information and knowledge from biomedicine data on the personal level (Fig. 1).**

Biology and **Medicine** are the two major sources of BMBD. **Research data** from these two sources are the axes connecting these two sources, each bearing cross-disciplinary attributes. **Life-omics data** is the indispensable foundation of modern life science, from which **Systems Biomedicine data** originated. The data from rapidly developing **microbiota and microbiome** is contributed from biology-related data (such as **metagenome**) and ecology/environment science-related **meta-data** (including societal data, especially that of epidemiology). The core of **biomedical research data** is derived from research and practice in **systems biomedicine, translational medicine and precision medicine**. It reflects the unique and fundamental focus of biomedical research –from the biological species of **“*Homo sapiens*”** population to the socially assembled nationalities and citizens of “**human being**” individuals. Naturally, BMBD closely links social/environmental information and health-related data, i.e., from the pure “**environmental data**” to the “**Exposome**”—agglomerated human facts in the environment. With the continuous development of big data technology and the progressive extension of the biomedical research scenario from laboratories to the real-world, research-derived BMBD must be, and

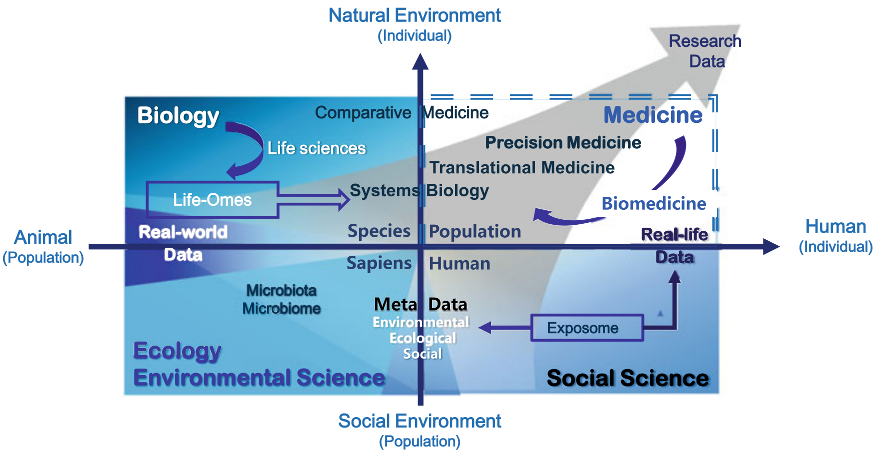


Fig. 1 Data sources/constituents of Bio-Medicine Big Data (BMBD): Their scientific connotations, Inter- and intra-relationships and implications for modern biomedical research and practice

can be, closely knitted with the macro environmental and populational "real-world data" and the micro environmental and individual "real-life data", to contribute to the well-being of mankind (individual and collected population).

1.3 Data Science Connotation of BMBD

As discussed above, the connotations of "BMBD" in biology and medicine are explicated by the analysis of two developmental trajectories: biology to life science, and medicine to biomedicine. To further explore the "data characteristics" of BMBD, two levels of BMBD may be postulated. Level1: **real-world data,** including biological data from animals, plants and microorganisms, as well as medical and health information collected from clinical records and physical examinations. These data sets have great research and application potentials. However, because they are massive in size and heterogeneous in content, a committed and long term well-maintained information platform is required to realize these potentials. The platform should have the capacities of systematic data collection, standardized data management, and comprehensive integration of research data sets. Level2: **research data,** the human-centric research, medicine and pharmacology data, including: (a) the data from human-subject systems biology research, much of which is supported by biomedical research programs with omics approaches; (b) the data from translational medicine research, including cohort and evidence-based medical studies, as well as those from comparative medicine using animal models for human disease and drug development research, and; (c) the data from drug R&D, generated from quantitative and systems pharmacology developed on the basis of new techniques such as ADME/T evaluation, pharmacometrics and pharmacokinetic metrics analyses, and the drug-target relationship discovered by multi-target drug research. Although the Level 2 data has a "limited sample size" compared to the Level 1 data, Level 2 data sets have the following superior features: higher in dimensions, better structured, designed and quality-controlled. Level 2 data sets are much more directly associated with human health, disease and other human-related medical research, and consequently bear more realistic and strategical importance for national security and socio-economic development. Therefore, Level 2 data sets are currently considered more important and sensitive than those of Level 1.

For more than half a century, biomedical data's "big data" characteristics became evident through multiple rapid transformations driven by revolutions of molecular biology and genomics. For example, (**a) the transformation from small data to big data:** The breakthrough of HTP experimental techniques, the digitalization of medical information, and the generation of real-world data have jointly pumped biomedical data from the magnitudes of PB (genomics) to EB (multi-omics, digitized health and medical records), and EB to ZB (real word data); (**b) the transformation from low dimension to high dimension:** Different types of omics data, together with various sources of clinical information such as those from medical imaging, in vitro diagnosis, continuous monitoring, clinical trial, and digital health records, are collected and integrated to make systematic analysis possible, providing BMBD with richer, deeper and more sophisticated content, and a constantly enriched data dimensions, and; (**c**) **the transformation from singular-scale to multiplex-scale:** The development of BMBD techniques and algorithms makes it possible to integrate data sets measured at different data granularities, such as the **multi-system granularities** of molecule, cell, tissue, organ, individual, the **multi-dimensional omics granularities** of multi-omics platforms, and the **multi-dimensional spatial/temporal granularities** of observations for population cohort, molecular epidemiology and real-world studies, the integration of which makes the holistic multi-scale analyses of the nature of life and disease possible. These transformations have created BMBD with 4V and 3H characteristics, which have posed a series of new data challenges to make full use of BMBD [7]. **The discovery paradigm shift to "conduct research with data approaches" and "analyze data with scientific methodologies"** [8] **are the data science connotation of BMBD.**

The above-stated rich connotations of BMBD in biology, medicine and data science have established the crucial role of BMBD in life science and medicine, from research to practice, and determined its value as a strategic resource for national social security and welfare. As such, research into and application of BMBD have attracted worldwide attention, giving rise to rapid development in recent years.

2 Current Status and Trends of BMBD

Currently, the value of BMBD has become the consensus view of all walks of life. Biomedicine has been developing into the "Fourth Paradigm" [9] era characterized by "Data Intensive Scientific Discovery". Consequentially, many planning and initiatives have been made to integratively manage, R&D and implement around BMBD both at home and abroad. Considerable experiences have been accumulated; at the same time, many lessons have also been introspectively learnt. For taking advantage of the transformations from data to information, subsequently to knowledge and the power to facilitate scientific research, four levels of synergies are required to enable serving individuals and society with more "accurate" health care and medical intervention. (**a**) **The synergy on data level**: to achieve the safe collection, storage, integration and management of data of different types by using technologies from data science and information technology. Standardized data quality management is still proven a challenge (especially for non-English-speaking countries). For the secure and high-efficient use of data, block chain digital identity is considered a promising underlining technique to make the shift from data "management" to data "governance". (**b**) **The synergy on information level**: to discover and extract data relationships by analytic algorithms and tools, so as to improve data validity and integrity and to enrich data content, for effective data dissemination and distribution. Standardized programmatic and user interfaces are required for stable services such as data submission, online analysis and user feedback. (**c**) **The synergy on**

knowledge level: to disseminate and casual-analyze biomedical information into precision medicine knowledge network. Unified biomedical thesauri, classifications and coding standards are the keys to improve data interoperability. Technical reference model also plays a critical role for the interoperability on the level of information framework, and ultimately the application of disease knowledge network to clinical decision support. (**d) The synergy on application level**: to conduct researches such as “deep patient”, for the advance of biomedicine science, development of health and medical products, and better supports for clinical, medical and public health practice and management. Only these 4 levels of synergies are realized, will BMBD be integrated by a unified platform, individual patient data be traceable, and the value of BMBD be truly maximized.

2.1 Current Status and Trends of BMBD R&D

(1) Current status and trends of BMBD R&D in Europe, United States, Japan and others

Since 1980–1990s, biological databases have accumulated a huge amount of life science data. They are instrumental not only to their hosting countries’ biological researches, but also for world-wide biological researches, and the build-up of global data-sharing infrastructure in life science. Some early players include (a) the National Institute of Genetics (NIG) of Japan: created DNA Data Bank of Japan (DDBJ) in 1986; (b) National Center for Biotechnology Information (NCBI) of the United States: created Genbank in 1988; (c) European Bioinformatics Institute (EBI):EMBL databank, now is known as ENA (European Nucleotides Archive), was first in service in 1980 at European Molecular Biology Laboratory (EMBL), hosted at EBI UK since 1992.The International Nucleotide Sequence Database Coalition (INSDC) was established in 1988 by these 3 nucleotide sequence databases, still providing active data services. INSDC plays an important role in promoting standardized collection, curation, dissemination, and distribution databases for nucleotide sequence data [10].

NCBI and EBI are globally recognized as two data centers providing the most comprehensive data services, not only covering basic biological data, but also start to offer genotype-phenotypic data services with individual phenotype information, providing more direct data support for translational medicine and precision medicine researches. For example, NCBI manages dbGap database that archives the studies of genotype and phenotype relationships in human, besides its long-time database offerings of basic biological data types, such as Refseq, Pubmed, PMC, NCBIGene, GEO and Pubchem; EBI manages EGA database that archives and distributes all types of personally identifiable genetic and phenotypic data from biomedical research projects, besides widely used UniProt, InterPro, ExpressionAtlas, PRIDE, Ensembl, ChEMBL databases.

In the area of biomedical data platform, EBI, along with its European partners, has built ELIXIR infrastructure for data and information sharing through the network built among the local nodes in EMBL participating countries. The medical informatization platform (eMedlab), initiated by British Medical Research Council (MRC), integrates and distributes heterogeneous medical records, images, pharmaceutical and genomics data.Meanwhile, it exchanges with NHS on medicare information, for comprehensively understanding health and disease progress. National Institute of Health Data Science [11] (HDR UK), closely related to eMedlab, was launched in 2017. It forms a close collaboration with research and medical institutions through the implementation of systems such as GogStack and SemEHR at participating hospitals.

In Europe, United States and some other western countries (regions), some close collaborations, between clinical practitioners and basic researchers, in translational medicine research and medical practice have achieved remarkable outcomes. They are pivotal for the effort in improving the quality of medicare, and the foundation for the inception of new biomedical knowledge system.

(2) Current Status and Trends of BMBD R&D in China

China's science and technology communities have been long recognizing the importance of the collection and sharing of scientific data. Three academicians, Guanhua Xu, Shu Sun and Honglie Su, were among the first to appeal for geological data sharing. In 1999, Academician Bailin Hao put forward a proposal for the creation of "National Bioinformatics Center" to the State Council [12]. In order to implement the standardized management and efficient utilization of scientific data resources, China in 2002 began to fully plan scientific data sharing, and funded "**National Scientific Data Sharing Project**" for the period of 2003–2005. As an important part of National Science and Technology Infrastructure Program, this project was tasked to build a data management and sharing service system with a tertiary structure of "principal database, scientific data center or network, and data gateway". Outline of the National Medium- and Long-term Planning for Development of Science and Technology (From 2006 to 2020) emphasizes that "the construction of Platforms of Scientific and Technological Fundamental Conditions should be strengthened by the joint-force of large-scale scientific engineering and facilities, scientific data and information platform, and natural science and technology resource service platform" [13].

In the period of "the 11th Five-Year Plan" to "the 13th Five-year Plan", various national ministries and commissions, as well as some research institutes and medical institutions, started to fund or participate the construction of biological and medical health-related data centers.

During the period of "the 11th Five-Year Plan", relevant departments of the state organized the construction of the National Platforms of Scientific and Technological Fundamental Conditions; scientific data was one of the six fields. Chinese Academy of Medical Sciences (CAMS) was taking the lead for the construction of the Chinese Medical Science Data Sharing Network, Chinese Academy of Sciences (CAS) is responsible for the construction of the Chinese Life Science Data Sharing Network. National science and technology development plan for the 12th Five-Year period developed by the Ministry of Science and Technology also proposed to "further improve the construction of scientific databases in different fields and industries, expand the pilot program for data collection and promote scientific data sharing" [14]. Since "the 13th Five-Year Plan", the construction of "national scientific data centers", including the "National Genomics Data Center" by Beijing Institute of Genomics, CAS, the "National Microbiology Data Center" by Institute of Microbiology, CAS, and the "National Population Health Science Data Center" of CAMS, has been accelerated.

In 2015, more than 30 academicians and experts in bioinformatics accentuated "the urgent need for the construction of national biological information center in China" after recapitulating over 20 years' experience and lessons. In 2016, the proposal of "National BMBD Infrastructure" (NBMI), jointly put forward by Shanghai Institutes for Biological Sciences (SIBS), CAS and Shanghai Municipality, was officially listed by the National Development and Reform Commission (NDRC) as one of the five backup projects for the "National 13th Five-Year Plan for Major Scientific and Technological Infrastructure Construction" [15]. In the same year, the Biomedical Big Data Center of SIBS was created to execute the 1st phase of NBMI pilot in the collaboration with "Zhangjiang Laboratory". The 2nd phase of pilot was funded by Shanghai Municipality and launched in 2018. At the end of 2019, CAS initiated the construction of "National Biological Information Center".

Starting from 2016, a "$1 + 7 + X$" master plan for the application and development of healthcare big data was initiated by National Health and Family Planning Commission, China. The model features one national data center, seven regional centers in the provinces (cities) of Fujian (Fuzhou, Xiamen), Jiangsu (Nanjing, Changzhou), Shandong (Jinan), Anhui and Guizhou, and several application centers from these provinces. Among these provincial centers, Jinan has the largest planned investment; Nanjing and Fuzhou are quick in implementation, with Nanjing Center putting its focus in the construction of gene database, and Fuzhou Center prioritizes in the collection and storage of hospital medical data. Other provinces and cities are also gearing to the master plan at different planning and implementation stages.

Other initiatives and plans put into action by universities, hospitals, industries, science and technology associations and national research institutes include: "National Engineering Laboratory of Medical Big Data Application Technology", jointly built by Wonders Information Co., Chinese PLA General Hospital and Central South University; "National Institute of Health and Medical Big Data", build collaboratively by The University of CAS, Chinese Center for Disease Control and Prevention, and Chinese Institute of Health Information and Healthcare Big Data; and "National Institute of Health and Health Data in Beijing University", by the partnership between University of CAS and Peking University.

Although great progress has been made in China in setting up data centers for BMBD, the data, however, are still largely in a decentralized state, and the cognition and actions for the complex connotations of BMBD in biology, medicine and data science are far from sufficient. They are the underlining factors for the lack of quality data services, standardization system, and practical applications. Better operational mechanism, and capacity- and team-building are required.

2.2 *Current Status and Trends of BMBD R&D by Medical and Health Institutions*

(1) Current Status and Trends of BMBD R&D by Medical and Health Institutions in US, UK and Germany

Medical big data is the collection of massive, real-world and continuous medical information covering those of diagnosis and treatment, health record, electronic medical record (EMR), medical image, as well as that of medical insurance, among which EMR is the kernel to be used to implement big data in health and medical institutions. In 2007, Health Level Seven International (HL7) published the *Electronic Health Record System Functional Model* (EHR-S FM), which was also approved by American National Standards Institute (ANSI). In 2009, the United States introduced the *Health Information Technology for Economic and Clinical Health* (*HITECH*) *Act* to encourage clinicians and hospitals to actively use EMR systems. Afterwards, many hospitals or institutions have accelerated the integration and application of clinical data. For instance, Beth Israel Deaconess Medical Center participated in the Doctor Medical Record Sharing Project (Open Notes) from 2010 [16]; Mayo Clinic etc. have been making massive investment in the infrastructure for big data collection and standard development since 2011.

The UK started the construction of a medical big data platform "care.data" in 2013, to collect medical records from hospitals and family doctors in the hope to realize data integration and utilization. In 2016, NHS stopped this plan. By learning the lessons of failed "care.data", the NHS puts its effort in the implementation of EMR, hoping to gain more power in the era of medical big data.

Germany has been endeavored to promote medical digitization process in recent years. In 2015, it passed the *E-Health Act* to accelerate the use of EMR. In 2019, the *Digital Care Act* [17] was passed, allowing doctors to offer video consultations to patient, prescription by mobile phone apps, and promote the uses of electronic prescriptions, EMRs and electronic sick-leave certificates. These measures greatly facilitate the collection, integration and management of medical data.

The collection, storage, integration and management of medical data represented by EMR are just the 1st step to utilize them. Further analysis and computation are required to transform such data to valuable information and knowledge. The exploration and utilization of the value of medical data are in general sub-optimal, so are the system integration and analysis of medical, biological, environmental and behavioral data.

(2) Current Status and Trends of BMBD R&D by Medical and Health Institutions in China

Many provinces and cities in China have started since 2006 the construction of regional health information platforms, by integrating clinical data from local hospitals, grass-roots clinics and public health centers, forming an individual-centered electronic health archives. The new-round of medical and health system reform started in 2009 further speeds up the capacity building for national medical information.

In 2006, Shanghai Shenkang Hospital Development Center launched "hospital-link project" in Shanghai. It has built a system recording patients' standardized electronic medical history data, and exchanging and sharing diagnosis and treatment information cross-hospitals at real-time. At present, the "hospital-link project" system has been implemented in 38 Shanghai municipal-level public hospitals linking with 16 grass-root clinics. The goal of Phase II of the project will be (a) to build structured EMR system that meets first-line clinical needs, and reaches the higher national standardization requirements, with more comprehensive coverage of municipal-level hospitals in Shanghai; (b) by using new technologies such as internet of things and edge computing, to build information system to manage key equipments and medical resources for Shanghai municipal-level hospitals,and the internet big data platform interconnecting the entire medical management ecosystem.

Beijing Tiantan Hospital established a unified data standard for cerebrovascular disease and a registry-based clinical study cohort—Chinese National Stroke Registry Studies, by using the common data elements from US NIH/NINDS. At present, high-quality of clinical research big data, including community cohort, clinical cohort, multi-center clinical trials and clinical image database, have been collected and managed. The most representative is the National Stroke Registry Study III, with a cohort of over 15,000 cerebrovascular patients, and a data collection of over 5,000 clinical phenotypes, high-resolution images and omics data.

As one of the informatization pilot units of National Center for Disease Control, Ningbo Yinzhou District Health Commission started to archive regional public health information and electronic health records in 2006, with the filing rate reached 96%. Up to 2016, it has completed the construction of "Health Big Data Platform" covering whole district. Yinzhou District Center for Disease Control also collaborates with its national counterpart in setting up an intelligent resident health index evaluation system based on the big data platform to achieve real-time automatic collection, processing, summarization and presentation of major health indicators.

2.3 Current Status and Trends of BMBD Development and Utilization by IT Enterprises

(1) Exploration by foreign IT enterprises: Layout and application scenarios of enriching BMBD

Foreign information technology companies explore the value of BMBD from various directions:

Medical informatization: Epic System and Cerner have a market-leading advantage in US at the development of EMRs. In last few years, they started to enter into cloud services and artificial intelligence business on the basis of their integrated patient and medical data, in the hope to transform data into more valuable information and knowledge.

Consumer health products and services: Alexa, a health product developed by Amazon using artificial intelligence, provides intelligent voice services in reminding the elderly to take medicine, managing blood pressure, and providing medical information services for hospitalized patients to get key information of medical terms, medical treatments, drug dosage and common diseases [18]. Verily, a life-health company owned by Google Alphabet, focuses on the development of AI-based medical solutions. Its smart watch has been approved by FDA.

Data application: In order to infuse big data into medical research, Google has formed broad alliances with pharmaceuticals such as Novartis (NVS), Otsuka, Pfizer and Sanofi, and academics such as Duke University and Stanford University.

Standard development: Apple is a major promoter for HL7 "Fast Healthcare Interoperability Resource (FHIR)" specifications. FHIR establishes a set of standards for different data elements in helping developers to build application programming interfaces (APIs) to access data sets from different systems, for solving data interoperability problem.

Server facilities and services: International Business Machine (IBM) has always been a strong provider of medical data services. Amazon and Google provide scientists with genomic data storage and analysis services, recently also speed up their planning and investment in health-related data collection services.

(2) Acceleration of BMBD development and application by domestic IT enterprises

Driven by the application demand, and the technology/tool advances in data analysis, as well as the support of engaging policies, Chinese IT companies also gradually entered BMBD business. By forming the alliance with biomedical and pharmaceutical enterprises, efforts are actively made to develop biomedical big data applications.

Wonders Information Co. has been committed to the field of "Three-Medical Linkage" for many years. Its health business covers 20 provinces in China. As an example, its Shanghai Health Information Network Project has achieved information mutual connectivity, mutual recognition and mutual validation between nearly 600 public medical institutions. Shanghai Sunshine Medical Procurement All-In-One, developed by the same company, provides the information support for national "4 + 7 Drug Procurement Platform Project". Health Cloud is the main gateway for provincial and municipal "Internet + medical health", providing closed-loop management services for millions of patients with chronic diseases.

Digital China Health serves various types of medical institutions, having an in-depth business layout in four core areas, i.e. health and medical big data, medical cloud service, medical and health informatization and precision medicine. It provides overall solutions for the next generation of medical informatization, including health and medical big data platform, cloud image platform, hospital information integration platform and precision medical platform.

Huawei is carrying out systematic technology research and development of a “fully connected medical” ecosystem by the provision of medical solutions, such as digital hospitals, regional health informatization, tiered diagnosis and treatment, as well as the development of wearable devices [19].

Ping An Medical Technology Co. and Institute of Medical Information CAMS jointly developed Chinese Medical Knowledge Atlas, an integrated platform with the comprehensive coverage of core medical concepts and knowledge in the medical ecosystem.

Tencent has not only invested in medicine-related firms Tencent Trusted Doctors, Micro Medical Group and HaoDF, but also created a big data and AI-based oncology joint-laboratory, the first in China, in the collaboration with Fudan University Shanghai Cancer Center [20].

Aliyun focuses service offering on the “basic infrastructure” of health big data. It developed ET Medical Brain 2.0, together with Ali Health, covering the application scenarios including clinic, medical research, medical training and teaching, hospital management and future urban medical brain.

2.4 Current Status and Trends of BMBD Policy Management

As the indisputable and unequivocal value of BMBD, concomitant fields, such as network security, data interoperability, information reliability, cloud infrastructure, integrative analysis, predictive modeling, tools for information management, as well as the public participation and information privacy, have attracted wide attentions.

(1) Current Status and Trends of BMBD Policy Management in the US and Europe

In recent years, a series of policies on medical data management have been released in countries of North America and Europe. The 21st Century Cures Act signed into law in 2016 added a section to the *Federal Food, Drug, and Cosmetic Act*. Pursuant to this section, U.S. Food and Drug Administration (FDA) has created a framework for evaluating the potential use of real-world evidence (RWE) to help support the approval of a new indication fora drug which was already approved. The US FDA issued the “*Framework for Real World Evidence”* in late 2018 recommending using RWE, giving it a full role in making regulatory decisions [21].

In the same year, FDA issued the “*Use of Electronic Health Record Data in Clinical Investigations Guidance for Industry*”, encouraging sponsors and clinical investigators to work with entities such as health care organizations to use EHR and EDC (Electronic Data Capture) systems as a data source for clinical research, improving the data accuracy and the efficiency for clinical trial.

National Institutes of Health (NIH in the United States formulated “*Data Sharing Policy and Implementation Guidance”* in 2003, that requires investigators submitting a research application requesting $500,000 or more in any single year to NIH to include a plan for sharing final research data, or state why data sharing is not possible

[22]. The 2003 policy was updated in 2019, grantees holding any NIH-funded grant would need to submit a detailed plan for sharing data, including steps to protect the privacy of research subjects [23]. In order to facilitate the implementation of *Precision Medicine Initiative* and to ensure information security (IS), an interagency working group that was co-led by the White House Office of Science and Technology Policy, the Department of Health and Human Services, and the NIH developed the "*Privacy and Trust Principles"* [24] to guide the use of medical data.

In May 2018, the *General Data Protection Regulation* (*GDPR*) became enforceable beginning in the EU. This regulation provides the following rights for individuals: the right to be informed; the right of access; the right to rectification, the right to erasure; the right to restrict processing, the right to data portability, the right to object; rights in relation to automated decision making and profiling [25]. It makes personal data more secure, patient records more intact, and data subjects (patients) in more control of their data.

(2) Current Status and Trends of BMBD Policy Management in China

China has also introduced a series of policies in the management of BMBD. In 2016, the General Office of the State Council issued the "*Guiding Opinions on Promoting and Regulating the Application and Development of Big Data in Health and Medical Services"*, mandating to build a national open application platform for tiered medical and health information by 2020, to realize the cross-sector and cross-regional sharing of basic data resources in population, legal personas and spatial geography, and to achieve significant data fusion in medicine, pharmaceuticals, medical insurance and other health-related fields. The "*Notice on Further Promoting the Informatization Construction for the Informationization of Medical Institutions with Electronic Medical Records as the Core*", published in 2018 by the National Health Commission, stresses the values of big data, calls for data inter-connectivity, and enforces rules of information security and health care data confidentiality.

In the same year, the National Health Commission issued the "*Administrative Measures on the Standards, Security and Services of National Health and Medical Big Data (Trial)"*, aiming to manage, develop and utilize health-care big data in the basic framework to guarantee citizens' right to know, to use and to protect their personal privacy. The standing committee of the 13^{th} National People's Congress (NPC) introduced in September 2018 the "*National Data Security Law"* and the "*Personal Information Protection Law"* into its five-year legislative plan. These two legislations provide strong legal basis for the development of digital economy, and harbinger a new era in the protection of personal information in China. The "*Notice on Issuing the Administrative Measures for the Application Level of Electronic Medical Record System* (*Trial*) *and Evaluation Standards* (*Trial*)*"* issued by the National Health Commission makes further requirement that "by 2020, all tier-3 hospitals should implement level 4 or above, all tier-2 hospitals reach level 3 or above" according to the 9-level evaluation standards.

The "*Standards and Norms for the Construction for the Informationization of National Primary-level Medical and Health Institutions* (*Trial*)" jointly issued in

2019 by the National Health Commission and the National Administration of Traditional Chinese Medicine, clarifies the content details and requirements of informatization construction for primary medical and health institutions. The "*Hospital Smart Service Grading Evaluation Standard System* (*Trial)*" issued by the general office of the National Health Commission in 2019 provides the classification standard for scientific and standard construction of smart hospital.

2.5 Application of BMBD in the Prevention and Control of COVID-19: Potential and Prospect

In 2003, the SARS coronavirus infected more than 8,000 people worldwide. Its spread was quickly brought under control because of the accumulated genome research data that enabled the sufficient understanding the disease. As the appreciation of "convergence" research and "precision medicine" developed afterwards, individual health and social security driven by "big data" have become a common pursuit of biomedical communities. The sudden occurrence of COVID-19 at the end of 2019 and its rampant spread to more than 200 countries by early April 2020, causing cumulative confirmed cases worldwide to exceed the level of millions, the virus has brought us an unprecedented challenge. In the process of anti-COVID-19 pandemic, genomic technology provided the full genome sequence of the virus almost instantaneously at the very beginning of medical identification [26, 27], a solid genetic basis for diagnosis [28, 29] and epidemiological analysis [30]. The genome sequences of similar or "related" viruses from bats and pangolins [31] provided the reference data for the provenance search of the virus. On this epic occurrence of pandemics, "big data", especially BMBD, has been instrumental to the functional annotations of viral genomes and biomedical studies and practices in epidemic identification, the delineation of the nature of viral infection, decision-making in prevention and control of viral spread, as well as the implementation of guidelines for virus detection, disease diagnosis and treatment. While BMBD system has presented its great application potentials in these application scenarios, new challenges in the utilization of big data have also become evident, highlighting the new prospect of development.

At the beginning of anti-COVID-19 epidemic, big data played an important role in providing timely data and information in epidemic monitoring, close-contact screening and epidemiological investigation. The Chinese Center for Disease Control and Prevention released technical solutions and literature reports, as well as dynamically updated domestic epidemics and latest control measures from WHO in its COVID-19 column [32]. The Department of Epidemiology and Biostatistics of Fudan School of Public Health formed a task force to build an epidemic-prediction model to dynamically predict the epidemic trend and possible regional risk of virus spread, and to provide advices to government, with the support of epidemiological real-time data, and the population flow data from the departments of public security and telecommunication. The 1st Affiliated Hospital of University of Science and Technology of China, along with iFLYTEK Health, has screened out suspected population with COVID-19 infection from more than 5 million community/grass-roots case records with the help of big data and intelligent voice-related technologies. It also used intelligent voice calling system tailored to educate and audio-follow-up the vulnerable population, and reconstructed transmission chain from some of them [33]. Data companies are also active in the development of relevant products. The mini program from Alibaba Intelligent Community Epidemic Prevention and Control is among more than 10 of such products listed by the Ministry of Civil Affairs for COVID-19 epidemic community prevention and control.

With the development of the pandemic, the searches for virus origin, its evolution and disease transmission route, epidemic characteristics, and the evaluation of reliability of disease prevention and control measures have been put on the agenda, which are not possible without the support of rich virus data, particularly their genomes. The 2019 Novel Coronavirus Resource database, provided by Beijing Genomics Institute (the National Biological Information Center) CAS, covers dynamically released COVID-19 genome sequences, their variation analysis data and relevant literature, etc. [34]. The Automated Identification Platform for Novel Coronavirus Genomewas jointly developed by Biomedicine Big Data Center (BM-BDC) of Shanghai Institute of Nutrition and Health, and the Institute Pasteur of Shanghai. Supported by Huawei Cloud technology and by working directly on genomic raw data, it has comprehensive and automatic processing and analytical functions from data quality control, splicing site and composition analyses, and online analysis of relative virus load [35]. BM-BDC has also developed a topological real-time analysis platform for thousands of viral genomes by leveraging machine learning methods, as part of its effort to develop a complete solution from virus genome identification to analysis.

BMBD has played an active role together with clinics in intelligent medical imaging, telemedicine, and online diagnosis and treatment. In the national health care information platform project, Huawei and its collaborators are working together to build an infrastructural cloud platform and application support platform, providing comprehensive supports to core businesses in public health, medical services, medical security and drug security; and to the basic medical care databases covering electronic health records, electronic medical records, and population information. **The most critical bottlenecks hard to rectify for BMBD are in the collection, management and analytical research of patient data from clinical diagnosis and treatment.** The 1st Affiliated Hospital of University of Science and Technology of China obtained the licenses from Oxford University to use standardized clinical epidemiological study protocol and associated case registration form (CRF) [36], introduced clinical research trial database system REDcap [37], and develop standardized operational procedures for clinical study execution (eSOP), by which 881 COVID-19 patients' clinical data were collected by 8 collaborators.

The application of big data has also accelerated the screening of "drug repositioning" for the treatment of COVID-19 infected disease. For example, Tianjin University of Traditional Chinese Medicine used the traditional Chinese herb ingredient database to screen for effective ingredients, and identified two "hopeful" herbs for the treatment. Another example came from Shanghai University of Science and Technology and its collaborators. Assisted by AI virtual drug screen platform, the team screened more than 2,900 drug molecules and over ten thousands of traditional Chinese herb ingredients.

Meanwhile, **the government has established a specialized research information exchange platform with universities and research institutions.** Tsinghua university and China Knowledge Center for Engineering Science and Technology have jointly built COVID-19 open data resource AMiner for the amalgamation of data, information and knowledge from epidemic, scientific research, media and policy-makers. The Ministry of Science and Technology (MOST), the National Health Commission, the China Association for Science and Technology and the Chinese Medical Association (CMA) have jointly built COVID-19 academic exchange platform to continuously update and harmonize academic resources, and promote significant scientific achievements.

The Chinese government has actively introduced supporting measures to combat the pandemic with BMBD. The State Administration of Health and Medical Administration released the "*Notice on Issuing the Novel Coronavirus Infection-Related ICD Code"* [38] in a timely manner, providing the semantic support for accurate and effective collection of patient clinical data, and efficient integration and analysis of clinical diagnosis and treatment information.

Although BMBD, together with genomics technology, has fully demonstrated its great scientific potential and social impact in pandemic prevention and control, its full value, however, is yet been realized. The bright prospect of BMBD can be only realized by overcoming the current obstacles facing to us today. **The core challenges, in midst of early diagnosis, early determination of virus transmission, early decision in disease prevention and control, early implementation of the measures for clinical intervention and disease prevention and control, as well as the early involvement of biomedical researches, lie in our ability to formulate sound scientific judgment and hypothesis based on the early occurrence of pandemics. Only these obstacles were overcome, could hypothesis be validated and optimized smoothly in "practical studies".**

There are two cognitive sources for us to comprehend new things or events such as emerging infectious diseases: **one is the current actual conditions (information), and the another is the summarization of the lessons learned from the past (knowledge). "Information" is represented by the connections between "data" items, while "knowledge" is the extraction of intrinsic and mechanical "interaction" between a large amount of "information" units.** If the data from public health, clinical medicine and scientific research are stored *in silo* at different hosting bodies, the lack of unified data access interfaces and standardized data collection and management, and the ambiguous data ownership and responsibility in data administrative departments, data interconnectivity and integrity would be greatly compromised. Ultimately it will be limited to a great extent of its value to support comprehensive data analysis and sound decision-making for social, medical and scientific bodies and governmental departments.

At present, the epidemic has not yet reached its end. There are still many open questions in the areas of virus pathogenesis, clinical diagnosis, treatment and prognosis of the disease, vaccine design and testing, and the development of new drugs, all of which call for joint in-depth studies by basic researchers, clinicians and epidemiologists, to provide scientific interpretation and solutions to the pandemic. **As the foundation to support all these works, a system with the capability to quickly collect real-time and accurate data comprehensively and continuously, and to systematically analyze and delineate data (raw data) and information (information at all levels) has its paramount importance, not only for scientific research and clinical practice, but also for timely scientific decision-making. Decisive steps are thus strongly suggested to take, as summarized in the following chapter, to guide the right direction to combat the pandemic that has never been experienced in the history of mankind.**

3 Policy Analysis and Recommendations on the Development of BMBD in China

The development of BMBD will propel life science research into a new paradigm of data-intensive science (including synthetic biology and convergence researches) and will revolutionize medicine (translational and precision medicines) and the healthcare industry (nutrition, pharmacy, health management and intervention). The potential impact of BMBD in improving human health as well as social harmony and development could never been over-estimated. In order to facilitate the development of BMBD in China, the authors have postulated four key recommendations in order to confront and overcome the identified BMBD challenges.

3.1 Construction of a Basic Science and Technology Service Platform for BMBD with an Integrative Infrastructure Comprised of Nationwide Distributed Special Nodes

The necessity of setting up an integrative "bioinformatics center" at the national level has been proven through 30 years of international practice to be essential for integration and sharing of biological big data and BMBD. Its establishment in China is quickly evolving. The Chinese government has made its investment on the creation of "national major scientific infrastructure", and is actively planning the "national laboratory" of such nature. Along with these commitments, however, a more in-depth consensus and higher level of strategic planning are required than ever before to establish a **BMBD basic science and technology service platform** (hereinafter referred as "service platform") with the framework of an integrative central infrastructure and distributed nationwide special nodes, by joining the efforts and resources of the "national bioinformatics center", "national data centers" and "biomedical data infrastructure". The insights, summarized from the above comprehensive review of BMBD, will be demonstrated as follows.

(1) **The service platform framework of "combination of integrative center with nationwide distributed special nodes" is an essential requirement due to the complex dual-nature (biological & medical) property of BMBD**. BMBD can be structurally divided into two categories: the unstructured or less structured data from the "objective world", and the more structured data from the human-influenced "research world". This dual-nature property of BMBD determines the dual-natured missions of the platform: data-oriented science/technology research and development, and data services driven by the best practices of engineering. It also determines the architectural frame work of the platform: a centrally integrated core facility focusing the governance of data from research world, and the distributed nodes focusing the data at the objective world for medical and social applications.

(2) **Biological big data is an inseparable component of BMBD**. Both the core value and the key challenges of biological big data are mainly derived from and connected to medical and pharmaceutical data. The modern medical, pharmaceutical and health industries also highly depend on the integration of these data sets with biological big data and related applications.
In the past 30 years, a tight bond of data from biology and medicine has been gradually developed, thanks to the rapid accumulation of BMBD and its great application potentials in research and medical practice. Obviously, it is unwise to artificially impose segregation between these two data sets. Besides, due to much less scientific and social complexities in other biological data-rich fields, such as agriculture, ecology and the environment, etc., the experience and know-how accumulated through tackling BMBD can be readily applied. Unfortunately, various types of "data centers" or "bioinformatics centers" built recently at different institutional levels or in different scientific domains in China are nevertheless largely separately operated and managed to this date. The bottleneck caused by the disparity of being "united in name, divided in practice" is one of the main reasons for the under-performance of BMBD. Therefore, it is necessary, also highly possible, to transform the "weakness" of current incomplete self-contained systems into the "strength" of integrating the distributed specialized nationwide nodes with the central infrastructure as a whole system of the national service platform, so as to lay a high-quality and efficient data foundation for long-term biomedical development.

(3) **The efficient and advanced support services by integrative central infrastructure to specialized nodes will be the cohesive force to develop this "integration and distribution" BMBD service platform**. Current BMBD practices have move gradually from "data island" to "data chimney", a positive move from "data *silo*" (complete isolation) to "data *systēma*" (targed/feature dintegration, but isolated from its sister domains). The equivalent data *systema* of INSDC are Genbank, ENA and DDBJ, each of which has developed its own

data format and unique data content with added values. The collaboration of them is a set of INSDC services covering data exchange and release policy, data format standard, and guidelines to ensure the data interoperability and integrity across the consortium and beyond. As international biological data centers, NCBI also hosts databases such as Taxonomy and Refseq that are *de facto* standards for biological databases, EBI excels in genome structural and functional annotations (Ensembl) and unique protein resources such as UniProt and Pride, microbiome and knowledge eatlas. Without the "integrative" role of INSDC, it is impossible for its member databases to be inter-connected with other data types such as proteins, transcripts and genomes. Therefore, data standards and exchange mechanisms should be developed and maintained at the central platform that providing central and basic data services with flexible interfaces and stable services for easy implementation, while each of the distributed nodes should develop databases, or data *systēma,* in compliance with published standards, and should also be encouraged to develop their unique resources that complement the whole. **The harmony of "integrated" center and "distributed" nodes is an essential requirement of any planned future BMBD service platform**. The challenge is that BMBD features multi-scale, high-dimension and high-heterogeneity data while the requirements for the platform services targeting research and application scenarios are extremely diverse, each with its own unique spatial–temporal characteristics. In addition, the paradox between high-speed data growth in distributed regions and the need for rapid data processing capability is becoming evident. **A cloud-based, grid BMBD supporting architecture at the national level, being in compliance with integrated and distributed principles, is thus undoubtedly the solution of choice for the implementation of the BMBD service platform**.

(4) **"Data service" is the most fundamental and important mission of the BMBD platform**. Big data is generally considered holding a high value. However, due to the intrinsic data complexity, non-standardized data generation, and fragmented data collection and storage, the value density of BMBD is relatively low. Therefore, good data services at a state-sponsored unified basic platform, being "non-profit" and high-tech, is crucial to raise the value of BMBD via increasing its value density. The service portfolio of the platform should be governed by the principles of "security management, information sharing, technology innovation, value-added standardization, respect for property rights and efficient utilization". Of course, in order to win the trust of their peers and users, the managers of the platforms, especially the integrative core of the platforms, do not, and should not, put their efforts to explore BMBD pursuing their own interests as their main mission. They should instead act and function as trustworthy stewards for data management.

(5) **The core scientific foundation for the effective application of BMBD is to realize its standardized integration, interactive sharing and intelligent analysis and mining**. This should be built into the platform's basic scientific and technological capability enabling it to provide quality services. It can be implemented into an integrated system with 3 key components: (a) a big data

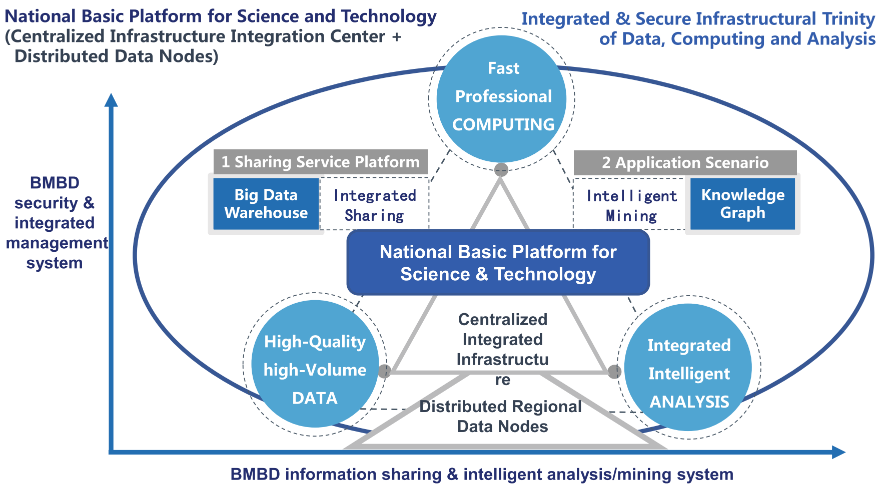


Fig. 2 The basic scientific and technological service platform for BMBD: Integrated core facilities with geographically distributed specialized nodes

warehouse built by standardized and secure integration; (b) an interactive shared network system supported by rapid professional computing facilities and; (c) intelligent application scenarios for integrated analysis of high-quality big data and knowledge atlas. This system should also be well connected with various nodes of scientific fields and regions, in order to ensure high-efficiency servers (Fig. 2).

3.2 *Optimization of Data Integration with Security Management and Data Sharing with Interactive and Efficient System for the BMBD Governance*

China is the world's most populated nation and has recently become the world's second largest economy. As a result, it is expected that its BMBD management system will have a significant influence worldwide. In recent years, some BMBD management standards and administrative guidelines have been established at various levels and/or provincial areas of data governance authorities. **The optimization of these standards and guideline, however, is required within the next few years** in order to accommodate the emerging needs from rapid development, and to remedy the historical weakness in providing professional technical support from service platform.

The BMBD management system should be built for the efficient and standardized utilization of data in the fields of biomedical research and application. BMBD is closely related to personal privacy, social and economic stability and development, as well as national security. Therefore, the usage of BMBD is naturally complicated and requires the careful and conscious enactment of data security measures at the individual level, as well as that of the state. **The secure and responsible use of BMBD is, therefore, an endeavor requiring major effort and a long-term commitment.**

This secure data management service platform can only be realized by the standardized integration of biomedical core research data and basic clinical data. At the same time, the service platform will only win trust with the successful dispatch of a high-security system, so that standardized data integration will be realized. **The safe and efficient utilization of data is therefore guaranteed based on the performance of the platform from both technical and administrative levels.**

To accomplish the above objectives, strong and comprehensive technical and engineering supports are highly desired. A committed centralized leadership is also required to govern and coordinate service providers at the different levels of BMBD platforms. It should also be emphasized that, without such a leadership, the successful implementation of government's security policies and regulations for data services will be severely compromised, and the utilization of data will be hindered due to non-technical reasons.

State laws and regulations are the safe-guards for data security management. Data security is accomplished by issuing policy specifications for secure integration of BMBD, as well as the policy framework to resolve possible conflicts between data sharing and privacy protection. A policy framework eliminates non-security factors in the integrative process of BMBD, encouraging efficient, interactive data sharing while ensuring necessary data security. **Chinese central and local governments have made some significant efforts in legislating data security management. However, the effect of these efforts has been limited due to social sensitivity, compounded by other complex issues arising during implementation. "Stepwise progression" and "regional trial" approaches are suggested based on the above observations, by which citizens, society, government and legislative bodies can work together to implement and optimize the data security legal system in the process of reform.**

Related to the above suggestions, a hierarchical, cohesive and stable governance system, built from central governmental administrative and trade offices and their local counterparts, must be established and maintained long term. Biosecurity and biosafety mechanisms should be devised through close collaboration with data platforms to facilitate data-related legislation and implementation for the process of data collection and integration, to ensure quality service in data stewardship and utilization of the platform.

3.3 Improvement of the Mechanisms for the Transformation of BMBD from Scientific Innovation to Application R&D

BMBD is developing at an unprecedentedly rapid pace. Because of its inherent complex 4 V/3H characteristics, BMBD possesses a series of practical and theoretical challenges for standardized data integration and efficient data use in the areas such as imagology, clinical laboratory science, data science, computer science and information technology. Innovative research and technological integration are the answers to these challenges. In the data service arena, engineering challenges have proven prominent as well. The reality is that systematic and theoretical breakthroughs are unlikely to emerge in the early years of practice. Instead, targeted research and development directed to real application scenarios have been popular among clinical medical institutions, the health survey industry, as well as the small/medium start-up information sector. For some biomedical giants, more resources for powerful computing have been invested in the hopes of gaining a competitive edge. The drug industry has a strong need and motivation to use BMBD to speed up their drug discovery process. However, these effort shave often been frustrated in practice due to considerable obstacles in data sharing, hampered either by ideological resistance or policy restrictions.

Therefore, regardless whether it is a national BMBD platform for basic science and technology, or a national BMBD system for security management, the core mission should be to provide quality services **to fulfill the needs of R&D and application in a broad scope. The transformation mechanism that links research, development and industrial application should be to build the platform and system on the pillars of sound engineering and legal framework for bio-security and bio-industry.** In addition to using cloud interfaces to partition data storage and data utilization, and using blockchain technology to resolve conflicts between data sharing and intellectual property protection, we can build a **medical terminology system**. The system integrates the medical system nomenclature-clinical terminology (SNOMED CT), the unified medical language system (UMLS), and the general framework of medical language, encyclopedias and generic naming of terms (GALEN). It should be highly compatible with clinical practices, facilitating the collation and unification of key medical terms, classifications and codes, realizing and gradually expanding the scope and the level of data sharing in an orderly and controlled manner. Finally, it should also encourage collaborative development and data mining to improve data utilization.

Meanwhile, a regulatory monitoring process should also be established for the development of BMBD products and applications, to gradually optimize validation criteria for the effectiveness and safety of applications, and to expedite collaborative innovation.

3.4 The Building Up of a BMBD Team at the Nationwide Level Requires Multi-disciplinary Talents and Engineering Professionals

Due to the strong data science properties of BMBD, and the growing opportunities in data business development from the continuous advance of machine learning and artificial intelligence technology, the demand for professionals far exceeds the existing talent pool of computational biology and bioinformatics. A large number of engineers and technicians are urgently needed—from data cleansing to data service providing—the recruitment of whom has proven challenging in the biomedical fields. In addition, professionals in medicine and public health should be encouraged to learn skills to comprehend big data, to prepare them to work and collaborate at this interdisciplinary converging point of biology, medicine and data, and jointly foster disciplinary development.

The **sharp discords between the supply and demand for talent** in the building and development of qualified teams has become a burning issue. Research institutes, higher and intermediate educations should be tasked with producing scientific and engineering talents at different levels, in quality and quantity, to fill the gap of the skill sets in data management, information transformational computing and medical data interpretation, and ease the constraint of current high demand. **The policies favoring the growth of such talents and teams,** e.g. appropriate accountability and evaluation incentive mechanisms designed specifically for certain talents, **are at the core of the success in solving this problem.**

The mechanism by which talents in different fields can work together, exchange their specialized experiences and share their research and development results should be established. The mechanism should emphasize collaborative achievements, rather than some counter-productive metrics such as certain performance "rankings". An exchange and cooperation platform should be provided for research teams from different disciplines to strengthen information exchange, and to nurture the development of shared cognitive methodology and research systems. Advanced data analysis and processing technologies should be used to solve clinically significant problems, to improve the efficiency of medical practice, and to promote the breadth and depth of medical research. It is also important to build an innovative entrepreneurship ecosystem that encourages talent teams to rigorously and realistically explore the intrinsic rational and scientific value of BMBD.

In conclusion, the aforementioned policy recommendations are the result of authors' rational thinkings and summaries of global and domestic BMBD events and activities from academic "discipline" to application "field", and from "national governance planning" to "social organization activities". Because biomedicine is still undergoing a fast-paced development and evolution, the "generalness", "objectiveness" and "feasibility" of our recommendations about BMBD are inevitably subject to the limitations of our knowledge and interpretation. Therefore, we sincerely hope that our readers, either from BMBD administratives, research and application institutions, or stakeholders and the

vast number of participants and users, will benefit from the use of this review, and also send us their valuable critiques and suggestions. We believe that only when such a scientific open discussion is started, will solutions be found to break bottlenecks, facilitate the healthy development of BMBD for the overall interests of the state and the society. China's engagement in BMBD will, and should, make a solid contribution to life science research and medical practice for the community of mankind.

Acknowledgements We would like to express our thanks to Yiming Bao, Runsheng Chen, Na He, Ping He, Boyang Ji, Yong Jiang, Xia Jin, Ruifeng Jing, Xu Lin, Guangya Li, Ye Li, Xiao Liu, Zefeng Wang, Jianping Weng, Hongfei Yang, Guoqing Zhang, Jingyi Zhang, Luxia Zhang, Haokui Zhou, Kaixin Zhou, Weimin Zhu and other experts who have been engaged in the R&D and management of biomedicine big data. They have provided information or valuable comments and suggestions for the writing and modification of this review.

References

1. Sagiroglu S, Sinanc D (2013) Big data: a review. IEEE Int Conf Collab Technol Syst (CTS), 42–47
2. Kang N, Ting C (2015) Big data for biomedical research: current status and prospective. Chin Sci Bull 60(5):534-546
3. Hastings A, Arzberger P, Bolker B et al (2005) Quantitative bioscience for the 21st century. Bioscience 55:511–517
4. Vinson V, Purnell BA, Zahn LM et al (2012) Does It compute? Science 336(6078):171
5. Lander ES1, Linton LM, Birren B, et al (2001) Initial sequencing and analysis of the human genome. Nature 409(6822):860–921
6. Bourne PE, Lorsch JR, Green ED (2015) Perspective: sustaining the big-data ecosystem. Nature 527(7576):S16-17
7. Guoqing Z, Yixue Li, Zefeng W, Guoping Z (2018) New challenges and trends in bio-med big data. Bull Chin Acad Sci 33(8):852–860
8. Gaoyan Ou, zhanxing Z, Dong B, Weinan E (2017) Introduction to data science. Higher Education Press
9. Tansley S, Tolle K (2009) The fourth paradigm: data-intensive scientific discovery. Microsoft Research, Redmond, WA
10. Stevens H (2018) Globalizing genomics: the origins of the international nucleotide sequence database collaboration. J Hist Biol 51(4):657–691
11. Health Data Research UK. About health data research UK [EB/OL]. https://www.hdruk.ac.uk/, 12 Jan 2020
12. Bailin H (2000) Proposal for establishing national biomedical information center as soon as possible. Bull Chin Acad Sci 15(2):133–134
13. State Council of the People's Republic of China. Outline of the national medium- and long-term plan for scientific and technological development (2006–2020) [EB/OL]. https://www.gov.cn/jrzg/2006-02/09/content_183787.htm, 31 Dec 2019
14. Ministry of Science and Technology of the People's Republic of China et al. China's 12th five-year plan for scientific and technological development [EB/OL]. https://www.gov.cn/gzdt/2011-07/13/content_1905915.htm, 13 July 2011
15. National Development and Reform Commission et al. National 13th five-year plan for major scientific and technological infrastructure construction [EB/OL]. https://www.ndrc.gov.cn/xxgk/zcfb/ghwb/201701/t20170111_962219.html, 11 Jan 2017

16. Walker J, Darer JD, Elmore JG et al (2014) The road toward fully transparent medical records. N Engl J Med 370(1):6–8
17. German Digital Care Act: Industry Experts Examine The New Law's Impact In 13th MedTech Radar. [EB/OL] https://www.htgf.de/en/german-digital-care-act-industry-experts-examine-the-new-laws-impact-in-13th-medtech-radar/, 30 Dec 2019
18. Alexa [EB/OL] https://www.alexa.com, 31 Dec 2019
19. Huawei. Huawei's Serving Big Health with Fully Connected Medical Solutions. https://www.huawei.com/cn/press-events/news/2015/09/huaweiyiquanlianjieyiliaofuwu, 22 Sept 2015
20. Chun W "Internal Strength" shown by China's first tumor laboratory based on AI big data [EB/OL]. https://www.xinhuanet.com/tech/2019-02/28/c_1124172577.htm, 28 Feb 2019
21. U.S. Food and Drug Administration. Framework for Fda's Real-World Evidence Program [EB/OL]. https://www.fda.gov/media/120060/download 07 Dec 2018
22. National Institutes of Health (NIH). NIH Data Sharing Policy and Implementation Guidance [EB/OL]. https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm, 31 Dec 2019
23. Kaiser J Why NIH is beefing up its data sharing rules after 16 years, NIH Data Management and Sharing Activities Related to Public Access and Open Science [EB/OL]. https://www.sciencemag.org/news/2019/11/why-nih-beefing-its-data-sharing-rules-after-16-years, 11 Nov 2019
24. Precision Medicine Initiative: Privacy and Trust Principles [EB/OL]. https://allofus.nih.gov/protecting-data-and-privacy/precision-medicine-initiative-privacy-and-trust-principles, 31 Dec 2019
25. European Union (EU). General data protection regulation (GDPR) [EB/OL]. https://gdpr.eu/, 31 Dec 2019
26. Fan W, Su Z, Bin Y (2020) A new coronavirus associated with human respiratory disease in China. Nature 579(7798):265–269
27. Xintian X, Ping C, Jingfang W (2020) Evolution of the novel coronavirus from the ongoing Wuhan outbreak and modeling of its spike protein for risk of human transmission. Sci China Life Sci 63(3):457–460
28. World Health Organization (2020) Instructions for submission requirements: in vitro diagnostics (IVDs) detecting SARS-CoV-2Nucleic Acid. https://www.who.int/diagnostics_laboratory/200228_final_pqt_ivd_347_instruction_ncov_nat_eul.pdf?ua=1 28 February 2020
29. Xiaolu T, Changcheng W, Xiang L et al (2020) On the origin and continuing evolution of SARS-CoV-2. Natl Sci Rev, 7(6):1012–1023
30. Peng Z, Xinglou Y, Xianguang W et al (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579(7798):270–273
31. Tommy Tsan-Yuk Lam, Marcus Ho-Hin Shum, Huachen Z et al (2020) Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. Nature (online)
32. Chinese Center for Disease Control and Prevention. COVID-19 Column [EB/OL]. https://www.chinacdc.cn/jkzt/crb/zl/szkb_11803/. 01 March 2020
33. iFLYTEK Health. iFLYTEK Health—Serving Healthy China with AI [EB/OL]. https://www.iflytek.com/health, 01 March 2020
34. CNCB/BIG, CAS. COVID-19 Information Base [EB/OL]. https://bigd.big.ac.cn/ncov/, 01 March 2020
35. Virus Identification Cloud (VIC) [EB/OL]. http://www.ecas.cas.cn/xxkw/kbcd/201115_128157/ml/xxhcxyyyal/202003/t20200306_4554740.html, 17 Feb 2020
36. International severe acute respiratory and emerging infection consortium. COVID-19 Clinical Research Resources [EB/OL]. https://isaric.tghn.org/protocols/clinical-characterization-protocol/, 30 March 2020
37. Harris PA, Taylor R, Minor BL, et al (2019) The REDCap consortium: building an international community of software platform partners. J Biomed Inform 95(103208).
38. National Health Commission of the People's Republic of China. Notice on Issuing the Novel Coronavirus Infection-Related ICD Code [EB/OL]. https://www.nhc.gov.cn/yzygj/s7659/202002/dcf3333b740f4fabad5f9f908d1fc5b4.shtml, 30 March 2020

Guoping Zhao is a molecular microbiologist and currently the Professor and Chairman of the Advisory Committee of CAS-Key Laboratory of Synthetic Biology of Shanghai Institute of Plant Physiology and Ecology (SIPPE), Chinese Academy of Sciences (CAS), after serving as its founding director for the term of 2008–2016. He has been the Chief Scientist of the Big Data Center for BioMedicine at the Shanghai Institute of Nutrition and Health (SINH), CAS since 2016, and Director of the Department of Microbiology and Microbial Engineering at the School of Life Sciences, Fudan University since 2004.

Guoping ZHAO has been working on microbial physiology and metabolic regulation since he received his B.S. degree of Biology from Fudan University in 1982, followed by his postgraduate studies in the PUB Program of Purdue University, Indiana, USA and receiving the Ph.D degree in 1990. He returned to China in 1992 as the founding production manager of Shanghai Promega Biological Products, Ltd, a subsidiary joint venture of Shanghai Research Center of Biotechnology (SRCB), CAS. He started working at the Shanghai Institute of Plant Physiology (SIPP, currently SIPPE), CAS in late 1994 as the Professor and Director of the microbiology laboratory. He served as the Director of SRCB from 1997 to 1999 and the Vice-President of Shanghai Institutes for Biological Sciences from 1999 to 2002, where he organized and participated the Human Genome Project of CAS (1998–2001). He worked as the Executive Director of the Chinese National Human Genome Center at Shanghai during 2002–2016 and the Director of the National Engineering Center for BioChip at Shanghai during 2001–2015. Along with his scientific duties in program/project organization and institutional administration, he focused his research activities on genomics and systems biology, largely for microorganisms, of which, he devoted and contributed to a few important projects including the molecular evolution study of SARS-CoV during the 2003–2005 SARS epidemic.

Guoping Zhao was elected to the CAS in 2005 and as the President of the Chinese Society for Microbiology for the term of 2006–2011. He was elected to the Third World Academy of Sciences in 2011. He was awarded a Doctor of Agriculture honoris causa, at Purdue University of the USA in 2014.

The Progress of High Performance Computing in China and the Development Trend of International High Performance Computing

Zuoning Chen

Abstract The development of high-performance computing comprehensively reflects the scientific and technological strength of a country, and is also an important part of the national innovation system. China has promoted the development of high performance computing by deploying a variety of national science and technology projects and funding plans, and gradually formed a national high performance computing service environment with a certain scale. Based on the landscape of the world TOP500 high performance computers, this paper firstly analyzes the latest progress of high performance computing in China; reviews the main shortcomings of HPC R&D in China; then explains the technological breakthroughs and applications of Sunway TaihuLight supercomputer; and finally discusses the R&D upsurge of Exascale computers in the world. In addition, this paper predicts the future development trend of high performance computing technologies.

Keywords High performance computing · Sunway TaihuLight computer · Gordon bell prize · Exascale computing

1 Introduction

High performance computing (HPC) is a national strategic high-tech technology and an important means to solve various challenging problems such as national security, economic construction, and social development and so on. It is regarded as "the pillar of the nation" that leads the development of scientific and technological innovation. HPC has become the main technological battlefield of high-tech competition and big countries' power-game in the information era.

Over the years, China has deployed a variety of national science and technology projects and funding programs to promote the development of high performance computing, successfully developed a number of cutting-edge high performance computers, and gradually formed a national high performance computing

Z. Chen (✉)
Chinese Academy of Engineering, Beijing, China
e-mail: znchen@tsinghua.edu.cn

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_3

service environment with a certain scale. For example, a series of high performance computers such as “Sunway”, “Tianhe” and “Sugon” supported by the national “863” projects have entered the leading ranks of the world [1].

With the arrival of the era of Big Data and Artificial Intelligence (AI), various industries and fields raise new requirements and challenges for computing power, which boosts the continuous development of high performance computing technology. The United States, Japan, Europe and other developed countries have made development plans to grasp Exascale computing power around 2021, some related R&D work has also been started. There was no doubt that the international high performance computers are entering a new stage of development.

2 The Latest Development of High Performance Computing in China

China adheres to the independent innovation road with Chinese characteristics in the development of high performance computers, and has made great progress in computer architecture, system software, microprocessors and application software through long-term efforts and hard work. In recent years, China has successively developed and deployed the world-leading Tianhe-2 and Sunway TaihuLight high performance computers, indicating that China’s high performance computing industry has entered a rapid development stage.

2.1 China’s HPC Achievements in the Recent TOP500 Lists

The international top500 supercomputers list (TOP500 [2]) released twice a year in June and November reflects the highest level and overall development status of high performance computers in the world from many aspects. Through the analysis of TOP500 lists in the past two years, we can see the latest development status of high performance computing in China.

(1) **China’s top system has stepped into the leading ranks of the world**

Since the Sunway TaihuLight high performance computer developed by the National Research Center of Parallel Computer Engineering and Technology (NRCPC) has been put into operation in 2016, it has ranked first in the TOP500 list for four consecutive times with the Linpack performance of 93.01Pflops, and led the global high performance computing into the 100P era. Until June 2018, the Summit supercomputer developed by IBM in the United States won the TOP500 championship with the Linpack performance of 122.3Pflops, and raised the performance to 143.5Pflops in November 2018. Table 1 shows the top five fastest supercomputers in the TOP500 lists in November 2017, November 2018, and November 2019. It can

Table 1 The top5 high performance computers in the TOP500 lists in November 2017, 2018 and 2019 (performance: Rmax/Rpeak)

Rank	November 2017	November 2018	November 2019
1	Sunway TaihuLight, NRCPC, 93.01/125.44Pflops	Summit, IBM, 143.50/200.79Pflops	Summit, IBM, 148.6/200.79Pflops
2	Tianhe-2, NUDT, 33.86/54.90Pflops	Sierra, IBM, 94.64/125.71Pflops	Sierra, IBM, 94.64/125.71Pflops
3	Piz Daint, Cray, 19.59/25.33Pflops	Sunway TaihuLight, NRCPC, 93.01/125.44Pflops	Sunway TaihuLight, NRCPC, 93.01/125.44Pflops
4	Gyoukou, ExaScaler, 19.14/28.19Pflops	Tianhe-2, NUDT, 61.44/100.68Pflops	Tianhe-2, NUDT, 61.44/100.68Pflops
5	Titan, Cray, 17.59/27.11Pflops	Piz Daint, Cray, 21.23/27.15Pflops	Frontera, Dell, 23.52/38.75Pflops

be seen that the top high performance computer systems developed in China have stepped into the leading ranks of the world.

(2) **China has the largest number of high performance computers in the world**

The number of systems in the TOP500 list reflects the usage breadth of a country's high performance computer systems. For a long time, the number of TOP500 computers in the United States has steadily ranked first in the world. In November 2017, the number of high performance computers in China was up to 202, greatly surpassing 143 in the United States for the first time and ranking first in the world. It shows that in China there is a strong demand for high performance computing power. Table 2 lists the top five countries that owned high performance computers in the TOP500 lists in November 2017, November 2018 and November 2019.

(3) **The strength of Chinese high performance computer manufacturers is strong**

Table 2 The top five countries that owned high performance computers in the TOP500 lists in November 2017, 2018 and 2019

Rank	November 2017		November 2018		November 2019	
	Country	Number of systems	Country	Number of systems	Country	Number of systems
1	China	202	China	227	China	228
2	US	143	US	109	US	117
3	Japan	35	Japan	31	Japan	29
4	Germany	21	UK	20	France	18
5	France	18	France	18	Germany	16

Table 3 The top five manufacturers that developed high performance computers in the TOP500 lists in November 2017, 2018 and 2019

Rank	November 2017		November 2018		November 2019	
	Manufacturer	Number of systems	Manufacturer	Number of systems	Manufacturer	Number of systems
1	HPE	122	Lenovo	140	Lenovo	174
2	Lenovo	81	Inspur	84	Sugon	71
3	Inspur	56	Sugon	57	Inspur	66
4	Cray	53	Cray	49	Cray	35
5	Sugon	51	HPE	46	HPE	35

For a long time, American companies such as IBM, Cray and HPE have been in a monopoly position in the field of high performance computer manufacturing. However, in recent years, in addition to making breakthroughs in top systems, the number of high performance computers manufactured by Lenovo, Inspur and Sugon in China has increased rapidly. In November 2018, all the three companies surpassed American companies, becoming the top three high performance computer manufactures in terms of number of machines developed in the TOP500 list. In the TOP500 list in November 2019, Chinese companies manufactured a total of 325 high performance computers, accounting for 65%. Table 3 shows the top five manufacturers that developed high performance computers in the TOP500 lists in November 2017, November 2018 and November 2019.

2.2 Main Deficiencies in the Development of High Performance Computing in China

China's high performance computers have developed rapidly and made great progress in recent years. However, from a deeper level, there are still some shortcomings and deficiencies in the research, development and application of high performance computers in our country.

First of all, the industrial foundation of China's high-end chip manufacturing is weak, and some core processes, devices and equipment are still subject to other countries. At present, most of the high performance computers developed in China are dependent on foreign imported chips, so the US embargo on high-end chips will have a certain impact on the future development of high performance computers in China, especially in terms of development cycle, research and development cost. In addition, China's self-designed high-end processor chips cannot be manufactured in China due to that we do not master the relevant advanced technology.

Secondly, there is more integrated innovation while less original innovation in the key core technologies. The main reason is that China lags behind the developed countries such as the United States and Japan in the basic research and education of high performance computers, the teaching level of related majors such as computer architecture, parallel algorithms and high performance computing is not high, and there is a lack of innovative talents. As a result, the original innovation ability of our country in the core technology of high performance computing is not strong. Although China's high performance computers are far ahead in the number of TOP500, there are few systems entering TOP100 and TOP50, and they have not yet formed a quality advantage.

Thirdly, computational science, which leads the development of high performance computers, still lags behind. Aiming at application models and physical models, we study and establish mathematical models and algorithm models to form corresponding algorithms, and then we can use high performance computers to solve complex practical problems. Although the number of TOP500 computers in China has surpassed that of the United States, and we have also won two Gordon Bell Prizes with the Sunway TaihuLight supercomputer. However, there is still a large gap compared with the United States, Japan and Europe in more extensive application modeling and algorithms, which hinders the overall application depth of high performance computers in our country.

Finally, high performance computing software and ecological development are relatively falling behind. Different from the relatively balanced development and accumulation of software and hardware in the United States, Japan and other countries for decades, the development of high performance computing software in China lags behind the development of hardware, and a mature software ecological environment has not been established yet. As a bridge between hardware and users, the lag of high performance computing software and ecological development has become an important factor restricting the application of high performance computers in China, and it is also an important problem to be solved urgently. Thus it is necessary to strengthen the research and development of parallel middleware software according to the characteristics of independent research and development chip architecture, and establish a perfect domestic high performance computer software ecology.

3 The Technical Breakthroughs and Applications of Sunway TaihuLight

The Sunway TaihuLight high performance computer is funded by the National 863 Program and developed by the National Research Center of Parallel Computer Engineering and Technology, which is deployed in the National Supercomputing Center in Wuxi. The system is the first high performance computer with peak performance of more than 100Pflops in the world. It is the first high performance computer that made entirely using domestic processors in China and ranked first in the world [3].

3.1 System Overview

Sunway TaihuLight is a very large-scale parallel processing computer system, which adopts a high performance architecture based on high-density elastic super-nodes and high-traffic composite network architecture for multi-objective optimization. The system is composed of high-speed computing system, secondary computing system, high-speed computing interconnection network, high-speed computing storage system, secondary computing storage system and corresponding software system (as shown in Fig. 1). High-speed computing systems and secondary computing systems are managed uniformly through the cloud management environment to provide users with a unified system view.

SW26010 heterogeneous many-core processor used by Sunway TaihuLight is developed by Shanghai High Performance Integrated Circuit Design Center with independent technology, using 64-bit independent SW instruction set, with 260 cores in the chip and a standard operating frequency of 1.5 GHz, and it has a peak performance of 3.168Tflops [4].

The peak performance of the Sunway TaihuLight high-speed computing system is 125.44Pflops, the total memory capacity is 1024 TB, the total memory access bandwidth is 4473.16 TB/s, the symmetrical bandwidth of the high-speed interconnect network is 70 TB/s, the aggregate I/O bandwidth is 341 GB/s, the Linpack performance is 93.015Pflops, the total system power consumption is 15.37 MW, the

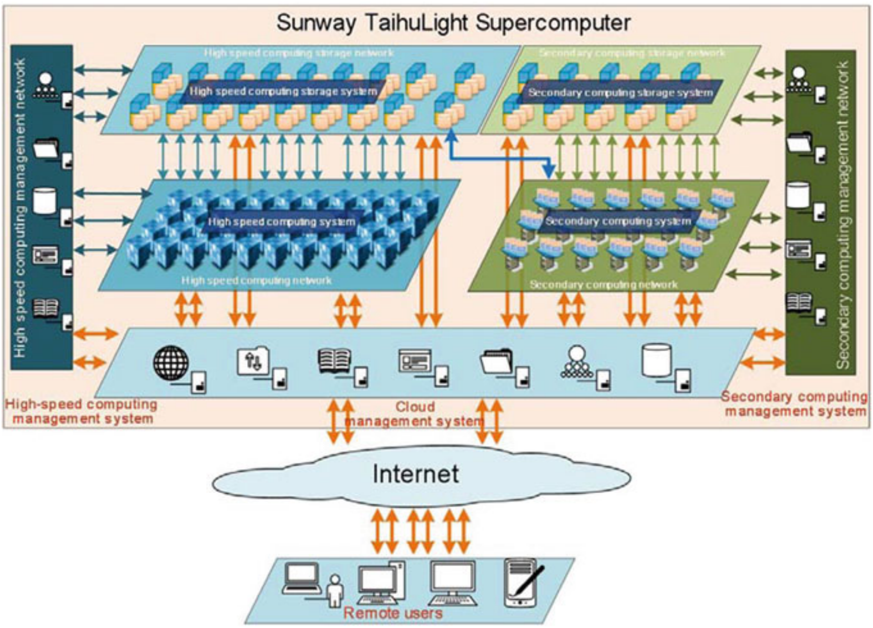


Fig. 1 The overall architecture diagram of Sunway TaihuLight system

performance and power consumption ratio is 6.05Gflops/W. As for the secondary computing system, the peak performance is 1.09Pflops, the total memory capacity is 154.5 TB, and the total disk capacity is 20 PB.

3.2 Major Technological Breakthroughs

(1) **For the first time, we completely used domestic processors to build the world's top high performance computers**

Based on domestic processors and using self-developed efficient scalable architecture, we designed and implemented a high-density computing tightly coupled elastic super-node structure. In the super-node, 256 CPUs are fully cross interconnected without cable. Between the super-nodes, the super-node elastic scalability under the full system scale is realized by the resource pool hot backup technology, which supports the efficient parallel operation of massive computing cores. It can meet the needs of computing-intensive, communication-intensive and I/O-intensive projects. A high-traffic scalable composite network structure composed of super-node network, shared resource network and central switching network is proposed, which realizes high-bandwidth and low-latency communication of 40,960 computing nodes and 240 I/O nodes.

(2) **For the first time, we independently developed and implemented the world's leading many-core processor**

We independently developed and implemented the 1.5 GHz SW26010 many-core processor, and the key innovations are shown as follows: (1) a heterogeneous many-core architecture based on on-chip computing array cluster and distributed shared memory is proposed to improve computing power and data sharing efficiency; (2) fixed-point and floating-point reuse and logic operation reconfiguration technology are proposed to implement an efficient and simplified core structure and improve the energy efficiency of the processor; (3) the technologies of register-level data communication, multi-mode asynchronous data flow transmission and fast synchronization of operation array are adopted to improve the cooperative execution efficiency of the operating cores; (4) the physical design technology of on-chip data full-path error correction, storage interface protection and on-chip hot spot noise suppression and isolation are employed to improve the basic reliability of the processor.

(3) **We realized advanced low-power design and control system in the world**

Multi-level power reduction measures are adopted, such as chip-level and component-level low-power design technology, system-level power management technology based on the combination of software and hardware, and compiler-guided application-layer power optimization technology. Low-power foundational support design is carried out in domestic CPU, computing system, network system, power supply and cooling, high-density assembly, etc.; fine-grained power detection system,

system state awareness system, external task-driven multi-level low-power control system are established; a hierarchical collaborative system-level power consumption control and management system is also established, which can effectively reduce the operation power consumption and achieve efficient green computing without affecting the system performance and usage mode.

(4) **We established a high-concurrency software system for tens of millions of cores**

Firstly, multi-layer and multi-granularity parallel job control, multi-strategy resource scheduling technology in heterogeneous environment are adopted to support efficient management of tens of millions of cores. Secondly, compilation optimization techniques such as heterogeneous fusion and efficient foundational compilation support framework, multi-level adaptive data layout, and data-driven multi-mode memory access optimization are adopted to improve application performance. Thirdly, message model-oriented runtime awareness and program anomaly diagnosis techniques are developed to reduce the cost of large-scale debugging. Fourthly, a data reuse method based on many-core array direct communication is proposed to improve the adaptability of memory access-intensive and communication-intensive problems. Besides, irregular rectangular static load balancing algorithm and multi-granularity dynamic task evaluation mapping algorithm are proposed to solve a number of major application problems.

(5) **For the first time, our applications are shortlisted and one of them was awarded the highest international prize for high performance computing applications**

The Gordon Bell Prize, which aims to recognize the major achievements in international parallel computing applications, is known as the Nobel Prize for High performance Computing. In the year that Sunway TaihuLight was put into use, three applications in the fields of atmosphere, ocean and materials were shortlisted for the 2016 Gordon Bell Prize for the first time in nearly 30 years. Finally, the application "10 M-Core Scalable Fully-Implicit Solver for Non-hydrostatic Atmospheric Dynamics [5]" won the prize, achieving a breakthrough in this prize and breaking the monopoly of western developed countries. In 2017, the application "18.9-Pflops Nonlinear Earthquake Simulation on Sunway TaihuLight: Enabling Depiction of 18-Hz and 8-m Scenarios [6]" won the Gordon Bell Prize again. These achievements indicate that the performance advantage of Sunway TaihuLight has been successfully transformed into real application advantage.

3.3 Application of the System

Since it was put into use, Sunway TaihuLight has completed the computation of hundreds of users' application projects, involving 20 application fields, such as weather and climate, aerospace, marine environment, biomedicine, shipbuilding engineering, and so on. Extreme-scale parallelism of millions of cores have been realized, including 22 full-system applications (10 million cores), 12 applications with more than half-system scale, and more than 40 applications with more than one million cores [7].

According to the international classification standards for scientific and engineering computing applications [8], we classify the applications as follows: (1) Dense linear algebra, such as LINPACK, large-scale fluid–solid coupling and fluid-acoustic coupling computation, full-directional sound scattering characteristics of submarine transceivers, etc. (2) Sparse linear algebra, such as numerical simulation of hypersonic aircraft, stall characteristics of C919 large plane, etc. (3) Spectral methods, such as direct numerical simulation of turbulence based on FFT, BNU_ESM earth system mode, etc. (4) N-body methods, such as molecular dynamics GROMACS, MD simulation of micro-channel diffusion process, etc. (5) Structured grids, such as aircraft numerical simulation, compressible boundary layer turbulence direct numerical simulation, earth system model, seismic simulation, etc. (6) Unstructured grids, such as hypersonic vehicle numerical simulation, pollution emission simulation, etc. (7) MapReduce, such as Monte Carlo simulation option pricing, BLAST gene sequence alignment, tokamak runaway electron behavior simulation, etc. (8) Combinatorial logic, such as AES, MD5, etc. (9) Graph traversal, such as social network analysis, etc. (10) Dynamic programming, such as accurate gene sequence alignment analysis, etc. (11) Backtrack and branch and bound, such as SAT algebraic attacks, etc. (12) Graphical models, such as deep neural networks, hidden Markov models, etc., (13) Finite state machine, such as network protocol analysis, etc. The above 13 types of applications have all finished large-scale parallel computing on Sunway TaihuLight, and have achieved good application benefits.

Through the establishment of joint laboratories with national research institutions in various fields in China, we can provide an open environment and platform for talents in the field of science and engineering computing to work together, and encourage the enthusiasm of cooperation among different disciplines, different units and different personnel, to improve the depth and breadth of high performance computing applications in China. Besides, we encourage domestic enterprises to use domestic high performance computers through high-quality computing services and price preferential policies, and provide technical support for enterprise innovation and development, transformation and upgrading, and promote the transformation from "made in China" to "created in China". In addition, we build a sound ecology of domestic high performance computer software, research and develop efficient parallel industry application software, intelligent manufacturing cloud service platform and industry application computing service APP, to improve system usability, and take full advantage of the ultra-high computing power of domestic high performance computers. By strengthening the support to artificial intelligence, quantum computing, brain-like computing and other cutting-edge technology fields, we can realize the subversive innovation of some cutting-edge computing methods. It can be seen that high performance computers are playing a key and decisive role in the informationization of scientific research in China.

4 Upsurge of Exascale Computers R&D in the World

Exascale computer is the next milestone in the development of international high performance computing. At present, the United States, Japan and Europe have all formulated Exascale computer plans and accelerated research and development work. The reason that Exascale computer research and development is so hot is mainly driven by the following changes in application requirements.

Firstly, it lies in the classic scientific computing applications. The current practical complex application system is moving toward the direction of multi-spatial and temporal scale, strong nonlinear coupling and three-dimensional real model, which contains a large number of computational problems of multi-scale and multi-model, and there are multi-granularity, multi-dimension and multi-level parallelism. Faced with the computational simulation of the full system, the full physical process, true 3D and natural scale, there is an urgent need for the support of computing performance of Exascale and above.

Secondly, it lies in big data applications. The rapid development of big data research has expanded the application field of high performance computers, and there is a trend of integration of "big computing" and "big data". Classic scientific computing has a new entry point under the condition of big data, and some applications are undergoing changes from computing-intensive to data-intensive or hybrid-intensive, which is inseparable from the support of stronger high performance computing power.

The third lies in artificial intelligence applications. The rise of a new generation of artificial intelligence technology has brought new opportunities for the development of high performance computers. The deep learning applications based on tensor operations is still computing-intensive, but it does not pursue high-precision computing in algorithm principle, and puts forward higher requirements for single processor performance, memory access performance and network performance of high performance computers, therefore it needs supercomputing power support.

4.1 American Exascale Computers

At present, there are three Exascale computers under development by the United States: Aurora, Frontier, and El Captain.

(1) **Aurora supercomputer**

The system will be jointly developed by Intel and Cray, with an investment of more than $500 million dollars and a target performance of more than 1Eflops. It is scheduled to be delivered in 2021 and deployed at the United States Department of Energy's Argonne National Laboratory. It will be the first Exascale high performance computer in the United States. Aurora will facilitate Intel's next-generation Xeon scalable processors, Xe GPU, Optane DC memory and One API software, as well as

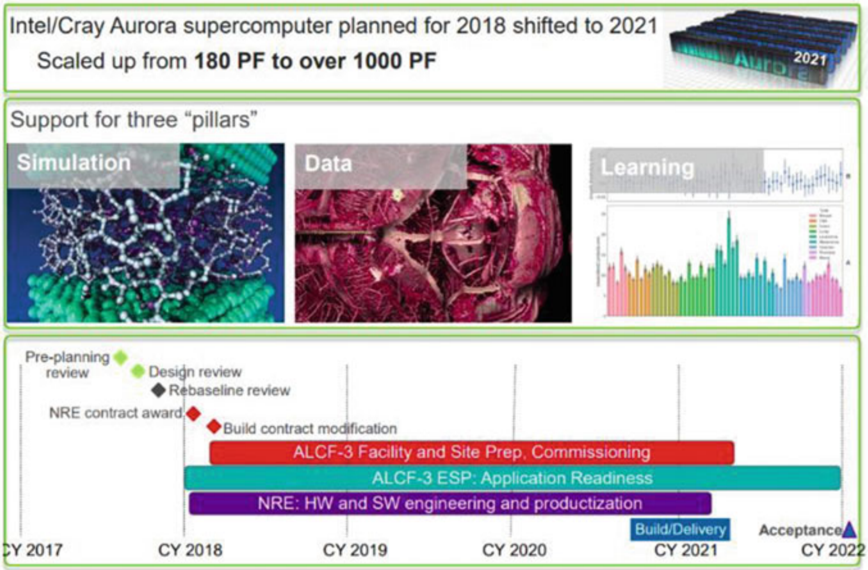


Fig. 2 The development schedule and expected applications of Aurora system [10]

Cray's Shasta architecture for Exascale computing and the Slingshot interconnection system. Aurora will deeply combine classic HPC and AI, and its applications include cosmic simulation, new drug research and development, climate modeling, health care and so on [9]. Figure 2 shows the development process and expected application of the Aurora supercomputer.

(2) **Frontier supercomputer**

The Frontier Exascale computer will be jointly developed by Cray and AMD, with an investment of more than $600 million and a target performance of more than 1.5 Eflops. It is scheduled to be completed by the end of 2021 and deployed in the Oak Ridge National Laboratory of the United States Department of Energy. The system will be based on Cray Shasta architecture and Slingshot interconnect, using AMD next-generation EPYC CPU and Radeon GPU with high-bandwidth low-latency Coherent Infinity Fabric interconnect in between, and the power consumption of the system is lower than 40 MW. The Frontier will be used in the application simulation of nuclear energy systems, fusion reactors and precision drugs. Table 4 lists the system specifications of the Frontier supercomputer [11].

(3) **El Captain supercomputer**

The El Captain Exascale computer will also be developed by Cray and AMD, with an investment of more than US $600 million and a peak performance of about 2Eflops, scheduled to be put into operation in 2023 and deployed at the United States Department of Energy's Lawrence Livermore National Laboratory, and its main mission is nuclear weapon simulation. The system will also be based on the

Table 4 System specifications of Frontier Exascale supercomputer

Peak performance	> 1.5Eflops
Footprint	> 100 cabinets
Node	1 AMD EPYC CPU + 4 AMD Radeon Instinct GPU
CPU-GPU Interconnect	AMD Infinity Fabric; Coherent memory across the node
System Interconnect	Slingshot dragonfly network, 100 GB/s bandwidth
Storage	2–4 × performance and capacity of Summit's I/O subsystem

Cray Shasta architecture and Slingshot network, with a power-consumption between 30 to 40 megawatts. El Captain will use Shasta computing blades with four nodes, each of which will be equipped with 4 AMD Radeon GPU and an EPYC CPU. Cray is also exploring the integration of optical technology into interconnect to transfer data more efficiently and improve the power efficiency and reliability of the system [12].

4.2 *Japanese Exascale Computers*

The Exascale computer under development in Japan is the Fugaku system, which is being by Fujitsu with the support of the Flagship 2020 plan launched by the MEXT of Japan, with an investment of about US $1 billion and a target performance of 1Eflops, and will be deployed in Japan's RIKEN/Institute of Physics and Chemistry [13]. Fugaku will use Fujitsu's independently developed ARMv8 SVE (code name A64FX) processor, integrate 48 dedicated computing cores and 4 secondary cores, use 7 nm process production, contain 8.786 billion transistors, equipped with high performance HBM2 memory, 16 PCIe 3.0 channels; use 6D Mesh/Torus interconnect. The system power consumption is expected to be 30–40 MW. The R&D schedule of Fugaku is shown in Fig. 3. According to the latest news, Fugaku supercomputer will be released in June 2020 with the peak performances of 537Pflops for 64bitFP and 1.07Eflops for 32bitFP, which is expected to grab the top spot in the newest TOP500 list.

4.3 *European Exascale Computers*

In 2017, a total of 20 European countries jointly signed the EuroHPC Declaration to jointly promote the development and application of HPC in Europe [15]. According to the EuroHPC Declaration, Europe plans to develop two pre-Exascale systems and two

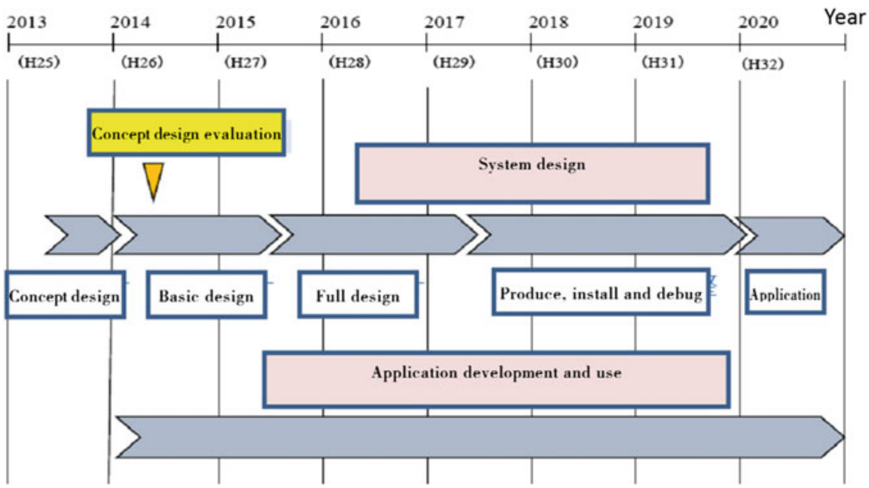


Fig. 3 R&D schedule of Fugaku supercomputer in Japan [14]

Exascale systems. The total cost of each Exascale system is estimated at 500 million euros. In September 2017, the European Union officially launched the EuroEXA project for Exascale computing. According to the plan, the EuroEXA project will develop high performance ARM and Xilinx FPGA chips, complete prototypes with new memory and cooling systems by 2020, and deploy Exascale computers around 2023. The German Forschungszentrum Juelich has installed and completed the first module of the JUWELS high performance computer manufactured by BULL in April 2018, laying the foundation for the future development of Exascale computers. Figure 4 shows the roadmap of the deployment of European Exascale systems.

4.4 Chinese Exascale Computers

China launched the 13th Five-year Plan key research and development project on high performance computing in 2016, with an implementation cycle of five years, according to three innovation chains: the development of Exascale high performance computers, the research and development of high performance computing applications, and the research and development of high performance computing environment, and a total of more than 20 key research tasks being deployed. With the support of this special project, the National Research Center of Parallel Computer Engineering and Technology, National Defense Science and Technology University and Sugon Company were approved to develop Exascale prototypes at the same time to explore possible technical routes for realizing Exascale systems. By October 2018, the Sunway Exascale prototype, Tianhe-3 Exascale prototype and Sugon Exascale prototype have all been developed and put into practice as planned. During the 14th

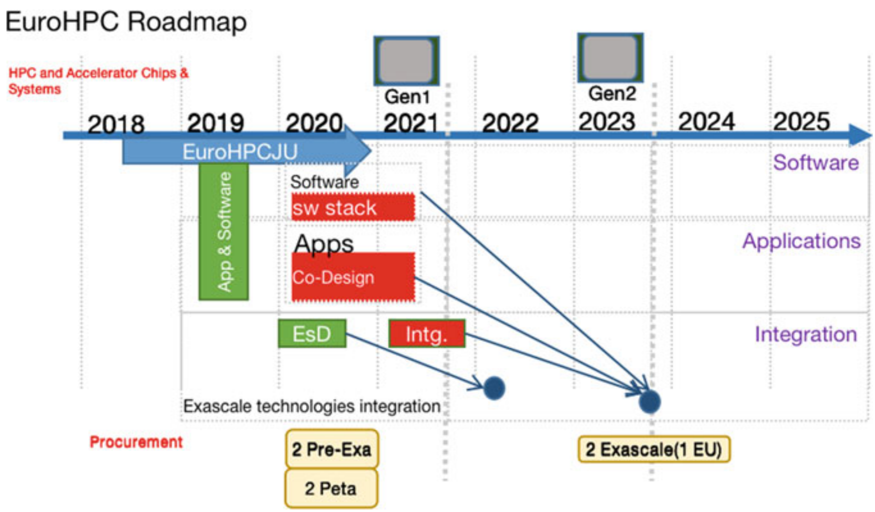


Fig. 4 EuroHPC roadmap [16]

Five-year Plan period, China's Ministry of Science and Technology will continue to support the research, development and application of Exascale high performance computers.

5 Development Trend of International High Performance Computing

With the gradual slowdown of Moore's Law, the process progress cannot meet the expectations. At present, the development of high performance computing continues to encounter obstacles such as power wall, programming wall, reliability wall and so on. The classic high performance computing technology faced many ceiling problems. In the near future, we need to continue to improve the problem-solving ability of high performance computers through innovative computing forms and computing models. In the long term, we should explore subversive alternative computing technologies to promote the development of high performance computers in the post-Moore era.

XPU + high performance computer refers to the integration of a large number of customized chips and general-purpose chips such as CPU and GPU into a heterogeneous computing system to meet the application needs of different fields. This customized special specification integrated circuit includes NPU (neural network processing unit) [17], DPU (deep learning processing unit) [18] and so on. For example, TPU (tensor processing unit) [19] is developed by Google to accelerate the computing power of deep neural networks. It uses 8-bit or 16-bit low-precision operations, places a huge amount of memory on the chip, closely adapts to deep learning algorithms, and has a significant improvement in efficiency and power consumption compared with general-purpose chips.

Big data + high performance computer means that high performance computers change from simply pursuing computing speed to paying attention to system throughput at the same time, transforming the design concept from the traditional computing-centric to data-centric, improving the large-scale computing efficiency of computing-intensive and data-intensive problems. The close connection between high performance computing and big data in the industrial ecological chain can better promote the reform and development of the organization mode of information resources [20]. The National Strategic Computing Plan of the United States has put forward the idea of the integration of high performance computing and big data, which will affect future scientific inventions, national security and economic competitiveness [21]. Cray announced the Unified Platform Strategy, regarding high performance computers as big data infrastructure, and established the technical route of analysis/big data/high performance computing [22].

AI + high performance computer refers to the deep integration of high performance computing and artificial intelligence. On the one hand, high performance computers can provide powerful computing support for artificial intelligence applications. The third wave of artificial intelligence marked by deep learning requires extremely strong computing power. High performance computer systems have been widely used in the field of deep learning research and engineering practice. The proportions of deep learning high performance computers in the international TOP500 lists are increasing rapidly. On the other hand, artificial intelligence technology also helps to solve many challenges faced by the development of high performance computers, and has a good prospect in system fault handling, scheduling resource use and online optimization decision-making. In the future, artificial intelligence applications will develop in the direction of higher dimensions and more complex models, so it is necessary to design and develop new architectures and core chips to explore new ways to deal with large-scale and complex artificial intelligence applications. In addition, the Deep500 benchmark is being developed to measure the deep learning performance of high performance computers [23].

Brain-like computing is a new ultra-low-power computing technology which is based on the principle of neuromorphology. Brain-like computing draws lessons from human brain information processing, breaks the shackles of von Neumann architecture, and can process unstructured information in real time with learning ability. IBM and Intel have designed TrueNorth [24] and Loihi [25] brain chips with complete functions respectively, and have developed high performance computing prototype systems based on 64 TrueNorth and 64 Loihi chips respectively [26, 27]. Intel says it will launch a larger brain-like computing system in the near future, which is expected to have more than 100 M neurons, 1 trillion synapses, 768 chips and 1.5 trillion transistors, providing unprecedented performance and efficiency [27]. Brain-like computing technology is likely to make a breakthrough and enter a period of rapid development in the next ten years.

Analog computing can break through the bottleneck of conversion between physical system information and binary information, and has great advantages and application potential. The United States Department of Defense has carried out theoretical research and practice on analog computers represented by probabilistic computing [28], exploring ultra-high-speed, ultra-low power consumption, high-precision and low-complexity computing technologies from algorithm design and computing structure to special chips. Under the same process conditions, the probabilistic computing processor can achieve 1000 times higher energy efficiency than × 86 processors [29]. The Defense Advanced Research Projects Agency (DARPA) has announced a program called Accelerated Computation for Efficient Scientific Simulation (ACCESS) [30]. The goal is to develop a new digital-analog hybrid computing architecture to simulate extremely complex systems in a scalable way.

In the long run, with the physical limit of classic computing technology approaching, researchers are trying to find supernormal computing methods beyond silicon-based CMOS, including quantum computing, biological computing, superconducting computing, optical computing and so on. At present, important progress and breakthroughs have been continuously made in these non-traditional computing fields. In particular, quantum computing technologies are developing most rapidly, and even some of them have moved out of the laboratory and turned to engineering design. For example, IBM has launched its new 53-qubit quantum computer, supporting online access in the cloud [31]; and Google has used its 53-qubit Sycamore chip to achieve "Quantum Supremacy" in solving the random circuit sampling problem [32]. But overall, there is still a long way to go before developing a full-sized and full-featured practical quantum computer. The effective combination of new concept technologies and traditional architectures to form a hybrid computing model, give full play to their respective advantages, and achieve a super-linear growth in the overall performance of solving complex problems, will be an important development trend of high performance computers in the foreseeable future.

6 Conclusions

Through long-term efforts, China has made great progress in the research and engineering of high performance computers, and has already achieved a world-class status and posture. However, it should be noted that the competition in this high-tech field is very fierce. We still have many deficiencies and shortcomings, which require continuous long-term investment and research. In the coming Exascale computing era, in addition to developing advanced high performance computer systems, efforts should be made to build a high performance computing ecological chain that covers system hardware (especially high-end CPU), system software, development tools, application software, and even talent. In this way, all aspects of research results can be transformed and integrated in time, and ensure the long-term sustainable development of high performance computing in our country.

References

1. Xianghui X, Hu S (2015) 863 plan of development of high performance computing in China. Technol Appl Sci Res InfIzation 6(4):3–10
2. TOP500 Organization. TOP500 Lists[EB/OL]. http://www.top500.org/, 18 Aug 2019
3. Fengbin Q (2017) Sunway TaihuLight supercomputer. China Comput Fed Commun 13(10):16–22
4. Zheng F, Li HL, Lv H, et al (2015) Cooperative computing techniques for a deeply fused and heterogeneous many-core processor architecture. J Comput Sci Technol 30(1):145–162
5. Yang C, Xue W, Fu H, et al (2016) 10M-core scalable fully-implicit solver for nonhydrostatic atmospheric dynamics. In: Proceedings of the international conference for high performance computing, networking, storage and analysis. Salt Lake City, USA, pp 57–68
6. Fu H, Yin W, Yang G, et al (2017) 18.9-Pflops nonlinear earthquake simulation on Sunway TaihuLight: enabling depiction of 18-Hz and 8-meter scenarios. In: Proceedings of the international conference for high performance computing, networking, storage and analysis. Denver, USA, pp 1–12
7. Xin L, Heng G, Rujun S, Zuoning C (2018) Large-scale application characteristics analysis and exascale scalability research on sunway TaihuLight. Comput J 41(10):2209–2220
8. Asanovic (2006) The landscape of parallel computing research: a view from Berkeley, vol 2. Technical report UCB/EECS-2006–183, EECS department, University of California, Berkeley, USA
9. US Department of Energy. US department of energy and intel to build first exascale supercomputer [EB/OL]. https://www.energy.gov/articles/us-department-energy-and-intel-build-first-exascale-supercomputer, 20 Aug 2019
10. US Department of Energy. Advanced Scientific Computing Research [EB/OL]. https://science.osti.gov/-/media/ascr/ascac/pdf/meetings/201709/HellandASCAC170926.pdf, 21 Aug 2019
11. Oak Ridge Leadership Computing Facility. Frontier Spec Sheet [EB/OL]. https://www.olcf.ornl.gov/wp-content/uploads/2019/05/frontier_specsheet_pdf, 20 Aug 2019
12. LLNL.GOV. LLNL and HPE to partner with AMD on El Capitan, projected as world's fastest supercomputer [EB/OL]. https://www.llnl.gov/news/llnl-and-hpe-partner-amd-el-capitan-projected-worlds-fastest-supercomputer, 12 March 2020
13. RIKEN Center for Computational Science. About the Project [EB/OL]. https://www.r-ccs.riken.jp/en/postk/project, 22 Aug 2019
14. MEXT. フラッグシップ2020プロジェクト (ポスト「京」の開発) について. [EB/OL]. https://www8.cao.go.jp/cstp/tyousakai/hyouka/kentou/flagship/151211/siryo5_1.pdf, 22 Aug 2019
15. EuroHPC Web Site. EuroHPC Declaration Sgned [EB/OL]. https://eurohpc.eu/declaration, 22 Aug 2019
16. e-IRG. The EuroHPC roadmap to exascale by 2022 in Europe [EB/OL]. https://e-irg.eu/news-blog/-/blogs/the-eurohpc-roadmap-to-exascale-by-2022-in-europe, 21 Aug 2019
17. ResearchGate. Neural processor in artificial intelligence advancement. [EB/OL]. https://www.researchgate.net/publication/328585193_Neural_processor_in_artificial_intelligence_advancement, 24 Aug 2019
18. Xilinx. Deep learning processing unit (DPU) architecture: an overview. [EB/OL]. https://www.xilinx.com/xilinxtraining/assessments/portal/ml-edge/dpu-over-view/story_html5.html, 24 Aug 2019
19. Jouppi NP, Young C, Patil N, et al (2017) In-datacenter performance analysis of a tensor processing unit. In: 2017 ACM/IEEE 44th annual international symposium on computer architecture (ISCA). [S.l.]. IEEE, pp 1–12
20. Guoliang C, Ye MRC (2015) High performance computing and related innovative technologies. Shenzhen Univ J 32(1):25–31
21. The White House. National strategic computing initiative strategic plan [EB/OL]. https://www.whitehouse.gov/sites/whitehouse.gov/files/images/NSCI%20Strategic%20Plan.pdf, 22 Aug 2019

22. Wire HPC. Cray lays out vision for HPC-Big data convergence [EB/OL]. https://www.hpcwire.com/2015/12/03/cray-lays-out-vision-for-hpc-big-data-convergence, 22 Aug 2019
23. Deep500. Deep500: an HPC deep learning benchmark and competition [EB/OL]. https://www.deep500.org, 22 Aug 2019
24. Semantic Scholar. TrueNorth: a high-performance, low-power neurosynaptic processor for multi-sensory perception, action, and cognition. [EB/OL]. https://pdfs.semanticscholar.org/4450/05167bef3cf7050b72fca805ba7032d37bfb.pdf, 24 Aug 2019
25. ResearchGate. Loihi: a neuromorphic manycore processor with on-chip learning. [EB/OL]. https://www.researchgate.net/publication/322548911_Loihi_A_Neuromorphic_Manycore_Processor_with_On-Chip_Learning, 24 Aug 2019
26. IBM. Lawrence livermore national laboratory and IBM collaborate to build brain-inspired supercomputer [EB/OL]. https://newsroom.ibm.com/2016-03-29-lawrence-livermore-national-laboratory-and-ibm-collaborate-to-build-brain-inspired-supercomputer, 24 Aug 2019
27. Wire HPC. Intel debuts pohoiki beach, its 8M neuron neuromorphic development system [EB/OL]. https://www.hpcwire.com/2019/07/17/intel-debuts-pohoiki-beach-its-8m-neuron-neuromorphic-development-system, 22 Aug 2019
28. Inside HPC. DARPA sees UPSIDE of analog computing [EB/OL]. https://insidehpc.com/2012/08/darpa-sees-upside-of-analog-computing, 22 Aug 2019
29. eWEEK. Lyric takes aim at intel, AMD with probability processing [EB/OL]. https://www.eweek.com/networking/lyric-takes-aim-at-intel-amd-with-probability-processing, 24 Aug 2019
30. DARPA. Accelerated computation for efficient scientific simulation (ACCESS) [EB/OL]. https://www.darpa.mil/program/accelerated-computation-for-efficient-scientific-simulation, 24 Aug 2019
31. Wire HPC. IBM opens quantum computing center; Announces 53-Qubit Machine [EB/OL]. https://www.hpcwire.com/2019/09/19/ibm-opens-quantum-computing-center-announces-53-qubit-machine, 25 May 2020
32. Arute F, Arya K, Babbush R, et al (2019) Quantum supremacy using a programmable superconducting processor. Nature 574:505–510. https://doi.org/10.1038/s41586-019-1666-5

Zuoning Chen, female, born in 1957, was graduated with a master's degree from the Computer Application Technology Department of Zhejiang University. Chen Zuoning is currently the Vice President of the Chinese Academy of Engineering. In 2001, she was elected the academician of the Chinese Academy of Engineering. Chen Zuoning is an academic leader in the system software and architecture fields of the IT area. She has presided and participated in many national major science and technology special projects, and made great contribution to helping home-made computer systems pursue and exceed the world's most advanced counterparts, and to developing home-made core hardware and software. At present, she is the principal of the "Cloud Computing and Big Data" project in the National Key R&D Program of China, guiding a team to build an independent cloud application ecosystem, leading the cloud computing application and industry in our country into the frontier of the world. Chen Zuoning won twice the grand prize of the National Prize for Progress in Science and Technology, twice the first prize of the National Prize for Progress in Science and Technology, and the QiuShi Award and Chinese Young Scientist Award from the China Association for Science and Technology.

The Informationized Application of On-orbit Data Processing of the Dark Matter Particle Explorer

Jin Chang, Jingjing Zang, and Liang Liu

Abstract DAMPE is the first satellite-based observatory of China targeting on astronomical objects. During its on-orbit observation, over 6 billion of high energy cosmic rays (CRs) have been detected and analyzed, providing several hundred of terabyte (TB) of date for various scientific purposes. Such rich observations have largely benefited the research of electrons (positrons), protons and gammas within the high energy CRs. In this paper, we would take a deep insight into DAMPE. Firstly, the structure and features of its data processing software DAMPESW is introduced. Then constructions of DAMPE infrastructures is explained that are built on the demand of high data processing performance. Finally, in hope of providing reference for other similar projects in the future, we summarize the informationized application of data processing on DAMPE.

Keywords DAMPE · On orbit operation · Cosmic rays · DAMPESW

1 Introduction

1.1 Overview of DAMPE

DArk Matter Particle Explorer (DAMPE) is a space satellite aiming at high energy CRs observations. Benefited by its high resolution and wide energy range of detecting high energy CRs and gamma ray emissions, DAMPE is used in the indirect search of dark matters through measuring or constraining physical properties of dark matter (DM) particles, including mass, annihilation cross section, decay lifetime and so on. Meanwhile, the detection of TeV electrons and heavy nucleuses would provide valuable clues for investigating the origin of CRs, and the observation of high energy gamma rays would both discover and exam the quantum gravity effects with high precision. In conclusion, the scientific goals of DAMPE are:

J. Chang (✉) · J. Zang · L. Liu
Purple Mountain Observatory, Chinese Academy of Sciences, Nanjing, China
e-mail: chang@pmo.ac.cn

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_4

(1) The first priority of DAMPE is the search of dark matter particles. The high spatial resolution and wide dynamic energy range allow DAMPE to search and study DM particles indirectly by detecting high energy electrons and gamma rays. These observations could not only provide measurements or precise constraints of the properties of DM particles (such as mass, annihilation cross section, decay lifetime), but can also limit the spatial distribution of them. Therefore, the observation of DAMPE is extremely valuable for making breakthroughs on DM researches.
(2) Investigating the origin of high energy electrons and nucleuses in CRs by detecting TeV particles.
(3) Obtaining achievements on gamma ray astronomy by detecting high energy gamma ray emissions.

1.2 Summary of DAMPE Payloads

In order to achieve scientific goals described in last section, 4 detectors are integrated within DAMPE as sketch shown in Fig. 1. From top to bottom they are the plastic scintillator detector (PSD), the silicon tungsten tracker (STK), the BGO calorimeter (BGO), and the neutron detector (NUD) made of plastic scintillator units, respectively.

(1) The plastic scintillator detector
The main scientific tasks of PSD are: determine the direction of incident particles, identify photons and electrons, and discriminate high energy heavy ions

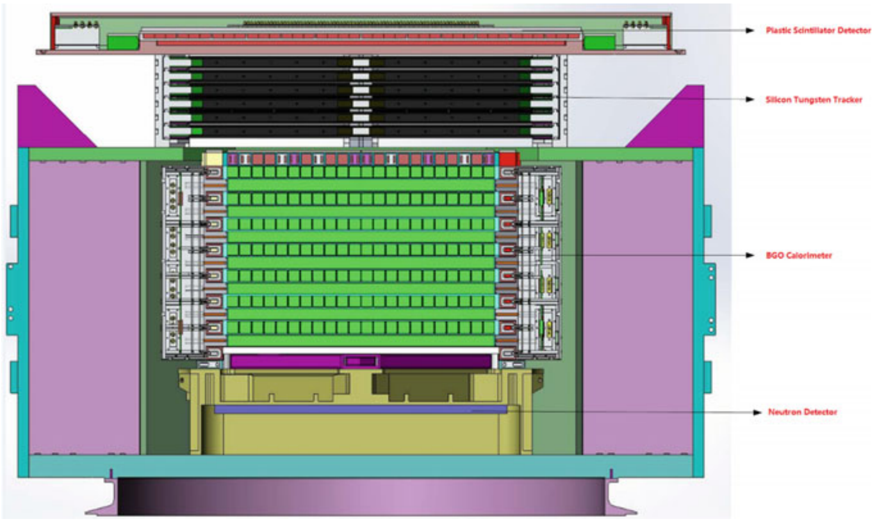


Fig. 1 The sketch of DAMPE payloads

categories (Z = 1~20). The PSD, has effective detecting area of 820 × 820 mm^2, is consisted of two layers, in which a total of 82 plastic scintillator units are placed perpendicularly in X and Y directions. Among them 78 units have identical sizes of 884 × 28 × 10 mm^2, and the other 4 have sizes of 884 × 25 × 10 mm^2. For next processing steps, photomultiplier tubes (PMTs) are set connecting to both ends of those units to convert optical signals to electrical ones.

(2) The silicon tungsten tracker
The STK is made of silicon micro-strip sub-detectors which are very sensitive to the position of incident particles. The main scientific tasks of STK are measure the direction of incident particles, identify electrons (charged particles) and gamma rays (non-charged particles), as well as measure high energy nuclides (Z = 1~26). Specifically, there are 6 layers of silicon micro-strip detectors in STK, each layer is consisted of two sub-layers of XY paralleled silicon micro-strip detectors to locate position (X and Y). The area of STK is 80 × 80 cm^2. Three tungsten plates are placed as conversion medium of gamma rays to meet the configuration requirement of STK. In order to identify charged and non-charged particles, especially gamma rays and electrons, tungsten plate do not place on the top of the first layer of silicon micro-strip detector. Instead, they are placed between the first and second layers, the second and third layers, and the third and fourth layers. Note that the outputs of STK are charge signals.

(3) The BGO calorimeter
The BGO is a fully absorption electromagnetic calorimeter. Its main scientific task is detecting CRs, especially the energy of high energy electrons and gamma rays (5 GeV~10 TeV). In the meantime, it can identify electron and hadrons based on different horizontal and vertical growths of showers, so that the high energy hadron (mainly protons) background can be excluded. Structurally, the BGO calorimeter is made of 308 BGO crystals and its area is approximately 60 × 60 cm^2. There are 7 major layers within the it and each of them are mad of two XY paralleled sub-layers, located by X and Y axis. Within each sub-layer, there are 22 identical detector units (BGO crystal) which have sizes of 2.5 cm × 2.5 cm × 60 cm. PMTs are used connecting to the both ends of units to convert optical signals to electrical ones.

(4) The neutron detector
NUD is a comprehensive detector of neutron moderation and detection. Its main scientific task is detecting secondary neutrons generated in the interaction between CR hadrons (mostly protons) and its top materials. Based on energy deposition of these neutrons, NUD could identify incident particles independently, which further supplement the identification of protons and electrons of BGO. Specifically, NUD is consisted of 4 independent squares cut from a 693 mm × 693 mm plastic scintillator detector, which is 1-cm thick and boron-doped (BC454, produced by Saint Gobain corporation). A corner has been removed for each square in order to couple the PMT, from which a compete detecting surface could be read out.

During the operation period of DAMPE, there are two observation modes for every detector listed above, calibration mode and observation mode. According to the initial design of DAMPE, the calibration mode of detectors is divided into three work settings: pedestal calibration, Digital-to-analog converter (DAC) calibration, and Multi-band Imaging Photometer calibration (MIPs). The purpose of observation mode is detecting CRs, such as high energy electrons, gamma rays, nuclides and so on. Correspond to CRs types chosen to measure, different triggering settings should be applied in order to rule out background events (mostly protons) as many as possible. During flight, detectors would keep repeating following steps: calibration → optimize parameters (high pressure, trigger delay, trigger threshold, etc.) → observation → calibration.

1.3 On Orbit Operation of DAMPE

At the present, the on-orbit operation of DAMPE is accomplished by the cooperation of ground supporting system from National Space Science Center, Chinese academy of Science, scientific application system from Purple Mountain Observatory, Chinese academy of Science, as well as the satellite telemetry and control of many related units. From 2015-12-20, the very day of receiving the first frame of scientific data, to 2019-08-05, the cumulative data receiving from DAMPE are 20208 orbits, which including 1324 days of on-orbit operation and 7 times of full sky survey. A total of 6.67 billion of high energy particles are detected and processed, generating 1B data 19.92 TB, 1F data 12.53 TB, 2A data 112.43 TB. Shown in Figs. 2 and 3 are the total events and average events per day, respectively. In the past three years of

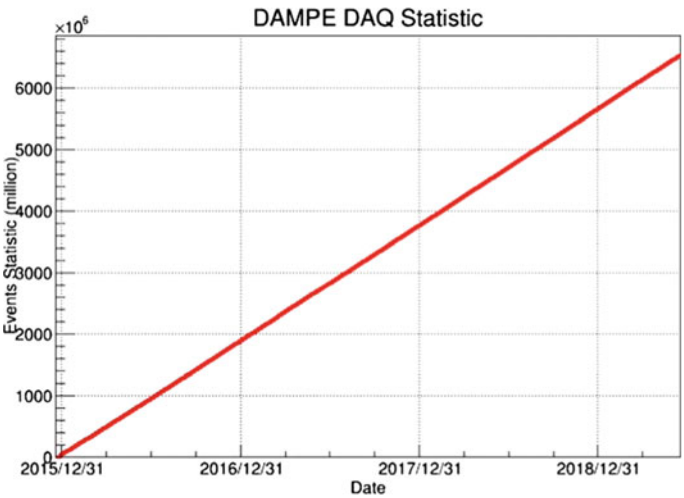


Fig. 2 Accumulated event data of DAMPE

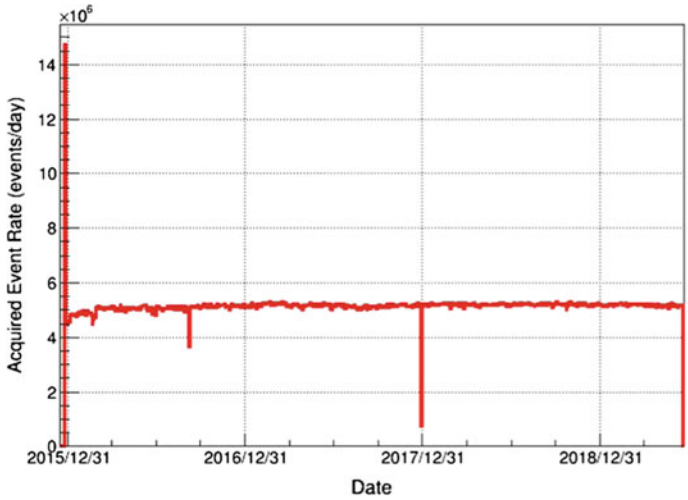


Fig. 3 The evet rate of DAMPE

on-orbit operation, multiple issues appeared (the reset of payload data manager, bad cluster in flash storage of data tube, single evet effect on high voltage power supply cabinet, etc.) and are adequately disposed. Compared to the period when satellite was just launched, no new bad tracks of electronic channels (70 thousand and more) are founded, and the whole satellite is still maintaining in a good condition, suggesting that design of satellite platform and payloads are very reliable.

At this moment, the most important task for DAMPE is continuously cumulate the observation data and release them to scientific groups for making scientific achievements.

1.4 Present Research Status

Currently, there are a number of international satellites also aiming at CRs detections, such as the Large Area Telescope instruments on board on Fermi Gamma-ray Space Telescope (Fermi-LAT), the Alpha Magnetic Spectrometer mounted on International Space Station (AMS02), as well as the CALorimetric Electron Telescope (CALET). Adopted to different properties and scientific goals of these projects, different data processing pipelines and software are developed.[1] For instance, the FermiTools is developed for the analysis of Fermi-LAT data (https://github.com/fermi-lat/Fermitools-conda/) and it has been continuously updated to meet new requirements, the latest release is v11r5p3. AMS02 also has its own offline data analysis software AMSsoft. However, the version of this software is unknown since it is not open to

[1] http://fermi.gsfc.nasa.gov/ssc/data/analysis/software.

the publics. Therefore, toward DAMPE observation, it is necessary to develop a data analysis software for scientific purposes.

1.5 The Advancement of DAMPESW

Borrowing data analysis experience from other similar satellite projects, multiple concepts, such as Kernel, Algorithm, Service and Event, are introduced into DAMPE Software (DAMPESW) that make it not only functional on the basic data processing, but also accessible, as a unified processing platform, for all users. Therefore, software development can be made between users without interference, making it possible to be continuously update.

The core and mainly body of DAMPESW are coded by C++. Boost.Python is employed linking C++ with Python so that the two languages could identify and coordinate with each other. Python, as a well-known programming language, is used for configuring user-customized algorithms to help them work properly. Meanwhile, the scripting language Bash is also used for the environmental configuration of software.

So far, DAMPESW has integrated three programming languages, C++, Python, and Bash, and it can also cooperate with ROOT, which is a data analysis software commonly used in particle physics. Consequently, DAMPESW is very convenient for new users that can process data easily and quickly.

To summarize, all the analysis threads of DAMPE observation data are unified in DAMPESW, meaning that it is convenient for independent users to compare their outcomes and improve their analytical techniques through communications. In other word, such a software would guarantee that outcomes are reliable, and DAMPESW could improve the efficiency scientific outcomes as well.

2 Processing of Scientific Data

2.1 The Scientific Application System of DAMPE

In engineering, the scientific application system of DAMPE is responsible for its operation and scientific data processing.

The system is also responsible for making observation plan, monitoring the status of payloads on board the satellite, generating advanced data products, calibrating payload, and organizing scientific research.

In order to ensure the achievement of scientific goals and the accomplishment of the overall task, the scientific application system is divided into three sub-systems (considering the function and workflow). They are the scientific operation sub-system, the scientific data & user managing sub-system, and the advanced data products processing sub-system. Each of those sub-systems are built to implement

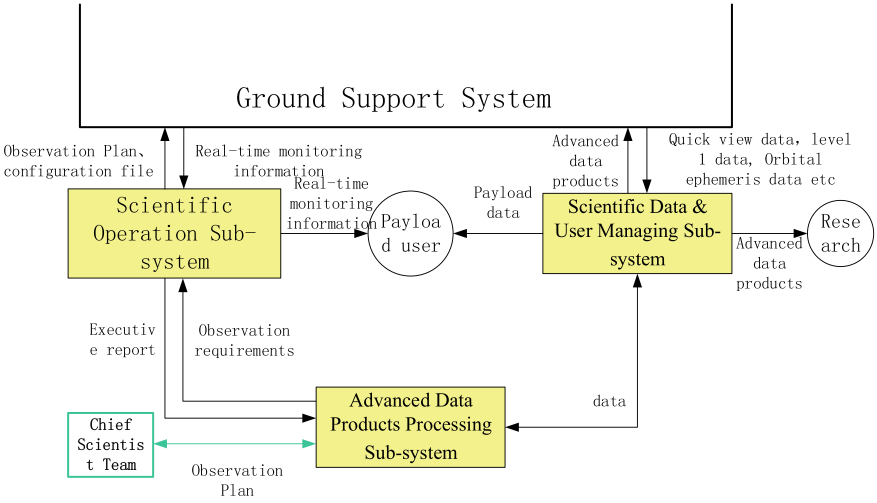


Fig. 4 Composition of the scientific application system

different functions corresponding to different stages. Among them, the scientific operation sub-system is responsible for building an effective mechanism for the operation and management of the satellite, including making short or medium-long term observation plan for payloads, monitoring the status of payloads, and determining the uplink parameters of satellite. The scientific data & user managing sub-system is mainly responsible for the management of data, including data intactness (guarantee data is intact during transmission), data transformation, data classification, storage of huge volume data, data calling and long-term preservation, calibrating and unifying time of scientific data, data archiving, backup and releasing, organizing scientific research and so on. Meanwhile, the advanced data products processing sub-system would take charge for generate advanced data products, analysis the background, identify particles, and build data base.

The internal I/O between these three sub-systems and the ground supporting system are present in Fig. 4.

2.2 Scientific Data Processing Software DAMPESW

2.2.1 The Facing Challenges of DAMPE Scientific Data

DAMPE is a multifunctional high energy particle detector with multiple component. Its main body is made of four sub-detectors and each of them is consist of a large number of sensitive detecting units. Same incident particle interacting with detector have various information that are recorded in every unit. From those units, the read-out information would be integrated, forming into the so-called Event at data level. Therefore, the data processing of users basically is to combine information and pick up interested events by adopting appropriate computing method. Such method, the so-called the Algorithm, will go through every event and perform identical operation to each information successively. Generally, in order to obtain a specific physical property of an incident particle, multiple algorithms are required. Also, at the different stages of data processing, algorithms should be adjusted based on different level of processing. Hence, to achieve one single goal, different algorithms need to be connected in some particular orders, and the interaction of outcomes between different algorithms would occurs. For instance, in the case of tracing particles, the track could be obtained in the cluster of STK through Kalman filtering, in which shower axis is needed (as the seed of track) that is reconstructed by BGO calorimeter. The reconstruction then requires the knowledge of the shower center of gravity of BGO calorimeters within each layer, in which the center of gravity of shower could only be calculated on the basis of knowing the energy of each BGO crystal. Therefore, various algorithms are needed to reconstruct the track of incident particles and they should be connected in a certain sequence.

The data processing of events demand for many physical algorithms which is developed by numerous of scientific groups based on different requirements. During processing, the execute of algorithms need auxiliaries from various functional modules which may also be demanded in other algorithms. Thus, the functional module, which is called as Service, should be callable and open to algorithms. Taking I/O for instance, data input and output are used in every algorithm and a I/O module (I/O Service) would be useful. Note that I/O service will only take charge for I/O. The only difference between service and algorithm is that service don not analysis and process the actual data of events, it existed only for providing assistant to algorithms in order to reduce pressure of algorithm developments.

DAMPESW is developed exactly for the requirements described above, the detailed functions are as follows:

1. Provide a unified development platform
   First of all, interfaces (virtual functions) are offered for developers to access the core of DAMPESW. Therefore, developers only need to focus on the specific functions of those interfaces without concerning other things, like the order of software drivers and the management of data (interaction and storage). They would run by the core.
   Secondly, all users' data processing would under the same software framework, which is convenient for communicating and sharing between users. This is good for the division and cooperation of developing software.
2. Provide unified processing mechanism
   All algorithms are executed by user defined python scripts (a text file) for flow control. Users need to load their algorithms and services into the core, then the software would be executed by following logics:

   Step 0. Configurations (formulating algorithms, inputs and outputs, etc.).
   Step 1. Initialization (DmpCore:Initialize(), like customizing histogram, define variables, etc.)

Step 2. Execution (DmpCore::Run(), go through every events).
Step 3. Cleaning and exiting current work (DmpCore::Finalize(), writing data to output files)

3. Provide flexible scalability
Depending on specific requirements, users could create their own algorithms, event classes, and services.

2.2.2 Components of DAMPESW

DAMPESW is consist of 5 main functional modules and 5 auxiliary modules. Those 5 main functional modules are RawDataConversion module, Reconstruction module, Calibration module, Simulation module, and Visualization module. The 5 auxiliary modules are Kernel module, Analysis module, Event module, EventOrbitSimulation & OrbitSimulation module, and Geometry module, shown in Fig. 5.

The functions of main functional modules are as follows:
RawDataConversion module: Converting raw data (binary) to TTree format could be recognized by ROOT, meaning that convert 1A/1B scientific data product to 1E/1F data product.

Reconstruction module: Converting the read-out values of electronic channels of 1E/1F data product to physical objects (energy, direction, particle types, etc.) through complicated calibration algorithm. Meaning of converting scientific data product of 1E/1F to 2Q/2A.

Calibration module: Analyzing calibration 1E/1F data products and generating calibration coefficients.

Simulation module: Mainly used for simulating the response of every sensitive detecting units inside detectors after particles (in the radiational space environment) entered detectors.

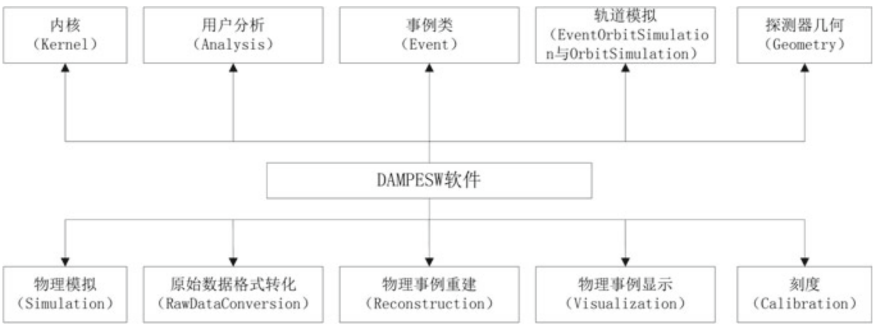


Fig. 5 Software structure of DAMPESW

Visualization module: Read 2Q/2A scientific data product and exhibit energies of detective units on the 3D geometric model intuitively, using colors to represent the energy level.

The functions of auxiliary modules are as follows:
Kernel module: Kernel module is the core of DAMPESW. It is responsible for executing and driving of all other modules, and provide unified I/O format and data interaction mechanism (between modules for all modules in the software). It would also provide base class for the extension of services and algorithms.

Analysis module: Integrated module of user-defined algorithms.

Event module: Definitive module for event data format.

EventOrbitSimulation & OrbitSimulation module: Module that simulate the environment of satellite orbit.

Geometry module: Module that storage the geometry data of detectors.

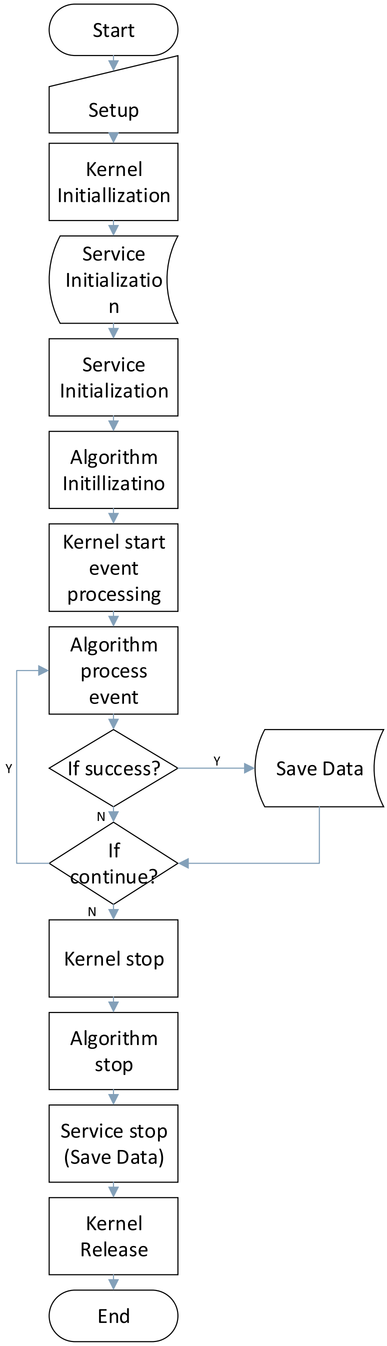
2.2.3 DAMPESW Control Flow

The software control flow is presented in Fig. 6. Before actual running of the software, the configuration of services and algorithms need be conducted at first. Then the initializing shall start on the sequence of core to services to algorithms. After completing services and algorithms initializing, the event processing procedure will be started by the core. Algorithms will analysis an event and determine whether the event is processed successfully. If so, the flow will continue after this event been saved. If not, the flow will continue without saving the event. Next, if the conditions (already finished all event processing or achieve the manual termination conditions) are satisfied, the event processing procedure will be ended and the flow will go to next stage. Otherwise the flow will go into a loop and prepare for processing the next event. The core will be terminated after the end of event processing procedure. Then services will be ended after algorithms terminated. By the time of services stopped, all data will be writing into hard disk. And after all algorithms and services are stopped successfully, memory will be freed up and flow will be finished.

2.2.4 Release of DAMPESW

The development of DAMPESW is independent. The speed of software version updates will be proportional to the number of developers involved. From the practical perspective, engineering management is not an appropriate approach for scientific data analysis. However, in order to guarantee the quality and progress of development, we have taken a series of measures. Firstly, the development is undertaken by various of international cooperation groups. Online group meeting is arranged weekly for task assignment, progress discussion, and schedule making. Secondly, version control repository is built on the computer platform of Purple Mountain Observatory, who is responsible for the construction of scientific application system. As demonstrate

Fig. 6 Software control flow



in Fig. 7, development codes from every cooperation groups would be submitted to DAMPESWSVN version control repository and SVN would manages version info of codes automatically. So that the version control issue can be solved and development quality can be ensured.

Name ▲	Size	Rev	Age	Author	Last Change
../					
DmpSoftware-1-0-0		417	6 years	andrii	
DmpSoftware-1-1-0		684	6 years	andrii	Last release of non-GDML non-EventClass? DAMPE software.
DmpSoftware-2-0-0		945	5 years	andrii	First stable release of GDML & EventClass? based DAMPE offline software
DmpSoftware-2-1-0		999	5 years	andrii	Last release of old-framework DAMPE offline software. Next releases will ...
DmpSoftware-3-0-0		1107	5 years	andrii	First release of the new framework
DmpSoftware-3-0-1		1134	5 years	andrii	STK raw data conversion based on trbread2.c script
DmpSoftware-3-1-0		1261	5 years	andrii	Last stable release before introducing the combined stable kernel (by Chi ...
DmpSoftware-3-2-0		1264	5 years	andrii	Combined stable Kernel from Chi and Andrii.
DmpSoftware-3-2-1		1271	5 years	andrii	First stable version of global Raw Data Converion (provided by JingJing? ...
DmpSoftware-3-2-2		1302	5 years	andrii	RDC for shanghai cosmics from JingJing?, only STK part is modified by ...
DmpSoftware-3-2-3		1400	5 years	andrii	Frozen releaze between moving to combined RDC for the beam test
DmpSoftware-4-0-0		1471	5 years	andrii	Frozen release of software for v1 beam-test data production
DmpSoftware-4-0-1		1488	5 years	andrii	bug fixes for the v1 production. This is the last release bofore moving to ...
DmpSoftware-4-1-0		1508	5 years	andrii	v2 beam test data production
DmpSoftware-4-1-1		1564	5 years	andrii	v2 beam test data production (bugfix for the STK raw data converion)
DmpSoftware-4-1-2		1689	4 years	andrii	Last release before getting to 2015 MArch beam test processing
DmpSoftware-4-2-0		1803	4 years	andrii	v0 reprocessing of March 2015 beam-test data
DmpSoftware-4-3-0		1856	4 years	andrii	stk-calib, stk-recon, stk-compare -- tools for quick STK monitoring, aimed ...
DmpSoftware-4-3-1		1859	4 years	andrii	Fully functional stk-compare, stk-calib and stk-recon tools
DmpSoftware-4-3-2		1868	4 years	andrii	First reprocessing of April 2015 cosmics
DmpSoftware-4-3-3		1865	4 years	andrii	v0 reprocessing of March 2015 beam-test data (STK RDC bug fixed): request ...
DmpSoftware-4-3-4		1882	4 years	andrii	stk-* commands tested by Valentina
DmpSoftware-4-4-0		1989	4 years	andrii	v0 reprocessing of May 2015 Shanghai cosmics data of DAMPE
DmpSoftware-4-4-1		2004	4 years	andrii	v0 reporcessing of June 2015 beam-test data
DmpSoftware-4-4-2		2016	4 years	andrii	V1 reprocessing of March 2015 beam-test data: fixed PSD reconstruction ...
DmpSoftware-4-4-3		2022	4 years	vgallo	REC3 reprocessing of cosmics April 2015 clusterseed 4
DmpSoftware-4-4-4		2042	4 years	andrii	v0 reporcessing of June 2015 beam-test data with DAMPE+AMS merging
DmpSoftware-4-4-5		2048	4 years	andrii	reprocessing of DAMPE intergration-test data, June 25, 2015
DmpSoftware-4-4-6		2152	4 years	andrii	Calibration of magnet for beam test, June 2015
DmpSoftware-4-4-7		2154	4 years	andrii	Production: satellite inflight simulation data, August 25-28, 2015
DmpSoftware-4-4-8		2206	4 years	andrii	bug fix in the STK raw mode decoding reported by Xin
DmpSoftware-4-4-9		2231	4 years	andrii	Last release before moving to the Kernel update: enable multiple input ...
DmpSoftware-4-5-0		2251	4 years	andrii	bux fix in STK DLD package unpacking (problem of duplication in .txt ...
DmpSoftware-4-5-1		2254	4 years	andrii	STK package (frame) size information added to STK metadata
DmpSoftware-4-5-2		2299	4 years	andrii	processing1 of aging test data in Geneva, November 2015
DmpSoftware-4-5-3		2316	4 years	andrii	processing1 of aging test data in Geneva, November 2015 (bug fixes: ...
DmpSoftware-4-5-4		2383	4 years	andrii	v0 processing, beam test November 2015
DmpSoftware-4-5-5		2712	4 years	andrii	Last stable release before merging rep1 and rep in a single Orbit ...
DmpSoftware-4-5-6		2788	4 years	andrii	Last release before moving to the new STK calo-seeded tracking: 1D+1D ...
DmpSoftware-5-0-0		2870	4 years	pmo	svn r2869, DMPSW v5.0.0
DmpSoftware-5-1-0		3111	3 years	pmo	release r3110 as gamma version 5.1.0
DmpSoftware-5-1-1		3152	3 years	pmo	Tag: put software version number to 5.1.1
DmpSoftware-5-1-2		3220	3 years	pmo	Tag: 5.1.2
DmpSoftware-5-1-3		3419	3 years	pmo	Tag: version 5.1.3
DmpSoftware-5-1-4		3448	3 years	pmo	Tag: v5.1.4
DmpSoftware-5-1-5		3537	3 years	pmo	tag: v5.1.5
DmpSoftware-5-2-0		3835	3 years	pmo	Tag: 5.2.0
DmpSoftware-5-3-0		4425	3 years	pmo	Tag: 5.3.0
DmpSoftware-5-3-1		4543	3 years	pmo	Tag: 5.3.1
DmpSoftware-5-3-2		4735	3 years	pmo	Tag: 5.3.2
DmpSoftware-5-3-3		5084	3 years	pmo	Tag: 5.3.3
DmpSoftware-5-4-0		5141	3 years	pmo	Tag: 5.4.0
DmpSoftware-5-4-1		5884	2 years	pmo	Tag: 5.4.1
DmpSoftware-5-4-2		6047	2 years	pmo	Cherry-pick : DmpSoftware?-BugFix?/5-4@c6046 : Modify minor bug in ...
DmpSoftware-6-0-0		6417	2 years	pmo	Tag: 6.0.0
DmpSoftware-6-0-1		7674	16 months	pmo	Tag: 6.0.1
DmpSoftware-6-0-2		7866	14 months	pmo	Tag: 6.0.2
DmpSoftware-6-0-3		7957	13 months	pmo	Tag: 6.0.3
DmpSoftware-6-0-4		8214	10 months	pmo	Tag: 6.0.4
DmpSoftware-6-0-10		8517	3 months	pmo	Tag : 6.0.10
ReleaseNote		8519	3 months	pmo	Add : Release note 6.0.10

Fig. 7 SVN version control repository of DAMPESW releasing

2.3 Satellite Data Processing Flow

The flow diagram of DAMPE data is presented in Fig. 8.

After DAMPESW receiving the order of data processing, the RawDataConversion module will read-in file from the order which contains 1A/1B data product, and unpack it to verify the content. The verification including remove head and tail of FITS data packet, separate different types of data (observation data, pedestal calibration data, Digital-to-analog converter calibration (DAC) data, MIPs calibration data and pedestal update data), and convert all of them to TTree format (1E/1F scientific data product) that could be recognized by ROOT. Observation data will be reconstructed to physical objects by Reconstruction module, others will be converted to calibration coefficient files through a series of complicated calculating by Calibration module.

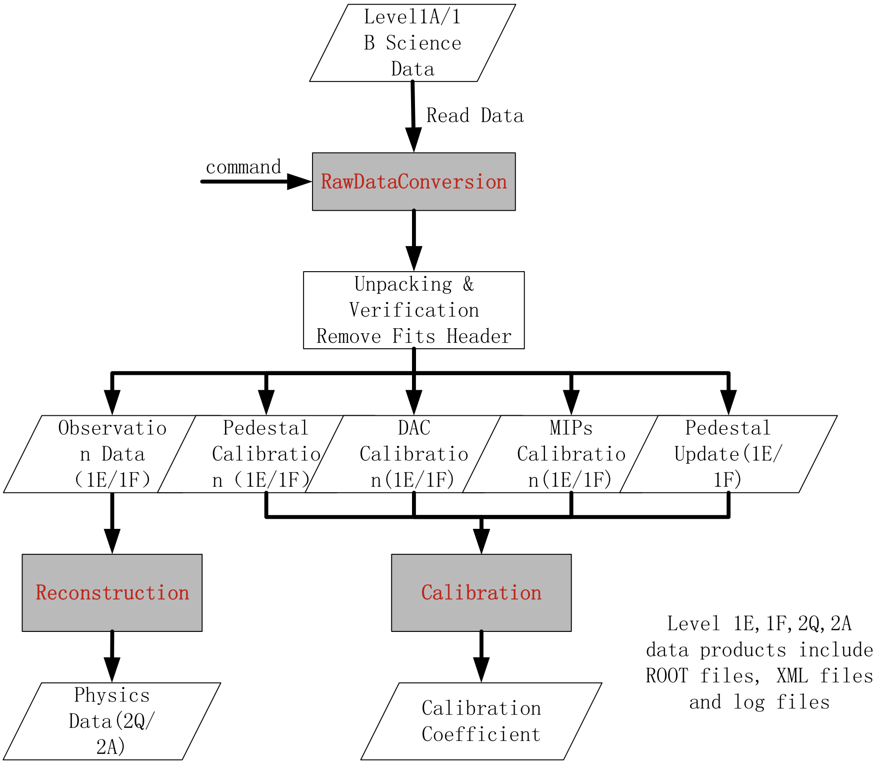


Fig. 8 Workflow of satellite data processing

2.3.1 Workflow of RawDataConversion

As shown in Fig. 9, the first step of flow is read-in 1A/1B data product. Then head and tail of FITs data packet will be removed. After that the inputting data will be split. Data packet of one single event will be read-out into a series of caches (BGO, PSD, NUD, STK, and TRG), and module will judge whether event in STK cache is intact. If intact, event will be filled in TTree. Otherwise module will continue to read next event. In the case of event been successfully filled in TTree, module will

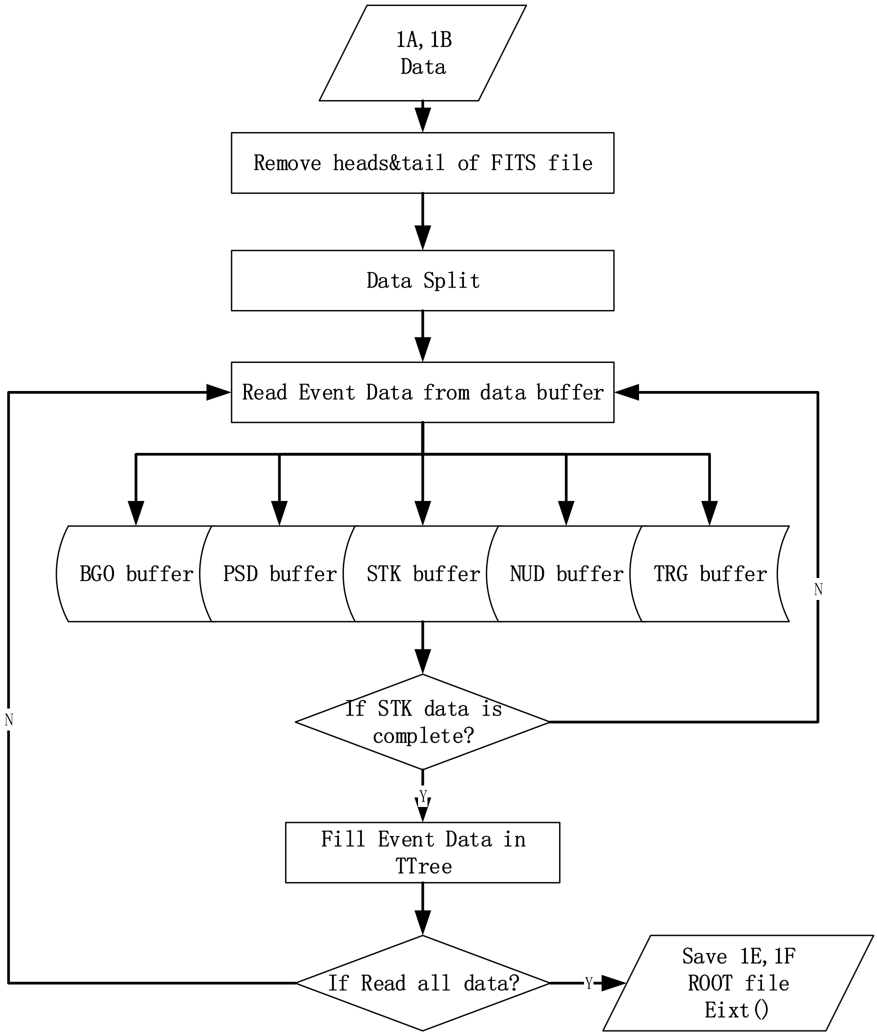


Fig. 9 Workflow of raw data conversion

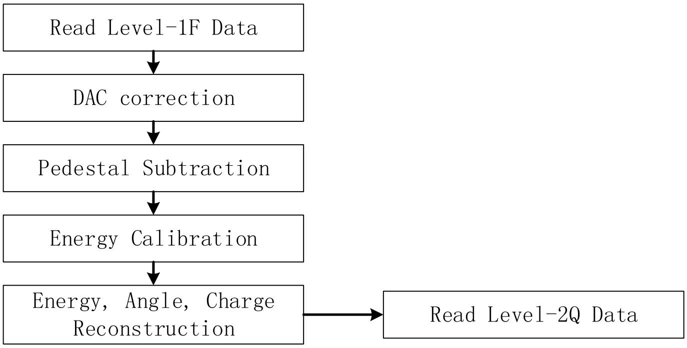


Fig. 10 Workflow of reconstruction

judge whether the generated ROOT file should be terminated. If not, module will keep processing the next event, otherwise TTree file will be saved and exited.

2.3.2 Workflow of Reconstruction

Reconstruct is started with the input of 1E/1F data product, which should be saved in ROOT format produced in other modules. Generally, reconstruct is a series of procedures that convert observational signals to physical signals on the help of calibration coefficients. Figure 10 is the flow diagram of the process.

The first step is DAC correction. Next, pedestal should be deducted and appropriate dynodes will be selected and read out. Through dynode ration transform, measurement of one single detective unit should be converted to the readings of ADC channels of the eighth dynode. Hence, energy of a single detective unit can be obtained by normalizing peak value of MIP. Based on that, advanced energy, direction, and charge are reconstructed.

2.3.3 Workflow of Calibration

Calorimeter is a type of device that absorb energy of incident particle by making it interact with the device. So that part of its energy would be released in a recordable way. To fully understand the performance of this device, the conversion factor between detected signal and incident particle energy is crucial. And the process of obtaining it is the so-called calibration. There are four major parts of calibration, pedestal calibration, MIPs calibration, dynode relative coefficient calibration, and Digital-to-analog converter calibration. All of them have inputs of 1E/1F data product in ROOT format, and the output should be calibration coefficients.

Pedestal calibration:
In the case of no inputting signals from electronic front-end chip, the output should be the readings of ADC channels, which would yield a normal distribution. Hence, the calibration coefficients should be the arithmetic mean and variance of the normal distribution.

MIPs calibration:
If detector is working under sampling mode, the trigger shall be set at MIPs mode that ADC spectrum of MIPs response can be obtained. In which case calibration coefficients should be the peak position and width of the most probable distribution derived from fitting the spectrum with Gaussian-convoluted Landau function.

Dynode relative coefficient calibration:
The detected signal would be output through the second, fifth, and eighth dynodes of PMTs. After electronic processing, the readings of ADC channels should yield a linear correlation. Therefore, the calibration coefficient should be the gradient and intercept derived from linear fitting.

Digital-to-analog converter calibration:
After digital-to-analog (DA) conversion, orders, which is digital, send from ground base are converted to analog signals which would act as electronic front-end input. Between the output, readings of ADC channels, and the amplitudes of the input, an approximate linear relation existed. Its deviation from linear reflect the non-linear feature of electronic.

2.3.4 Workflow of Data Simulation

By assuming various circumstances that detector has, physical simulation should refer to the process of computing emitted signals from the interaction between incident particles and the detector. The purpose of physical simulation is to create simulated data that are identical to observed data, so that it can be analyzed by DAMPESW following the same method used in observed data.

Figure 11 is the flow diagram of generating simulated data.

Multiple DAMPESW modules are responsible for generating simulated data, which are Geometry, Simulation, OrbitSimulation, EventOrbitSimulation modules.

The physical simulation and electronic simulation are integrated in Simulation module. These two simulations are performing simultaneously for BGO, PSD, and NUD (electronic simulation of STK would be introduced latterly). Meaning that energy deposit would be convert to readings of ADC channels by electronic response function as soon as it was produced by simulating interactions of incident particles

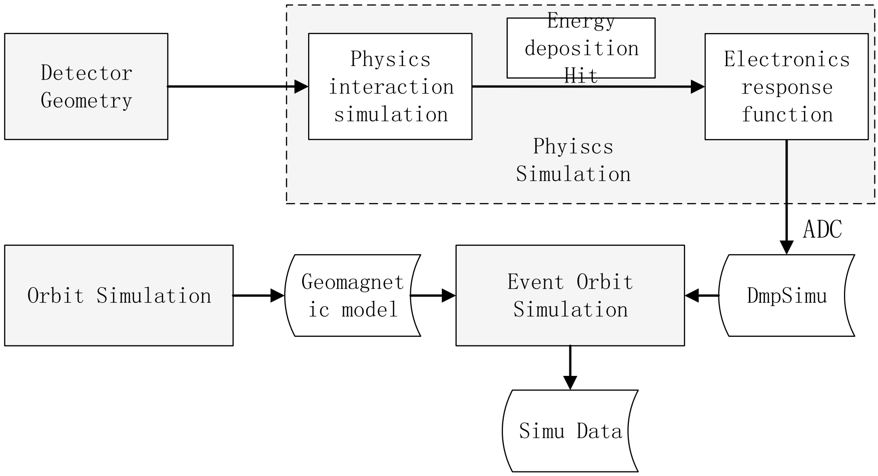


Fig. 11 Workflow of data simulation

and detector. Module will produce root date file (hereafter called DmpSimu) that have same format of data level 2.

The simulation of satellite orbiting environment is divided into two steps: Orbit-Simulation module is responsible for computing the satellite orbit and magnetic field distribution at corresponding altitude based on ephemeris. Root data file (OrbitSimu) will be produced in this step.

Figure 12 shows maps of magnetic fields and vertical geomagnetism cutoff at 500 km altitude.

Figure 13 shows the satellite orbit and experienced geomagnetism during one day EventOrbitSimulation module takes DmpSimu and OrbitSimu as input. It would decide whether the incident particles are originated from outer space using a method called "backward tracing". If true, a time will be given for each particle and new root data file (EventOrbitSimu) will be generated, which can be used for estimating event rate and trigger rate (with respect to various observation mode) at any satellite positions. The so-called "backward tracing" is the method of judging incident particle position, that is to numerically solve the equation of motion of particle inside magnetic field by inverse its charge and direction. In fact, we only interested in particles came from outer space so that particles from earth are excluded during event rate calculating.

As mentioned before, STK working different from BGO, PSD, and NUD so that its electronic simulation needs to be done separately. Firstly, during physical simulation, every energy deposit on silicon micro-strip should be recorded. After finished one event, whether a silicon micro-strip should be read-out or suspended is depending on the previous record. Then energy will be reassigned to other neighboring read-out silicon micro-strips in order to calculate the number of electron-hole pairs can be produce in such an energy. Finally, depending on the monitoring of respond curve

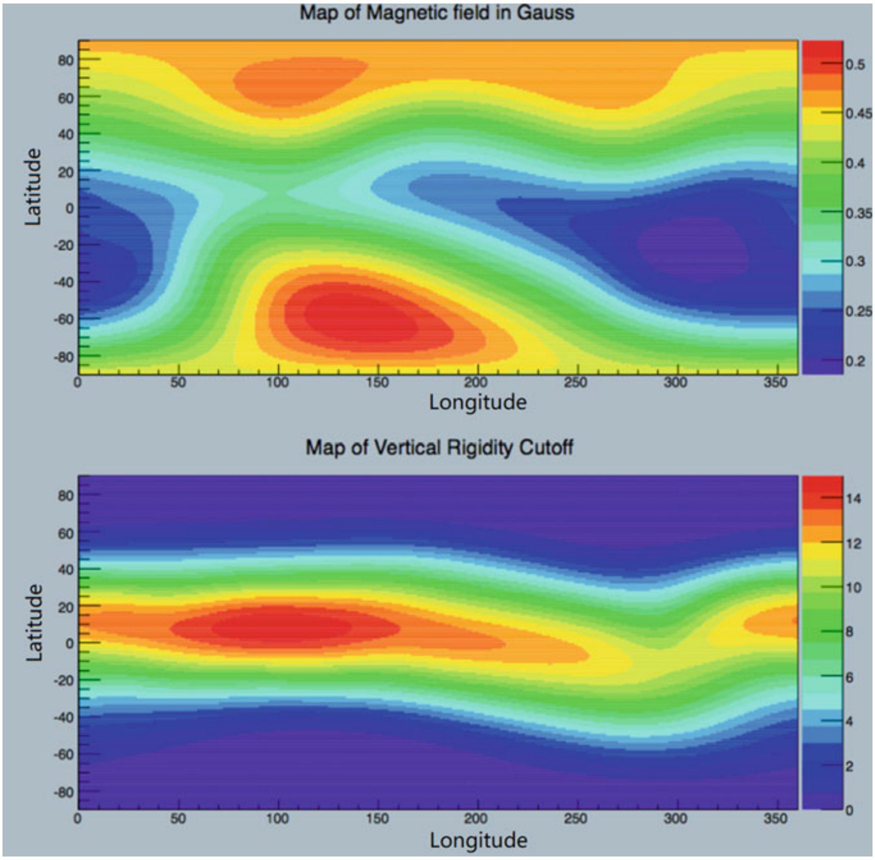


Fig. 12 The map of magnetic fields and vertical geomagnetism cutoff at 500 km altitude

of VA chip within silicon micro-strip (shown in Fig. 14), energy would be converted to ADC readings.

2.4 DAMPESW Workflow

The whole workflow of processing 1A/1B data products (by the advanced data product managing subsystem) into advanced 1E/1F and 2A/2Q data products is complete on data processing server. During that, all data processing procedures are driven by python scripts. Thus, the whole process requires a continuously use of different python scripts for different stages, presented in Fig. 15.

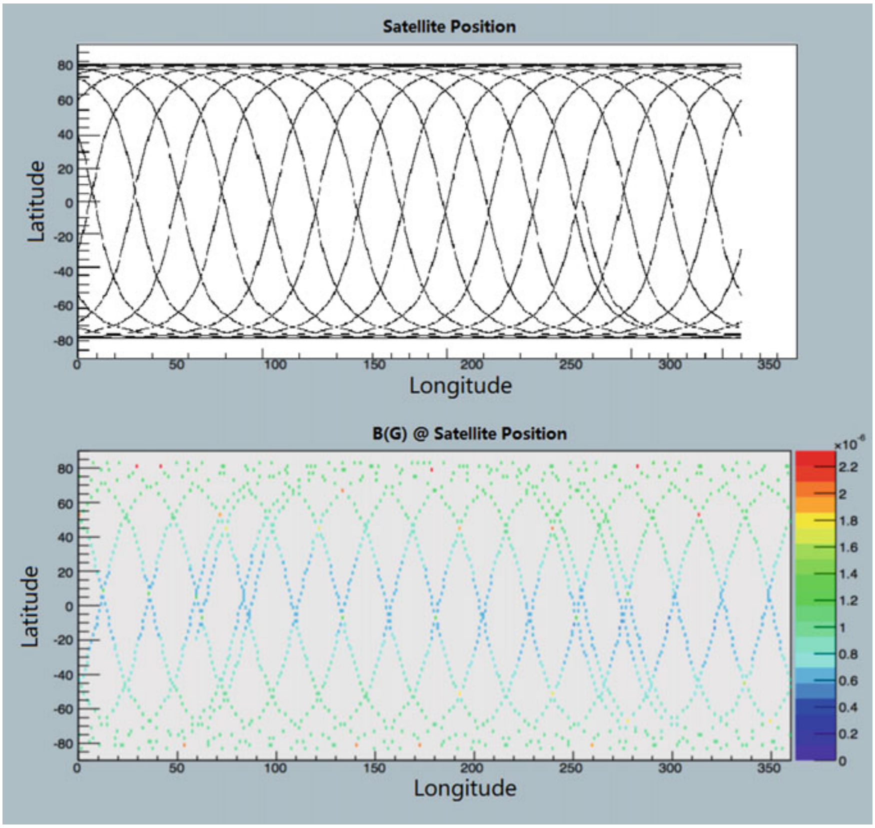


Fig. 13 The satellite orbit and experienced geomagnetism during one day

(1) Converting FITS data file into another format: Script JobOpt_Fits2Frd.py take file name as input. Its function is to read the head information of FITS file and convert file to frd format. The new file will be saved in a temporary directory.
(2) RawDataSplit. The original data file of remote observation contains data under many observation modes. Processing of those data requires different procedures that correspond to observation modes. Frd file created in the last step need to be split to multiple files. Each of them should contain data of a single mode with respect to each time period, and they should be marked accordingly (OBS for observation data packet, PED for pedestal calibration data packet, DAC for DAC calibration data packet, MIP for MIPs calibration data packet, UPD for pedestal update calibration packet).
(3) RDC. Decoding and converting the original data packet of remote observation for each observation modes and produce 1E/1F data product. During its operation, the first step is checking the original data, including CRC check, length check, data packet serial number continuity check, data packet trigger number

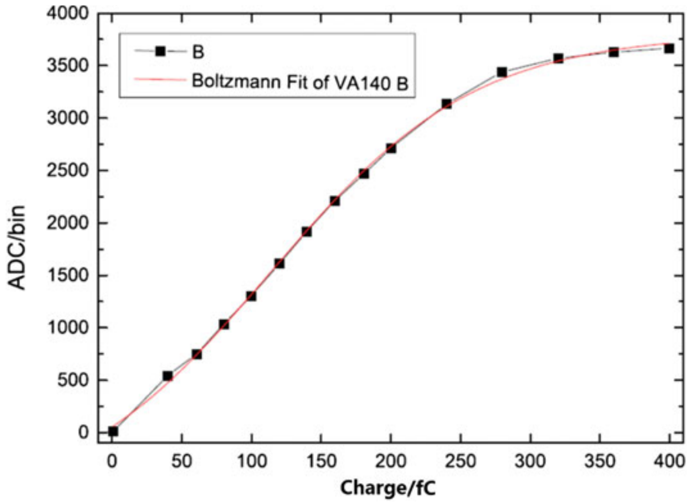


Fig. 14 The response curve of VA chip

continuity check, etc. And the error info will be saved in a file having same name and directory of the data packet. Then trigger data and sub-detectors data will be read-out and written into corresponding EventClass, which will be saved as ROOT file with name consistent with the naming conventions. The key information during data generation will be automatically preserved in a description of advanced data product for scientific data management and user sub-system storage and analysis.

(4) Calibration. Obtaining calibration coefficient for calibration files (MIP, DAC, PED) through a series of complicated computation. The results will be saved in corresponding directory with a specific format.

(5) Reconstruction. For data after RDC, it is necessary to reconstruct all events and carry on follow-up scientific analysis. For 1E/1F data product marked as OBS, the most recently used calibration coefficients will be loaded to reconstruct incident particle energies and tracks in every detective unit. The result will be saved as ROOT file in the corresponding directory, named under 2A/2Q naming conventions.

The flows described above would be executed by shell script automatically, including load global variables required for system operation, generate file names based on corresponding naming conventions, manage directories that storage various of temporary files and calibration coefficient files, automatic judge observational mode of data, output error info generating during operation, etc. Again, all those procedures will be automatically executed that all needed is to run shell scripts in data processing server with appropriate input parameters.

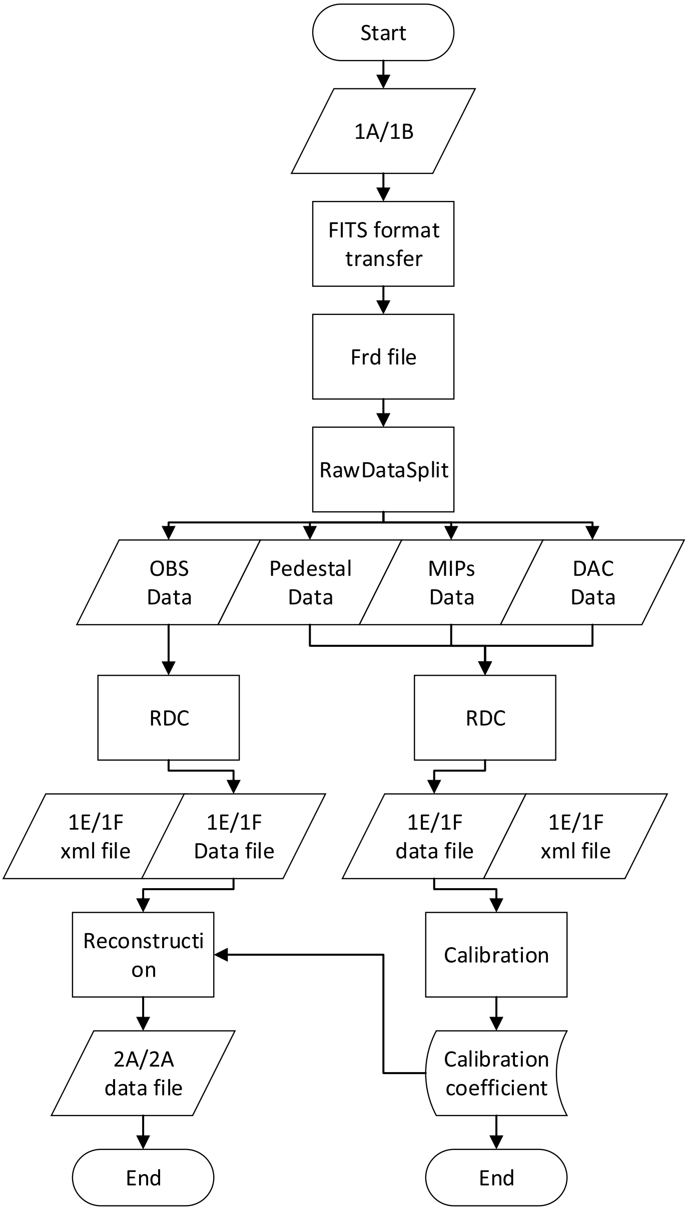


Fig. 15 Workflow of advanced data product sub-system

3 Hardware Components and Network Structure

In order to meet the needs of processing satellite scientific data, a platform is built in DAMPE scientific application system. There several sections of the platform, calculation section, main storage section, hot data storage section, user data storage section, data base section, data backup section, data interaction and network security section, and monitoring hall section, showing on Fig. 16. To be more specific, the calculation section is made of 128 high density blade servers, which have more than 3000 CPUs. The main storage section adopts EMCX410 of high-performance

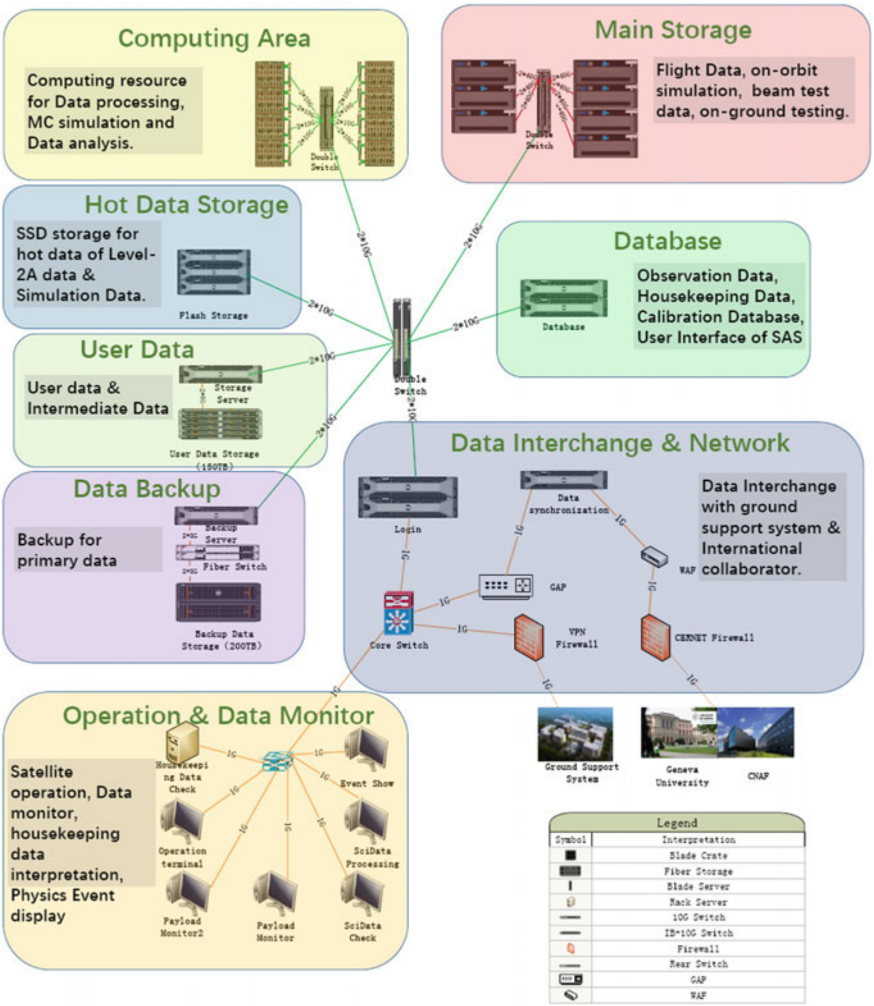


Fig. 16 Sketch of DAMPE scientific data processing platform

parallel storage that have in total 7 storage nodes with a volume of 450 TB, of which the highest data throughput reaches 40 Gbps. It would be used for storage all scientific data products relative to the satellite. The hot data storage section uses flash memory for data storage which has a volume of 175 TB that used for store 2A data product which frequently accessed by users. Large capacity disk storage is used in user data storage section, which store user's personal data generated during data processing. Its volume is 130 TB. Similar to user data storage section, the data backup section (used for backing up primary data advanced data, and simulation data) also adopts disk storage which has 200 TB. The data interaction and network security section are connected to HuaiRou space science center and European cooperation groups through 100 Mbps of private VPN network and 100 Mbps of science and technology network, respectively. While the network security and isolation are guaranteed by firewall, WAF, and gateways. Finally, the monitoring hall is consisted of a series of minor terminals, taking charge of the operation and monitoring data processing procedures.

4 Conclusion

DAMPE is the first Chinese satellite-based observatory built for astronomic purpose. The way to handle its observation data is unprecedented. Internationally, the raw data processing of other same type of satellites are exclusive and hard to be used on processing DAMPE data because of the different instrument structure. In this paper, therefore, we systematically introduced DAMPE data processing flow, methods, software, and hardware based on the perspective of software requirements. We hope that this paper would provide technical reference for future projects.

Jin Chang was born in July, 1966. He graduated from University of Science and Technology of China with master's degree in June, 1992, and his doctor's degree is received in July, 2006. At the present, he is the director of Purple Mountain Observatory, the deputy director of the National Astronomical Observatories (concurrently), chairman of the Key Laboratory of Dark Matter and Space Astronomy, director of the School of Astronomy and Space Science of the University of Science and Technology of China, chairman of research department of dark matter and space astronomy. He obtained the title of academician of Chinese Academy of Sciences in 2019.

Jin Chang have long been engaged in the detections and researches of gamma ray emission and high energy charged particles that come from outer space, and is one of the pioneers for studying space astronomy in China. He created and developed a new technique of detecting high energy cosmic ray electrons. This technique was used in a long-term balloon experiment (ATIC experiment) of America aiming at detecting cosmic rays in Antarctica. Based on the technique, he leaded the

DAMPE (also called as “WuKong”) project as the chair scientist. The “WuKong” was successfully launched in December 17, 2019 and it is the very first satellite of China used for astronomical research. Also, this launch marked that China is leading the world at several key techniques and breakthrough on detecting the spectrum of cosmic ray electrons and protons. Furthermore, in order to response for the needs of the major national strategic, Jin Chang and his team successfully developed gamma ray spectrometers for ShenZhou II, Chang’e I and II. His major awards are: the second prize of National Science and technology progress in 2004 (4th to accomplish), the second prize of National Natural Science Award in 2012 (accomplished independent), special prize of National Science and Technology Progress Award in 2012(40th to accomplish), National innovation award in 2017, Ho Leung Ho Lee Foundation Award for Scientific and Technological Progress Award in 2018 (astronomy), Zhang Yuzhe award of China Astronomical Association in 2018, outstanding scientific and technological achievement award of Chinese Academy of Sciences in 2018, ten advances in Chinese Science in 2018, science and technology award of China Academy of Space Science (first), etc.

Atomic-Scale Investigations on Water Science Based on Information Technology

Runze Ma, Jinbo Peng, and Ying Jiang

Abstract Water–solid interfaces play important roles in a wide range of scientific fields and technique processes and it is a crucial scientific issue to understand the hydrogen-bonding network and relevant dynamics of interfacial water at molecular or sub-molecular level. Of the many surface science techniques, scanning probe microscopy has become one of the most powerful tools to study the water–solid interfaces due to its capability of imaging with atomic resolution. In recent years, our research group has achieved atomic-scale imaging and spectroscopic characterization of water–solid interfaces utilizing the atomic resolution capability of scanning probe microscopy and the imaging techniques of e-Science. Further combined with simulation and calculations based on supercomputer techniques and e-Science, our group has achieved important progress in the microscopic structure of the hydrogen-bonding network of water, the dynamics of proton transfer, nuclear quantum effect on the strength of hydrogen-bond and water-ion interaction.

Keywords Water–solid interfaces · Scanning tunneling microscope · Atomic force microscope · Hydrogen-bonding network

1 Introduction

As one of the most common and important substances, water plays a crucial role in various physical and chemical processes of nature. Such processes include, metabolism, body temperature regulation and material transportation within living creatures; rain, snow, frost and other meteorological phenomena in daily lives; washing, lubrication, corrosion, catalysis and other applications in industrial productions. These are all related with the unique properties of water such as strong solubility and heat shrink, despite the extreme simplicity in its chemical structure. The complexity of water mainly arises from its hydrogen-bonding structure, which is

R. Ma · J. Peng · Y. Jiang (✉)
International Center for Quantum Materials, School of Physics, Peking University, Beijing, China
e-mail: yjiang@pku.edu.cn

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_5

listed as one of the 125 most challenging scientific problems in the 125th anniversary issue of Science magazine. To attack this problem, it is essential that we study the water systems carefully at atomic scale.

The most commonly studied water system is the water–solid interface, where lots of important physicochemical processes take place, such as dissolving, lubrication, corrosion, electrochemistry, heterogeneous catalysis, etc. [1–4]. These complex processes are closely related with not only the interactions between water molecules but also the interactions between water molecules and solid surfaces. The competition between these two interactions results in very rich phases of interfacial water structure and interesting dynamics, such as proton transfer, abnormal transport for restricted water systems, dissociation of water molecules, etc.

Conventional methods for studying interfacial water are mostly spectroscopic ones, including nuclear magnetic resonance [5], helium diffraction [6, 7], X-ray diffraction [8], low energy electron diffraction [9], infrared absorption spectrum [10], etc. However, the limitation of their spatial resolution from hundreds of nanometers to several micrometers makes it rather difficult to probe the interfacial water structures at atomic scale, thus researchers have to rely on complicated theoretical calculations and simulations. On the other hand, scanning tunneling microscopy (STM) and atomic force microscopy (AFM) have been widely applied to the studies of hydrogen-bonding network structures for interfacial water systems at single molecule level, which greatly advance our microscopic understanding of interfacial water structure and dynamics. However, due to the small mass and size of the hydrogen atoms, these studies are mostly focused on molecular scale, lacking the intramolecular information of water, which is also crucial to resolve the hydrogen-bonding structure.

In recent years, our group has achieved sub-molecular resolution of water molecule to locate the hydrogen atoms within interfacial water for the first time, using a combined system of STM and AFM based on the technology of automated control and digital imaging. Along with the first principle calculations based on supercomputer and information technology, we were able to conduct atomic-level studies on the interfacial water systems. So far, we have achieved some progresses at water–solid interfaces, such as hydrogen-bonding topology, proton transfer dynamics, nuclear quantum effects on hydrogen bond strength, ion-water interactions, low-dimensional ice etc. In this article, we will focus on the recent works on interfacial water from our own group, to summarize the achievements of information technology in the research of interfacial water systems at atomic scale. The purpose is to emphasize the importance of information technology and pave the way for the further development and application of information technology in the fields of surface science, high-resolution imaging and recognition, large-scale system simulations, etc.

2 Sub-molecular Imaging for Interfacial Water

2.1 Digital Imaging and Automated Control

Using STM and AFM for the study of interfacial water, requires processing and digital imaging for large quantities of data in order to generate straightforward images for researchers to analyze. These common digital imaging technologies can be viewed as the earliest application in the STM/AFM systems and has been successfully commercialized. However, the rapid developments of science resulted in specific requirements, which can no longer be fulfilled with commercial software. Thus, programming languages such as LabVIEW came up, serving as a necessary supplement for the commercial software. Customized digital imaging and automated control were easily achieved using these languages, providing great convenience for the study of interfacial water systems.

Based on Nanonis software, our group has programmed VIs using LabVIEW language, such as the one shown in Fig. 1. These VIs can acquire dI/dV, dI/dZ,

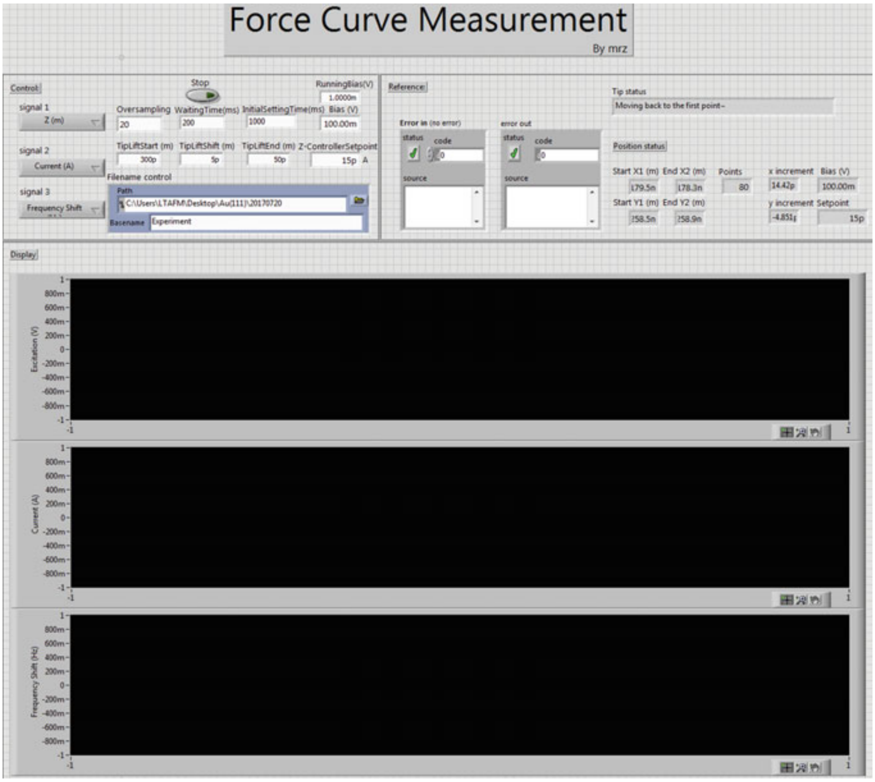


Fig. 1 Automated control program using LabVIEW for different height force curve measurements

LCPD (local contact potential difference) and other signals together with the original STM current, AFM frequency shift signals simultaneously, providing straightforward image data for interfacial water researches. Some of the VIs, such as the force curve measurements for different heights shown in Fig. 1, are able to collect data automatically, reducing the mistakes in repetitive operations and greatly enhancing the experimental efficiency.

2.2 Sub-molecular Orbital Imaging

The first way to achieve sub-molecular imaging for water molecules is the STM orbital imaging. Based on the imaging principle for STM, the current signal of STM is directly related with local electron density of states for both tip and sample, so the imaging of water molecular orbital structure is practical [11]. However, no sub-molecular orbital imaging had been achieved for water molecules and all the STM images for single water molecules are bright protrusions [12–14]. This is mainly because of the following two reasons. Firstly, most of the previous experiments are using metal substrates. Surface states usually exist at metal surfaces, resulting in high electron density of states. Due to the strong coupling between the metallic states and water molecule, the orbital information for water is buried in the density of states from the metal substrate. Secondly, pure water is a good insulator, the highest unoccupied molecular orbital (HOMO) and lowest occupied molecular orbital (LUMO) are far away from the conducting fermi level. The imaging of the orbitals would require higher bias than the stability threshold for water molecules, causing them to diffuse away or even dissociate.

In order to solve these problems, our group [15] has deployed a bilayer of NaCl thin films to decouple the interaction between the water molecules and the electronic states from the metal substrate. The decoration of the tip with Cl^- induces stronger coupling between the tip and water molecules and shifts the orbital frontiers of the water molecules closer to the fermi level, making it possible to obtain high-resolution orbital images for HOMO and LUMO of water molecules, see Fig. 2a, b. In the experiments, we also used LabVIEW programming to acquire dI/dV data with feedback loop on at every data point. As shown in Fig. 2c, double-lobe structure confirms that the HOH plane of water is perpendicular to the surface (Fig. 2e, f). When changing to a negative bias voltage as in Fig. 2d, the double-lobe structure almost fades out and an egg-shaped lobe appears within the nodal plane of the HOMO. Based on the correspondence between the water molecular orbitals and water molecular structure, we can determine the adsorption structure and orientation for single water molecules, as well as the hydrogen-bonding directions within water tetramers (Fig. 2g-i).

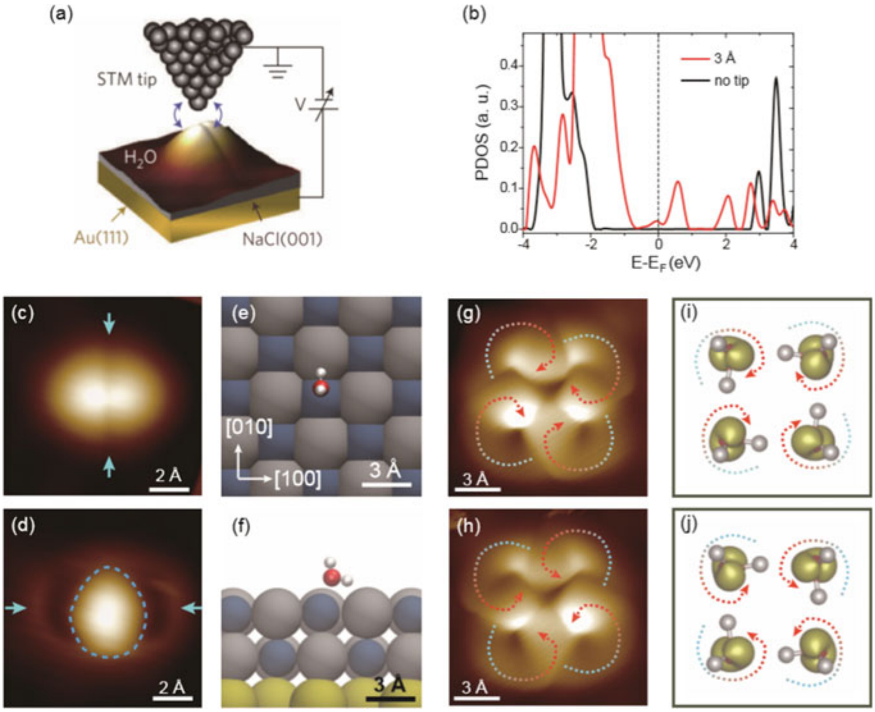


Fig. 2 Single water molecule on NaCl(001) surface and sub-molecular STM imaging for water tetramers

2.3 *Sub-molecular Atomic Force Imaging*

Besides STM techniques, we can also use AFM techniques to obtain sub-molecular imaging for interfacial water systems. As mentioned above, due to the complexity of electronic orbitals, the STM orbital imaging requires a lot of theoretical calculations to establish the relations between electronic orbital structure and the molecular structure for the identification of hydrogen atoms. In comparison, the imaging principle for AFM technique is simpler and more direct. Due to the strong polarization of water molecules, oxygen and hydrogen atoms within water molecules carries a decent amount of negative and positive charges, respectively. Theoretically, using a charged AFM tip, one can distinguish the hydrogen atoms from oxygen atoms via the electrostatic force interaction between the tip and water molecules. However, the long interacting distance of the electrostatic force decreases the variation of detecting force with respect to tip heights, limiting the spatial resolution of AFM. One of the ways to solve this problem is to deploy high order electrostatic force.

Based on the Q-plus sensor AFM shown in Fig. 3a, our group [16] has successfully obtained sub-molecular imaging for water molecules by decorating AFM tips with CO molecules, as shown in Fig. 3. For large height AFM imaging of the water

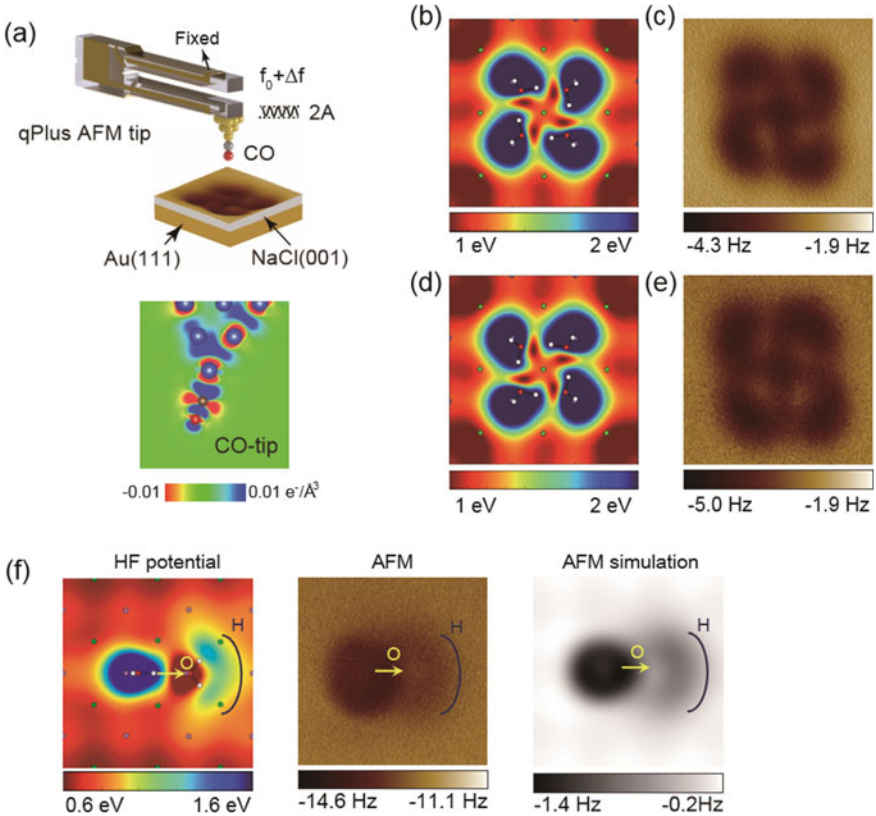


Fig. 3 Sub-molecular AFM imaging for water tetramers and dimers

tetramers, clear inner structures are shown in Fig. 3c, e, which is very similar to the electrostatic potentials of the water tetramers in Fig. 3b, d. Compared with theoretical calculations, it is obvious that the high-resolution imaging at large tip heights can be attributed to short-ranged high order electrostatic force between the quadrupole of the CO-tip and water molecule. Meanwhile, we also used the program for force curve measurements at different tip heights shown in Fig. 1 to characterize the minimum atomic force required to have such atomic resolution. The derived force is very weak, which reveals the almost non-perturbative nature of the CO-tip AFM technique. This technique makes it possible to image many weakly-bonded water clusters and other metastable states (Fig. 3f), which may be a powerful tool for the researches into interfacial water systems.

Breaking the limitation of scanning probe microscopy, the CO-tip AFM technique may open up the possibility of researching into the structures of interfacial water/ice, ion hydrates and biological water. It can not only provide information of the spatial distribution of electrostatic force, but also determines the position of hydrogen atoms and topological details of hydrogen bond structures, which is crucial to understanding the interaction and dynamics of hydrogen bonding between water molecules.

3 Sub-molecular Researches for Interfacial Water Systems

3.1 Proton Transfer Dynamics in Hydrogen-Bonding Network

Proton is the lightest atom nuclei in water. When disturbed by heat or local electric field, it can get rid of the binding of nearby oxygen atoms and transfer, which contributes to the complexity of the water structure. The phenomena of proton transfer along hydrogen-bonding networks play essential role in many physical, chemical and biological processes [17, 18]. Theoretically, proton dynamics is susceptible to quantum tunneling, which typically involves many hydrogen bonds simultaneously, leading to correlated many-body tunneling. However, no direct experimental evidence of such correlated many-body tunneling of proton had been presented. This is mainly because the correlated many-body tunneling of proton is very sensitive to the local interaction with the environments due to the requirement for tunneling coherence.

In order to elucidate the above phenomenon with direct experimental evidence, our group [19] directly visualized the concerted tunneling of four protons in an individual hydrogen-bonded water tetramer (Fig. 4). To investigate the effect of the atomic-scale environment on the tunneling process in a well-controlled manner, we functionalize the STM tip apex with a single chlorine anion and then tune the Cl–proton electric coupling in three dimensions with picometre precision. We found that symmetric coupling between the Cl anion and the four protons considerably enhances the tunneling probability, whereas asymmetric coupling easily destroys the quantum coherence and the cooperativity of the four protons, resulting in rapid quenching of the concerted tunneling. Our experimental results are corroborated by density functional theory (DFT) calculations.

To further verify the quantum nature of such proton transfer, we conduct isotope experiment by changing one water molecule of the tetramer from H_2O to D_2O. Large quantities of current jumping curves were acquired using automated LabVIEW programs. With the help of machine learning, we were able to identify the characteristics of the current jumping edge as well as different current plateaus. Statistically, the switching rate for the chirality of water tetramers was greatly reduced at 5 K, proving the concerted tunneling nature of the chirality switching. This indicates that correlated many-body tunneling is of crucial importance in the hydrogen-bond dynamics inside water systems. The observed concerted tunneling turned out to be much easier than single-particle tunneling within the water tetramer, which may be a

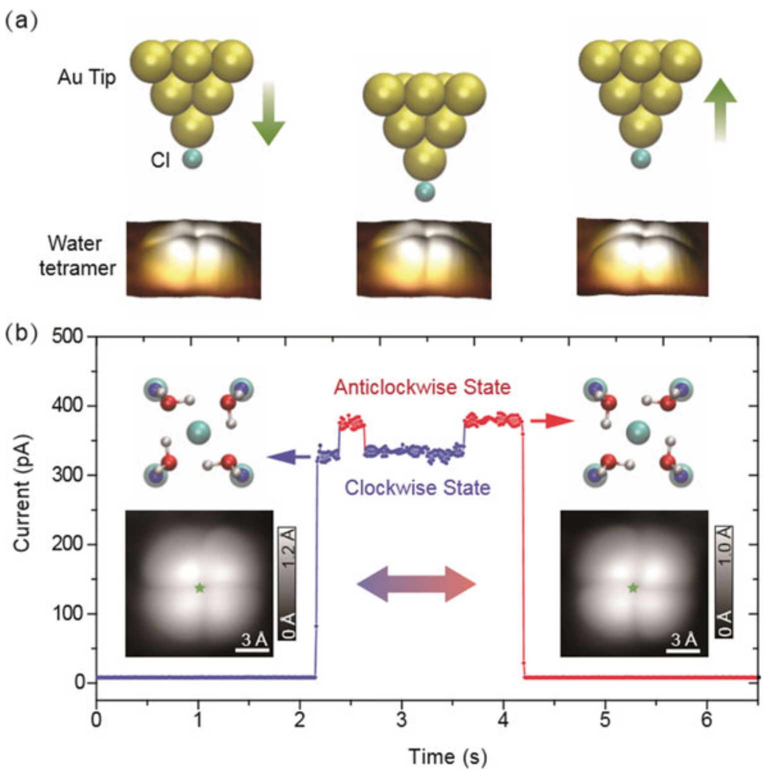


Fig. 4 Chirality switching of water tetramers

generic phenomenon for hydrogen-bonding systems. These findings may be important to understanding the phase transitions in ice and organic ferroelectric materials as well as signal transfer inside biological systems.

3.2 *The Influence of Nuclear Quantum Effect on Hydrogen-Bond Strength*

One of the main reasons for the complexity of hydrogen-bond interaction in water is the nuclear quantum effect (NQE) of the proton. Nuclear quantum effects of protons play important roles in the structure, dynamics, and macroscopic properties of hydrogen-bonded materials [20–22]. Despite enormous theoretical efforts toward pursuing proper treatment of the nuclear motion at a quantum mechanical level, accurate and quantitative description of NQEs on the H-bonding interaction has proven to be experimentally challenging. Conventional methods for probing the

NQEs are based on spectroscopic or diffraction techniques. However, those techniques have poor spatial resolution and only measure the average properties of many H bonds, which are susceptible to structural inhomogeneity and local environments. The spatial variation and interbond coupling of H bonds lead to spectral broadening that may easily smear out the subtle details of NQE.

One promising technique to overcome these challenges is inelastic electron tunneling spectroscopy (IETS) based on a scanning tunneling microscope (STM), which combines sub-angstrom spatial resolution and single-bond vibrational sensitivity. In a conventional IETS regime, the electron-vibration coupling is very weak compared with the elastic scattering, leading to extraordinarily weak IET signals, less than 10% of the total signal. This limitation particularly affects closed-shell molecules such as water, whose frontier orbitals are located far away from the Fermi level. Thus, we [23] developed a tip-enhanced IETS technique to achieve a sufficiently high signal-to-noise ratio for water (Fig. 5a-c), which allowed us to quantitatively reveal the quantum component of the H bonds and the important role of the local environment in dictating the NQEs. In experiments, we used a Cl-terminated STM tip to "gate" the HOMO of water closer to the Fermi level via tuning tip-water coupling. At a large tip-water separation, the spectra were featureless (middle curve), simply following the background (bottom curve). Once the tip-water separation was decreased by ~0.8 Å (top curve), additional kinks arose in the dI/dV spectrum. In the corresponding d^2I/dV^2 curve, these kink features are further magnified as peaks and dips point-symmetric on the two sides of the zero bias. Comparison with density functional theory calculations allowed us to assign the spectral features to the frustrated rotational (R), bending (B), and stretching modes (S), respectively.

High-resolution vibration spectra for single water molecules can also be obtained via tip enhanced IETS, which can be used to distinguish OH and OD as shown in Fig. 5d, e. Isotopic substitution experiments combined with quantum simulations reveal that the inharmonic quantum fluctuations of hydrogen nuclei weaken the weak hydrogen bonds and strengthen the relatively strong ones (Fig. 5f, g). The quantum component of the H bond can account for up to 14% of the bond strength, which is much greater than the thermal energy contribution, even at room temperature. This implies that the NQEs of a H bond are extremely sensitive to coupling with the local environment, which is, at present, inaccessible by macroscopic spectroscopic methods. Not only do these findings substantially advance our understanding of the quantum nature of H bonds, they also open up a new route for spectroscopic studies of H-bonded systems at the single-bond level.

This work verified the Ubbelohde effect [24] observed back in 1950s at atomic scale. This effect includes the microscopic phenomenon of hydrogen-bond length changes with respect to isotope, as well as the macroscopic isotope phenomenon, where the melting (boiling) point of D_2O is higher than H_2O by 3.82 K (1.45 K), the triple point of D_2O is lower than H_2O by 3.25 K. This indicates that researches into NQE of water at atomic scale provide possibilities of understanding isotope effect and unique properties of water systems.

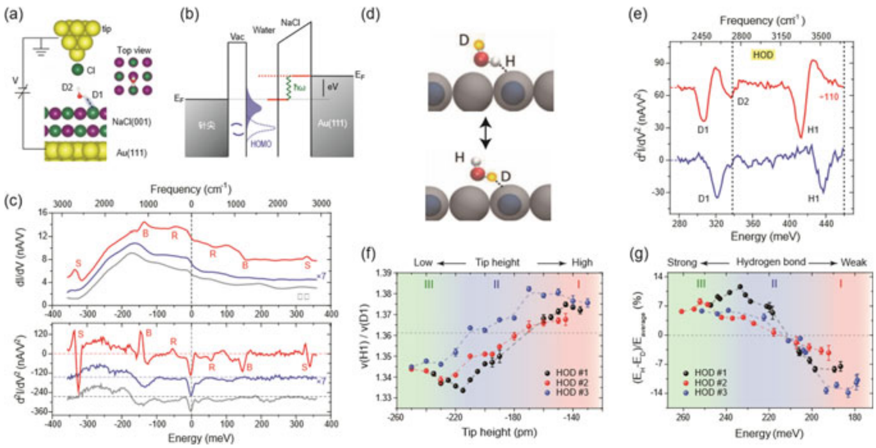


Fig. 5 Illustration for tip-enhanced inelastic electron tunneling spectroscopy and the influence of nuclear quantum effect on hydrogen-bond strength

3.3 Structure of Ion Hydrates and the Magic Number Effect in Ion Hydrate Transportation

Ion hydrate is one of the common structures on water–solid interfaces. In these hydrate structures, due to the interactions between ions and water molecules, the ions have important impact on not only the hydrogen-bonding network, but also vibrations, rotations, diffusions and proton transfer dynamics in water. Reversely, water molecules can form hydration shells around ions, shielding the electric field of ions, affecting the dynamics of ions, such as transportation and conductivity. Interactions between water molecules and ions are relevant to a wide range of applied fields and natural processes, such as salt dissolution, ion transportation in biological systems, corrosion, nucleation of aerosol, desalination, etc. [25–29]. Researches into water-ion interactions at atomic scale is of crucial importance to understanding these processes.

Ion hydration and transport at interfaces are relevant to a wide range of applied fields and natural processes but the structure and dynamics of ion hydrates has been in dispute for decades. As early as late 1800s, systematic researches into ion hydration had begun. Despite the effort of more than 100 years, there hasn't been settled solutions for many ion hydration problems such as hydration cell numbers for ions, number and structure of water molecules at each hydration cell, the influence of ion hydrates on hydrogen-bonding structures, impact factors on the transportation of ion hydrates, etc. Things are even more complex when it comes to ion hydrates on interfaces and restricted systems, due to inhomogeneity of surfaces, variety of lattices as well as interaction between ions, water molecules and surfaces. Determination of molecular-level details of hydration processes remains a great challenge, both experimentally and theoretically. Various spectroscopic techniques have been used to identify the structure and dynamics of solvated ions or molecules through vibrational fingerprints. However, all these techniques suffer from poor spatial resolution and the difficulty of spectral assignment. Molecular simulations have also become powerful tools with which to investigate atomic-scale hydration properties, but the reliability of the results depends critically on many tunable factors, such as interaction potential, how to deal with long-ranged interactions, size and timescale for the simulations, etc. Thus, it still requires detailed molecular-level characterization to correlate atomic structure with the transport properties of hydrated ions, both the interfacial inhomogeneity and the complex competing interactions among ions, water and surfaces.

Our group [30] constructed individual Na^+ hydrates on a NaCl(001) surface by progressively attaching single water molecules (one to five) to the Na^+ ion using a combined system of STM and AFM. With the help of DFT calculations and AFM simulations, we were able to identify the structure for every type of Na^+ hydrates on NaCl surface in real space. During the identification process, the CO-tip AFM technique mentioned in previous sections is playing an important role. On the one hand, it generates higher resolution than STM technique, e.g. the darkest spot in AFM images in Fig. 6 correspond to the positively charged Na^+ ion, the nearby bright spots correspond to negatively charged oxygen atoms in water molecules while the dark ring feature originates from hydrogen atoms. Obviously, this sub-molecular imaging technique contributed greatly to the identification of the hydrate structure. On the other hand, it can avoid the disturbance on the ion hydrates, providing an effective method to characterize the metastable state of the hydrates.

Next we explored the transport of those hydrates. To activate their diffusion at low temperature (5 K), we used the inelastic electron tunneling technique by injecting ‘hot’ electrons/holes into the Au substrate, which transport along the surface and transfer their energy to the hydrates, as shown in Fig. 7a. When comparing the mobility of different Na^+ hydrates, we found an interesting magic number effect (Fig. 7b): the Na^+ hydrates with three water molecules have the highest mobility, higher than other types of hydrates by the order of 1–2 and even higher than in the bulk. To gain insights into the diffusion pathway of those hydrates, we need the support of DFT calculations and MD simulations. Due to the large number of atoms in the ion hydrate systems as well as the large time scale for simulations, these calculations need lots of computing resources. DFT calculations and MD simulations supported by the supercomputer TianHe II provided crucial evidence for the research results. The DFT calculations prove that it is the degree of symmetry match between the hydrates and the surface that causes the magic number effect. To be specific, the hydrates with one, two, four and five water molecules have a better lattice match with the fourfold NaCl surface, resulting in a much tighter binding and less mobility; while the ones with three water molecules cannot have a perfect match, resulting in lots of metastable states. Besides, the collective rotation of water in the metastable state is rather easy, leading to a small barrier (about 80 meV, as shown in Fig. 7c). To investigate the surface diffusion at finite temperatures, especially close to room temperature, we carried out classical MD simulations (Fig. 7d). The results hold for a large temperature range, even at room temperature. More generally, the magic

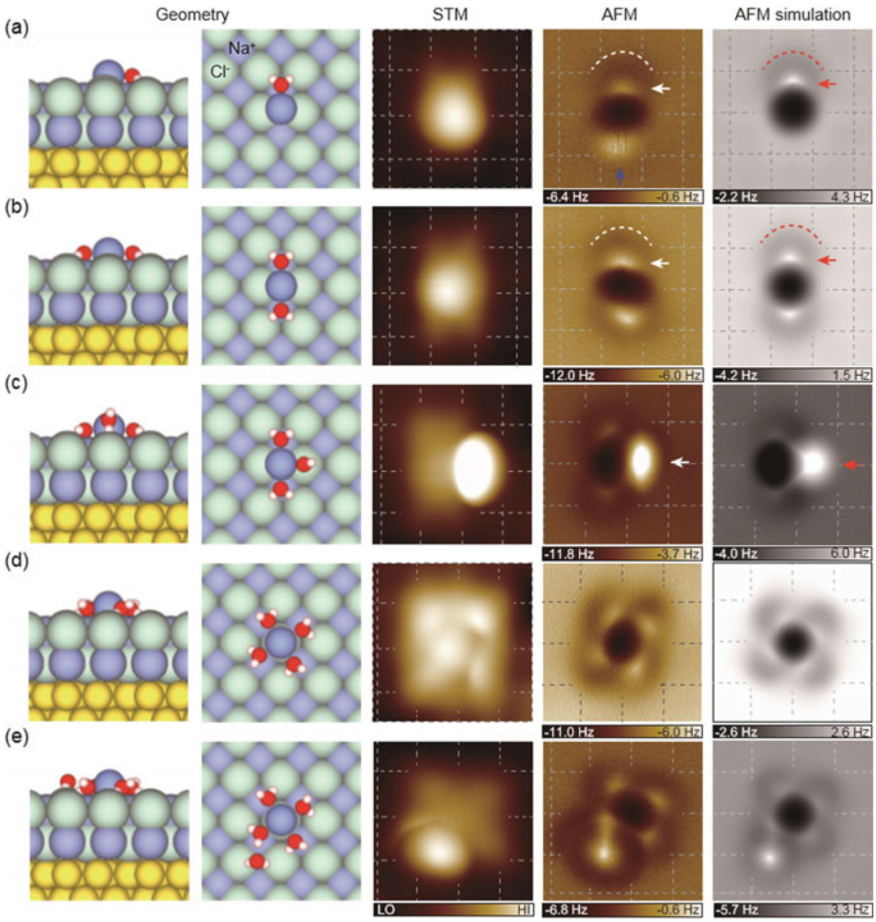


Fig. 6 The structure of Na^+ ion hydrates with high resolution STM/AFM images

number effect observed in this work may also exist for other salt ions (Li^+, K^+, Cl^-, and so on), but the hydration number can be different depending on the size and the hydration asymmetry of positive and negative ions.

For a long time, the continuum model has been widely used to explain the ion transportation in water solutions, neglecting the microscopic details of the interactions between ions, water molecules and surfaces. Our results point out a new way to control the ion transport in nanofluidic systems by interfacial symmetry engineering, which can have potential applications in salt dissolution, ion batteries, electrochemistry, anti-corrosion, desalination. The techniques developed in this work can easily be extended to different ions and other hydration systems, opening up the possibility of studying various hydration processes down to atomic scale, such as protein hydration.

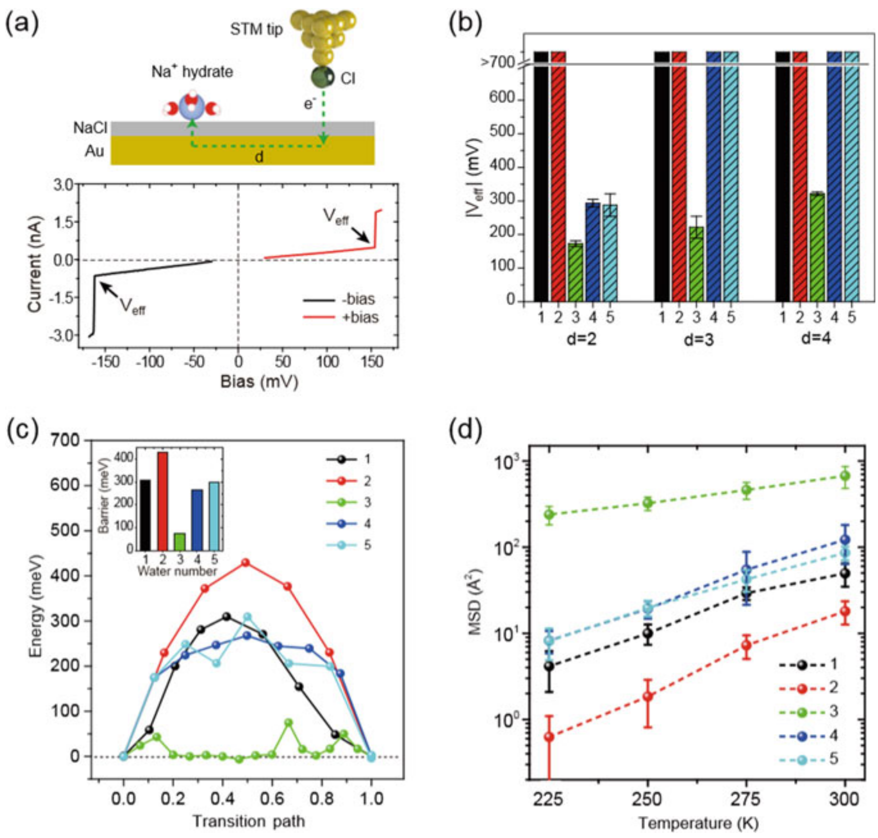


Fig. 7 Magic number effect of Na^+ hydrate transportation on NaCl surface

4 Summary and Expectations

By combining the scanning probe microscopy based on the technology of automated control and digital imaging with the supercomputer-supported simulations and calculations from information technology, we have gained in-depth understanding of the structure and properties of interfacial water. Using the digital processing and imaging technology, it becomes possible to acquire and process multiple signals/data simultaneously. The automated control also greatly enhances the accuracy and efficiency of the scanning probe microscopy studies. With the help of information technology, we were able to further improve the resolution of scanning probe microscopy from single molecular level to sub-molecular level, allowing us to access the degree of freedom of hydrogen atoms spatially and energetically. In the previous studies, we performed detailed investigations into proton transfer dynamics within water tetramers, the influence of nuclear quantum effects on hydrogen-bond strength as well as the structure and dynamics of ion hydrates, resolving several long-standing debates in water science. These specific studies elucidate the important applications of information technology in the fields of surface science, high-resolution imaging/recognition, large-scale system simulations, etc. In future work, we will continue to deploy the LabVIEW programming to increase the automation level of digital imaging processes in the experiments. Hopefully, with the help of machine learning technology, we may be able to simplify the operations of STM and AFM, making them easier yet more powerful tools for more researchers in other fields.

References

1. Thiel PA, Madey TE (1987) The interaction of water with solid-surfaces fundamental aspects. Surf Sci Rep 7(6–8):211–385
2. Henderson MA (2002) The interaction of water with solid surfaces: fundamental aspects revisited. Surf Sci Rep 46(1–8):1–308
3. Al-Abadleh HA, Grassian VH (2003) Oxide surfaces as environmental interfaces. Surf Sci Rep 52(3–4):63–161
4. Hodgson A, Haq S (2009) Water adsorption and the wetting of metal surfaces. Surf Sci Rep 64(9):381–451
5. Ernst JA, Clubb RT, Zhou HX, Gronenbom AM et al (1995) Demonstration of positionally disordered water within a protein hydrophobic cavity by NMR. J Cell Biochem 71
6. Braun J, Glebov A, Graham AP, Menzel A et al (1998) Structure and phonons of the ice surface. Phys Rev Lett 80(12):2638–2641
7. Glebov AL, Graham AP, Menzel A (1999) Vibrational spectroscopy of water molecules on Pt(111) at submonolayer coverages. Surf Sci 427:22–26
8. Sun TJ, Lin FH, Campbell RL, Allingham JS et al (2014) An antifreeze protein folds with an interior network of more than 400 semi-clathrate waters. Science 343(6172):795–798
9. Doering DL, Madey TE (1982) The adsorption of water on clean and oxygen-dosed Ru(001). Surf Sci 123(2–3):305–337
10. Nakamura M, Ito M (2000) Monomer and tetramer water clusters adsorbed on Ru(0001). Chem Phys Lett 325(1–3):293–298
11. Repp J, Meyer G, Stojkovic SM, Gourdon A et al (2005) Molecules on insulating films: scanning-tunneling microscopy imaging of individual molecular orbitals. Phys Rev Lett 94(2):026803
12. Michaelides A, Morgenstern K (2007) Ice nanoclusters at hydrophobic metal surfaces. Nat Mater 6(8):597–601
13. Carrasco J, Hodgson A, Michaelides A (2012) A molecular perspective of water at metal interfaces. Nat Mater 11(8):667–674
14. Mitsui T, Rose MK, Fomin E, Ogletree DF et al (2002) Water diffusion and clustering on Pd(111). Science 297(5588):1850–1852
15. Guo J, Meng XZ, Chen J, Peng JB et al (2014) Real-space imaging of interfacial water with submolecular resolution. Nat Mater 13(2):184–189
16. Peng JB, Guo J, Hapala P, Cao DY et al (2018) Weakly perturbative imaging of interfacial water with submolecular resolution by atomic force microscopy. Nat Commun 9:122
17. Benoit M, Marx D, Parrinello M (1998) Tunnelling and zero-point motion in high-pressure ice. Nature 392(6673):258–261
18. Horiuchi S, Tokunaga Y, Giovannetti G, Picozzi S et al (2010) Above-room-temperature ferroelectricity in a single-component molecular crystal. Nature 463(7282):789–792
19. Meng XZ, Guo J, Peng JB, Chen J et al (2015) Direct visualization of concerted proton tunnelling in a water nanocluster. Nat Phys 11(3):235–239

20. Paesani F, Voth GA (2009) The properties of water: insights from quantum simulations. J Phys Chem B 113(17):5702–5719
21. Tuckerman ME, Marx D, Klein ML, Parrinello M (1997) On the quantum nature of the shared proton in hydrogen bonds. Science 275(5301):817–820
22. Voth GA, Chandler D, Miller WH (1989) Rigorous formulation of quantum transition-state theory and its dynamical corrections. J Chem Phys 91(12):7749–7760
23. Guo J, Lu JT, Feng YX, Chen J et al (2016) Nuclear quantum effects of hydrogen bonds probed by tip-enhanced inelastic electron tunneling. Science 352(6283):321–325
24. Ubbelohde AR, Gallagher KJ (1955) Acid-base effects in hydrogen bonds in crystals. Acta Crystallogr A 8(2):71–83
25. Cohen-Tanugi D, Grossman JC (2012) Water desalination across nanoporous graphene. Nano Lett 12(7):3602–3608
26. Gouaux E, MacKinnon R (2005) Principles of selective ion transport in channels and pumps. Science 310(5753):1461–1465
27. Klimes J, Bowler DR, Michaelides A (2013) Understanding the role of ions and water molecules in the NaCl dissolution process. J Chem Phys 139(23):234702
28. Payandeh J, Scheuer T, Zheng N, Catterall WA (2011) The crystal structure of a voltage-gated sodium channel. Nature 475(7356):353–358
29. Sipila M, Sarnela N, Jokinen T, Henschel H et al (2016) Molecular-scale evidence of aerosol particle formation via sequential addition of HIO_3. Nature 537(7621):532–534
30. Peng JB, Cao DY, He ZL, Guo J et al (2018) The effect of hydration number on the interfacial transport of sodium ions. Nature 557(7707):701–705

Ying Jiang, Boya Distinguished Professor of Physics at International Center for Quantum Materials of Peking University. He received his PhD from Institute of Physics, Chinese Academy of Sciences (CAS) in 2008. After working as a Postdoctoral Associate in University of California, Irvine (2008-2010), he joined International Center for Quantum Materials, Peking University as a tenure-track assistant professor, and was promoted to full professor in 2018.

Ying Jiang mainly focuses on condensed matter physics and chemical physics. He is an expert in advanced scanning probe microscopy/spectroscopy and focused on the atomic-scale properties and ultrafast dynamics in single molecules and low-dimensional materials. He has published over 50 papers, including 14 in Science and Nature journals. He serves on the Editorial Boards of many international journals, including Journal of Chemical Physics, Chemical Physics, and Advanced Quantum Technologies. He was awarded Nishina Asia Award (Nishina Memorial Foundation), Fellow of the American Physical Society, Outstanding Research Award (1st class) (Ministry of Education of China), Young Global Leaders (World Economic Forum), Tan Kah Kee Young Scientist Award, China-Youth Science and Technology Award (The China Association for Science and Technology), Emerging Leaders (JPhys, Institute of Physics). His researches have been selected as Top-ten Science Advances in China twice (2016 and 2018).

Opportunities and Challenges for Biometrics

Zhenan Sun, Qi Li, Yunfan Liu, and Yuhao Zhu

Abstract Biometrics refers to the science and technology of automatic identification achieved by computers through acquiring and analyzing physiological and behavioral characteristics of human body. The purpose of biometrics research is to give computers advanced intelligence to automatically detect, capture, process, analyze, and identify digital biometric signals, that is, make machines "can see and hear". This is one of the basic functions of machine intelligence as well as one of the most significant challenges in theoretical and applied research human beings face. In conclusion, biometrics research is important in terms of both academic significance and practical value. In recent years biometrics has become an important part of national strategies such as the "Internet + Action Plan" and the "Development Plan on the New Generation of Artificial Intelligence". At the same time, it has already become a new growth point for strategic high-tech and electronic information industry in the field of national and public security. This paper introduces research progress of several common biometric modalities such as face, iris, fingerprint and gait, summarizes development trends and opportunities of current biometrics technology, and analyzes main challenges on the road to the development of a new generation of biometrics. Finally, this paper provides some suggestions regarding the future development of biometrics.

Keywords Biometrics · Face recognition · Iris recognition · Fingerprint recognition · Gait recognition

1 Introduction

Identification is the key technology to protect national and public safety, to maintain economic and social order, and to ensure the security of personal information. Traditional identification methods are based on specific knowledge (such as passwords, codes, questions, answers, etc.) and physical objects (such as keys, ID cards,

Z. Sun (✉) · Q. Li · Y. Liu · Y. Zhu
National Laboratory of Pattern Recognition, Center for Research on Intelligent Perception and Computing, Institute of Automation, Chinese Academy of Sciences, Beijing, China
e-mail: znsun@nlpr.ia.ac.cn

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_6

USB shields, etc.). They have inherent defects such as being cracked, forgotten, and stolen, and it is difficult to meet the requirement on reliability, security, and convenience. Biometric identification (referred to as Biometrics) is based on the identification of physiological or behavioral characteristics of individuals. It has the unique advantage of being distinct, stable, reliable, and portable. Common biometric modalities include fingerprints, iris, face, palm print, hand shape, veins, handwriting, gait, voice, etc. (Fig. 1). Biometrics is an interdisciplinary field: acquisition devices at the hardware level involve optical engineering, mechanical engineering, and electronic engineering; recognition algorithms at the software level involve key problems in fields of pattern recognition, machine learning, computer vision, artificial intelligence, digital image processing, signal analysis, cognitive science, neural computing, human–computer interaction, and information security.

Biometrics refers to the science and technology of automatic identification achieved by computers through acquiring and analyzing physiological and behavioral characteristics of human body. The purpose of biometrics research is to give computers advanced intelligence to automatically detect, capture, process, analyze, and identify digital biometric signals, that is, make machines "can see and hear". This is one of the basic functions of machine intelligence as well as one of the most significant challenges in theoretical and applied research human beings face, which is of great scientific significance.

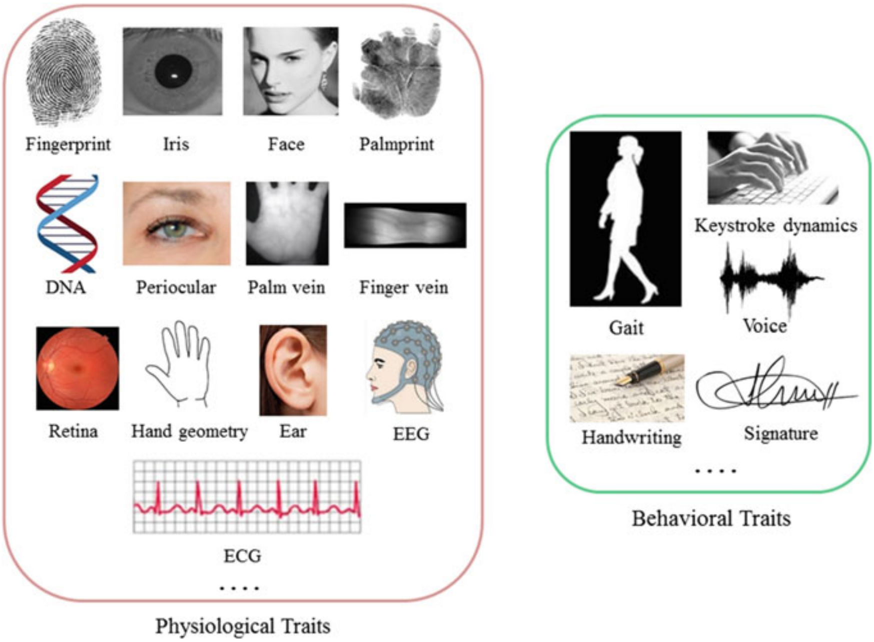


Fig. 1 Representative modalities of biometrics

Biometrics has complex patterns (points, lines, regions, textures, etc.), various types (faces, irises, fingerprints, gait, handwriting, etc.), diverse categories (up to tens of millions of people per class), rich information (statistical and structural information, local and global information), and subtle differences. They could be reflected and described in various signal forms such as images, videos, and speech, so biometric recognition is a typical and complex problem involving pattern recognition, computer vision, and neural computing, setting a challenging goal for these disciplines, building a good basic experimental platform for related researchers for trying out new methods, validating new theories, and explaining new phenomena. Since the 1960s, basic problems of biometric recognition have always been inspiring the development of disciplines such as pattern recognition, computer vision, cognition and neural computing. Biometric recognition has also been at the forefront of the development of disciplines such as pattern recognition. Therefore, the in-depth research and final solution to the problem of biometric recognition can greatly promote the maturity and development of these disciplines.

Alibaba DAMO Academy claims that "digital identity will become the second ID card" is one of the top ten technology trends in 2019 [1]. From mobile phone unlocking, community access control to cashiers of restaurants and supermarkets, and then to stations of high-speed rail and airports, the era of face recognition and iris recognition is accelerating to come. We commented on the statement in major media such as the Xinhua News Agency, saying that "biological features, such as faces and irises, will become a key for people to enter the connected world and enjoy digital life" [54]. It can be seen that biometric recognition, such as face recognition and iris recognition, is a popular technology at the moment, and has attracted much attention from fields of political, industrial, academic, and research applications. The "Internet +" action plan [2] and "new-generation artificial intelligence development planning" [3] and other national strategies have clearly proposed to focus on supporting the development of biometric identification technology. Biometrics is not only an important academic frontier of pattern recognition and computer vision, but also one of the main directions for the fastest implementation of artificial intelligence and the largest commercial market. Biometric technology has been widely used in important areas of the country such as public security and anti-terrorism, financial payment, social security certification, and security clearance, creating a market with a size of tens of billions of dollars. In addition, face recognition involves issues of public interest such as privacy, ethics, and law, which has also caused widespread concern in the community.

2 The Development Status of Biometric Recognition Technology

2.1 Overview

The research of computer-based automatic biometric recognition started in the 1960s. Fig. 2 shows the development history of biometric identifications. With the development of basic disciplines including pattern recognition, computer vision, digital image processing, and signal processing, the research level of biometric recognition algorithms has rapidly improved; with the development of biometric sensor technology and computer technology, rapid implementation and low-cost promotion have become possible; with the increasing emphasis on security management and the increasing demand for automatic identity authentication, biometric systems have been widely used in households, workplaces and public areas.

Therefore, in recent years, biometric identification has become a popular direction in both academia and industry. Driven by the broad prospect of practical applications, biometric recognition has become a hot research topic in the fields of pattern recognition, image processing, and computer vision. New modalities, new devices, new theories, and new methods have emerged endlessly to guide and promote the rapid development of related disciplines.

The National Science Foundation (NSF) funded biometrics as a key direction, and also held biometric disciplinary development seminars. The Center for Identification Technology Research, an institute dedicated to researching biometric recognition, was established in conjunction with 14 universities, including Michigan State University, West Virginia University, Rice University, and the University of Chicago. The National Science and Technology Council (NSTC) is a cabinet-level committee chaired by the president, members of ministers, and heads of major federal agencies related to science and technology policy, which is in charge of the decision-making process related to scientific and technical issues.

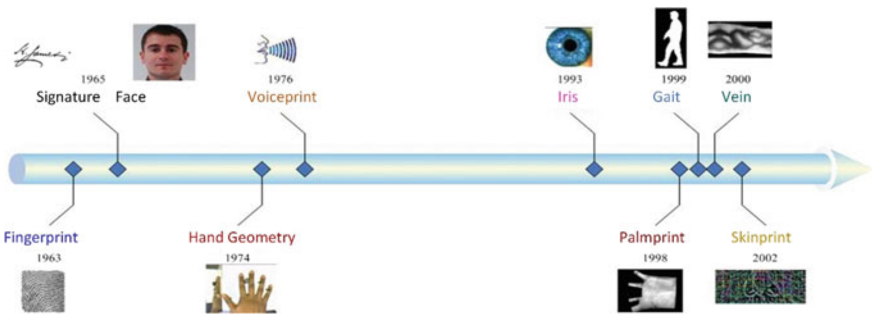


Fig. 2 The history of biometric identification

NSTC attaches great importance to promoting the development and application of biometric identification technology in the United States. It has specifically established a biometric identification committee and has released a series of biometric identification technology development planning reports, such as "Biometrics in Government Post 9–11-Advancing Science, Enhancing Operations, NSTC Policy for Enabling the Development, Adoption and Use of Biometric Standards, The National Biometrics Challenge, and more.

In order to enhance the ability to protect US defense and important civilian facilities and prevent attacks from terrorists, the US Department of Defense Advanced Research Projects Agency (DARPA) launched the Human ID (Human Identification at Distance) program to develop multi-modal long-range identification of human biological characteristics. It aims to realize the technology to detect, classify and identify individuals or groups of people in all weather conditions, and provide early warning for military protection, national defense, combating terrorist acts, preventing criminals and other man-made sabotage activities.

The EU framework plans take biometrics as the focus of funding, and have organized a series of large-scale research projects such as BIOSEC, BIOSECURE, TURBINE, ACTIBIO, HIDE, MOBIO, HUMABIO, 3DFACE, MTIT, BITE, etc., collaborating with some key European universities to study critical issues in the field of biometrics, such as basic databases, test platforms, mobile biometrics, multimodal biometrics, etc.

A large number of academic papers on biometrics has been published in conferences in fields of pattern recognition, image processing, signal processing (such as CVPR, ICCV, ICB, ICPR, ICIP) and journals (such as IEEE Transactions on PAMI/IP/IFS, PR, IVC, CVIU). Some systematic in-depth works, such as Handbook of Biometrics, Handbook of Multi-biometrics, Handbook of Face Recognition, and Handbook of Fingerprint Recognition have also been published. Many universities and research institutes have engaged in the study of biometric recognition, and new biometric identification products are also emerging.

Biometric identification is not only a major appliable technology oriented to the national strategic needs, but also at the forefront of the development of pattern recognition disciplines. Therefore, the sources of biometrics research power include universities, research institutes, companies, and government agencies.

Due to the importance of biometrics for homeland security, financial security, and cybersecurity, government departments in many countries are very concerned about the research progress of biometrics. FBI and NIST have been researching fingerprint recognition technology since 1967, and have built the largest fingerprint database of tens of millions of people in the world. The United States Department of Defense (DOD) established a biometric management office and biometric fusion center in 2000, and started the face recognition scientific research project (FERET) from 1993.

The US government organizes the Biometrics Consortium every year and organizes various algorithm evaluations through NIST. Moreover, the US Department of Homeland Security, the Central Intelligence Agency, and some military research institutions are also organizing the development of large-scale biometric identification projects. As biometric identification is a major challenge to the traditional social lifestyle, it will cause many new problems to be solved, such as privacy, ethics, law, etc. Judicial departments in many countries are also tracking the research and application status of biometric systems to guide the development of relevant laws and policies.

The development of standards for biometrics has started. Specialized agencies have already been established, such as the joint technical committee established by the ISO (the International Organization for Standardization) and the IEC (the International Electrotechnical Commission). There is a branch, SC 37 Biometrics, which is responsible for formulating relevant standards for biometric technologies. Some developed countries have also established designated departments to develop biometric standards.

The following sections mainly introduce the research progress of main biometric recognition technologies, including human face, iris, fingerprint, and gait.

2.2 Face Recognition Technology

Human face is one of the most traditional and intuitive biological features. It has received full attention and has important applications in a variety of identity authentication scenarios due to its good user acceptance and simplicity of collection. In addition, the analysis and synthesis of human faces and the emerging problems of liveness detection in face images and videos have also attracted increasing attention from both academics and industry.

In the early stage of the development of artificial intelligence, face recognition was studied as a general pattern recognition problem. At this stage, face recognition technology is mainly implemented by expert systems that rely on manual features. Its robustness and generalization are poor, and it has not achieved widespread practical application.

With the rapid development of statistical learning, many face recognition algorithms based on statistical models have been proposed. The "EigenFace" model proposed by the Massachusetts Institute of Technology is one of the most important research results at this stage [4]. This method introduces a statistical-based machine learning method to the face recognition task, and derives a series of subspace analysis methods and improved strategies based on kernel learning, such as FisherFace [5]. The introduction of statistical learning theory has also improved the performance of classifiers. Among them, Support Vector Machine (SVM) has become the first choice for classifiers due to its simple and intuitive theoretical basis and excellent adaptability to high-dimensional large sample data. Feature extraction techniques have also been developed rapidly during this period. For example, both Gabor filters and LBP filters were proposed at this time. Therefore, early face recognition methods mainly depended on the combination of artificially designed features and machine learning technology. During this period, manually designed features were not able to cope with different changes (lighting conditions, occlusions, etc.) in unconstrained environments. Since 2011, with the rapid development of deep learning theory and the widespread application of related algorithms in the field of computer vision, deep neural network-based face recognition technology has become the mainstream. With its hierarchical structure characteristics, deep neural networks could mine higher-level information that is more representative, so that it could achieve much better discriminative performance than shallow classifiers, and could surpass the accuracy of human recognition on commonly used databases.

During this period, studies on face recognition focused on the design of loss functions. Facebook published its research results in 2014, where the DeepFace algorithm based on the softmax loss function [6] had reached 97.35% on the LFW (Labeled Face in the Wild) landmark database, reducing the error rate of existing best face recognition by 27%. Google proposed the FaceNet algorithm based on the triplet loss function in 2015 [7], relying on millions of training data, and achieved 99.63% accuracy on LFW. After that, center loss [8] and large margin softmax loss function [9] were proposed to train face recognition networks. The above-mentioned loss functions usually adopt the Euclidean distance metric. SphereFace [10] first proposed the use of cosine distance instead of Euclidean distance for face recognition training and testing, and its performance surpassed the previous method. Both post-CosFace [11] and ArcFace [12] use an improved cosine distance.

In the process of moving from the laboratory to practical applications, adaptability to large-scale face data is the primary challenge face recognition technology confronts. In 2015, the research team of the University of Washington proposed the MegaFace dataset containing millions of face images, and held a recognition competition [13], aiming to improve the recognition accuracy in the case of large-scale face data. As of 2019, research results from companies such as Tencent, SenseTime, and Sogou have topped the competition. The United States National Bureau of Standards and Technology has restarted FRVT face recognition evaluation in recent years [14], and many companies in China, such as Yitu, have achieved good performance in FRVT. In addition to the scale of the dataset, lighting conditions and poses are also important factors affecting the performance of face recognition technology. To solve these problems, 3D face recognition and cross-modal face recognition have gradually become the main research hotspots in recent years.

Face generation technology is another research hotspot that comes with the rising of big data, which could help solve the problems of large poses and age gaps in face recognition. In addition, editing attributes of face pictures, such as age, expression, hair color, etc., have great practical values in webcasting and photo retouching. Making artificial intelligence algorithms to create faces with various styles like painters also has great application prospects in the entertainment industry. High-quality artificial intelligence paintings are also considered to have high artistic value, and there have been cases of auctions in the international market. The fundamental theory behind face generation technology is Generative Adversarial Networks, which have made great breakthroughs in face frontalization [15] and face aging [16].

As one of the most common biological characteristics, face recognition has become the most widely used method of identity authentication due to its advantages such as high accuracy and low requirements on user cooperation. The FaceID technology, released by Apple in 2017, uses structured light to capture three-dimensional information on the user's face. It can successfully solve the problems in face recognition technology such as lighting and posture, and has been applied to various identity authentication services across the entire product line. The facial recognition technology of domestic companies such as CloudWalk, Megvii, SenseTime, and Yitu has been widely used in many aspects, including face payment, bank intelligent outlets, security control in key areas, and intelligent medical care. Based on face recognition, technologies such as face liveness detection and pedestrian trajectory analysis have also emerged in smart finance and intelligent transportation.

Although face recognition technology is now developing rapidly, it is still a problem that has not been completely solved. For example, accurate face recognition in surveillance scenes and large scale data, for people with dark skins, and for face images with large age gaps. There is a lot of room and potential for face recognition technologies to be improved.

2.3 Iris Recognition Technology

The industry is the main research force of iris image acquisition devices. LG, Panasonic, IrisGuard, IrisKing and other companies have designed a series of close-range iris image acquisition equipment. In order to improve the convenience of iris imaging and to expand the application range of iris recognition, an increasing number of institutions have started research on long-range iris image acquisition. The InSight system of AOptix in the United States implements clear imaging of iris from 1.5–2.5 and 2.4–3 m away. Carnegie Mellon University is currently developing a device with an imaging distance of 12 m. OKI's IRISPASS-M and Panasonic's BM-ET500 use PTZ gimbals to adjust the camera's pitch angle to accommodate users of different heights. At the same time, the iris imaging device gradually has become more and more lightweight to be used in practical situations. In 2013, AOptix developed a mobile phone external iris, face, fingerprint image acquisition module, which can be seamlessly connected with the iPhone. In May 2015, Fujitsu released a smartphone that can use iris recognition to unlock, log in to online accounts, and pay. IrisKing officially released the first domestic iris recognition mobile phone in early 2016. In 2017, Samsung S8, Note7 and other smartphones began to add iris recognition modules.

As shown in Fig. 3, two main steps for iris recognition on given iris images are iris region segmentation and iris texture feature analysis. The iris region segmentation can be roughly divided into two categories: boundary localization based methods [17, 18] and pixel classification based methods [19, 20]. The iris texture feature analysis includes feature expression and comparison. The feature expression method extracts discriminative information from complex texture images that can be used for identity recognition. Representative feature expression method include Gabor phase-based methods [17], multi-channel texture analysis-based methods [21], correlation filters-based method [22], and sequential measurement-based method, etc. [23]. The stability and discrimination of eigenvalues are the main factors affecting the accuracy of feature comparison.

Traditional iris recognition algorithms mostly use artificially designed logic rules and algorithm parameters, resulting in poor generalization performance of the algorithm, which cannot meet the need of large-scale application scenarios. Data-driven machine learning methods automatically learn the optimal parameters from a large number of training samples, which can significantly improve the accuracy, robustness and generalization performance of the iris recognition algorithm [24]. Large-scale iris recognition applications have brought many new challenges. Fast retrieval of iris features [25], and robust recognition of multi-source heterogeneous iris images [26] have become current research difficulties and hot issues of iris recognition.

The current mainstream iris imaging methods fail to consider subsequent image processing steps, while emerging computational imaging methods (such as wavefront coding and light field cameras) consider both imaging and image processing at the same time, which is expected to break through the existing technology bottleneck and significantly increase the imaging range. In terms of algorithms, the research team of the Institute of Automation, Chinese Academy of Sciences, inspired by the human visual mechanism, proposed the use of sequential measurement filters to describe the local texture of the iris, and designed multiple feature selection methods to determine the optimal parameters of the filter [27]. For the first time, they applied deep learning to iris recognition, and proposed a multi-scale full convolutional neural network-based iris segmentation method [28] and a convolutional neural network-based iris feature learning method [24]. They also explored the complementary relationship between deep learning features and sequential measurement features [29].

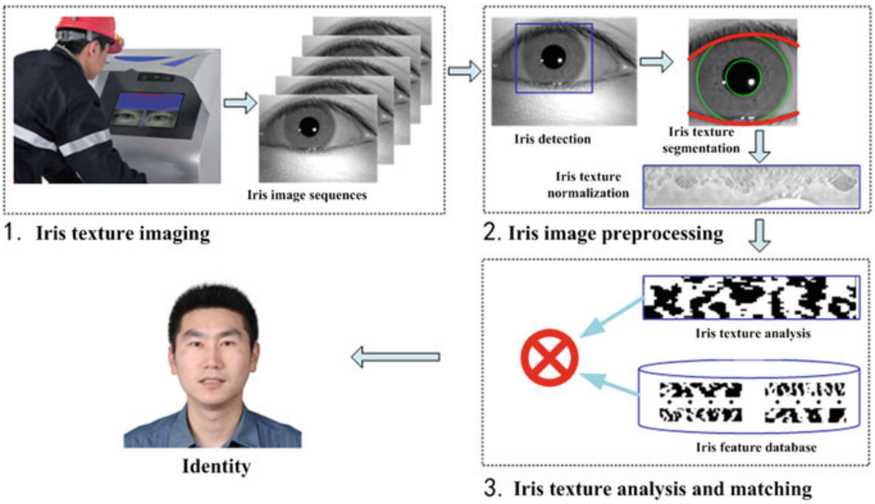


Fig. 3 The pipeline of iris recognition

Fig. 4 The development path of Chinese iris recognition technology

In addition, they systematically studied the iris image classification method based on a hierarchical visual dictionary, which significantly improved the accuracy of iris feature retrieval, ethnic classification and living body detection [25]. The development roadmap of iris recognition technology of the Institute of Automation, Chinese Academy of Sciences is shown in Fig. 4.

2.4 *Fingerprint Recognition Technology*

Fingerprint recognition technology is one of the most common and earliest biometric identification technologies for civilian use. In 1892, Galton and others pointed out that as a unique and stable biological feature, fingerprints can be used for identity authentication, marking the beginning of fingerprint identification technology.

Fingerprint recognition technology mainly includes three aspects, namely fingerprint image acquisition, fingerprint image enhancement, and fingerprint feature extraction and matching. After the electronic computer was invented, optical-based fingerprint collection devices have replaced traditional inks, which greatly improved the efficiency of fingerprint collection, identification, and storage. Subsequently, capacitive sensor-based fingerprint collectors were invented [30], and have been widely used in user identity authentication systems for mobile terminal devices such as Apple phones. There are mainly two types: pushing based and scratching-based. In addition, fingerprint acquisition technologies based on temperature sensors, ultrasonic waves, and electromagnetic waves have also been proposed, each with its own strengths. In recent years, non-contact 3D fingerprint acquisition systems have also been proposed to improve user experience and recognition accuracy [31].

Fingerprint image enhancement mainly includes image smoothing (de-noising and texture stitching), image binarization (separation of foreground and background), and refinement (fingerprint skeleton acquisition). Traditional image processing methods such as frequency domain filtering, Gabor transform, and matched filter [32] can effectively remove noise from fingerprint images. They can also detect, complete and refine breakpoints in fingerprint lines. With the development of deep learning, deep convolutional networks have been widely used in fingerprint image enhancement related problems, such as warped fingerprint image correction [33] due to their strong feature extraction capabilities.

Fingerprint image feature extraction and matching methods can be broadly divided into two categories: directional field-based methods and feature point-based methods. The directional field depicts ridges and valleys of the fingerprint image, which is an important basis for fingerprint image matching. Many methods have been proposed to reduce the influence of noise on the calculation of the directional field, and improve the operation efficiency. Feature points refer to the common patterns of fingerprints, including arches, account bows, left loops, right loops, and thread patterns. The characteristics of regional distribution and rotation invariance of feature points are also often used to improve the robustness of recognition algorithms. With the application of fingerprint recognition technology in different scenarios, the quality of collected fingerprint images is uneven, and sometimes even complete fingerprints cannot be obtained. Therefore, partial fingerprint image recognition problem is currently a research hotspot [34]. In addition, for protecting the safety of users' personal property, the problem of liveness detection in fingerprint recognition technology is also a research focus. In order to solve this problem, on the one hand, additional sensors can be added to the fingerprint acquisition system from the hardware perspective to detect the evidence of liveness, such as the temperature of the finger [35], color, and blood flow. On the other hand, the quality of capture fingerprint images could be considered as an evaluation metric to obtain high-quality live fingerprint data [36].

2.5 Gait Recognition Technology

Gait can be defined as a combination of action cycles that lead to motion [37]. Under this definition, biped or quadruped walking, running, climbing, jumping, and swimming can be considered as a kind of gait. In the current stage of research, the research on gait is more focused on the scope of "human walking". Gait recognition aims to use people's walking gestures to recognize identities, which is a very challenging research topic. In view of the long-distance and non-aggressive perceptual characteristics of gait, the research of gait recognition has a good application prospect and research value, especially for long-range and large-scale visual surveillance applications. Since the United States Defense Advanced Research Projects Agency (DARPA) proposed the Long-range Human Identification Program (HumanID) in 2000, a large number of research institutes and universities worldwide have joined the research wave of "gait recognition", including Massachusetts Institute of Technology (MIT), Carnegie Mellon University (CMU), Georgia Institute of Technology (GIT), Chinese Academy of Sciences Institute of Automation (CASIA) and other world-class research institutions.

Generally, a gait recognition system takes a gait sequence as input, obtains gait features through a feature extractor, and compares the extracted features with features in the gait database to give the identity of the target person. As the key of gait recognition systems, the innovation of gait feature extraction algorithm has always been the focus of researchers. The earliest gait recognition algorithm was based on the analysis of spatiotemporal information [38]. By establishing the XYT spatiotemporal coordinate system, the sagittal angle of the target's direction is considered as the feature for identification. In [39], the optical flow method was introduced for gait feature modeling, and the best recognition results at the time were successfully obtained. Finally, the feature space transformation (EST) [40] was proposed, terminating the study on small-scale gait datasets with a 100% recognition rate. With the proposal of more gait datasets with much larger scales, modern gait feature extraction algorithms can be divided into two broad categories by whether explicit kinematic modeling is performed.

For gait feature extraction algorithms that do not require a kinematics model, most of them rely on the analysis of the spatiotemporal information of the contour of the target person. Almost all gait feature extraction algorithms without kinematics models require preprocessing such as background extraction and contour extraction. The Institute of Automation of the Chinese Academy of Sciences (CASIA) has proposed a gait recognition algorithm that analyzes contour edge information [41]. The introduction of Dynamic Time Warping (DTW) and Hidden Markov Model (HMM) [42] further improves the recognition accuracy of algorithms without kinematic models. The proposition of Gait Energy Image (GEI) [43] and Gait Flow Image (GFI) [44] has once again pushed forward the research in this direction, and a series of steps based on this method has been derived. Gait feature extraction algorithms based on kinematics use more accurate spatiotemporal information to perform motion analysis on human joints. The focus of its research is to accurately model the structure of the human body. Compared with algorithms without kinematics models, this type of algorithm achieves better robustness and recognition accuracy at the expense of algorithm complexity. Earlier methods used 16 points to approximate the human body, and with the advent of Microsoft's depth camera Kinect, depth information helped to model the human body more accurately [45].

Deep learning has become the mainstream method of gait recognition in recent years and has made breakthroughs in cross-view gait recognition. CASIA has developed a framework based on deep convolutional neural networks to learn the similarity between paired GEIs and automatically learn valuable static and dynamic features for gait recognition, and the use of "positive–negative pair"-based training methods to expand the sample size can also achieve high accuracy on small sample training databases. This method based on dual-channel mid-level gait feature fusion solves the difficulty of cross-view gait features matching in traditional methods. It achieves an average accuracy rate of more than 94% across perspectives on large-scale gait database CASIA-B, reaching the international leading level. Related results have been published in IEEE-TPAMI [46] and IEEE-TMM [47]. Recently, the team further proposed an acceleration idea of full-graph segmentation based on convolutional neural network, which accelerated the previous algorithm by nearly 1000 times. They also built the world's largest outdoor gait database, which contains 760,000 gait sequences. The long-distance gait recognition system developed by the Institute of Automation, Chinese Academy of Sciences, won the second prize of Beijing Municipal Science and Technology Invention 2018, and it performed well on CCTV's "Smart Wisdom" program and was rated as the "wisdom pioneer". Embedded gait recognition has been applied in the field of home appliances in the United States. Gait recognition has been used in public security systems for more than 1,000 h and participated in the detection of more than 20 cases.

Although many breakthroughs have been made in gait recognition in recent years, the recognition performance is far from being saturated. In general, the current research hotspots focus on (1) improving the robustness of the algorithm in the case of clothing changes, different travel speeds, and perspective changes, (2) improving the algorithm's accuracy in long-interval gait recognition, and (3) improving the accuracy of person re-identification and other directions. Few researchers have been involved in studying the influence of factors such as race, injury, fatigue, weight-bearing, and self-control on gait recognition, which are difficult points in gait recognition research. In the next few years, the research on gait recognition will also continue the current research hotspots, and gradually expand the research direction into the above-mentioned difficult areas. As a bridge connecting current and future research fields, the research of gait-based pedestrian re-identification algorithms is expected to bring new opportunities and inspirations for research breakthroughs in gait recognition from multiple levels and perspectives.

2.6 Other Biometric Recognition Technologies

2.6.1 Palmprint Recognition Technology

Compared with other biological features, palmprint recognition has the advantage of convenient collection, good privacy, and high user acceptance. Palmprint recognition is widely used in high-level access control, public security criminal investigation, medical social security, network security, attendance and other fields. Palmprint recognition currently has two application areas: detection-oriented and civilian-oriented. Palmprint recognition based on law generally requires high-resolution palmprint images. Palmprint images based on commercial applications can generally be low-resolution grayscale images. In addition, related researches on three-dimensional palm print images and non-contact collected palm print images are also being carried out.

2.6.2 Voiceprint Recognition Technology

Voiceprint recognition is a technology that automatically recognizes a speaker's identity based on speech parameters that reflect the physiological and behavioral characteristics of the speaker in the speech waveform. Voiceprint recognition uses sound as a recognition feature and can be collected in a non-contact manner. The collection method is more concealed, the collection space is wider, and it is easier for users to accept. The technology was developed by Bell Labs in the late 1940s and is mainly used in the field of military intelligence. With the gradual development of voiceprint recognition technology, the technology was used in the fields of forensic evaluation and forensic evidence in the United States in the late 1960s. At present, voiceprint recognition technology has important applications in the fields of information, financial security, justice, security and document anti-counterfeiting, military and national defense.

2.6.3 Eyeprint Recognition Technology

Eyeprint recognition is a technology for individual identity verification through the unique vein pattern on the white region (sclera) of the human eye. People's eye conditions are not static, eye congestion could be caused by allergies, red eyes, or hangovers all night, but this will not affect the pattern of eyeprint and blood vessels in the eye. This shows that the eyeprint features are stable enough to be used for authentication. Compared with other biometric recognition technologies, eyeball reflection, blinking, eyelashes and other factors will seriously interfere with the accuracy of eyeprint recognition, resulting in high thresholds and challenges for its research and development.

In addition, the field of biometrics continues to explore emerging information modalities for identity verification, such as veins [48], knuckles [49], human ears [50], EEG/ECG signals [51], eye movements [52], screen swiping patterns [53], etc.

3 Development Trends and Opportunities of Biometrics

By analyzing the policy environment, market size, application platform and subject development of biometrics, we believe that biometrics are currently in an opportunity period of great strategic importance.

3.1 Regulatory and Policy Support

At present, the important role of biometric identification technology has been highly recognized by governments and the general public, providing a loose policy environment for the widespread application of biometric identification technology. The United States successively signed the Patriot Act, the Border Visa Act, and the Aviation Security Act after "911", all requiring the use of biometric technology as a guarantee of legal implementation, requiring biometric features such as fingerprints and irises to be added to passports. In 2006, ICAO mandated that facial features must be stored in electronic passports, with fingerprints and irises as options. Also, more and more countries and regions have added biometrics to their ID cards. For example, the Indian government has launched a national identification and management UID project and has collected more than 1.2 billion people's iris, face and fingerprint information.

Biometric identification technology has been highly valued by various sectors of society such as government, industries, universities, research institutes, and other industries. For example, the annual biometric forum in the United States brings the Department of Security, the Department of Justice, the Department of Defense, FBI, CIA, NIST, universities, research institutes, and thousands of companies together to discuss biometric technology, industry, and the unprecedented policy opportunities of it.

3.2 Growing Biometric Market

Bill Gates, the CEO of Microsoft, had predicted in 2004 that biometrics would become an important change in the IT industry in the next few years. In August 2012, Gartner published a report about technology hype cycles of the year 2012–2013 which shows biometric authentication has now entered a bright period and reached its peak as one of the 48 emerging technologies. Currently, biometric technology and products are now available for public security areas such as border clearance, residence permits, public security and justice, financial securities, e-commerce, social security benefits, information networks and other civil areas such as access control or attendance in schools, hospitals, venues, supermarkets, etc., which is to say biometrics has been widely used and formed a new industry of information technology.

Faces, irises and other identifying information have become the portal of the "Interne+" era and the intelligent era, which has broad development capacity. For example, iPhone X and Samsung S8 use face recognition and iris recognition technology as mobile phone login methods, respectively. According to a forecast from the BCC market research company in January 2016, the global biometrics market size will grow significantly in the next few years. It estimates that from 2015 to 2020, the global biometrics market will reach a compound annual growth rate (CAGR) of 22.7%, and the global biometrics market will reach 41.5 billion by 2020.

3.3 *Inter-Connected Application Platform*

Biometric technologies may be useful as long as there is a space for human existence on whether physical space or virtual space. At present, the development of mobile internet, IoTs, and social networks provide a new application platform for the development of biometrics. With the in-depth development of information technology in human society, the biometric sensors (audio and video) of mobile internet and IoTs are showing a general development trend. At the same time, the scale of audio and video data on the internet is growing explosively, which provides a new development opportunity for biometric identification technologies.

Many countries have deployed tens of millions of high-definition monitoring terminals all over the country where personal activities are the major monitoring content and biometrics is an important technology to determine the identity of individuals. The high-definition cameras on hundreds of millions of new smartphone and tablet computer growth each year have built a ubiquitous mobile visual perception platform. Such iPhone and other smartphones are equipped with voice interaction technology, which provides a new way to collect the biometrics of face, iris, voiceprint, and other modes; the booming social networking sites generate massive user data every day, in which a large number of images and videos involve biometric information such as faces, irises, and voices can be collected. In addition, some somatosensory game platforms, such as Kinect, can obtain depth and color image information at the same time, providing new opportunities for 2D and 3D face fusion recognition. Besides, augmented reality and virtual reality (AR/VR) scenes would use biometric technology either.

3.4 *Upgrading Technology*

After more than 40 years of development in the field of biometrics, researchers have accumulated a wealth of theories and methods, and the proposed technologies can basically correctly identify highly coordinated users under strictly controlled conditions. However, biometric images are subject to internal physiological changes and external environmental changes, in turn, the performance of biometrics declined sharply, which leads these technologies cannot meet the needs of identification in the complex environment of the real world. Many basic scientific problems need further study in terms of ease of use, robustness, real-time, security, and wide-area perception, which have severely restricted the subject progress, technology promotion and industrial development of biometrics. Therefore, with ubiquitous biometric sensors, biometric recognition technology is facing a historical opportunity from applied only at controlled conditions to complex real-world environments.

In order to meet the growing practical application needs in the field of information security, we must innovate existing biometric recognition models and propose a set of innovative technologies, systems, and applications to form a systematic solution, handling biometric image acquisition, living detection, pattern recognition, security protection; break through the various bottlenecks of the existing biometric recognition system in ease of use, accuracy, robustness, real-time, and security; and build convenience, automation, intelligence, networking, safe and reliable biometric identification system for massive users. Therefore, we are facing the historical opportunity to seize the commanding heights of the new generation of biometrics technology, which is of strategic significance for the scientific progress of biometrics, technology promotion and industrial development.

4 Major Challenges in Developing Next-Generation Biometrics

The core concept of the new generation of biometric recognition technology is "people-oriented", and it is necessary to break through many technical bottlenecks to achieve the leap-forward development from controlled and restricted scenarios to ubiquitous biometric recognition cases. For one object to be identified, a more intelligent biometric image acquisition device and computer vision software will realize a new model of biometrics that transitions from "humans cooperate with machines" to "machines actively cooperate with people": the requirements on location, distance, posture, and expression of this object will be more relaxed. Biometric systems can complete the identification process and complete visual monitoring tasks such as detection, tracking, and trajectory analysis in a relatively relaxed environment without explicit user's cooperation. Indeed, there are many technical challenges for the new generation of biometric recognition technology in every step from biometric information acquisition to information processing, from real-time response to security assurance and from physical space to cyberspace.

4.1 The Ease of Acquisition

The biometric information acquisition device is a front-end module of any biometric identification system and is also a major bottleneck that affects the popularization of biometrics. Therefore, the necessity and significance of developing a convenient biometric acquisition device are self-evident. The convenient acquisition of biometric information involves a series of core technologies such as human–computer interaction, target detection, quality evaluation, and integration of optical-mechanical–electrical computing. In terms of biometric information acquisition, potential research directions include long-distance iris-face integrated imaging, multispectral biometric imaging (such as visible light fingerprint/palmprint image, near-infrared finger vein/palm vein image acquisition, near-infrared human face imaging, etc.), non-contact fingerprint/palmprint imaging, facial depth image acquisition which is similar to Kinect technology, etc. In practical application scenarios where the environment has complex lighting conditions, various environmental backgrounds, crowd movement, mutual occlusion, and small biological targets such as iris, face, and palmprint, the key issue remains to be solved that how to automatically capture multi-modal organisms with dynamic changes over long distances. In addition, there are also many important challenges in obtaining high-quality voiceprint information in noisy environments.

4.2 Recognition Robustness

The algorithm part of biometrics is undoubtedly the core module of any biometric system. This part aims to analyze and extract robust biometrics from the biometric information collected by the biometric acquisition device to achieve accurate and reliable individual matching and recognition. This process involves preprocessing, feature analysis, feature extraction, and feature matching for identification as well as information fusion of multi-modal biological characteristics.

Due to the complexity of the imaging environment under complex conditions such as lighting condition, distance, pose, expression, blurriness, deformation, glasses, occlusion, noise, and other factors will cause large intra-class differences between biometric images collected at different times. It is a challenging pattern recognition problem that how to propose an image feature expression and analysis method which is robust to these external factor changes while ensuring inter-class separability. Recently, the academic community has proposed theories and methods such as sparse representation and deep learning as new tools for solving the robustness problem of biometric recognition.

Biometrics in complex environments includes both active and passive identity authentication. In the traditional active identity authentication mode, users need to cooperate actively and the image quality needs to be relatively good. Different from the active authentication, the identified objects in the passive identity authentication do not necessarily be willing to pass the authentication, and may even adopt resistance strategies, such as makeup and non-cooperation. In order to achieve passive identity authentication, we must study automatic biometrics in non-cooperative situations, and we must tackle key issues such as robust biometric identification.

In the case of face recognition, the accuracy rate of identity authentication currently exceeds 99% under controlled conditions. But in video surveillance scenarios, the accuracy rate quickly drops below 80%; in addition, the low-resolution facial images, incomplete fingerprints and palmprints pose a major challenge to the robustness of biometrics. Thus, there is still a long way to go for robust biometric recognition.

4.3 Real-Time Matching

With the declining cost of biometric imaging and the urgent need for governmental high-performance identity authentication technology, biometrics will be widely utilized. For example, biometrics has begun to be promoted in national and industry level applications such as electronic passports, ID cards, suspect investigations, identification of missing persons, banking, e-commerce, medical treatment, insurance, social welfare, etc. At this time, the scale of the biometrics template in the central database is bound to reach a huge amount (ten million or even hundreds of millions), and the cost of time to identification will be unbearable. This is one of the three major issues in the field of biometric recognition: the problem of scale. In complex application environments such as airports, stations, and docks, it is often necessary to screen certain preset special groups (such as those in the "blacklist" of the Ministry of Public Security and the Ministry of Security), which requires real-time efficiency large-scale biometric data retrieval technology breakthroughs.

4.4 Credibility and Security

A biometric system is a security system. Like other information security technologies, it may be attacked by various kinds of attacks, and threatened by hackers. Besides the forgery the biometric samples of others, many other possible attacks include: modifying the sample data on the communication link between the acquisition device and the server; modifying the recognition results; replacing the matching program; attacking the biometric template database, etc. The safety factor of the biometric system depends on the weakest part since each joint is equally essential. To make biometric technologies can be applied in the situation of high-security requirements, besides algorithm design, it is also very important to protect the security of the system itself and improve the resistance to various hacker attacks. In order to prevent malicious hackers from forging and stealing other's biometrics for identity authentication, the biometric system must have the functionality of living detection, that is, to determine whether the biometric images come from living individuals. For all kinds of potential attack methods against biometric systems in an open environment are unpredictable, it is very difficult to design a strict and foolproof security protection system for biometric systems. At present, the security of a biometric system has been highly valued by the academic and industrial communities. For example, Trusted Biometrics, the seventh framework research project of the EU, specializes in the security and anti-counterfeiting of biometric systems.

4.5 Coordinated Identification in Physical and Cyberspace

Cyberspace, as the expansion and extension of human society's living space in the era of information, is closely related to the traditional social space of human existence as a supplement. The rapid development of public transportation and information networks has made social individuals increasingly active in physical and cyberspace, placing huge challenges to public safety and social management. Therefore, how to realize coordinated identification in physical and cyberspace is an important problem with many challenges. For example, in recent years, Google and Facebook have acquired some face recognition companies and developed several biometric applications for the network environment, illustrating the development prospects of biometric recognition technologies. Basically speaking, it is a new research problem on how to make use of social network relationships, integrate the network audio and video information and the IoT monitoring biometric information to achieve a coordinated identification system across physical and cyberspace.

4.6 Evaluation and Certification

From a perspective of research, the performance evaluation of biometric recognition algorithms can follow the latest progress in academic research, explore the inherent potential of recognition algorithms, compare the pros and cons of different algorithms, and propose algorithm optimization solutions based on deficiencies, such as feature selection, parameter adjustment, etc., so as to push forward the research of biometric recognition algorithms. From the perspective of industry and applications, if there is no standard technical supervision and performance evaluation system in the field of biometric identification, the entire industry will be in a state of chaos and disorder, which will bring many fatal hidden dangers to the public and personal security systems.

In the face of the booming biometric recognition technology, countries around the world have set up specialized evaluation centers. For example, the United States has the National Institute of Standards and Technology NIST and the National Biometric Test Center relies on San Jose State University. In addition, the United States Department of Defense has established a Department of Defense & Federal Biometric System Protection Profile which conducts a rigorous evaluation of all biometric systems that enter the US military and federal government. The British government's information security technology authority CESG has proposed a Biometric Evaluation Methodology under the framework of the Common Criteria for general information security products. Korea established the National Biometric Test Center (K-NBTC), built a large-scale database, and carried out testing services for standards compliance, identification performance, and security performance.

In summary, there many problems with how to evaluate next-generation biometric technologies that are still unsettled, including evaluation data (how to quantify the representative, quality and scale of test data, etc.), evaluation models (how to statistically analyze test results and how to predict actual application performance based on test results, etc.) and evaluation software (how to build automatic, configurable, and integrated evaluation system).

5 Development Strategies on Biometric Technologies

Focus on the challenges encountered by biometric applications in complex environments, those many aspects like human–computer interaction, information acquisition, preprocessing, feature analysis, pattern matching, large-scale retrieval and comparison, multimodal information fusion, security, anti-counterfeiting, and application mode in biometrics need innovations on friendly interfaces, accurate identification, and reliable systems that can do real-time comparisons. The improved intelligence, automation, and informatization level of biometric identification system are required to achieve environments and users adaption, improve the user experience and satisfaction so as to complete the technical transition of biometrics from "humans cooperate with machines" to "machines actively adapt to people", realize the applied biometric technologies from a controlled environment to a complex environment, develop a biometric industrial cluster with independent intellectual property rights (including core chips, acquisition devices, recognition algorithms, and application systems), and create a large scale biometric technology and product certification system. Through technological breakthroughs of biometrics in complex environments, it will greatly expand the scope of biometrics in the real world, promote the revitalization of the biometrics industry, and increase the international market share of biometric identification products. Achievements in scientific research with independent intellectual property rights through collaborative innovation of industries, universities, and research institutions will be the core factor to enhance the biometric industries, to create domestic biometric brands, to satisfy the urgent public security needs.

5.1 Overall Solution for Large-Scale Biometric Applications

At present, many countries have determined to embed biometric information in the new generation of ID cards and e-passports. These ID cards are widely used in education, social security, finance, customs clearance, telecommunications, transportation, tourism and other fields by the supports of an inseparable national biometric infrastructure platform. Therefore, it is important and urgent to study architecture and key technologies of the ultra-large-scale biometric application system with independent intellectual property rights.

5.2 Biometrics' IntelliSense and Human–Computer Interaction

Focus on developing biometric imaging devices such as non-interfering face and iris image acquisition devices; non-contact fingerprint, palmprint, and vein acquisition devices at a medium-and-long distance; acquisition devices which can endure object motioning; and large depth of field acquisition devices based on the new photoelectric principles.

5.3 Robust Biometric Identification Method

The researches should focus on human face, gait, and iris recognition technologies in video surveillance scenes, incomplete fingerprints in the field of criminal investigations in public security, palmprint recognition technology, and multi-source heterogeneous biometric recognition methods.

5.4 Biometric System Security Protection

Construct a comprehensive biometric identification security protection system based on multispectral imaging, live detection, cryptography, information hiding and other technologies to solve the security problem of biometric identification.

5.5 Coordinated Biometrics Across Physical and Cyberspace

Focus on the researches of cross-camera identification technology in the large-scale monitoring scene of physical space, the identification technology based on social networks, and the identification technology of the fusion of physical space and cyberspace.

5.6 Biometric Evaluation and Certification Platform

The evaluation and certification of biometric performance are related to national security, industrial development, and academic progress. Therefore, it is necessary to study the standards and technical systems of biometric evaluation and certification, including not only biometric accuracy and robustness but also security tests. It also needs to establish an expert system that guides the optimization of algorithms based on the evaluation information and make better evaluations with better algorithms.

6 Conclusion

With the rapid development of artificial intelligence technologies, such as adversarial generative networks, automated machine learning (AutoML), neural network architecture search (NAS), etc., and emerging sensing devices, such as light-field cameras, 3D imaging, multispectral imaging, etc., biometric technologies and its applications could have a broad developing capacity. Biological features such as faces, irises, fingerprints, veins, voiceprints, and gait in the era of artificial intelligence are more convenient to collect and can be identified more accurately and safely in wider applications. Such biological features now have become the identity token for people to go through security checks, customs clearance, door opening, withdrawal, payment, social insurance, examination, medical treatment, and attendance in physical and cyberspace. New technologies, applications, and industries of biometrics have become new driving forces for social progress and human civilization.

References

1. Alibaba DAMO Academy (2019) The top ten technology trends in 2019 by Alibaba DAMO Academy [EB/OL]. https://damo.alibaba.com/events/50
2. The Central People's Government of the People's Republic of China (2015) Guiding opinions of the state council on vigorously advancing the "internet plus" action [EB/OL]. https://www.gov.cn/zhengce/content/2015-07/04/content_10002.htm
3. The Central People's Government of the People's Republic of China (2017) Notice of the state council on issuing the development plan on the new generation of artificial intelligence [EB/OL]. https://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm
4. Turk MA, Pentland AP (1991) Face recognition using eigenfaces. IEEE, The United States
5. Belhumeur PN, Hespanha JP, Kriegman DJ (1997) Eigenfaces versus fisherfaces: recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell 19(7):711–720.
6. Taigman Y, Yang M, Ranzato M, Wolf L (2014) Deepface: closing the gap to human-level performance in face verification. IEEE, The United States
7. Schroff F, Kalenichenko D, Philbin J (2015) Facenet: a unified embedding for face recognition and clustering. IEEE, The United States
8. Wen Y, Zhang K, Li Z, Qiao Y (2016) A discriminative feature learning approach for deep face recognition. Springer, Switzerland
9. Liu W, Wen Y, Yu Z, Yang M (2016) Large-margin softmax loss for convolutional neural networks. PMLR.
10. Liu W, Wen Y, Yu Z, Li M, Raj B, Song L (2017) Sphereface: deep hypersphere embedding for face recognition. IEEE, The United States
11. Wang H, Wang Y, Zhou Z, Ji X, Gong D, Zhou J, Li Z, Liu W (2018) Cosface: large margin cosine loss for deep face recognition. IEEE, The United States
12. Deng J, Guo J, Xue N, Zafeiriou S (2019) Arcface: additive angular margin loss for deep face recognition. IEEE, The United States

13. Kemelmacher-Shlizerman I, Seitz SM, Miller D, Brossard E (2016) The megaface benchmark: 1 million faces for recognition at scale. IEEE, The United States
14. National Institute of Standards and Technology (NIST) (2020) Face Recognition Vendor Test (FRVT) [EB/OL]. https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt
15. Huang R, Zhang S, Li T, He R (2017) Beyond face rotation: global and local perception gan for photorealistic and identity preserving frontal view synthesis. IEEE, The United States
16. Liu Y, Li Q, Sun Z (2019) Attribute-aware face aging with wavelet-based generative adversarial networks. IEEE, The United States
17. Daugman J (1993) High confidence visual recognition of persons by a test of statistical independence. IEEE Trans Pattern Anal Mach Intell 15:1148–1161
18. He Z, Tan T, Sun Z, Qiu X (2009) Toward accurate and fast iris segmentation for iris biometrics. IEEE Trans Pattern Anal Mach Intell 31(9):1670–1684
19. Proenca H (2010) Iris recognition: on the segmentation of degraded images acquired in the visible wavelength. IEEE Trans Pattern Anal Mach Intell 32(8):1502–1516
20. Tan CW, Kumar A (2012) Unified framework for automated iris segmentation using distantly acquired face images. IEEE Trans Image Process 21(9):4068–4079
21. Ma L, Tan T, Wang Y, Zhang D (2003) Personal identification based on iris texture analysis. IEEE Trans Pattern Anal Mach Intell 15(11):1148–1161
22. Thornton J, Savvides M, Kumar V (2007) A bayesian approach to deformed pattern matching of iris images. IEEE Trans Pattern Anal Mach Intell 29(4):596–606
23. Sun Z, Tan T (2009) Ordinal measures for iris recognition. IEEE Trans Pattern Anal Mach Intell 31(12):2211–2226
24. Liu N, Zhang M, Li H, Sun Z, Tan T (2016) DeepIris: learning pairwise filter bank for heterogeneous iris verification. Pattern Recogn Lett 82(2):154–161
25. Sun Z, Zhang H, Tan T, Wang J (2014) Iris image classification based on hierarchical visual codebook. IEEE Trans Pattern Anal Mach Intell 36(6):1120–1133
26. Liu N, Liu J, Sun Z, Tan T (2017) A Code-level approach to heterogeneous iris recognition. IEEE Trans Inf Forensics Secur 12(10):2373–2386
27. Sun Z, Wang L, Tan T (2014) Ordinal feature selection for iris and palmprint recognition. IEEE Trans Image Process 23(9):3922–3934
28. Liu N, Li H, Zhang M, Liu J, Sun Z, Tan T (2016) Accurate iris segmentation in non-cooperative environments using fully convolutional networks. IEEE, The United States
29. Zhang Qi, Li H, Sun Z, Tan T (2018) Deep feature fusion for iris and periocular biometrics on mobile devices. IEEE Trans Inf Forensics Secur 13(11):2897–2912
30. Marco T, Guerrieri R (1997) A 390 dpi Live Fingerprint Imager based on Feedback Capacitive Sensing scheme. IEEE, The United States
31. Galbally J, Bostrom G, Beslay L (2017) Full 3D touchless fingerprint recognition: sensor, database and baseline performance. IEEE, The United States
32. Lawrence O, Nickerson JV (1988) Matched filter design for fingerprint image enhancement. IEEE, The United States
33. Dabouei A, Kazemi H, Iranmanesh S, Dawson J, Nasrabadi N (2018) Fingerprint distortion rectification using deep convolutional neural networks. IEEE, The United States
34. Roy A, Memon N, Ross A (2017) Masterprint: exploring the vulnerability of partial fingerprint-based authentication systems. IEEE Trans Inf Forensics Secur 12(9):2013–2025
35. Komeili M, Armanfard N, Hatzinakos D (2018) Liveness detection and automatic template updating using fusion of ECG and fingerprint. IEEE Trans Inf Forensics Secur 13(7):1810–1822
36. Park E, Cui X, Nguyen T, Kim H (2019) Presentation attack detection using a tiny fully convolutional network. IEEE Trans Inf Forensics Secur 14(11):3016–3025
37. Boyd JE, Little JJ (2005) Biometric gait recognition. Springer, Switzerland
38. Niyogi SA, Adelson EH (1994) Analyzing and recognizing walking figures in xyt. IEEE, The United States
39. Little J, Boyd J (1998) Recognizing people by their gait: the shape of motion. Videre: J Comput Vis Rese 1(2):1–32.

40. Huang P, Harris C, Nixon M (1999) recognizing humans by gait via parametric canonical space. Artif Intell Eng 13(4):359–366
41. Wang L, Tan T, Hu W, Ning H (2003) Automatic gait recognition based on statistical shape analysis. IEEE Trans Image Process 12(9):1120–1131
42. Cuntoor N, Kale A, Chellappa R (2003) Combining multiple evidences for gait recognition. IEEE, The United States
43. Han J, Bhanu B (2006) Individual recognition using gait energy image. IEEE Trans Pattern Anal Mach Intell 28(2):316–322
44. Lam TH, Cheung KH, Liu JN (2011) Gait flow image: a silhouette-based gait representation for human identification. Pattern Recogn 44(4):973–987
45. Kastaniotis D, Theodorakopoulos I, Theoharatos C, Economou G, Fotopoulos S (2015) A Framework for gait-based recognition using kinect. Pattern Recogn Lett 68(2):327–335
46. Wu Z, Huang Y, Wang L, Wang X, Tan T (2017) A comprehensive study on cross-view gait based human identification with deep CNNs. IEEE Trans Pattern Anal Mach Intell 39(2):209–226
47. Wu Z, Huang Y, Wang L (2015) Learning representative deep features for image set analysis. IEEE Trans Multimedia 17(11):1960–1968
48. Zhou Y, Kumar A (2011) Human identification using palm-vein images. IEEE Trans Inf Forensics Secur 6(4):1259–1274.
49. Kumar A, Ravikanth C (2009) Personal authentication using finger knuckle surface. IEEE Trans Inf Forensics Secur 4(1):98–110
50. Yan P, Bowyer KW (2007) Biometric recognition using 3D ear shape. IEEE Trans Pattern Anal Mach Intell 29(8):1297–1308
51. Marcel S, Millan JDR (2007) Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation. IEEE Trans Pattern Anal Mach Intell 29(4):743–752
52. Rigas I, Komogortsev OV (2014) Biometric recognition via probabilistic spatial projection of eye movement trajectories in dynamic visual environments. IEEE Trans Inf Forensics Secur 9(10):1743–1754
53. Frank M, Biedert R et al (2013) Touchalytics: on the applicability of touchscreen input as a behavioral biometric for continuous authentication. IEEE Trans Inf Forensics Secur 8(1):136–148
54. http://www.xinhuanet.com/fortune/2019-01/02/c_1123938779.htm

Zhenan Sun received the B.E. degree in industrial automation from the Dalian University of Technology, China, in 1999, the M.S. degree in system engineering from the Huazhong University of Science and Technology, China, in 2002, and the Ph.D. degree in pattern recognition and intelligent systems from the Institute of Automation, Chinese Academy of Sciences, China, in 2006. He is currently a Professor with the Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, and also with the School of Artificial Intelligence, University of Chinese Academy of Sciences, China. He has authored/coauthored over 200 technical papers. His research interests include biometrics, pattern recognition, and computer vision. He is the Chair of Technical Committee on Biometrics, International Association for Pattern Recognition (IAPR) and an IAPR fellow. He serves as an Associate Editor for the IEEE Transactions on Biometrics, Behavior, and Identity Science.

Progress of e-Science in Major Projects

Big Data Promotes the Tibetan Plateau and Pan-Third Pole Earth System Science

Xin Li, Xiaoduo Pan, Xuejun Guo, Jun Qin, Baosheng An, Tao Wang, Qinghua Ye, Weimin Wang, Xiaojuan Yang, Xiaolei Niu, Min Feng, Tao Che, Rui Jin, and Jianwen Guo

Abstract The Pan-Third Pole includes the Tibetan Plateau and the northern intra-continental arid region of Asia, extending to the Caucasus Mountains in the west and the western Loess Plateau in the east. This region covers 20 million square kilometers and affects the environment inhabited by three billion people. Two special projects have been implemented to provide important scientific support for eco-environmental refining and sustainable economic and social development of the Pan-Third Pole region, with the Tibetan Plateau as its core: the Second Tibetan Plateau Scientific Expedition Program (a national special project) and the Pan-Third Pole Environmental Change and Construction of the Green Silk Road (hereinafter referred to as the Silk Road and Environment), a strategic pilot science and technology project (Category A) of the Chinese Academy of Science. The Pan-Third Pole big data system is an important data support platform for these two major research programs and has several purposes: the storage, management, analysis, mining and sharing of scientific data for various disciplines, such as resources, the environment, ecology and atmospheric science of the Pan-Third Pole; preparation of key scientific data products of the Pan-Third Pole; the gradual development of functions such as online big data analysis and model application; the construction of a cloud-based platform to integrate data, methods, models; and services for Pan-Third Pole research and promote application of big data technology in scientific research in the region. This

X. Li (✉) · X. Pan · X. Guo · J. Qin · X. Yang · X. Niu · M. Feng
National Tibetan Plateau Data Center, Key Laboratory of Tibetan Environment Changes and Land Surface Processes, Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China
e-mail: xinli@itpcas.ac.cn

X. Li · B. An · T. Wang · T. Che · R. Jin
CAS Center for Excellence in Tibetan Plateau Earth Sciences, Chinese Academy of Sciences, Beijing, China

B. An · T. Wang · Q. Ye · W. Wang
Institute of Tibetan Plateau Research, Chinese Academy of Science, Beijing, China

T. Che · R. Jin · J. Guo
Heihe Remote Sensing Experimental Research Station, Key Laboratory of Remote Sensing of Gansu Province,Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou, Gansu, China

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_7

paper demonstrates in detail various aspects of the Pan-Third Pole big data system, including the system architecture, data resource integration and big data analysis methods. The system improves big data processing capability in geoscience, serves as a new paradigm of geoscience research driven by big data and facilitates scientific research of the Earth system of the Pan-Third Pole.

Keywords Tibetan Plateau · Pan-Third Pole · Data integration · Big data analysis

1 Introduction

The Tibetan Plateau, with the highest elevation in the world, is known as the Third Pole of the Earth. Multiple major rivers in Asia originate from the Third Pole, which is also referred to as Asia's water tower. The Third Pole is considered one of the regions that have been subjected to the highest warming and environmental variation over the last 50 years. This region is a typical region interacting with the global climate system at multiple levels, and it is also a key region sensitive to global climate and environmental change. It plays a vital role in the global circulation of energy and water and has important influences on the global and regional climate [1, 2]. The Third Pole features a unique and yet vulnerable eco-environment, and the inhabitant species and ecosystem are highly sensitive to climate change. Consequently, the Second Tibetan Plateau Scientific Expedition Program (hereinafter shorted as STEP-2) was initiated in 2017. This program investigates variations in glaciers, the environment, lake and hydro meteorology, changes in biology and the ecosystem, the paleoecosystem and the paleoenvironment. The program probes the regularity of these variations, forecasts change scenarios and proposes countermeasure strategies while satisfying requirements for societal development in Tibet and the Belt and Road Initiative. This comprehensive scientific expedition is driven by dual considerations, the needs of the regional development and the demands of the scientific research frontier. Development in the Belt and Road Initiative has resulted in environmental changes that have attracted worldwide attention [3]. This 20-million-square-kilometer fan-shaped region has an extremely vulnerable ecological environment and intensive human activity. Moving westward, with the Third Pole as its center, this region covers the Tibetan Plateau, the Pamirs, the Hindu Kush, the Tianshan Mountains, the Iranian Plateau, the Caucasus and the Carpathian Mountains. Sustaining the resources of the Pan-Third Pole will provide critical scientific and technical support for the Belt and Road Initiative [4, 5].

A big data environmental system is being constructed for the Pan-Third Pole to solve regional environmental problems and enhance scientific research, thereby promoting the Belt and Road Initiative. This system performs strategically necessary functions for regional development: comprehensive summarization, long-term preservation, integrated management and sharing of national-level important scientific and technical data resources, and solution of major issues in economic & social development and national security.

The existing data of Pan third polar cryosphere, lake, ecology, hydrology, atmosphere, solid earth, etc. are characterized by massive but fragmentary, scattered and incomplete space–time coverage. On the one hand, a lot of resources are wasted, and different researchers need to repeatedly collect and preprocess the data when conducting research. On the other hand, they cannot fully and accurately be used to understand the integration of Pan third polar data Physical condition, did not play out the great potential. Until now, there is no scientific research big data platform for the Pan-Third Pole in either China or other countries. Such a platform for the Pan-Third Pole environment is urgently required for in-depth integration and rapid sharing of data to increase data utilization and scientific research efficiency. We develop a new big data platform and corresponding big data analysis methods based the development concept and experience of scientific resource integration and sharing of geoscience-related scientific data centers around the world [6–11]. A novel research paradigm is provided for scientific research of the Pan-Third Pole, and an example of big data-based innovation in geoscience is demonstrated. The platform can serve as a data support platform for STEP-2 and the Pan-Third Pole Environmental Change and Construction of the Green Silk Road project (a stragetic pilot project of the Chinese Academy of Science (category A), hereinafter referred to as the Silk Road and Environment), and the teleconnection analysis of global cryosphere as well.

2 Architecture of Pan-Third Pole Big Data System

A big data platform for geoscience research of the Pan-Third Pole has been constructed by integrating multisource heterogeneous data and combining big data mining and geoscience models. The platform is mainly used in five research fields: glaciers, lakes, ecological simulations, seismic monitoring and sustainable development. Efforts are made to explore and summarize big data driven new approaches for geoscience studies (Fig. 1).

2.1 Data Management and Sharing Platform

The big data system stores, manages, analyzes, mines and releases scientific data on the resources, environment, ecology, atmosphere, etc., of the Pan-Third Pole. Functions such as the online big data analysis and model application are gradually developed and enabled to integrate extensive data, methods, models and services for Pan-Third Pole science (https://data.tpdc.ac.cn/en/) (Fig. 2).

This system incorporates the submission, review and publishing of data resources in various forms, such as data uploaded by noninstitutional users, data from collaborative programs and institutions, journal data and platform data mining products. The system allows for dataset storage, quality control and the sharing of observation

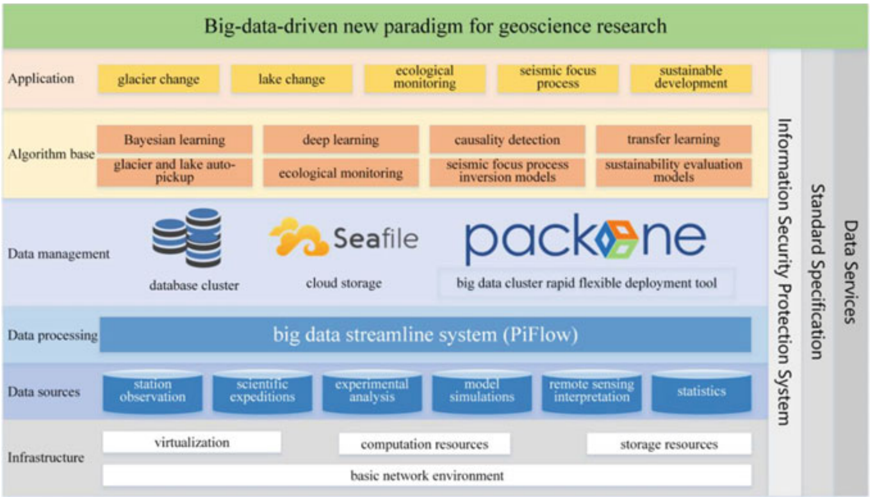


Fig. 1 Overall system architecture

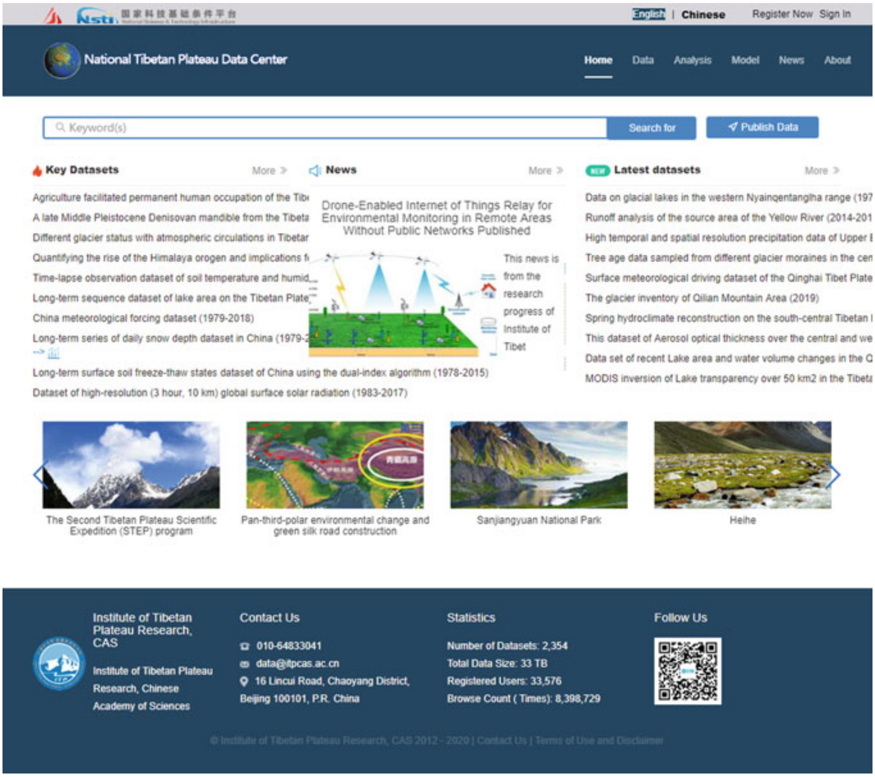


Fig. 2 User interface of Pan-Third Pole big data system

and measurement data from field stations, hierarchical data sharing and online data analysis, mining and visualization.

2.2 Construction of Information Infrastructure

Virtualization technology is used to realize dynamic allocation, flexible distribution and interterritory sharing of hardware resources, which increases the utilization efficiency of IT resources considerably. The Internet, storage and computation resources provided by China Science and Technology Cloud are fully exploited to develop applications such as cloud storage and online archives to facilitate the accumulation and sharing of scientific data. A PB-level data-intensive data storage and processing facility is constructed and embedded with big data processing software and tools (Ambari, PackOne, PiFlow, etc.). An environment is created to demonstrate big data applications for the cryosphere, hydrology and ecology, solid Earth science and regional sustainable development (Fig. 1).

2.3 Standardized System Regulation

The National Tibet Plateau Data Center (hereinafter referred to as the Data Center) is the interface of Pan-Third Pole big data system. The objectives of the Data Center are as follows: to fully implement the *Notice of the General Office of the State Council (of China) on Regulations of Scientific Data Managemen*t (GBF (2018) No. 17) and the "*Notice of the Chinese Academy of Science on Printing and Distributing "Regulations of Scientific Data Management and Public Sharing of the Chinese Academy of Science (Tentative)*" (KFB (2019) No. 11); to further strengthen and regularize scientific data management; to ensure scientific data security; to increase public sharing; and to sustain management of the Data Center. To fulfill these objectives, the Data Center has developed standardized specifications for administration and operation, data management, metadata standards, scientific data management, data-sharing services and IT and data security. Each process for data acquisition, integration and submission, storage, sharing and publishing has been regularized and standardized to ensure data security and protect the rights and interests of the data producer (Fig. 3).

A *Discipline-based Content Standard of the National Tibet Plateau Data Center* is under development. This standard is based on 102 national and sector standards and 35 research papers from 15 disciplines, including glaciers and permafrost, the paleoenvironment, geology, hydrology, soil, the atmosphere, atmospheric chemistry, ecology, remote sensing, bio-diversity and disasters.

Taking the discipline term of glacier as an example, the data content standard includes the following 10 sub-discipline: material balances, glacier terminus changes, a glacier inventory, black carbon content, isotopic oxygen and aerosol concentrations,

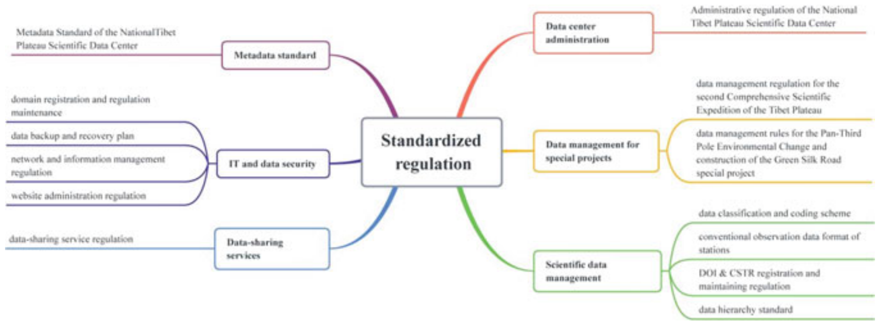


Fig. 3 Standardized system regulation of the National Tibet Plateau Scientific Data Center

glacier surface movement, glacier temperatures, glacier thicknesses, glacier meteorology and glacier streamflow. Each subclass has several indicators. The content and value standard of each indicator are described in detail by Chinese and English names and units, precision, granularity and data value ranges. Specific provisions in these standards provide thorough references for the Data Center for collection, processing, quality assessment and calibration of glacier data.

3 Pan-Third Pole Data Resources

3.1 Keywords of the Data Classification System for the Pan-Third Pole Scientific Data

A new list of keywords is created from the "Keywords of the Data Classification System for the Pan-Third Pole Scientific Data", the Third Pole data catalog and the "First-Level Discipline Orientation Classification and Keywords of Geoscience (Tentative Version, 2012)" of the National Natural Science Foundation of China. This list includes 11 1st-level classes, 62 2nd-level classes and 702 keywords. The 1st-level class contains the cryosphere, hydrology, soil science, the atmosphere, the biosphere, geology, paleoclimate/paleoenvironment, human dimensions/natural resources, disaster, remote sensing and basic geography.

3.2 Combination of Pan-Third Pole Data

The existing 148 scientific datasets of the Data Center are combined into over 2300 scientific datasets: STEP-2, the satellite-aerial-remote sensing-ground station integrated ecological monitoring and data platform of the Three-River-Source National Park, the Cold and Arid Region Scientific Data Center [5] and extensive research

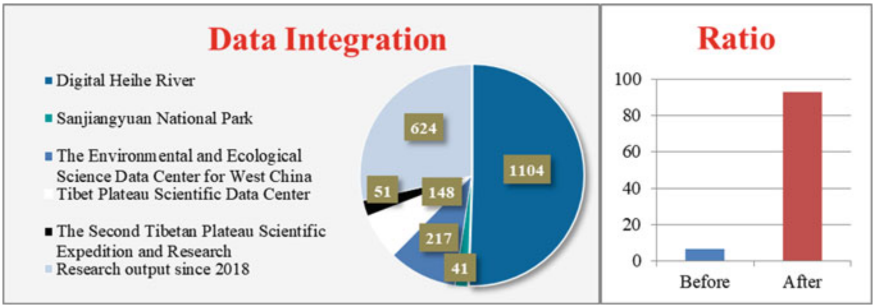


Fig. 4 Sources and quantities of integrated data resources for Pan-Third Pole

results (Fig. 4). The total data volume reaches 33 TB, and the relational data contain over 1.4 billion rows.

3.3 Integration of Pan-Third Pole Data

The metadata of the Tibetan Plateau are managed in accordance with ISO 19115. The cloud-based Pan-Third Pole environment big data system has been constructed to achieve multilevel, multiterminal and multilingual cloud sharing. The datasets originate from various platforms and are often multisource and heterogeneous. Thus, there is an urgent need both for data integration, to eliminate information redundancy among multisource data and to identify interactions between various elements in different layers of the Pan-Third Pole area affected by global warming, and for teleconnections with other areas. The overarching principles of data integration are to input heterogeneous spatial data of various classes into the dataset in a unified manner, implement strict quality control, provide complete metadata and data documents and achieve data sharing. The time sensitivity, completeness, principles and logicality of data are evaluated. In addition to conventional quality control methods such as the manual method, the metadata method and the geographic correlation method, considerable effort has been expended to introduce big data cleansing (quality control). Appropriate techniques such as mathematical statistics, data mining and predefined cleansing criteria are used to clean dirty data. The resulting data are transformed into high-quality data by computer-based automatic extract, transform and load (ETL) processing. Hence, the reliability (accuracy, integrity, consistency, effectiveness and uniqueness) and availability (timeliness, accessibility and satisfaction degree) of the Pan-Third Pole data resources are enhanced.

The quality of the metadata and the data are ensured by an online bilingual data submission system and a semi-intellectual review system. The submission system operates under Chinese-English linking, and drop-down lists for selection are used to the greatest extent possible to prevent manual typing errors. For example, three-level Chinese-English linked drop-down lists are used for discipline keywords,

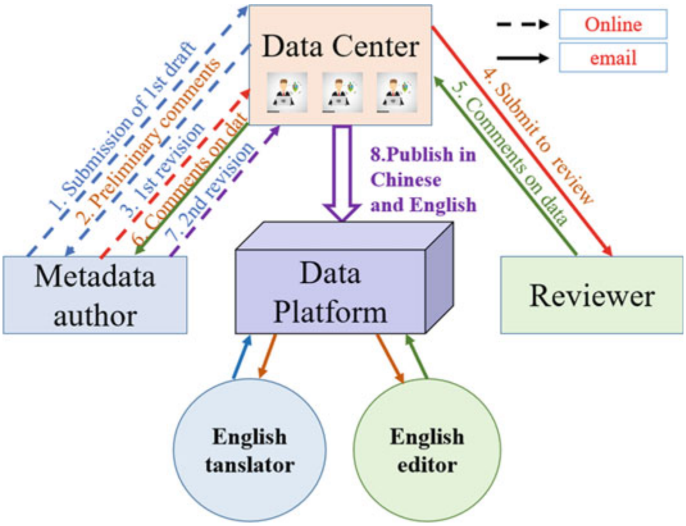


Fig. 5 Workflow of bilingual (Chinese-English editing) and metadata review in the Pan-Third Pole system (Solid lines represent workflows via email, and dashed lines represent workflows that are executed online)

topic keywords and others, and Chinese-English linked drop-down lists are used for temporal and spatial resolution. The semi-intellectual review system of the Pan-Third Pole big data system mainly manifests in the workflow when the system automatically emails a preliminary review notice to data service personnel upon uploading new data. The system invites expert reviewers according to discipline keywords via data review emails: the data are instantaneously open for sharing as soon as the reviewers recommend to publish the data; and if the reviewers require revisions, metadata producers are automatically notified. The workflows for English translation/editing of metadata and the semi-intellectual metadata review are presented in Fig. 5.

3.4 Key Datasets for Pan-Third Pole

The key datasets for Pan-Third Pole are developed from the enormous quantity of multisource and heterogeneous data of various classes and different disciplines through investigating, arranging, reconstructing and merging by using the same dataset system and geographic information system (GIS) platform in accordance with mechanisms related to data collection and submission, management, quality control, sharing and updating. The resulting key data resource of the Pan-Third Pole has considerable significance and high research value (Table 1).

Table 1 Classification of Pan-Third Pole core datasets

English Names of Key Datasets	Principles
Calibration and verification key datasets over Tibetan Plateau	Elements: meteorology, soil temperature and humidity, hydrology and eddy correlation (EC), based on principles of unified time span and time granularity (year, month, day and hour)
Cryospheric key datasets over Tibetan Plateau	Cryosphere elements: glaciers, glacier lakes, permafrost and snow cover over a unified area using a unified projection system
Basic geographic key datasets over Tibetan Plateau	Unified projection system, same spatial resolution and possibly synchronized data preparation
Near-surface atmospheric forcing datasets over Tibetan Plateau	Providing a group of near-surface meteorologically driven datasets with reliable quality and long time span
Scientific discovery key datasets over Tibetan Plateau	Mainly composed of scientific research findings

The permafrost thermal condition distribution map of the Tibetan Plateau (2000–2010) is considered as an example. The geographic-weighing regression model is used to integrate the following data: reconstructed temporal-spatial data, the moderate resolution imaging spectroradiometer (MODIS) surface temperature, the leaf area index, the snow cover ratio and the multimodel soil moisture prediction product of the National Meteorological Information Center, China. The map incorporates precipitation measurements from over 40,000 meteorological stations, the precipitation measurement product of Satellite FY2 and the average atmospheric temperature data of 152 meteorological stations from 2000 to 2010. The abovementioned data are used to simulate the long-term average atmospheric temperature data of the Tibetan Plateau with a spatial resolution of one kilometer. The thermal conditions of the permafrost classification system is used to identify the permafrost as very cold, cold, warm, very warm and likely to thaw. After deducting the areas of the lakes and glaciers, the total area of permafrost in the Tibetan Plateau is approximately 1.0719 million square kilometers, which demonstrates the high accuracy of the developed map and its ability to support planning and design of permafrost projects and environmental management. These results have been published in *The Cryosphere* [12].

An ecological and hydrological wireless sensor network in the midstream Yingke/Damanguan area of the Heihe River Valley is used to collect continuous observations from 50 WATERNET nodes over a 5.5 km * 5.5 km survey matrix. The resulting WATERNET observation and measurement dataset includes the soil moisture, the soil temperature, the conductivity, the complex permittivity, the surface infrared radiation temperature and the atmospheric infrared radiation temperature. The dataset has temporal and spatial continuity and can be used in several research areas: the remote sensing estimation of key water and thermal variables, remote sensing validation for a heterogeneous surface, ecological and hydrological studies and irrigation optimization. Relevant results have been published in Scientific Data [13–15].

3.5 Data Sharing: Principle and Mode

The Pan-Third Pole big data system follows the FAIR (findability, accessibility, interoperability and reusability) principle for data sharing [5]. The system offers user access to bilingual (Chinese-English) data with primary online and secondary offline services (Fig. 6). Both the online and offline sharing approaches guarantee the following data rights and interests: 1) a unique data identification code, that is, a digital object identifier (DOI); 2) data distribution protocol, the default being Creative Commons Attribution 4.0 International (CC BY 4.0), which retains author's copyright; 3) literature citations: one or two publications are suggested to be cited for use of the data; 4) data citations: users are encouraged to cite the shared data via DOI; and 5) feedback on the status of data browsing, downloading and citation is regularly sent to the data producer. The online and offline data-sharing approaches are in different procedure. The online data can be conveniently directly downloaded by users, obviating notification emails to the data producer, whereas the offline data allows data producers to easily specify data rights and interests in addition to those listed above. The offline data service, under the premise of securing the exclusive rights and interests of the data producers, creates an automatic interactive system between data users and producers to increase sharing efficiency. The bilingual service is an improved approach that highlights both Chinese and English contexts and supplies high-quality metadata and data entities in English. The following types of service are provided: regular service, specialized service, typical case service, push notification service and information service (Fig. 7). Moreover, the Pan-Third Pole big data system will develop a mobile platform APP, a website for mobile phones, a

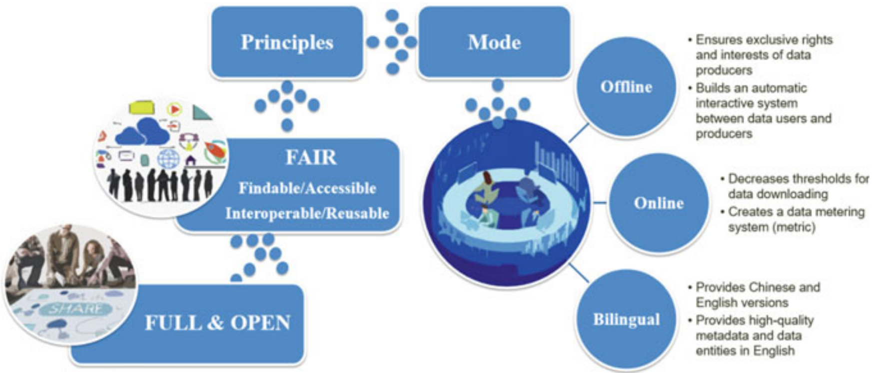


Fig. 6 Schematic showing sharing principles and mode (Pan et al., 2020, in review)

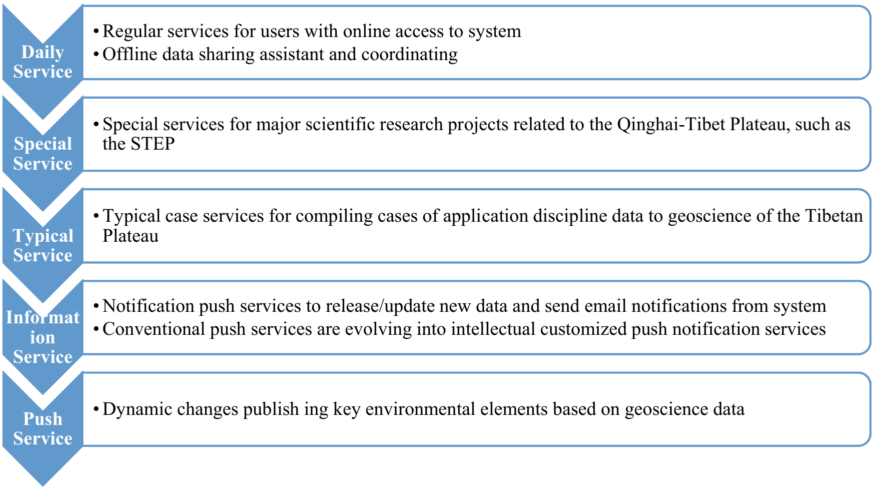


Fig. 7 Service mode of the Pan-Third Pole big data system

WeChat official account and a Weibo account. A newsletter will be regularly issued to promote the influence of the Pan-Third Pole big data system.

4 Intellectual Property and Data Publication

Big data has produced numerous changes in the information age. Increasing attention is being paid to intellectual property protection to provide security and sustainability for the big data industry. Data are products of both physical labor and intellectual activity. Therefore, data sharing requires intellectual property to protect the rights and interests of data producers and make data sharing sustainable. The Pan-Third Pole Data Center adopts multiple methods of securing data intellectual property rights.

4.1 Data DOI

The DOI is a unique system that was developed to address multiple links and the copyright transfer of digital resources over the Internet. The DOI system is used for Pan-Third Pole data resource development. This system can be applied to data sharing and enables tracking, citation, integration and networking. The DOI format of the Pan-Third Pole big data system is shown below:

10.11888/category.tpdc.metadataID.

where 10.11888 is a fixed prefix; 11888 is the registration number of the Research Institute of Tibet Plateau Research, the Chinese Academy of Science; category denotes the data type; tpdc is a fixed term; and metadataID denotes a six-digit data sequence number.

4.2 Data Distribution Protocol

The Pan-Third Pole big data system adopts Creative Commons 4.0 to retain author copyright, while allowing users to reprint, use and deduce data within the limited scope of an agreement and without informing the author. Six options are provided for data producers in the data submission system: CC BY 4.0 (Attribution), the default; CC BY-NC 4.0 (Attribution-NonCommercial); CC BY-ND 4.0 (Attribution-NoDerivatives); CC BY-NC-ND 4.0 (Attribution-NonCommercial-NoDerivatives); CC BY-SA 4.0 (Attribution-ShareAlike); and CC BY-NC-SA 4.0 (Attribution-NonCommercial-ShareAlike).

4.3 Literature Citation of Original Data

The submission module of the Pan-Third Pole big data system contains functions to create a data-related literature information entry and upload literature for reference and citation. Data users are thereby provided with a research background, data preparation, a processing method, quality evaluation and data application. The system provides the following notification to promote a benign academic atmosphere for scientific data sharing: "To use this data, you are suggested to reference the articles listed in the Required Data Citation section".

4.4 Data Citation

Data citation is a new concept that was proposed in the global publishing industry and data-sharing sector. Data are treated like article references and listed in the references section of a paper. Data citation enables (1) data usage to be tracked, (2) data services to be counted and (3) protects the intellectual property rights of data. The data citation format is typically provided by the Data Center as "Data producer list. Data title. Publishing/release organization of data, date of data publishing/release. Permanent address of data DOI." in the Pan-Third Pole big data system.

4.5 Data Protection Period

The *Notice of the General Office of the State Council (of China) on Regulations of Scientific Data Managemen*t (GBF (2018) No. 17) requires that all the scientific data in projects of science and technology programs (special projects, funds, etc.) at all government budget funding levels be submitted to the appropriate scientific data center by the leading organization of the project. In the interest of timeliness, it is recommended that data be submitted every year in accordance with the project tasks. Different data protection periods are set in the Pan-Third Pole big data system according to data acquisition method: (1) for data acquired in real time via the Internet of Things, quasi-real-time sharing is implemented; (2) for basic and automatic station data, no data protection period is set, in principle; (3) for incremental data generated in scientific research and field data acquired manually at elevations less than 4,000 m, a data protection period of no more than one year is planned, and direct sharing is encouraged; and (4) for field data acquired manually at elevations no less than 4,000 m, the set protection period should not exceed two years, and direct sharing is also encouraged.

4.6 Data Publication and Data Repository for Research Paper

The Pan-Third Pole big data system collaborates with prestigious journals to promote publication of data for the Tibetan Plateau and would like to increase the number of scientists sharing data. Considerable effort is being expended to ensure that the system will become a data repository certified by major international data publications such as *Scientific Data* and AGU journals and recommended by Earth System Science Data (ESSD). Researchers are being encouraged to publish and share their latest research results and related original data. Thus far, the Pan-Third Pole big data system has satisfied all the criteria for a data repository of major international data publications. A DOI and data-sharing system has been created to facilitate scientific data sharing and peer review of standardized metadata. Once the system becomes a data repository recommended by major international journals, original study data sharing can be promoted, data sources can be expanded, and the analysis of Pan-Third Pole scientific big data will be stimulated.

4.7 Permission of User Authority

The limited publication range and copyright of a dataset requires that user access be controlled under some conditions. After logging in, users are free to access datasets within the scope of authorization (including the authority to browse metadata and download specific data).

A user may be assigned multiple roles, each of which may have a customized data access range (that is, a mapping relation between the dataset and the user role). For example, special project users for the Silk Road environment project and STEP-2 program are authorized differently to control the dataset access range.

Control of user authority creates and maintains a correlation between roles/users and datasets rights and can be classified into role management and membership management. The role management module permits the viewing and downloading of metadata and datasets. The membership module creates membership, edits member information, reviews membership and manages member authority. In member authority management, a member can be assigned a role or provided customized access authority to specific datasets.

5 Big Data Analysis for Pan-Third Pole

5.1 Method Library Framework

Scientific data for the Pan-Third Pole region are characterized by high uncertainty and dimensionality. Various types of data result from multiple sources of observations and models and from the particularity and complexity of the Pan-Third Pole environment. The characteristics of big data are becoming increasingly clear. In geoscience, both big data-based information mining methods and spatial–temporal visualization methods are still in development. Many issues need to be resolved urgently. For instance, it remains unclear how to accurately analyze the overall change trend, detailed variation characteristics and the temporal evolution pattern of the Pan-Third Pole environment using geological big data that are characterized by different structures, decentralized storage and enormous volumes. It is also still difficult to maintain a global perspective on the interactions and synergy among multiple elements of the Pan-Third Pole environment while supporting the joint analysis of different temporal and spatial scales, which in turn makes mining, analysis and cognitive discovery of big data in the Pan-Third Pole environment challenging. The big data analysis system uses incremental integration and independent research and development to construct a method library for the Pan-Third Pole environment. The library comprises big data quality control, automatic modeling and analysis, data mining and interactive visualization. A tool library with high reliability, expandability, efficiency and fault tolerance is established in the meantime. Thus, collaborative analysis methods can be integrated and shared for big Pan-Third Pole environmental data with multi-source heterogeneity, multigranularity, multitemporal phases and long time series, thereby facilitating efficient and online big data analysis and processing. The demonstration and application of big data for critical surface processes in the Pan-Third Pole environment opens an overarching technical link for deep data mining. The six major categories of big data analysis in the method library are as follows: advanced

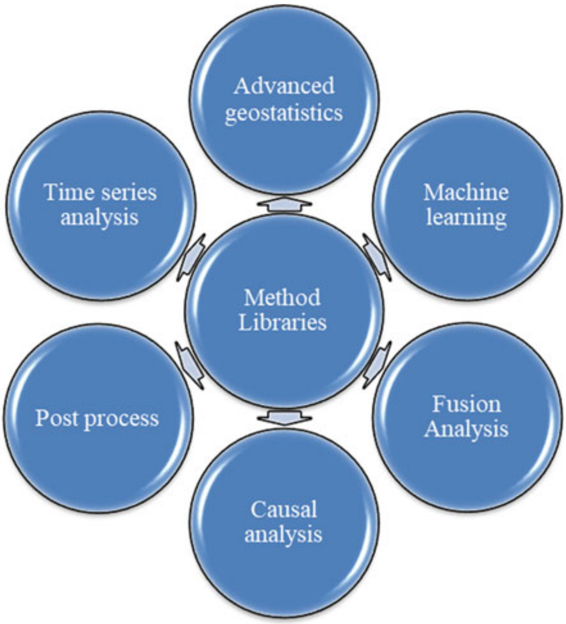


Fig. 8 Method library for big data analysis

geostatistics, time series analysis, traditional machine learning methods, postprocessing, model-observation fusion and causal analysis (Fig. 8). A code-sharing mechanism is established by using meta information to manage and perform intelligent searches/recommendations of these methods. The codes are hosted on GitHub.

5.2 Cases for the Use of Big Data Methods in Geological Research on the Pan-Third Pole Environment

The advent of the big data era has provided new opportunities for understanding and appropriately addressing water, ecology and environmental issues in the Pan-Third Pole. Big data science and technology is the product of a brand new revolution in scientific methods based on empirical, deductive and digital computing and has successfully predicted and analyzed the behavior of complex systems in many case studies. This technology is expected to dramatically change research in Earth system science. Multisource big data (such as from observations, remote sensing and simulations) can be combined using big analysis methods for use in automatic boundary extraction, inheritance of ecological network data and models and big data-driven

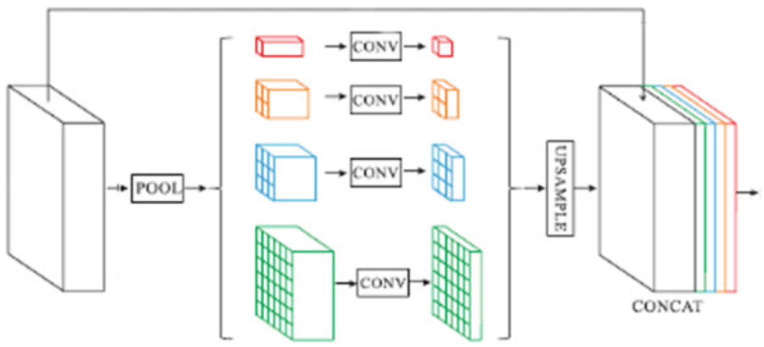


Fig. 9 Multiscale nested deep learning network

imaging of rupturing processes near the seismic source. Earth system science is thereby facilitated in the Pan-Third Pole environment.

(1) Deep learning-based ground object segmentation algorithm

In this study, an image pyramid-based concept is constructed in conjunction with a multiscale nested deep learning network to segment ground objects (Fig. 9).

Consequently, these networks require numerous convolution kernel structures, and in turn, high adaptability, during model optimization and calibration. The large number of required network parameters results in low training efficiency. In this study, the computing efficiency is increased by introducing a deep learning neural network with a variable convolution kernel structure to segment ground objects. A characteristic map pyramid is constructed by down-sampling and convolving the frontal-layer characteristic maps of the network. Up-sampling is then implemented, followed by merging and fusion with the back-layer characteristic maps. The proposed novel network exhibits a significantly improved response to ground objects at different scales than existing semantic segmentation networks.

(2) Algorithm for automatic extraction of glacier and lake vector data

The big data platform can be used to perform multisource data calculations that make full use of existing multisource remote sensing data and comprehensive data information. For instance, glacial data can be used for automatic glacier boundary extraction to accurately estimate glacial retreat on the Tibetan Plateau and surrounding areas under global warming and to predict changing trends. The deep learning neural network with a variable convolution kernel structure is used to segment ground objects based on a large quantity of data for typical lakes in the Tibetan region: Sentinel-1 wide-range synthetic aperture radar (SAR) data, Landsat remote sensing image data, high-resolution digital terrain model remote sensing data, glacier catalog vector and aerial data. Down-sampling and convolution are implemented on the frontal-layer characteristic maps to construct the characteristic map pyramid, followed by up-sampling and fusion with the back-layer characteristic maps. The network response to ground objects at different scales is improved. In addition, terrain shadows, which are easily confused with glaciers, are removed, and the effects of speckle noise, winds and waves on the lake are prevented. The boundaries of glaciers and lakes can be identified with an accuracy above 90%.

(3) Observation data for ecological observation network and integration of ecological models

The following multiplatform, multispatial-scale and long time series remote sensing data are integrated: top apparent albedo (TOA) data at a Landsat 30-m resolution for the Third Pole region from the past 30 years; MODIS zenith angle albedo products at a 500-m resolution from 2000 to 2017; and MODIS vegetation coverage products and terrain data (SRTM DEM) at a 250-m resolution from 2000 to 2017. Carbon and water flux data from Tibetan Plateau flux stations are employed to optimize the parameters of the ecosystem model. The Morris method is used to perform a sensitivity analysis of the carbon and water cycle processes in the model. The most sensitive parameters are selected for optimization using Bayesian assimilation technology in conjunction with the carbon and water flux data. The parameter optimization significantly improves the simulation accuracy of the model for the net ecosystem carbon exchange (NEE) and reduces the simulation errors of the model for the gross primary productivity (GPP) of the ecosystem and the NEE.

(4) Development of seismic observation big data-driven imaging system for seismic rupturing processes

This study develops a software system that automatically classifies, retrieves, evaluates and picks up far-field P and SH waveform data (Fig. 10). A large quantity of raw data is thus rapidly processed to obtain far-field body wave data for imaging rupturing processes at the seismic source. A program based on the Green function is developed to calculate the static displacement field by the generalized reflection and transmission coefficient matrix method. Velocity structure models of the crust and upper mantle can be automatically extracted along with a static displacement Green function that depends on the depth and epicenter distance. The far-field waveform data are used to rapidly image rupturing processes at the sources of recent earthquakes with magnitudes above 6.0 in the Pan-Third Pole region and above 7.0 throughout the world. The earthquake intensity can then be theoretically evaluated.

6 Conclusions

The advent of the big data era has produced new opportunities and challenges for understanding the mechanisms of environmental problems in the Tibetan Plateau and the Pan-Third Pole and enacting appropriate measures. The Pan-Third Pole big data system is an important data support platform for the STEP-2 program and the Silk Road and Environment project. This system has multiple functions: the storage, management, analysis, mining and sharing of scientific data for various disciplines

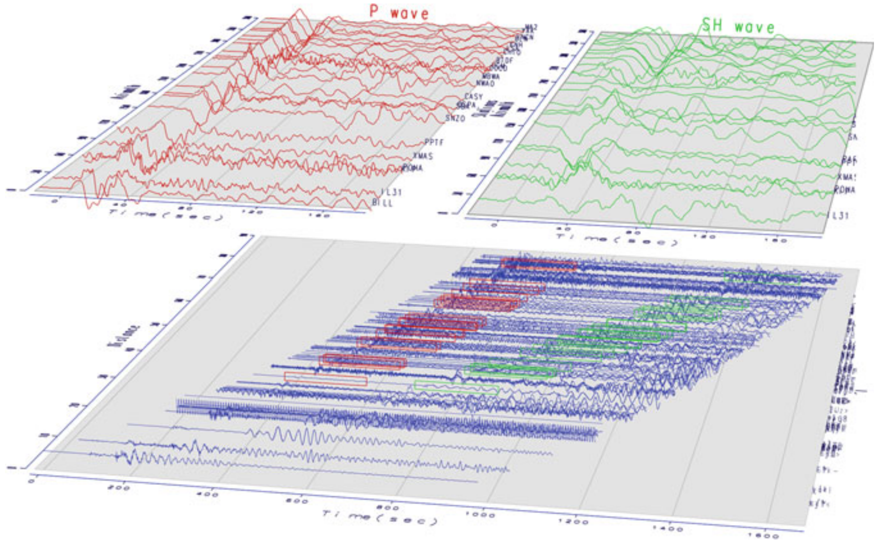


Fig. 10 Automatic selection and imaging analysis of far-field body wave data

of the Pan-Third Pole, such as resources, the environment, ecology, atmospheric science and solid-earth science; the refining of products of key scientific data of the Pan-Third Pole; the gradual development of functions such as online big data analysis and model application; the construction of a cloud-based platform to integrate data, methods, models; and services for Pan-Third Pole research; and the promotion of the application of big data technology to scientific research on the Pan-Third Pole. The system helps protect the eco-environment of the Pan-Third Pole region and facilitates the healthy, environmentally friendly and sustainable development of society and the economy. The system platform and the data resources of the Pan-Third Pole big data system are developed as follows: a standardized system regulation is created; the data catalog for the Pan-Third Pole is reviewed and organized; data resources of various platforms are combined and integrated; and the core dataset is compacted. Comprehensive data intellectual property protection measures for data rights and interests are adopted, and data protection periods are set to secure the rights and interests of data producers. The system follows the FAIR sharing principle and provides primary online and secondary offline services as sharing modes. The barrier for users to download data is lowered. The system has been proactively submitted for qualification as a data storage center for major international journals in an attempt to absorb more original data and energize Pan-Third Pole scientific research. The system offers a bilingual (Chinese-English) interface to provide scientific data resources from the Pan-Third Pole to research institutes and scientists worldwide. The Pan-Third Pole big data system is expected to deepen investigation of the "water-glacier-atmosphere-biology-human activity" multilayer interaction. The process and mechanism for environmental change in the region will be thereby elucidated, along with the impact on and response regularity to global environmental change. Forecasting, prewarning and mitigation capabilities of regional disasters will also be improved.

After nearly two years of development, the Pan-Third Pole big data system (https://data.tpdc.ac.cn) is in regular operation and is delivering services. Over 2,500 scientific datasets have been published, and primary online services and secondary offline services are adopted as sharing modes. Multiple measures for intellectual property rights of data are enforced. The online big data analysis methodology is gradually being developed. The Pan-Third Pole big data system is dedicated to providing data support for scientific research on the Earth system and regional environmentally friendly development.

References

1. Yao T (2019) Tackling on environmental changes in Tibetan Plateau with focus on water, ecosystem and adaptation [J]. Science Bulletin 64(7):417
2. Yao T, Xue Y, Chen D, Chen F, Thompson L, Cui P, Koike T, Lau WKM, Lettenmaier D, Mosbrugger V, Zhang R, Xu B, Dozier J, Gillespie T, Gu Y, Kang S, Piao S, Sugimoto S, Ueno K, Wang L, Wang W, Zhang F, Sheng Y, Guo W, Alikun YX, Ma Y, Shen SSP, Su Z, Chen F, Liang S, Liu Y, Singh VP, Yang K, Yang D, Zhao X, Qian Y, Zhang Y, Li Q (2019) Recent third Pole's rapid warming accompanies cryospheric melt and water cycle intensification and interactions between monsoon and environment: multidisciplinary approach with observations, modeling, and analysis. Bullet Am Meteorol Soc 100:423–444
3. Yao T, Chen F, Cui P, Ma Y, Xu B, Zhu L, Zhang F, Wang W, Ai L, Yang X (2017) From tibetan plateau to third pole and pan-third pole. Bullet Chin Acad Sci 32(9):924–931
4. Yang X (2017) Scientists from various countries discuss the Pan-Third Pole environment and the Belt and Road Initiative. Sci Technol Daily 12(3)
5. Li X, Niu XL, Pan XD, et al. (2020) National Tibetan plateau data center established to promote third-pole earth system sciences [EB/OL]. https://www.gewex.org/gewex-content/files_mf/1590612006May2020.pdf
6. Li X, Nan ZT, Cheng GD, Ding YJ, Wu LZ, Wang LX, Wang J, Ran YH, Li HX, Pan XD, Zhu ZM (2011) Toward an improved data stewardship and service for environmental and ecological science data in west China. Int J Digital Earth 4(4):347–359. https://doi.org/10.1080/17538947.2011.558123
7. Stall S, Yarmey L, Cutcher-Gershenfeld J, Hanson B, Lehnert K, Nosek B, Parsons M, Robinson E, Wyborn L (2019) Make all scientific data FAIR. Nature 570:27–29
8. Wang L, Li X (2019) Challenges and practice of geoscience data sharing: the example of the West China Ecological and Environmental Science Data Center. Science Press, Beijing
9. Wang L, Nan Z, Wu L, Ran Z, Li H, Pan X, Zhu Z, Li X, Ding Y (2010) Environmental and ecological science data center for western china: review and outlook. China Sci Technol Resour Rev 42(5):30–36
10. Pan X, Li X, Nan Z, Wu L, Wang L, Li H (2010) Research on data description documents. China Sci Technol Resour Rev 42(3):30–35
11. Li X, Che T, Li XW et al (2020) CASEarth poles: big data for the three poles. Bull Am Meteor Soc. 101(9):E1475–E1491. https://doi.org/10.1175/BAMS-D-19-0280.1
12. Ran YH, Li X, Cheng GD (2018) Climate warming over the past half century has led to thermal degradation of permafrost on the Qinghai-Tibet Plateau. The Cryosphere 12(2):595–608
13. Li X, Liu SM, Xiao Q, Ma MG, Jin R, Che T, Wang WZ, Hu XL, Xu ZW, Wen JG, Wang LX (2017) A multiscale dataset for understanding complex eco-hydrological processes in a heterogeneous oasis system. Sci Data 4170083. https://doi.org/10.1038/sdata.2017.83.

14. Jin R, Li X, Yan B, Li X, Luo W, Ma M, Guo J, Kang J, Zhu Z (2014) A Nested eco-hydrological wireless sensor network for capturing surface heterogeneity in the middle-reach of Heihe River Basin, China. IEEE Geosci Remote Sens Lett 11(11):2015–2019. https://doi.org/10.1109/LGRS.2014.2319085
15. Kang J, Li X, Jin R, Ge Y, Wang J, Wang J (2014) Hybrid optimal design of the eco-hydrological wireless sensor network in the middle reach of the Heihe River Basin, China. Sensors 14(10):19095–19114

Xin Li is currently a professor at Institute of Tibetan Plateau Research (ITP), Chinese Academy of Sciences (CAS) and the Director of National Tibetan Plateau Data Center at ITP/CAS. His primary research interests include land data assimilation, application of remote sensing and GIS in hydrological and cryospheric sciences, and integrated watershed study. He received the B.S. degree in GIS and Cartography from Nanjing University in 1992 and the Ph.D. degree in Remote Sensing and GIS from CAS in 1998. He was a member of WCRP GEWEX scientific steering committee, World Data Center for Glaciology and Geocryology at Lanzhou, and is a member of the International Science Advisory Panel of Global Water Futures programme. He is in the editorial board of Journal of Hydrology, Science Bulletin, Vadose Zone Journal, Remote Sensing, Big Earth Data and other international journals. He has published over 380 journal articles (SCI > 230) and coauthored 8 books. Total citations to these publications are 15,000+ and h-index is 60+. He is the lead scientist of WATER (Watershed Allied Telemetry Experimental Research, 2007-2010) and HiWATER (Heihe Watershed Allied Telemetry Experimental Research, 2012-2017), which are two comprehensive remote sensing ecohydrology experiments conducted sequentially in recent years in China.

Information Construction and Prospect of EAST Collaborative Experimental Platform

Feng Wang, Bingjia Xiao, and Jinyao Xia

Abstract EAST is the only experimental device in the world with similar conditions as ITER and have the best ability to achieve long pulse and high performance operation on the particle balance time scale, which attracts extensive international cooperation and generates a huge amount of experimental data. The construction goal of EAST collaborative experimental platform is to follow the concept of "collaboration", "efficiency" and "openness" and provide support and guarantee for carrying out high-level scientific research and personnel training. It covers the whole process of EAST experiment, including the collection and arrangement of proposals, Experimental process status monitoring, real-time data acquisition and storage, experimental log recording, remote joint experiment, Data access and analysis and so on, which involved in high-speed reflection memory network, virtual applications, virtual reality, high-speed acquisition and storage, big data processing and other information advanced technology. Moreover, it provides many support platforms such as unified account management, unified data access interface, document management and wiki internal communication platform and so on. The platform not only promotes the accumulation of research resources and improves the work efficiency, but also provides an open and shared way of academic exchange and promotes domestic and foreign cooperative and research. Until 2018 there are about 1000 registered users from more than 40 cooperative units, attracting 4000 foreign visitors per year, the accumulated experimental data has reached 1000 TB. Several countries have participated in joint experiments through the platform, the percentage of external proposals is as high as 57% in 2018. It is of great benefit for EAST research, helping EAST set the world record for 101.2 s of steady-state long-pulse high-confinement plasma operation, and obtained the completely non-inductive plasma with electron temperature of 100 million degrees for the first time in 2018. The next step is to realize an independent and controllable information data platform based on domestic hardware and software combined with big data and artificial intelligence technology, providing a preliminary study for control and information system of future fusion reactor.

F. Wang (✉) · B. Xiao · J. Xia
Institute of Plasma Physics, Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei, China
e-mail: wangfeng@ipp.ac.cn

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_8

Keywords EAST · Nuclear fusion · Collaborative · Virtual application · Real-time control

1 EAST Introduction

Nuclear fusion is an ideal energy source in the future. Tokamak is a ring container that uses magnetic constraints to achieve controlled nuclear fusion [1]. EAST (Experimental Advanced Superconducting Tokamak) is the first full Superconducting Tokamak fusion Experimental device to be built and put into operation on international [2], which is a National Mage-project of Science Research of China. EAST adopt the advanced technology with a long pulse high parameter operation capacity of over 1000 s, similar to international thermonuclear fusion reactor experiment (ITER). It is one of the few experimental platforms in the world capable of carrying out long pulse fusion plasma physics and engineering research under high parameter conditions in the next 10 years. It is open to both domestic and foreign countries and has attracted extensive international cooperative research.

The research target of EAST device is to study the physics and engineering techniques of the long pulse steady state operation of the tokamak and lay the engineering technology foundation for the future full superconducting tokamak reactor. Aiming at the research frontier of nuclear fusion energy, we will carry out joint experimental research at home and abroad on the basic physics and engineering problems of advanced tokamak fusion reactor in steady state, safe and efficient operation, so as to provide scientific basis for the design and construction of nuclear fusion engineering test reactor, and promote the development of other related disciplines and technologies in the field of plasma physics [3, 4]. Centering on the main line of high-constrained, steady-state operation research, the EAST device usually conducts experiments twice a year, each of experiment lasts for three months. Among every experiment, there will be a large number of upgrade to get stronger discharge operation ability to be carried out, and a variety of new plasma discharge experiment operation mode and method to be explored along with many related engineering and physical experiment research, achieving a series of breakthroughs, respectively, In 2012, 2016, 2017, the world records of steady-state discharge operation with high constraint modes of 30, 60 and 100 s were successfully broken [5, 6]. It is the world's first tokamak fusion experiment device with a high confinement mode plasma running time of 100 s.

2 The Demand of Information Construction for EAST Magnetic Confinement Nuclear Fusion Experiment

In the past decade, the experimental data of EAST has gone from GB to TB, and now it is moving from TB to PB. More and more users at home and abroad want to participate in the EAST experiment in different ways. How to improve the access rate and experiment efficiency of experimental data by using information technology and means, and promote the output of nuclear fusion research results, has brought new challenges to the information construction of EAST experiment. The whole process of EAST fusion experimental research mainly includes the collection and arrangement of experimental proposals, experimental control, monitoring and recording of experimental conditions, collection and storage of experimental data, analysis of experimental data, release and sharing of experimental data, etc. Each part has an urgent need for the information environment.

At first, the internal staff of the unit participated in the EAST experimental study, and the experimental arrangement was determined through internal discussion. As with EAST device is open and shared with domestic and foreign users, it is necessary to provide unified experimental application entrance that can be accessed quickly for users from different regions with the help of information platform. Through this platform, users can apply for experimental machines to conduct specific experimental verification or research, and relevant responsible persons can review the experimental proposals submitted by users in different places, so as to make experimental arrangements with the overall planning of the main line of the experiment.

During the operation of EAST experiment, it is necessary to release the experimental plan regularly with the help of information tools, track and record the experimental plan, monitor the experimental status in real time, and publish the experimental results, so that all experimenter can master the latest situation of the experiment. EAST has always maintained good cooperative relations with the United States, Russia, France, Japan, Korea, Germany, Britain, Denmark, ITER and other major fusion countries or organizations in the world. It has been listed by the U.S. department of energy as the preferred device for foreign cooperation. Therefore, more and more collaborators hope to realize remote control and carry out joint experiments with the help of advanced network technology.

The EAST experiment aims to generate a large amount of experimental data for research and analysis, including video data, raw data and processed data. All kinds of data need to be collected or calculated in real time and stored in the corresponding database or storage. With the gradual development of the long-pulse experiment, the experimental data show an increasing trend every year, up to now, the experimental data has been as high as 1000 TB. In addition, the data sources are diversified, with more than 7,000 signal data stored in different data servers. The increasing data size and the change of signal type lead to the increasing requirements on the information environment, such as more storage space, higher access speed and so on. Physical experts need to compare and analyze multiple types of signal data per shot, so integrated data browsing and visualization tools need to be developed.

The successful operation of the EAST experiment and the breakthrough progress are attributed to the cooperation and communication between various system departments. Therefore, an open and collaborative communication platform and a knowledge sharing platform are needed to create an open and collaborative information atmosphere and environment to help the rapid growth of young talents. In addition, EAST experiment cannot be separated from the efficient and excellent operation management of the management department, including document management, experimental results management, etc., which need to improve work efficiency by using information technology.

3 EAST Collaborative Experimental Platform Information Construction

The main goal of the information construction of EAST collaborative experimental platform is to provide support and guarantee for talent cultivation and the development of high-level scientific research and experiments. The architecture of EAST collaborative experimental platform is shown as Fig. 1. It mainly includes the collection and arrangement of the experimental proposals in the early stage of the experiment; State monitoring during the experiment, real-time display of experimental results and remote joint experiment; Data analysis and display tools in the later stage of the experiment and information display such as publication of results. At the same time, it provides unified account management, document management, conference release and other information support platform, with unified identity authentication, unified data access interface, open, collaboration and other characteristics.

The platform adopts unified account login to all systems, and establishes a unified account management system based on LDAP to handle account application, information update and account permissions, etc. The researchers only need one account

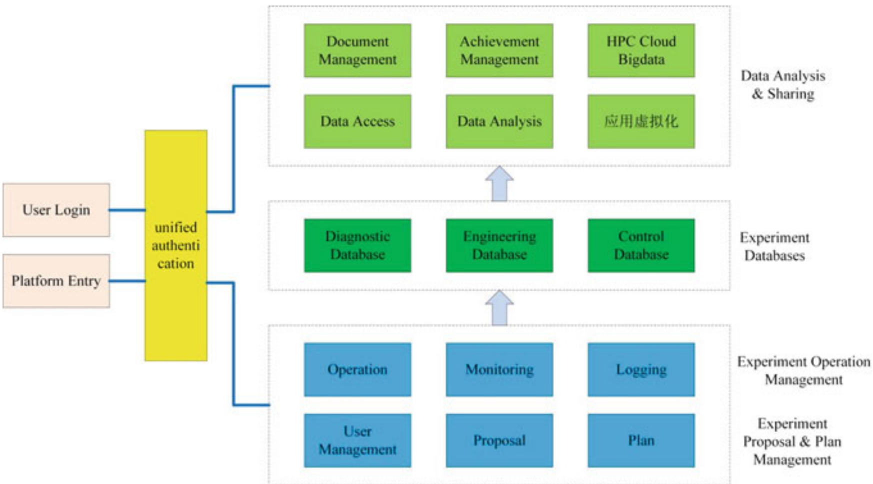


Fig. 1 The architecture of EAST collaborative experimental platform

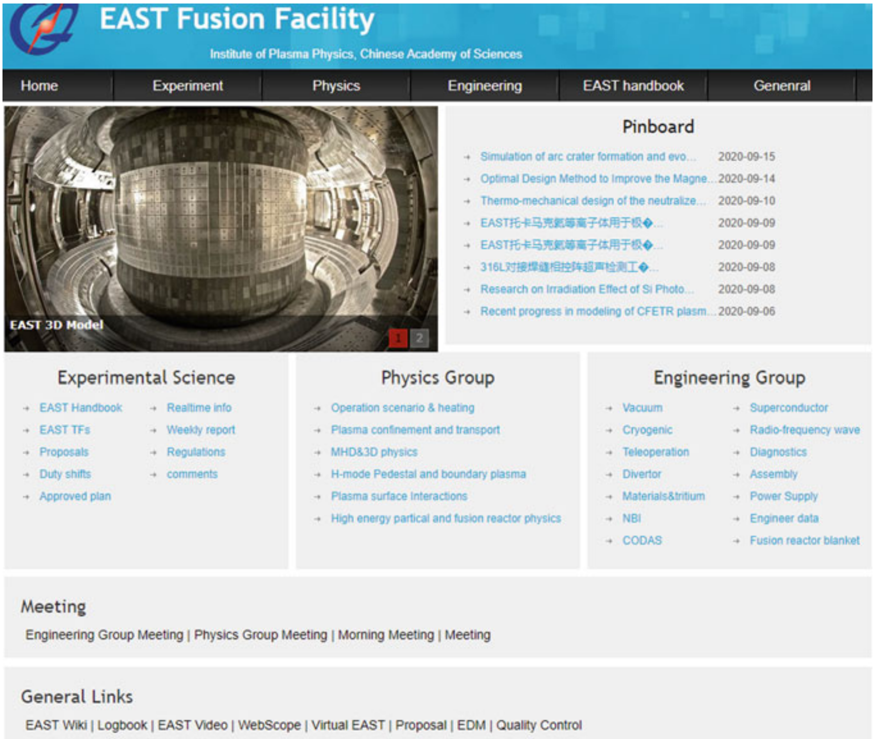


Fig. 2 Main entrance of EAST collaborative experiment platform

to browse the platforms with access rights. The main entrance of EAST collaborative experimental platform is shown as Fig. 2.

3.1 Experimental Proposal and Plan Management

3.1.1 Experimental Proposal Management System

The experimental proposal management system is an online application platform based on the browser/server architecture. It has the functions of writing, reviewing, querying and tracking the experimental proposal online. Before the experiment, domestic and foreign users learn about the application requirements of the experimental machine through the platform, and submit the experimental requirements and plans. The staff of EAST physics group organize the discussion of relevant experimental proposals, arrange the mainline proposals according to the annual operation target plan of the device, and determine the priority proposals or combined several proposals through the proposal system. The construction of this platform provides

Fig. 3 EAST proposal management system

users at home and abroad with a unified EAST experimental machine entrance for application, which saves time for proposal approval and improves the work efficiency of researchers (Fig. 3).

3.1.2 Experimental Plan Management System

The experimental plan management system is an open information sharing platform established on the basis of Mediawiki [7]. All logged-in users can edit the contents of the platform together, with multiple functions such as collaborative editing, permission management, and custom directory. The platform not only publishes the overall arrangement of the experiment, the recent experiment plan, the arrangement of the experiment duty and the tracking record of proposals, but also includes the introduction of each system. All the contents are edited by users in collaboration to help young researchers quickly master different systems and participate in experimental research as soon as possible. In addition, the EAST device is upgraded each year, and the platform has version control to facilitate users to trace historical information (Fig. 4).

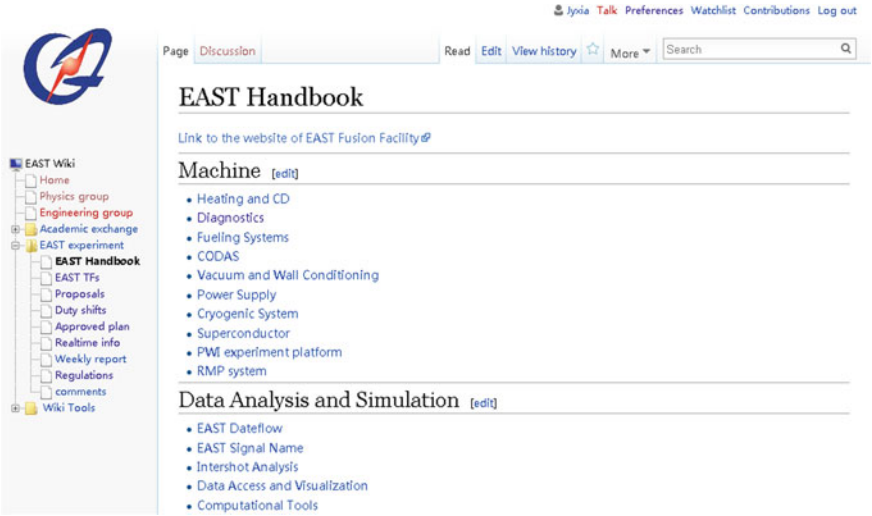


Fig. 4 EAST experimental plan management system

3.2 Experimental Operation Management

3.2.1 Experimental Operation System

EAST device is composed of dozens of subsystems. The purpose of the master control system is to coordinate the safe and stable work of each subsystem and ensure the effective operation of physical experiments. The system realizes the configuration and management of operation parameters of the subsystem based on PXI platform, and monitors the experiment logic in real time during the experiment to ensure the correct and orderly progress. The master control system runs 7 × 24 h and the inspection cycle is less than 1 ms. The snapshot of EAST experimental central control system is shown as Fig. 5.

At the same time, in order to protect the safe operation of important equipment on the device, a safety interlocking system is designed and developed. The system adopts Siemens PLC to realize system redundancy and centralized interlock monitoring of all important equipment. The fault response time is less than 4 ms and the fault log is accurate to 1 ms. After the fault occurs, an alarm will be issued to inform the experimentalists, so that the experimentalists can deal with it in time. In addition, the platform can be published as web pages to support remote monitoring of signal status. The snapshot of EAST safety interlock monitoring system is shown as Fig. 6.

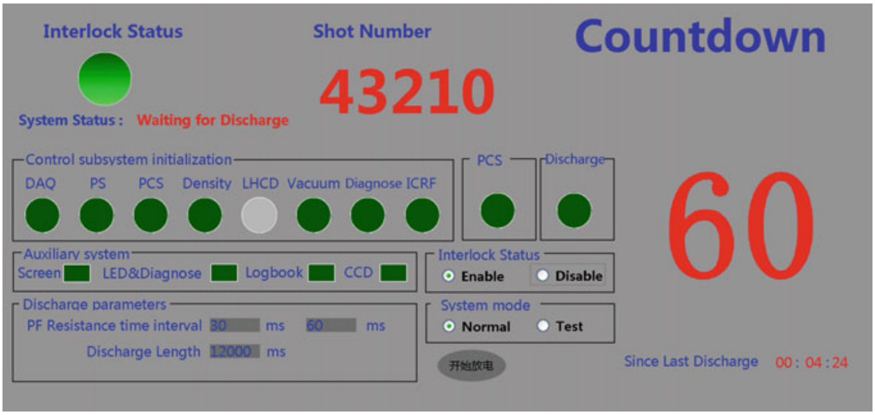


Fig. 5 The EAST experimental central control system (PS: power system; PCS: plasma control system; LHCD: low clutter current drive; ICRF: ion cyclotron heating; CCD: camera video diagnosis; PF: polar field)

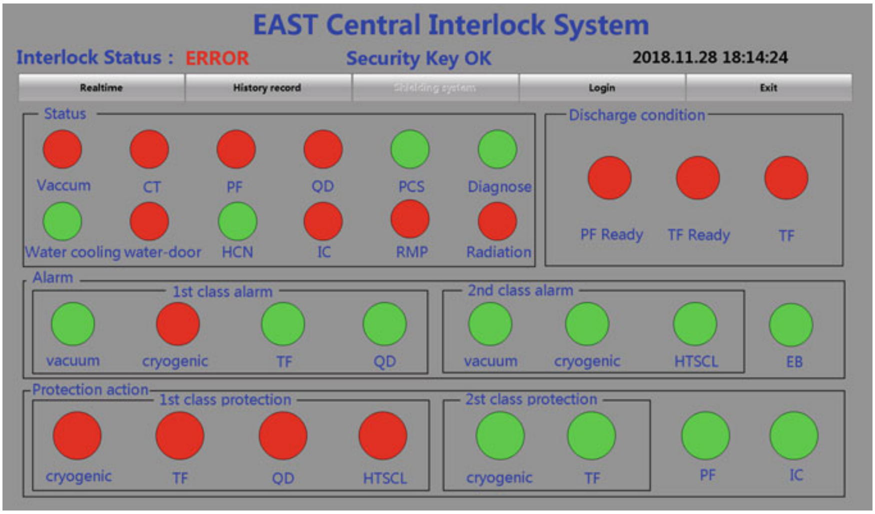


Fig. 6 EAST safety interlock monitoring system (PCS: plasma control system; HCN: plasma density; IC: fast control; RMP: resonance magnetic field disturbance; PF: polar field; TF: longitudinal field)

3.2.2 Experimental Monitoring System

The large-screen centralized monitoring system located in the control hall was formally put into use in the experiment in 2018. It is composed of LED full-color dot matrix and LCD liquid crystal splicing, with a total area of about 100 m^2, which is used to display all important parameters of the experiment, experimental status

Fig. 7 EAST large-screen centralized monitoring system

and monitoring status of each system. It includes the display of important parameters such as shot number and longitudinal field, as well as the state monitoring of all subsystems such as safety interlock, plasma shape and position display, video display, vacuum, low temperature and neutral beam injection, so that the experimenters in the control hall can master the experimental running state in time. It can control the free combination and distribution area of signals accessed by optical fiber. In addition, the platform can preset a variety of display schemes and switch the display scheme according to the occasion. The EAST large-screen centralized monitoring system is shown as Fig. 7.

(1) Engineering data monitoring

The engineering data display system is a real-time publishing platform based on the browser/server architecture, which can display the changes of key parameters of vacuum, magnet, power supply and other engineering systems in real time, and monitor the running status of each system. This platform mainly displays the changes of parameters in the corresponding positions of each system structure diagram, such as valve opening and closing, temperature change, etc., so as to clearly and intuitively explain whether the system is normal or not. If there is a fault, the function of fault alarm will be started.

(2) RTSCOPE

RTSCOPE is a real-time plasma shape display system designed and developed independently by using advanced technologies such as C/S architecture and reflective memory network. When the experiment is running, the server side reads the shape data sent by the plasma control system in real time from the high-speed reflection memory network and then distributes it to different clients to draw the display, which can realize the real-time refresh in milliseconds.

(3) Diagnostic system status display

EAST device contains dozens of diagnosis system, the normal operation of the various diagnosis system is essential to the experiment, thus the diagnosis system based on browser/server architecture status display system is established. It mainly includes the real-time display of each diagnosis status, the status display of the diagnosis valve, the diagnosis introduction, the diagnosis name list, the historical information, the operation and maintenance information and so on.

3.2.3 Experimental Logging System

The experimental logging system Logbook is a real-time publishing platform based on the browser/server architecture [8], which is used to publish the latest key experimental parameters such as shot number and longitudinal field current, and display the waveform of key experimental data. Through this platform, operators can record the experimental results of each shot in real time, including whether there is any breakthrough progress, whether there is a fault and the cause of the fault, etc. At the same time, the platform has access control, historical data view and other functions.

3.3 High Speed Acquisition and Storage of Experimental Data

3.3.1 Real-Time Network Communication System

The system users a distributed real-time framework of "multi-type network", which guarantees the experimental data of each subsystem is stored to the EAST server in real time and transmitted to PCS (plasma control system) to carry out complex calculation in control cycle (100 subtle) [9]. The system adopts reflection memory card (RFM) with high speed, good real-time performance and high reliability, MRG (Messaging, real-time and Grid) real-time operating system providing deterministic response time, and multi-threading technology with high concurrency to ensure the real-time performance and reliability of data transmission.

3.3.2 Long Pulse Real-Time Data System

According to the requirements of long pulse discharge experiment of EAST device and the characteristics of pulse discharge of fusion device, a real-time storage method of long pulse data is designed, and a long pulse real-time data system based on shard storage is realized [10]. Using the slice storage mechanism, the continuously collected 5 GBytes/s long pulse data stream is divided into several time slices, which are gradually added to the MDSplus database according to the slice index, which can meet the steady-state operation requirements of EAST device over 1000 s. At present, a complete experimental database has been established on the EAST device. It mainly includes engineering database, diagnosis database, high-speed camera database, simulation database, control operation database, analysis database, etc., which provides a perfect data access guarantee for researchers and a unified data platform for international cooperation and exchange.

3.4 Experimental Data Access and Sharing

3.4.1 Data Access Tools

(1) Data visualization software: WebScope

WebScope is mainly a software system for visual display and analysis of experimental data [11]. Users can easily view the data waveform of all experimental signals and conduct analysis through the browser. It adopts technologies such as spot sampling and thumbnail snapshot to greatly improve the data access speed and provide a good user experience. It mainly includes data acquisition, data display, and data analysis and data management four modules.

(2) Web-based camera diagnostic video publishing

The system converts the visible and infrared video collected by the hd camera into FLV video and publishes it to the front end of the Web in real time. The experimenter can view the plasma discharge video and analyze the video frame-by-frame, so as to grasp the running state of the plasma intuitively, which is also conducive to comparative analysis with other diagnostic data.

(3) Plasma shape and position view software: EAST Viewer

In order to see the result of EFIT magnetic surface balance, we designed and developed client software of C/S architecture, developed by Python and GTK+ toolkit. EAST Viewer is the software to display and analyze the result data of EFIT algorithm. It can obtain data from EFIT output file and MDSPlus database, and display plasma equilibrium parameters calculated by EFIT, plasma boundary, magnetic surface configuration, PCS control points, etc. At the same time, EFIT results of different shot number and time can be compared.

(4) The 3D virtualization system: Virtual EAST

Virtual EAST is a comprehensive platform integrating 3d visualization of EAST device model, engineering physical parameters and experimental data. Users can roam the vacuum chamber in EAST Virtual scene through a Web browser, and obtain the corresponding engineering physical parameters and experimental data through model parts and it also support browsing on mobile phone platform (Fig. 8).

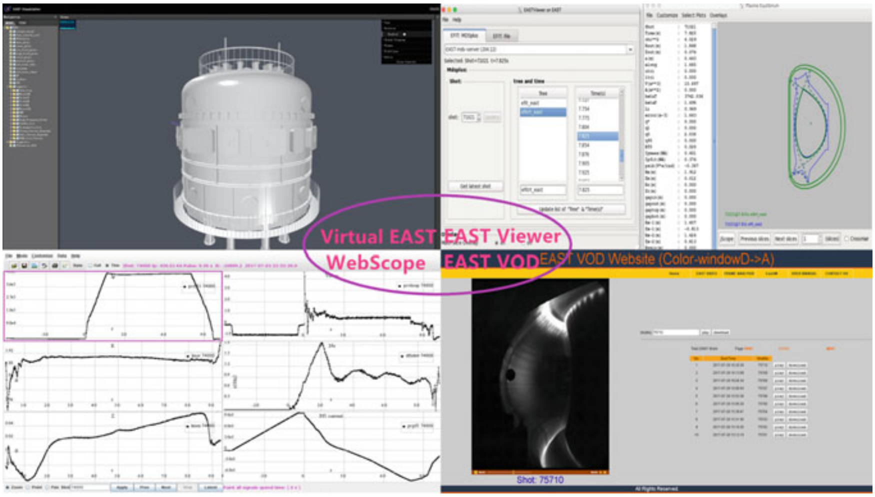


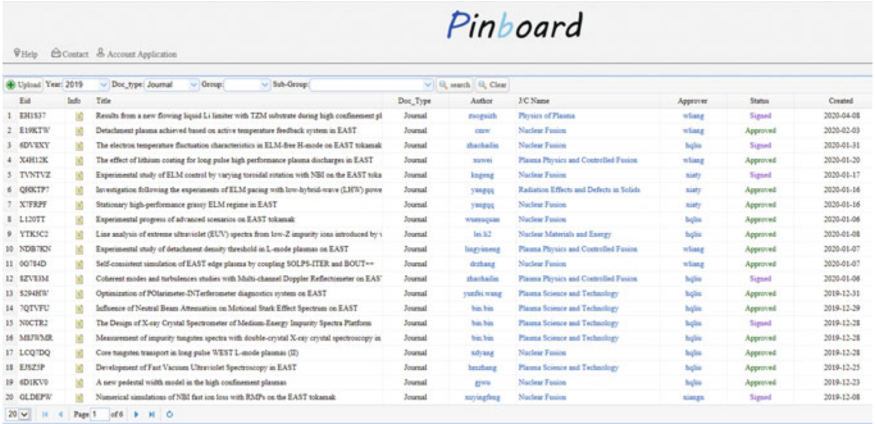
Fig. 8 EAST experimental data access analysis tools

3.4.2 Unified Data Access Interface Based on Virtual Application

The unified data access interface based on XenApp application virtualization can achieve cross-platform unified data access. Install and configure all the data access analysis tools and software on the high-performance server, users only need to install a plug-in in the browser to access all the tools and software, and do not need to install and maintain the relevant software separately in the local, not only to achieve cross-platform, and reduce the burden of users, which improves work efficiency.

3.4.3 Data Analysis and Processing

In order to provide a good data analysis environment for EAST experiment participants, an integrated data analysis and processing system was designed and implemented. It is mainly composed of a NoMachine login server cluster and a Linux computing server cluster. Users at home and abroad can log into the server through NoMachine or SSH client anywhere, and can directly run the commonly used programs and software of EAST data processing on the server, including Matlab, GDL, EASTViewer, jScope, reviewPlus, etc. Since the data server is located in the EAST core data room and is connected to the experimental network with 10 gigabit optical fiber, the data access and processing speed can be greatly improved. During each round of the EAST experiment, about 200 users have logged into the server for data analysis and processing. The system also interconnects with EAST and other supercomputing clusters to provide users with massively parallel computing services.



	Eid	Info	Title	Doc_Type	Author	J/C Name	Approver	Status	Created
1	EH1S37		Results from a new flowing liquid Li limiter with TZM substrate during high confinement pl	Journal	zuoguizh	Physics of Plasma	wliang	Signed	2020-04-08
2	E19KTW		Detachment plasma achieved based on active temperature feedback system in EAST	Journal	cmrw	Nuclear Fusion	wliang	Approved	2020-02-03
3	6DV8XY		The electron temperature fluctuation characteristics in ELM-free H-mode on EAST tokamak	Journal	zhaohailin	Nuclear Fusion	hqliu	Signed	2020-01-31
4	X4H12K		The effect of lithium coating for long pulse high performance plasma discharges in EAST	Journal	xuwei	Plasma Physics and Controlled Fusion	wliang	Approved	2020-01-20
5	TVNTVZ		Experimental study of ELM control by varying toroidal rotation with NBI on the EAST toka	Journal	kngeng	Nuclear Fusion	xiaty	Signed	2020-01-17
6	QHKTP7		Investigation following the experiments of ELM pacing with low-hybrid-wave (LHW) powe	Journal	yangqq	Radiation Effects and Defects in Solids	xiaty	Approved	2020-01-16
7	X7FRPF		Stationary high-performance grassy ELM regime in EAST	Journal	yangqq	Nuclear Fusion	xiaty	Approved	2020-01-16
8	L120TT		Experimental progress of advanced scenarios on EAST tokamak	Journal	wuenquan	Nuclear Fusion	hqliu	Approved	2020-01-06
9	YTK5C2		Line analysis of extreme ultraviolet (EUV) spectra from low-Z impurity ions introduced by v	Journal	lei.h2	Nuclear Materials and Energy	hqliu	Approved	2020-01-08
10	NDB7KN		Experimental study of detachment density threshold in L-mode plasmas on EAST	Journal	lingyimeng	Plasma Physics and Controlled Fusion	wliang	Approved	2020-01-07
11	0Q784D		Self-consistent simulation of EAST edge plasma by coupling SOLPS-ITER and BOUT++	Journal	dzzhang	Nuclear Fusion	wliang	Approved	2020-01-07
12	8ZV83M		Coherent modes and turbulences studies with Multi-channel Doppler Reflectometer on EAS	Journal	zhaohailin	Plasma Physics and Controlled Fusion	hqliu	Signed	2020-01-06
13	S294HW		Optimization of POlarimeter-INTerferometer diagnostics system on EAST	Journal	yunfei wang	Plasma Science and Technology	hqliu	Approved	2019-12-31
14	7QTVFU		Influence of Neutral Beam Attenuation on Motional Stark Effect Spectrum on EAST	Journal	bin bin	Plasma Science and Technology	hqliu	Approved	2019-12-29
15	N0CTR2		The Design of X-ray Crystal Spectrometer of Medium-Energy Impurity Spectra Platform	Journal	bin bin	Plasma Science and Technology	hqliu	Signed	2019-12-28
16	M8JWMR		Measurement of impurity tungsten spectra with double-crystal X-ray crystal spectroscopy in	Journal	bin bin	Plasma Science and Technology	hqliu	Approved	2019-12-28
17	LCQ7DQ		Core tungsten transport in long pulse WEST L-mode plasmas (II)	Journal	xdyang	Nuclear Fusion	hqliu	Approved	2019-12-28
18	EJSZ5P		Development of Fast Vacuum Ultraviolet Spectroscopy in EAST	Journal	hmzhang	Plasma Science and Technology	hqliu	Approved	2019-12-25
19	6D1KV0		A new pedestal width model in the high confinement plasmas	Journal	gjwu	Nuclear Fusion	hqliu	Approved	2019-12-23
20	GLDEPW		Numerical simulations of NBI fast ion loss with RMPs on the EAST tokamak	Journal	xuyingfeng	Nuclear Fusion	xiangn	Signed	2019-12-08

Fig. 9 EAST paper publicity system

3.4.4 Experimental Results Management

EAST experimental results need to be reported and discussed in the engineering group or the physics group before the paper is published, and published on the public platform after the paper is passed, and the complete experimental results need to be supplemented in the proposal system after receiving or winning relevant awards. The EAST paper publicity system is shown as Fig. 9.

3.4.5 EAST Document Management

The document management system (EDM) based on the browser/server architecture is established, which has the functions of version control, document approval and permission management, etc. (Fig. 10)

4 Achievements of Information Construction and Application of EAST Collaborative Experimental Platform

Through the continuous upgrading, improvement and optimization of the collaborative experimental platform, good application results have been achieved, mainly in the following aspects:

(1) It promotes the construction and accumulation of scientific research resources. Through the information construction of EAST collaborative experimental platform, experimental data resources with scientific and research value have been

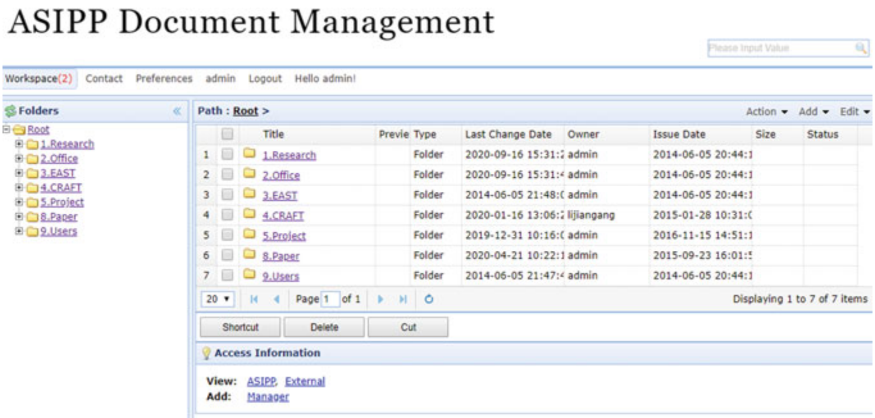


Fig. 10 EAST document management system

collected and sorted out, and different experimental databases has been built according to research experience and professional knowledge. For example, raw database, secondary database, diagnostic database, proposal database, experimental operation control parameter database, and so on. It has formed the basic information support for scientific research and will become an important data resource for scientific research in fusion field in the future.

(2) It improves the efficiency of researchers in studying and analyzing experimental data and creates an information-based scientific research environment. Scientific research informatization is the key driving force to promote scientific research methods and means. The information construction of EAST experimental platform provides researchers with data analysis and view tools on various platforms, and improves the utilization and reliability of servers and the working efficiency of researchers through virtualization technology. For example, the key physical and technical problems in the field of nuclear fusion are studied by using WebScope to view and analyze the waveform of experimental data and combining with the experimental discharge video.

(3) It provides an open and shared platform for researchers to exchange research activities. In recent years, Institute of Plasma Physics, Hefei Institutes of Physical Science, Chinese Academy of Sciences has strengthened the exchange and cooperation among researchers of the same research direction through the innovative mechanism of setting up task force. The construction of the EAST Wiki system makes the communication between research teams more smooth, and the resource sharing and collaborative work more convenient. In particular, young talents can independently learn information resources in Wiki system, including the introduction of physics and engineering systems, experimental arrangement and progress, and reports of experimental seminars, etc., so that they can learn lessons and grow rapidly. At the same time, the progress of experiments can be recorded in the EAST Wiki system through collaborative sharing.

(4) It promotes international cooperation and exchanges. EAST collaborative experimental platform is an open and shared international platform. Users at home and abroad can use the experimental information platform to participate in the experiment on the premise of following the experimental specifications and procedures, including applying for remote acquisition, analysis and view of experimental data and participate in the joint experiment remotely and so on. For example, in recent years, Institute of Plasma Physics, Hefei Institutes of Physical Science, Chinese Academy of Sciences has carried out many remote joint physics experiments with general atomics of the United States and achieved some very meaningful experimental results and attract about 4000 foreign scholars to visit our unit per year.
(5) It provides managers with information management means. Through the informatization construction of EAST experimental platform, the experimental process and cooperative participation mechanism are more standardized. The physics group and the engineering group perform their respective duties to provide experimental data or obtain experimental information in the information platform. The operation management personnel can timely grasp the experimental running situation, record the experimental running machine time and running state, and improve the efficiency of experimental management. On the other hand, the experimental information platform facilitates the statistics of scientific research results and the open and shared statistics. For example, the paper publicity platform shows all the papers to be published after the review, which promotes the normative management of scientific research results. The experimental proposal management system is connected with the major infrastructure sharing platform of the Chinese academy of sciences to directly extract the proportion of external users from the platform, and investigate the platform users' satisfaction with the implementation of the proposal and the experimental equipment, which realizing the information statistics of user evaluation. Finally, EAST experimental platforms, such as EAST account management, document management system, etc., provides great convenience to managers.

5 Future Planning and Prospect of EAST Collaborative Experimental Platform

China's tokamak research has made great progress after nearly 40 years of development. In the next five years, the near-core stage steady-state plasma experiment platform will be established to carry out high-level scientific experiments; key technologies of fusion engineering experimental reactor will be absorbed, developed and reserving; the design of fusion engineering experimental reactors will be improved and re-research on key components will be carried out, so as to lay a solid scientific and technological foundation for the independent development of China's fusion engineering reactors around 2020. As with the progress of science and technology, the demand for information construction of fusion experiments is also changing [3], which requires further planning of information platform of fusion devices in the future. It mainly includes the use of big data and artificial intelligence technology to conduct rapid intelligent data analysis and processing of experimental data in EAST, and the realization of independent and controllable information data platform based on domestic software and hardware, and carry out relevant pre-research for the future control and information system of China Fusion Engineering Test Reactor (CFETR).

(1) It is proposed to realize rapid intelligent processing and analysis of experimental data of EAST. According to the characteristics of the CFETR data transmission and in order to ensure the reliability and effectiveness of data transmission, the memory and local cache will be used to form two levels of data caching mechanism, combined with the data block identification technology to realize data breakpoint continuingly functions, at the same time the redundant data servers and data storage mechanism will be used to prevent the machine failure or network failure cases of data loss. Databases will be built for different data types and requirements. For the original data, the current mainstream fusion database such as MDSplus and HDF5 will be adopted to maintain data compatibility with other devices and facilitate as well as benefit for international cooperation and exchange. For computing processing and business database, it is planned to adopt big data technology architecture such as Hadoop and Spark to design unified database and data warehouse. It is proposed to adopt big data and parallel technology to design and support concurrent data reading and processing, so as to improve the efficiency of data access. The processed data will be saved to the appropriate relational database, enabling rapid data mining and intelligent data analysis, and the processing results will be presented in real time through the portal. The data access portal system is intended to adopt the open source interactive front end, load balancing web services, the modular data access interface, combined with the basic structure of the data encryption communication technology. It is proposed to improve the utilization rate of network bandwidth by establishing virtual private network and wide area network acceleration technology to solve the problems of data communication delay and fat network in international link.
(2) It is proposed to realize independent and controllable information data platform based on domestic software and hardware. The corresponding data model is established for the data of CFETR device engineering and diagnostic system, independent template-based device interface layer is designed, common device control interface API is developed, and interface extension is provided to ensure compatibility with third-party devices. It is proposed to adopt the domestic hardware board card and the domestic operating system as the basic hardware and software environment. At present, the domestic hardware board card design and development is also gradually improved, has a relatively comprehensive data acquisition board card product line, and the gap between the foreign imported hardware and the domestic hardware is getting smaller and smaller. In recent years Chinese computer processor and chip has also made great progress, such as domestic Sunway, Loongson, Phytium processor is also becoming more and

more perfect, and has been applied in a lot of control and security areas. Domestic operating systems have also been applied in various fields, including Deepin Linux, NeoKylin, and other general operating systems, as well as SylixOS and other real-time embedded operating systems, which can be used to replace foreign Windows and VxWorks operating systems.

(3) It is proposed to conduct pre-research for the future CFETR control and information system. The engineering, diagnosis and video data acquisition and storage methods will be studied in order to meet the steady-state operation requirements of CFETR fusion reactor, including designing the unified data acquisition hardware and software interface that supports multiple systems, realizing the flexible configuration of the template-based system, designing and developing the corresponding prototype test system. it is proposed to design and develop user transparent data storage and access interface that meets the continuous operation requirements of CFETR, realize unified storage and management of experimental data, develop data access client based on desktop, browser and mobile terminal, and study data analysis and processing algorithm based on big data and artificial intelligence. The access control and status monitoring of experimental data is also researched, it includes that realizing real-time monitoring of experimental network and data, ensuring data security and timely warning of illegal intrusion, and research on data security policies based on cloud computing and cloud storage. The data access web portal will be built by integrating CFETR all kinds of data resources. The data access portal Web portal integrating CFETR all kinds of data resources will be established, it provides the unity data access system to the domestic and foreign partners which supporting of the multiple systems and browsers.

It can realize user authentication and authorization management through the deployment of security policies and achieve efficient access remotely by combining the technique of virtual private networks and wide area network. The 3d simulation model of CFETR device will be established by using virtual reality technology, and the immersive device and component installation scheme will be designed and developed to realize the virtual component installation system. The augmented reality technology will be used to combine the real scene of experimental video with virtual scene and the 3d simulation results of experimental data to comprehensively display the experimental results of CFETR.

6 Summary

EAST information collaborative experiment platform provides support and guarantee for carrying out high-level scientific research experiments and talent cultivation through advanced real-time network technology, high-speed acquisition and storage technology, big data processing technology and virtualization technology. The platform covers the whole process of EAST fusion experiment, which not only promotes the construction and accumulation of scientific research resources and improves the working efficiency of scientific researchers, but also provides an open and shared communication platform and promotes collaborative and innovative research at home and abroad. As of November 2018, there are about 1,000 registered users from more than 40 domestic and foreign cooperation units, attracting about 4,000 foreign scholars to visit China every year. So far, several countries, including the United States, France, Italy and Japan, have participated in joint experiments through the platform, the proposal percent from outside is up to 57%. In 2018, the platform successfully boosted EAST to obtain the electronic temperature of 100 million degrees of completely non-induced plasma at first time, which greatly promoted EAST's nuclear fusion energy research. With the progress of science and technology, the demand for the information construction of fusion experiment is changing constantly, and it is necessary to further plan the information platform of fusion device in the future. It mainly includes the use of big data and artificial intelligence technology to conduct rapid intelligent data analysis and processing of EAST experimental data, and the realization of independent and controllable information data platform based on domestic software and hardware, so as to conduct relevant pre-research for the control and information system of Chinese fusion engineering test reactor in the future.

References

1. Wan B, Xu G (2015) EAST superconducting tokamak. Chin Sci Bull 2015(23):2157–2168
2. Xu GS, Li JG, Wan BN et al (2013) EAST superconducting tokamak. AAPPS Bull 23:9–13
3. Wan B, Xu G (2019) Advances in experimental research towards high confinement and steady state operation on the experimental advanced superconducting tokamak. Scientia Sinica Phys Mech Astron 49(04):47–59
4. Li J (2016) Present status and development of superconducting tokamak research. Physics 2016, 45(2):88–97
5. Gong X, Wan B, Li J et al (2017) Realization of minute-long steady-state H-mode discharges on EAST. Plasma Sci Technol 19:032001
6. Wan BN, Liang YF, Gong XZ et al (2017) Overview of EAST experiments on the development of high-performance steady-state scenario. Nucl Fusion 57:102019
7. Mediawiki [EB/OL]. https://www.mediawiki.org/wiki/MediaWiki. Accessed 10 Nov 2019
8. Yang F, Xiao BJ (2010) Web-based logbook system for EAST experiments. Plasma Sci Technol 12(5):632–635
9. Li C (2018) Study of EAST real-time network communication system. University of Science and Technology of China, Hefei
10. Yang F, Xiao B, Zhu Y (2011) Real-time data system in long pulse discharge experiments of EAST. Comput Eng 37(04):12–14
11. Yang F, Dang N, Xiao B (2009) WebScope: a new tool for fusion data analysis and visualization. In: Real time conference

Feng Wang, male, born in 1977, doctor degree, associate professor, deputy head of Computer Application Division, Institute of Plasma Physics, Hefei Institutes of Physical Science, Chinese Academy of Sciences (ASIPP). His main research are tokamak control, distributed data acquisition, real-time data storage and data processing. Now he is mainly in charge of EAST tokamak data and information technology system. He has completed HT-7 main control system, EAST data and IT system, and KTX data service system. As guest researcher he has worked and visited in Kyushu University in Japan, NIFS in Japan, ITER in France. He has published more than 30 papers and hosted several national research projects.

Chinese VLBI Network and e-VLBI Technology Applications in Chinese Lunar Exploration Program

Zhong Chen and Weimin Zheng

Abstract Very long baseline interferometry (VLBI) is derived from the synthetic aperture radio observation method invented in the last century. By combining a number of large aperture radio telescopes distributed over long distance, a virtual giant radio telescope is logically created. Global and regional VLBI observation networks has extremely high spatial angular resolution, which has been widely used in astrophysics, geodesy and deep space exploration aircraft tracking. In the Chinese Lunar Exploration Program (CLEP), domestic VLBI network adopted real-time e-VLBI technology as part of TT&C system has been providing fast and accurate orbit and positioning services during the Chang'E serie missions. It works for trajectories of Earth-Moon flying, orbiting, descending, and returning to earth, made great contributions for orbit measurement. This paper introduces the general work of China VLBI network and detailed e-VLBI technology application in the three phases of CLEP, gives the details about system developments, operations and performance of the VLBI observation system. Final is the outlook of VLBI system future plan in deep space exploration missions and scientific programs.

Keywords Chinese VLBI network · China lunar exploration project · e-VLBI · Real-time · Orbit determination

Z. Chen (✉) · W. Zheng
Shanghai Astronomical Observatory, Chinese Academy of Sciences, Shanghai, China
e-mail: chenzhong@shao.ac.cn

Shanghai Key Laboratory of Space Navigation and Positioning Techniques, Shanghai Astronomical Observatory, Chinese Academy of Sciences, Beijing, China

National Basic Public Science Data Center, Beijing, China

Key Laboratory of Radio Astronomy, Chinese Academy of Sciences, Nanjing, Jiangsu, China

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_9

1 Introduction

1.1 Very Long Baseline Interferometry

Four major discoveries of modern astronomy in the 1960s: pulsars, quasars, cosmic microwave background radiation, and interstellar organic molecules are all benefited from the invention and application of radio telescopes. Until now, 6 Nobel Prizes for astronomy have been awarded to radio astronomy related discoveries.

Very Long Baseline Interferometry (VLBI) technology is an important radio astronomy method derived from the 1960s, as shown in Fig. 1. It combines multiple globally distributed radio telescopes into a virtual giant synthetic aperture radio telescope. In order to maintain time and frequency synchronization, each VLBI station has high stability time-frequency precision reference source—atomic clock as the independent local oscillation. The VLBI network has extremely high-precision spatial angular resolution which greatly improved the ability of astronomical observations.

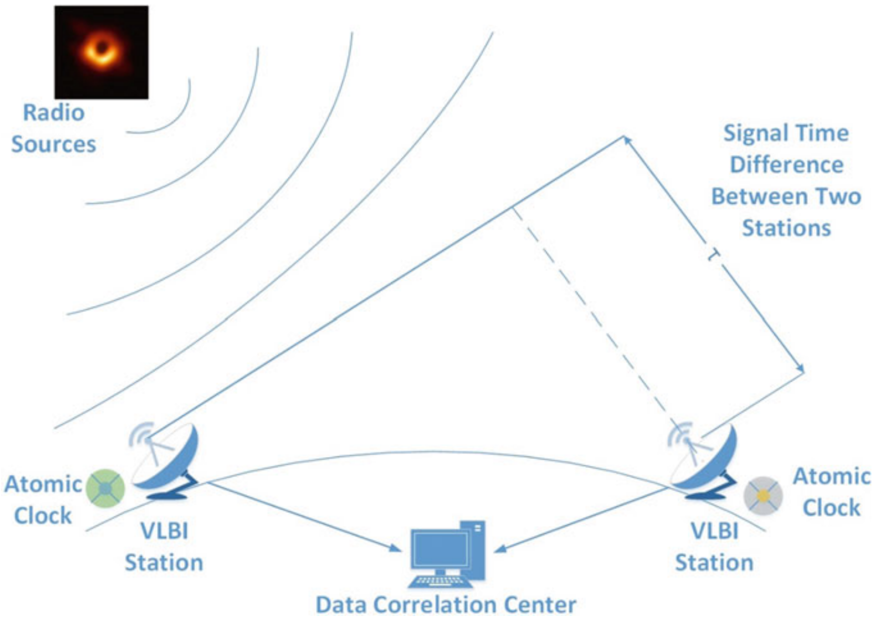


Fig. 1 VLBI schematic diagram

1.2 VLBI Application

The VLBI observation network itself is a "distributed" large scientific instrument. Observation data obtained from all stations is recorded synchronously according to the atomic clock time scale in the disk system. All raw observation data is collected to data processing center for correlation to obtain data products of time delay and the time delay rate, representing the precise time lags of radio sources signal reaching the different antennas. The accuracy of VLBI's delay measurement is directly proportional to the data recording bandwidth. Higher data rates usually lead to higher accuracy of results. Because of its extremely high space angular resolution, VLBI has an important role in astrophysics, astrometry, geodesy, and spacecraft orbit determination. For astrophysics, VLBI is an important tool for researches of universe origin, galaxies evolution, massive black holes in the galaxy centers, and interstellar organic molecules. For astrometry and geodesy, VLBI help for precise positioning of dense radio sources inside and outside the galaxy, establishment of celestial sphere reference system and earth reference system, measurement of earth rotation orientation parameter and precise coordinates of VLBI observation stations, and movement of continental crustal deformation. In field of deep space exploration, VLBI can provide high-precision measurements of the precise orbit and positioning of lunar and deep space probes, as well as relative motion between extraterrestrial planet surface probes [1].

1.3 Real-Time e-VLBI

Traditional VLBI experiments record the raw observation data into magnetic media, such as tapes and disks. After that, these media are delivered to special data processing center. It usually takes several weeks or months to obtain the final scientific data. The disadvantages of traditional operation mode are: the record data bandwidth of the magnetic recording equipment cannot meet the rapid increasing requirement of ultra-wideband observation, no real-time monitoring and adjustments of the observation instruments status, no easy way for new technologies and devices rapid testing and verifications. With the development of high-performance commercial computers and high-speed network data transmission, real-time electronic transmission VLBI (e-VLBI) technology based on digital VLBI data acquisition terminal, recording and high-speed network data transmission and processing emerges.

e-VLBI technology organically connect geographically distributed observation stations and VLBI data processing centers, by data transmission through high-speed cross-regional educational and scientific networks. With e-VLBI technology observation data collection and correlation processing can be done in real time. Broadband observation and distributed processing also become possible. Even more, scientists can check the data online when the observation is going on, that's kind of WYSIWYG. Depend on the bandwidth of networking, e-VLBI mode could be near real-time or real-time, both modes improve resource utilization of radio telescope network, and meet the need of deep space exploration probe by real-time tracking and measurement.

e-VLBI has three real-time operation modes: real-time, near real-time and post transmission. Real-time mode has the shortest data turnout time when doing transfer. Near real-time mode has observation data buffered and then transferring by not so quick or stable speed, with jitter time ranging from ten seconds to several minutes. The post-transmission mode has the raw data recorded in the disks and then played back to the data center through the narrow bandwidth network afterwards. Usually the real-time mode are widely used int e-VLBI observations.

Because of the advantages of e-VLBI, it has been developed rapidly in the past years and has played an important role in radio astronomy and deep space exploration fields. The Chinese VLBI Network (CVN) has successfully applied the e-VLBI technology in the orbit determination mission of China Lunar Exploration Project(CLEP).

2 Chinese VLBI Network

The Chinese VLBI network is under the lead of Shanghai Astronomical Observatory (SHAO) of Chinese Academy of Sciences (CAS), consists of 5 stations and 1 data center (Table 1). These five radio telescope stations are Tianma and Seshan Station of SHAO, Miyun Station of National Astronomical Observatory of China (NAOC), Kunming Station of Yunnan Astronomical Observatory (YNAO), Nanshan Station of Xinjiang Astronomical Observatory (XAO). The VLBI data processing center is located in SHAO's seshan campus.

Table 1 CVN stations information

Station name	Diameter (m)	Frequency (GHz)	Completed year	Dedicated bandwidth (Mbit/s)	Internet bandwidth (Gbit/s)
Seshan (Sh)	25	1.6, 2.3, 5, 6.7, 8.4, 22	1987	10,000	2
Nanshan (Ur)	26	1.6, 2.3, 5, 8.4, 22	1994	200	0.3
Miyun (My)	50	2.3, 8.4	2006	200	–
Kunming (Km)	40	2.3, 6.7, 8.4	2006	200	1
Tianma (T6)	65	1.6, 2.3, 5, 8.4, 15, 22, 32, 43	2012	10,000	2

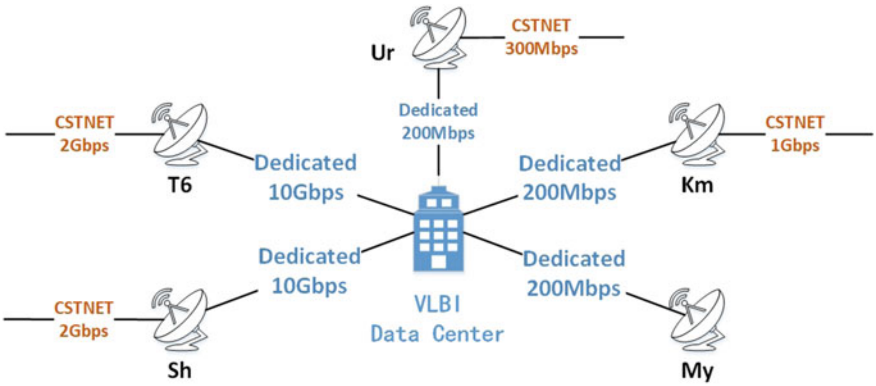


Fig. 2 CVN communication network topology

VLBI stations and the data center are connected by dedicated fiber network links, only used in real-time tracking and measurement system for Chinese lunar exploration probes (Fig. 2). Tianma, Seshan and Kunming stations are also connected to international stations and VLBI data centers through China Science and Technology Network of CAS (CSTNET), been carrying out international astronomy and geodesy e-VLBI data transmission.

In January 2019, Tianma station and 500-m spherical radio telescope (FAST) successfully performed the first joint experimental VLBI observation. The observation got successful clear result, verified the capability of FAST VLBI equipments performance and joint observation possibility between FAST and CVN station.

3 Application of CVN and e-VLBI in CLEP

3.1 Chinese Lunar Exploration Project

The first extraterrestrial planet that mankind explored is the Moon. Since the beginning of this century, China started its own lunar exploration project, proposed a three-step plan of "Orbiting, Landing, and Sample-returning", also known as the Chang'E Project, will completed trilogy in 2020 [2].

The first phase of CLEP began with the launch of Chang'E-1 probe (CE-1) in October 2017 and successfully flied to and orbited the Moon. The first attempt of lunar exploration mission successfully verified and mastered the lunar exploration technologies.

The second phase of CLEP included two missions, Chang'E-2 (CE-2) in 2010 and Chang'E-3 (CE-3) in 2013. The CE-2 mission tested new X-band telemetry, tracking and command system (TT&C) for the first time, later it became the first lunar probe transfer from Moon orbit to Lagrange L2 point in the world. After that it continued the overflight detection of the Toutites asteroid. The Chang'E-3 mission implemented in 2013 was the first time for Chinese lunar probe to soft landing on the Moon, with Yutu rover released for patrol detection of the Moon [3, 4].

The third phase of CLEP includes the Chang'E-5 lunar return reentry tester mission (CE-5T1) and the Chang'E-5(CE-5) mission. In 2014, CE-5T1 mission was successfully carried out and the key technology of the lunar-earth high-speed reentry return technology was verified, which will be put into use by CE-5 probe when returning from Moon.

In 2018, China finished the Chang'E-4 (CE-4) mission. The mission was divided into two periods: the CE-4 relay satellite was launched in May 2018 and successfully entered the Earth-Moon L2 point and circling with Halo trajectory, ready for deep space data relay service. The CE-4 probe, which was launched in December at the same year and became the first man-made probe landed on the far side of the Moon in January 2019 [5, 6].

3.2 VLBI Measurement System for CLEP

Since the start of CLEP, the VLBI system is been part of TT&C system. Before CLEP, the unified S-band measurement and control system (S-USB) was used for near-Earth satellite orbit measurement. The distance of the lunar probe to earth is hundreds of thousands of kilometers to millions of kilometers, which is beyond the system limit of the old USB system. In terms of short-arc orbit determination, the S-USB system has the advantages of fast ranging and speed measurement, but the angle measurement is not good enough at the distance between the earth and the moon. This became a challenge for the telemetry and tracking. With recommendation from Chinese Academy of Science with simulations and analysis proof , the project experts team decided to take advantage of Chinese VLBI network and together with the upgraded USB system, jointly provided tracking service for CLEP.

When the VLBI system has real-time measurement capabilities, it can work perfectly with S-USB system to get accurate orbit and position of the lunar probe. Therefore, since the CE-1 mission, the innovative combination uses of the S-USB + VLBI comprehensive measurement system has been the key technology for TT&C system [7].

The VLBI measurement system is developed and implemented under the lead of SHAO. The system consists of five stations and one command and data processing center.

(1) VLBI Stations

VLBI station have radio telescope, highly sensitive receiver, time-frequency reference system, data acquisition and recording system, monitoring system, control and communication network system and power supply system. The radio telescope reflector converges weak signals from outer space. Receiver system is for filtering

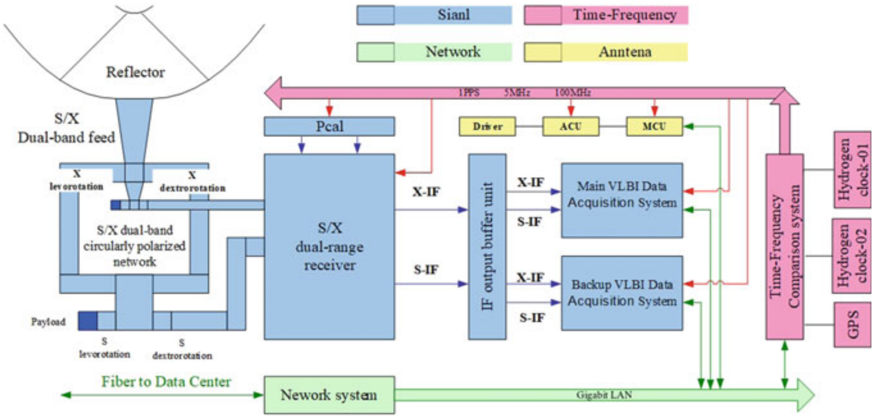


Fig. 3 VLBI station system composition structure diagram

noises and amplifying signals in a specific frequency range. Time-frequency reference system provides stable high-precision time-frequency reference signal source. Data acquisition and recording system is for collecting, recording or transmitting raw observation data with VLBI-specified universal data format. Station monitoring system is for unified collection and display of the overall systems status. The network system is responsible for data transmission and control messages communication. Power supply system provides essential electricity for the station devices.

VLBI station system composition is shown as Fig. 3.

The scene of the each CVN radio telescope are shown as Fig. 4.

During the Chang'E observations, once VLBI system received the observation tasks, data center prepared the schedule files and sent them to each station through the dedicated network. Each radio telescope automatically tracking the probe according to the control schedule files. Raw observation data was acquired and transmitted to Shanghai VLBI data center in real time.

Tianma, Seshan and Nanshan station are all traditional VLBI stations for astrophysics, astrometry, and geodesy observation. Miyun and Kunming station are the Ground Research and Application System (GRAS) of CLEP. GRAS's task is scientific data downlink reception from Chang'E probes. Both My and Km stations are also equipped with extra VLBI data system. With proper task planning, My and Km station are switched from time to time to serve the GRAS and VLBI jobs.

(2) VLBI Center Data Processing and Information System

The VLBI data processing center is the brain of the VLBI measurement system, controlling the overall systems scheduling and coordinated operations. It consists of command and control hall alone with server and network equipments room. During the lunar exploration missions, VLBI data center receives the commands from TT&C headquarter and uploads VLBI measurement results to the headquarter. VLBI center coordinates synchronous observation and data transmission from each station to data center, and then with two parallel data processing

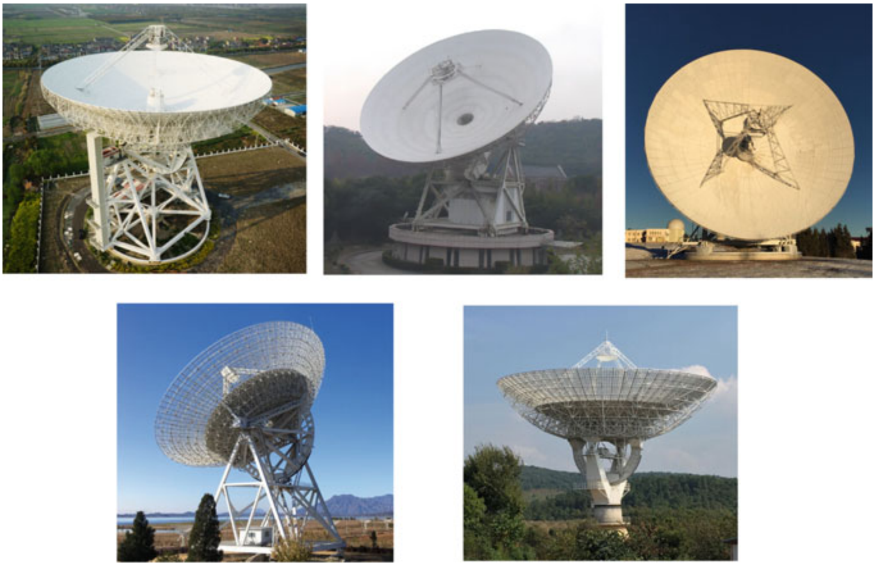


Fig. 4 CVN Station. Top left to bottom right: T6, Sh, Ur, My, Km

pipelines that backup each other the result data is generated in real time. The IP-based audio and video collaboration system connected the headquarter, VLBI center and stations was best for online real-time communication and consultation.

The VLBI center hardware platform is composed of pipeline data processing computing systems, data storage system, network communication system, distributed display system, command and dispatch system, and auxiliary support system like uninterrupted power system, cooling system and so on. The information system is the key component providing high-reliability and real-time computing power to digest data streams. The composition of the pipeline processing hardware platform of the VLBI center is shown in Fig. 6.

Fig. 5 VLBI data center and server room. Left: Centrally control hall, Right: Server and network device room

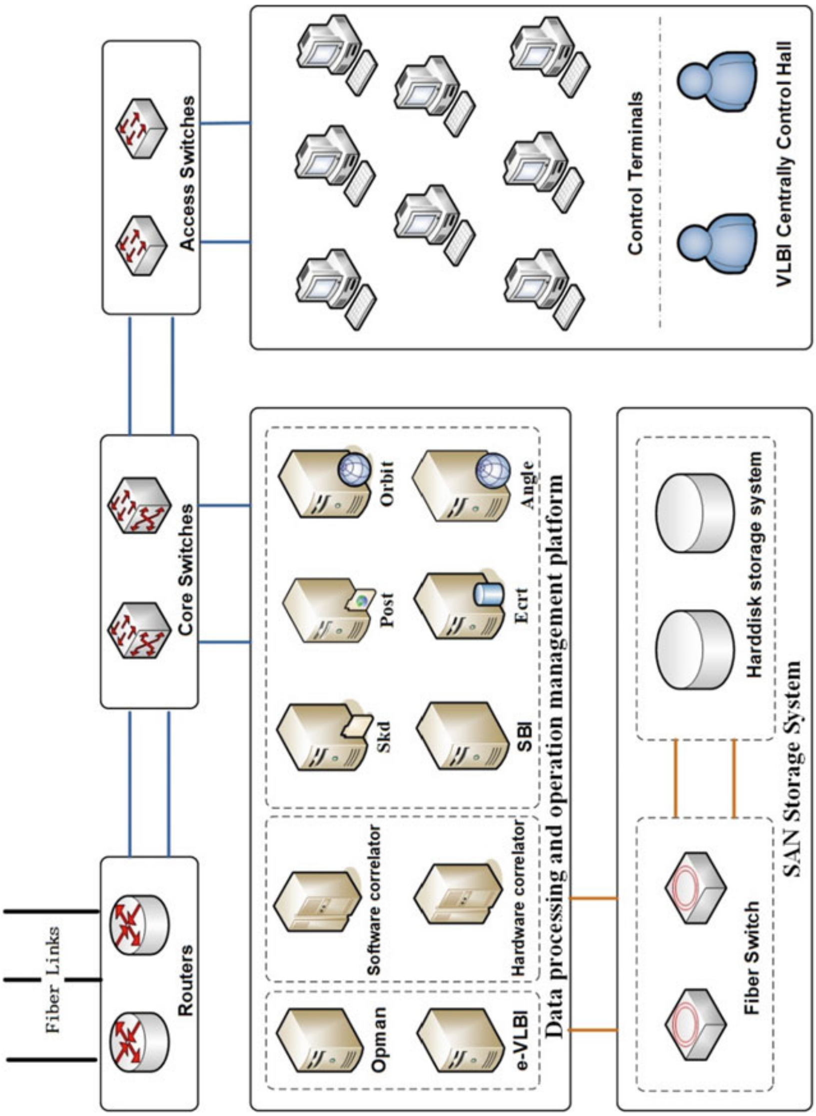


Fig. 6 VLBI center hardware system composition diagram

Data processing pipeline software system can be divided into the following configuration items: schedule, signal simulation, e-VLBI, software and hardware correlators, post-processing, phase delay calc, propagation medium error correction, orbit determination, angle positioning, station status monitoring, and operation manager. Among them, the e-VLBI, post-processing, and operation manager items are the most important software in data processing pipeline.

The functions of each software item of the VLBI center are as follows.

- Schedule (Skd): prepare vex (VLBI Experiment Definition) files needed by stations observing and data center processing
- Signal simulation: signal simulation for lunar and deep space exploration aircraft and calibration radio sources, produce simulation raw VLBI data set
- e-VLBI: control the data acquisition of multiple stations and synchronously transmit data streams to the VLBI center in real time and distribute them to software and hardware correlators
- Correlator: the core of VLBI data processing system. Its function is correlating multi-station raw observation data to obtain single-channel delay, delay rate and phase information of radio source and satellite. There are two types of correlator: software and hardware. Software correlator uses special high-performance data processing software based on MPI and OpenMP parallel computing techniques, running on a high-performance computer cluster platform. The hardware correlator uses a dedicated signal processing platform based on Field Programmable Gate Array (FPGA) chips for parallel data processing. Software and hardware correlators back up each other, providing high reliable real-time correlating capabilities
- Post-processing: its input source data is correlators products, with propagation medium error correction models and using bandwidth synthesis method to get accurate delay and delay rate results
- Same beam processing (SBI): receiving the signals of the two space targets in one radio telescope same beam, and with this special kind data it calculates the differential phase of two targets to eliminate various propagation medium errors when these radio signals reaches the telescope
- Propagation medium error correction (Ecrt): generate the models of signal propagation errors introduced by the earth's atmosphere, ionosphere and other medium. This information is necessary for other data processing software items and is crucial for final data accuracy
- Orbit determination: Obtain precise lunar probe orbit information during the flight to or circling around the Moon
- Angle positioning: Obtain the precise position information of the lunar lander on the Moon surface
- Station status monitor: remotely monitoring and centrally display various devices status of VLBI observation stations. It dynamically shows operation parameters and status of radio telescopes, receivers, data terminals and time-frequency equipment, etc.

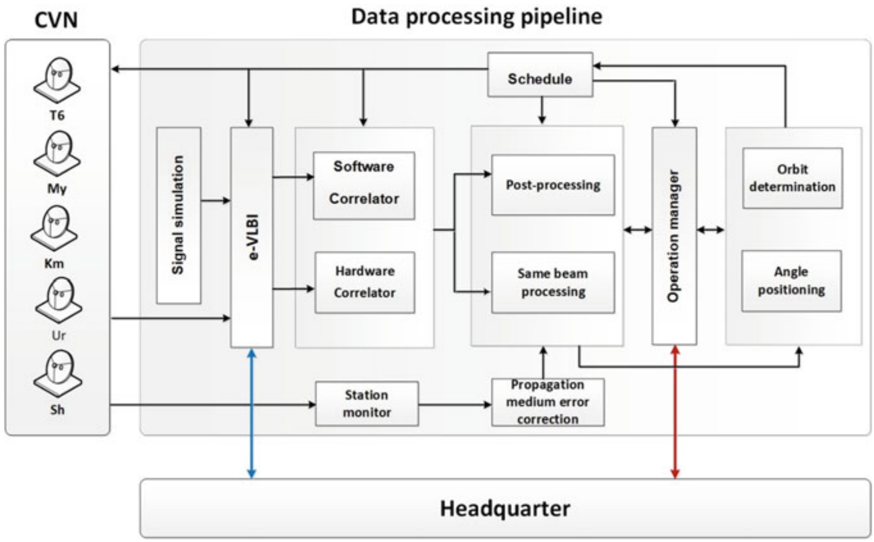


Fig. 7 Data processing pipeline and operation diagram

- Operation manager: Receive the task information for VLBI system including observation plans and other instructions issued by the TT&C headquarter, and send the VLBI data products to the headquarter

The data processing pipeline of the VLBI center is shown as Fig. 7.

(3) Real-time e-VLBI system

Due to timeliness requirement of space probe tracking and measurement, it's necessary to upgrade the traditional VLBI astronomical observation network with real-time capability. Real time e-VLBI sub-system is the key for real-time system upgrading. Recent years with serveral technological upgrading such as VLBI acquisition terminal transformed to digital system, ×86 computer based platform of raw observation data collection, transmission and distribution, software correlator based-on high-performance computing technology, real-time data stream pipeline processing technologies, CVN has completed the technical innovations and was capable for real-time lunar mission [8].

Concerning e-VLBI digital data terminal development history, in the first phase of CLEP, CVN kept using the traditional VLBI analog signal terminal and developed a new hard disk recording and transmission system named CVNHD (CVN Highspeed Disk System) and deploied them to each CVN station. So CVN had near real-time tracking and observation capability with 16 Mbps/station data rate. In the second phase of CLEP, a new VLBI standard high-speed disk recording system Mark5B was adopted. And the raw data rate increased to 32–64 Mbps/station. In the third phase of CLEP and the CE-4 missions, a more advanced broadband digital terminal and ×86 server-based recording and transmission system developed, which further increase e-VLBI data rate to 128 Mbps/station.

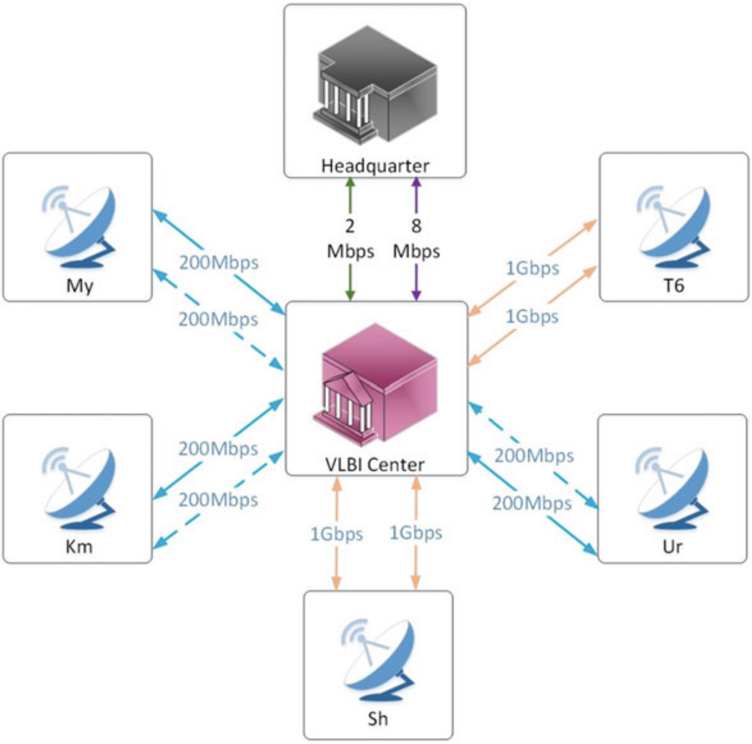


Fig. 8 e-VLBI network link topology

In order to ensure the stability and reliability during the long term run of Chang'E missions, each station was established two dedicated fiber links to VLBI data center to make sure uninterrupted e-VLBI data stream synchronized transmission, as shown in Fig. 8.

In the network device layer of communication system, the Open Shortest Path First Interior Gateway Protocol (OSPF) and Bidirectional Forwarding Detection (BFD) protocol are used together to provide an active-standby backup way to assure high-reliability. The network link switching time is reduced from the seconds to the milliseconds if one link was broken, which greatly reducing the jitter time of e-VLBI data retransmission, made no obvious time delay among stations data streams.

3.3 Application in CLEP Phase I

Before involved in CLEP, CVN only had 2 + 1 observation stations (Sh and Ur Station, Yunnan Kunming Mobile Station) and two-station FX-type hardware correlator. The FX-type hardware correlator was based on special designed ASIC-chip technology, at that time the general computer performance was far more enough for VLBI correlation. The data exchange between the station and correlator center depended on high density tapes. There was no possibility to do near real-time or real-time VLBI observation by tape system. In order to meet the requirements of quick turnout of Chang'E probe observation, CVN developed a number of key technologies including: VLBI space probe tracking observation technology, real-time data acquisition, recording and transmission system, real-time correlator, S-band post-processing, lunar probe orbit determination and positioning, etc.

Here is the key components of system construction and technologies developments.

(1) CVN shared use with two big telescopes of GRAS My and Km station, together with existed stations, comprised a reasonable layout of east-west and south-north baseline in China. Its longest baseline length was more than 3200 km. This network increase the observation time slot and space resolution than the previous CVN network.
(2) Due to the low speed and relative high bit error rate of the magnetic tape system, it is not suitable for high-precision real-time VLBI data recording and transmission. To solve this limitation, SHAO developed a near real-time data recording, transmission and playback system CVNHD. It's a harddisk system with special design based on × 86 industry computer platform. CVNHD used commercial high-speed A/D sample card for VLBI raw data acquisition, which improved the data rate and eliminated the instability of tape system. CVNHD system provided essential conditions for carrying out various observation experiments and tests before CE-1 mission. By CVNHD system, 4 VLBI stations were successfully connected to VLBI data center by networking data stream, which transformed radio observation network to aerospace observation capable network. In order to solve the bottleneck of data transfer rate, the default Linux network stack was adjusted and tuned to maximize network data throughput. By increasing the kernel memory buffer of TCP and UDP protocol, utilizing Ping-Pong memory buffer mechanism in data transfer software, carefully choosing kernel congestion control algorithm, e-VLBI system parallelly and synchronously transfered multiple station data to VLBI center. In the first phase of CLEP, the real-time performance of the e-VLBI system was averaged by 40s.
(3) The core of VLBI data processing system is the correlator, it's a special designed data processing system for VLBI. In the system design period, VLBI data center decided to use twotype solutions based on hardware and software platform. The software correlator was a stand-alone version of the multi-station data correlating system with fringe searching function integrated in. The fringe searching function is for precise correlator model reconstruction come into play when the lunar probe orbit maneuver or the initial theory orbit is too coarse. The software correlator system was developed by C language, using multithreading and OpenMP libraries to achieve multi-core parallel computing acceleration. In the CE-1 mission, the software correlator played a key role in the data processing pipeline of the VLBI center. The hardware correlator was developed on FPGA chip and was the another correlating system leaded the second data processing pipeline as backup.

(4) Post-processing obtains the multi-channel delay and delay rate data results in the S frequency band, which is one of important data products sent directly to the TT&C headquarter. The post-processing software is the most sophisticated scientific computing software in VLBI data process. It adopted dynamic configuration methods to achieve flexibility while still capable of interactive analysis. A real-time graphical display interface helps instantly display data results as the observation was going on, provided operational status at the data quality level.
(5) In the CE-1 mission, TT&C system used USB and VLBI systems for joint orbit determination. The orbit determination software in VLBI center use both data sets of USB and VLBI to do joint orbit calculation for each observation segments such as Earth-Moon midway trajectory correction, lunar orbit capturing, and circumlunar trajectory. Orbit determination software also can give near term prediction orbit based on historic accumulated data before next probe observation.
(6) In the pipeline data processing system of the VLBI center, rapid data exchanging between heterogeneous software systems rely on a special tuned Network File System (NFS) with fine control of access permissions. It simplifies software data interface design and provides intermediate data storage.

By attended the CE-1 mission, CVN broke through the limitation of traditional system and mastered the lunar probe tracking and measurement technologies, especially established data processing system for space aircraft observation. The VLBI system got real-time performance of 6–7 min which was better than the design target of 10 min for VLBI tracking system. That's the first giant leap for CVN.

3.4 Application in CLEP Phase II

(1) Chang'E-2 mission

The launch of Chang'E-2 was the begging sign of second phase of CLEP. It's a pilot test probe for CE-3, aimed for key technologies verification. VLBI system attended the new X-band telemetry and tracking technology verification included differential doppler signal (DOR) interferometry technique [9].

X-band DOR signal has higher frequency than the S-band, by applying a higher-precision propagation medium error correction model to new adaptability improved data processing software, VLBI system contributed the probe detection accuracy to better than 1 mm/s and 100 m, and all data products were obtained within 10 min. Through the software optimization and hardware upgradation on CE-1 VLBI system, the raw data rate increased to 32 Mbps/station.

After successfully supported CE-2 formal task, the VLBI system continued to participate in the expansion experiments periods, including transfer from the lunar orbit to the Lagrange L2 point between Sun and Earth, and thereafter flew to the Tutattis asteroid to verify the TT&C capability with maximum distance of 80 million kilometers. The VLBI system had showed the very important tracking and measurement ability for the space distance more than 80 million kilometers from earth.

(2) ChangE-3 mission

The CE-3 mission was the first time for China to soft land on an extraterrestrial planets surface. This very challenging mission posed new requirements for VLBI system in the short arc trajectory measurement. During the system developments for CE-3 mission, VLBI broke through a series of key technologies and built some new systems, such as new generation digital baseband VLBI data terminal, ΔDOR (Delta Differential One-Way Ranging/Doppler) measurement technology, newly built Tianma 65-m radio telescope, real-time VLBI data processing, same-beam phase delay measurement for two nearby lunar targets, high-precision orbit determination and positioning [10].

The brand new digital VLBI data terminal replaced the old analog ones, which is capable of multibit sampling. ΔDOR observation and corresponding data processing system with new models and algorithm was developed for higher precision measurement. Same beam observation and data analysis system specified created for relative motions detection of the CE-3 lander and Yutu-1 lunar rover. The new big telescope joined into CVN network was Shanghai Tianma radio telescope which greatly improved the whole CVN sensitivity and resolution for lunar spacecraft tracking. VLBI data center moved to new building and updated new IT hardware platform which was conducive to software performance. VLBI phase reference mapping method in radio astronomy was also creativity applied to get the accurate location of the CE-3 lander on Moon mare surface.

The real-time e-VLBI system upgraded as well. First, with the new generation VLBI digital baseband converter system CDAS (Chinese Data Acquisition System) wider frequency band sampling and higher rate data collection was possible. High-speed real-time data collection, local storage and transmission system based on the commercial ×86 COST (Commercial Off The Shelf) computers was developed and put in use with VLBI standard Mark5B data format support. New e-VLBI system had the capability of synchronous data collection and dual-channel distribution from 5 stations to VLBI data center, and also had the functions of data format decoding and error correction on the software correlator front side. The real time performance of data stream transfer of new e-VLBI system was reduced from 30 to 1s. At the same time, the parallel data processing pipelines in the VLBI center were upgraded both in architecture and algorithm to meet 1-min real-time requirement for the whole VLBI system. To improve data exchange efficiency, new NAS (Network Attached System) and NFS protocol had been carefully adjusted to shorten the directories and files caching time and latency, minimized the attribution modification operations so as to help real-time data processing systems coupled more tightly. The overall VLBI system real-time performance for CE-3 was finally boosted from 10 to 1 min [11].

ΔDOR observation is one kind of differential observation way for deep space probe tracking. When observing, radio telescope shift pointing targets between the spacecraft and nearby reference radio sources on celestial sphere. This observation mode can eliminate common errors exist in signal receiving chain of each station. Corresponding computing systems for ΔDOR observation data in VLBI center were also setup. Key units such as DOR data correlating system, post processing dedicated for DOR signal, joint orbit determination and positioning based on X-band USB and VLBI data, are successful developed and passed the system proficiency testing before actual use. For precise relative motion detection of CE-3 lander and rover, T6 telescope is the most suitable for same beam interferometry observation and new kind of SBI data processing system joined VLBI data processing pipeline. SBI software could extract the differential delay and delay rate of two lunar objects and capable of real-time accurately monitoring relative motion [12, 13].

Technological innovation and system upgradation happened in stations and data center was a big leap for CVN. These improvements are not only better served the CE-3 mission but also laid an important foundation for the subsequent lunar and deep space exploration missions. Because of the VLBI system's work and high accuracy results during CE-3 traskcing and measuring, TT&C system headquarter wrote a letter phrasing VLBI system for its "Outstanding Contribution".

3.5 Application in CLEP Phase III

The third phase of CLEP is the CE-5T1 test mission and CE-5. The main purpose of the CE-5T1 mission was to test the lunar return spacecraft and verify the high-speed earth orbit reentry return technology.

The VLBI system accurately tracked the CE-5T1 probe as usual and provided high-precision orbit results during the high-speed reentry when the probe returned from Moon. This technical verification mission provided guarantees and paved the way for CE-5 official mission. On 24 November 2020, CE-5 was lunched and started its way to the Moon.

3.6 Chang'E-4 Mission

The CE-4 probe, a back-up for CE-3, the scientific goal was adjusted to land manmade detector on the South Pole-Aitken Basin, far side of the Moon [14]. There is no direct way to do communicate with the ground station from back side of the Moon. So before CE-4 lander and rover mission, a relay satellite "Que Qiao" was launched May 2018 and sent to orbit around the Lagrange L2 point between earth and Moon for relay communication. During the CE-4 soft landing on the lunar far side, relay satellite was providing continuous relay communication services for CE-4 probe and ground station. So the accurate orbit of Que Qiao was crucial as it's the bridge for telemetry and communication. For VLBI system, a time-sharing working mode was created to successive observation of CE-4 probe and relay satellite. The switching time of two types observation is within 70 min. The CE-4 and the relay satellite observation still use the radio source calibration as system calibrator. The observation source change strategy is “5 min radio source—5 min CE-4 probe or relay satellite”.

From May to June 2018, VLBI system successfully served the trajectory measuring task of the relay satellite. From December 2018 to January 2019, VLBI system successfully provided tracking service for CE-4 probe landing.

For the CE-4 mission, the main innovations of the VLBI system includes: the first time S-band ΔDOR VLBI observation and signal data processing, the first time achieved the precise Halo orbit of the lunar relay satellite at Lagrange L2, realized rapid switching of the time-sharing observation and supported two different space probes.

3.7 Technical Performance Statistics

(1) Real-time performance

The traditional radio astronomy observation network has upgraded from none-time-sensitive system to real-time tracking system for CLEP. The e-VLBI data system achieved high reliability of real-time data transmission with zero data lost when the network link or devices failure occurred. The overall real-time performance is better than requirement assigned by TT&C headquarter [15]. Table 2 summarizes the statistics of the real-time work of previous tasks.

Table 2 e-VLBI operation status for CLEP

Mission	Transfer mode	Station to center network link	VLBI data rate	Link switching	Real-time target	Real-time actual
CE-1	Near real-time	SDH 34 Mbps, IP-VPN (100 Mbps)	16 Mbps	Manual >30 m	<10 m	<6 m
CE-2	Near real-time	2 × MSTP (100 Mbps)	32 Mbps	Manual >15 m	<10 m	<4 m
CE-3	Real-time	2 × SDH 155 Mbps	64 Mbps	Automatic <3 s	<1 m	<40 s
CE-5T1	Real-time	2 × SDH 155 Mbps	64 Mbps	Automatic <3 s	<1 m	<40 s
CE-4	Real-time	2 × MSTP 200 Mbps	128 Mbps	Automatic <3 s	<1 m	<40 s

(2) Data products quality

The direct measurement results of the VLBI network for the spacecraft is signal delay and delay rate. Measurement methods covers △VLBI, △DOR and SBI. The quality of VLBI data products are shown in Table 3.

(3) Data volume

After long term servings for CLEP, VLBI system has accumulated lot of data product. It's not only for real-time tracking task, but also could be used to carry out other scientific researches such as interstellar medium and solar wind. VLBI data products statistics is showed as Table 4.

Table 3 Measurement accuracy of previous lunar missions

Mission	Frequency	Obs. mode	Delay target (ns)	Actual delay (ns)
CE-1	S	△VLBI	12	6
CE-2	S	△VLBI	12	5
CE-3	X	△DOR, SBI	4	1
CE-5T1	X	△DOR	4	1
CE-4 relay probe	S	△DOR	5	1
CE-4 lander&rover	X	△DOR	3	0.6

Table 4 Statistics and data volume of data product

Mission	Data products	Data volume	Realtime work time	Extended work time
CE-1	Skd, Correlator, Post-process, orbit, angle, Ecrt	17 TB	1 month	6 months
CE-2	Skd, Correlator, Post-process, orbit, angle, Ecrt	31 TB	1 month	26 months
CE-3	Skd, Correlator, Post-process, sbi, orbit, angle, Ecrt	6.8 TB	1 month	4 months
CE-5T1	Skd, Correlator, Post-process, sbi, orbit, angle, Ecrt	1.8 TB	2 weeks	6 months
CE-4 relay probe	Skd, Correlator, Post-process, sbi, orbit, angle, Ecrt	2.3 TB	1 month	2–8 years
CE-4 lander&rover	Skd, Correlator, Post-process, sbi, orbit, angle, Ecrt	2.2 TB	1 month	/

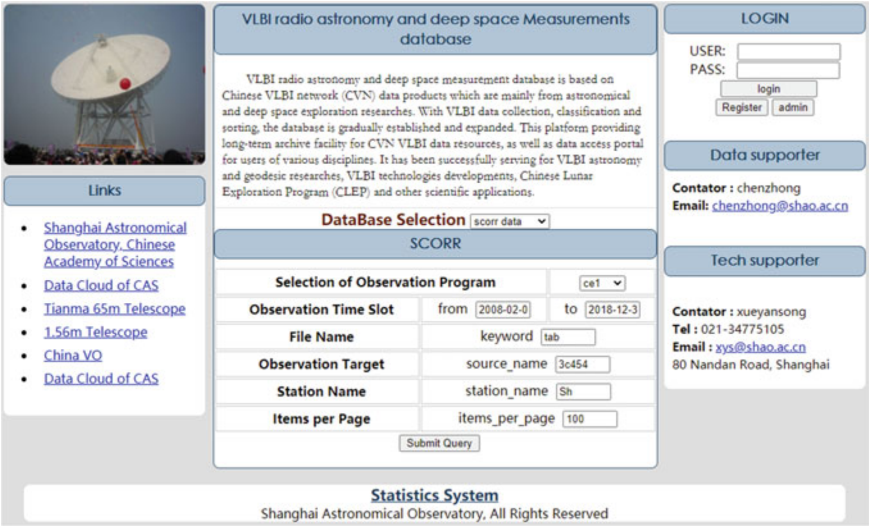


Fig. 9 VLBI radio astronomy and deep space exploration database

After the data protection period, VLBI data products will be public available on web site (https://www.vibi.csdb.cn) (Fig. 9). The web hosted on the national basic scientific data sharing service platform, SHAO is responsible for web site maintenance and data sharing service.

4 Future Plans

Following three phases of CLEP, China will continue its aggressive lunar and deep space exploration programs. Next one is Mars exploration program and it was launched at the end of July 2020. Following may be asteroid exploration plan. The fourth phase of CLEP is under discussions. For the long term there are some proposals about Mars sampling and return program. CVN and VLBI measurement system will continue to provide real-time observation and precise orbit and angle determination services for these space exploration programs. With different space missions being carried out the time slots would be very likely overlapped. The number of existing observation stations and VLBI data center capabilities is not enough for multiple deep space exploration missions, especially in need of parallel tasks execution. VLBI has made plan to establish two new 40-m radio telescope in Tibet Autonomous Region and JiLin province, northeast of China. Also planning transfer a 40-m antenna in ShanXi province to be VLBI observation capable. With totally 7 domestic stations, two subnets could be configured for concurrently tracking two targets at the same time. The optimized configuration of the observation subnet in the nurture is shown as Fig. 10.

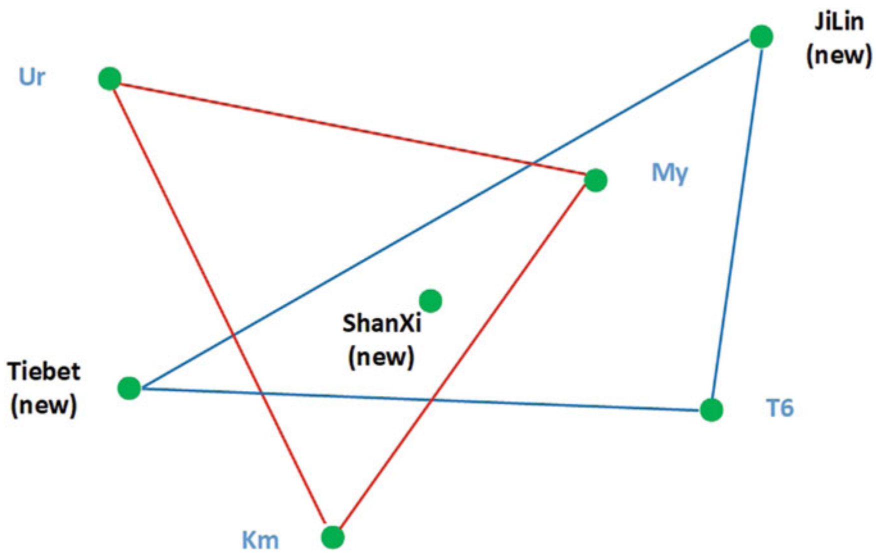


Fig. 10 Future plan of VLBI two subnet configuration

The VLBI data center also needs to expand the data processing system capacity and develop new software pipeline system suitable for dual subnet data processing. To improve the level of automation and intelligent operation, multi-task and multi-target cooperative operation system (MTMOS) is essential for automatic operation, especially in the circumstance of the complexity of two subnet and two pipelines. With MTMOS, manual operations could be replaced by efficient machine operations with fewer people. The preliminary architecture design of the multi-task multi-target VLBI system is shown in Fig. 11.

There is another plan to use the 4.2m-diameter antenna on the lunar relay satellite to carry out joint observation between the lunar orbit VLBI telescope and ground station to test the lunar-earth VLBI ultra-long baseline experiments.

SHAO is also proposing a space low frequency VLBI program. In its plan two 30-m (or even larger) space radio telescopes will be launched with an orbital altitude of up to 90,000 km (The baseline of the telescope to the Earth is up to 10,000 km). The space telescope operating frequency is between 30 MHz and 1.7 GHz (including the four main bands of 30 MHz, 74 MHz, 330 MHz and 1.67 GHz). The highest resolution is 2mas at 300 MHz and 0.36mas at 1.7 GHz. The space low frequency radio telescope array program is not only a big step on the basis of the ground VLBI array, but also a giant improvement in angular resolution and sensitivity by several folds. The unique low-frequency band is an area has not been touched by previous space VLBI projects. With these space areas, scientific breakthroughs in fields of cosmology, gravitational waves, and exoplanets is expected to be achieved. All international cooperation are welcome.

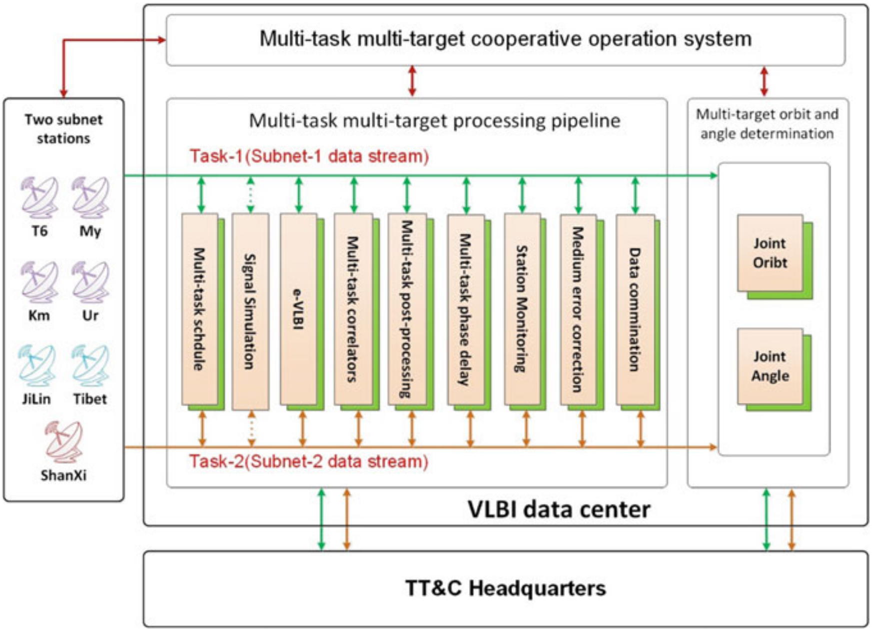


Fig. 11 Preliminary architecture design of the multi-task multi-target VLBI system

Acknowledgements Thanks to National Natural Science Foundation of China-Joint Foundation for Astronomy of Chinese Academy of Sciences (U1931135), the Chinese Academy of Sciences Youth Innovation Promotion Association, the Chinese Academy of Sciences Key Laboratory of Radio and Shanghai Space Navigation and Positioning Laboratory and Fundamental Science Data Sharing Platform (DKA2017-12-02-09) for their support of this work.

References

1. Qian Z, Li J (2012) Application of very long baseline interferometry in deep space exploration. China Science and Technology Press
2. Ouyang ZY (2004) Scientific objectives of Chinese lunar exploration project and development strategy. Adv Earth Sci 19(3):355–357
3. Yu D et al (2016) Overview of the development of lunar exploration engineering technology in China. J Deep Space Explor 3(4):307–314
4. Dong G et al (2018) Construction and technology development of China's deep space measurement and control system. J Deep Space Explor 5(2):99–114
5. Wu W et al (2019) Chinese lunar exploration program. J Deep Space Explor 6(5):405–416
6. Ye P et al (2014) A preliminary study on the development history and experience of China's lunar probe. Sci China: Tech Sci 44(6):543–558
7. Zheng W et al (2012) e-VLBI applications of Chinese VLBI network. In: International VLBI service for geodesy and astrometry 2012 general meeting proceedings, pp 99–102
8. Chen Z et al (2015) Current situation and development trend of software correlator. Astron Adv 3(4):489–505

9. Wu W et al (2013) High precision VLBI technology based on DOR signal. Sci China: Inf Sci 43(2):185–196
10. Wu W et al (2013) Engineering and technology of deep space TT&C communication system. Science Press
11. Zheng W et al (2014) Real-time and high-accuracy VLBI in the CE-3 mission. In: International VLBI service for geodesy and astrometry 2014 general meeting proceedings, pp 466–472
12. Huang Y et al (2014) Orbit determination of ChangE-3 and positioning of the lander and the rover. Chin Sci Bull 59:29–30
13. Liu Q et al (2010) Precise relative positioning of lunar rover based on VLBI technology of ultra-high precision multi-frequency and same beam. Sci China: Phys Mech Astron 40(2):253–260
14. Ye P et al (2019) Mission design of Chang'E-4 probe system. Sci China: Tech Sci 49(2):124–137
15. Chen Z et al (2015) Establishment and application of e-VLBI network in China. Annu J Shanghai Astron Obs 36:136–147

Zhong Chen, Ph.D., Radio Astronomy Science and Technology Laboratory, Shanghai Astronomical Observatory, Chinese Academy of Sciences. Main research field is very long baseline interferometry technology and its applications. In the VLBI system of TT&C for Chinese lunar exploration program, is responsible for the development and construction of the VLBI data center, real-time e-VLBI system and distributed network communication system. During the second and third phases of CLEP, served as the chief designer of the VLBI data center and networking system. Also acted as chief designer of the VLBI overall technical group for the Chinese first time Mars exploration program. He is a member of the Youth Innovation Promotion Association of the Chinese Academy of Sciences. Awarded as "Outstanding Contributor in Chang'E-3 mission of CLEP" honour. A member of two times of First Prize of Shanghai Science and Technology Progress (Group).

International Cooperation Program for Major Microbial Data Resources: Global Catalogue of Microorganisms (GCM)

Juncai Ma, Linhuan Wu, and Jianyuan Zhang

Abstract Established in the 1960s, WFCC-MIRCEN World Data Centre for Microorganisms, hereinafter referred to as WDCM, is the most important physical resource data platform in the field of microorganisms throughout the world. Getting hosted by the Institute of Microbiology, Chinese Academy of Sciences in 2010, WDCM is the first world data center in the field of life sciences in China. Taking WDCM as a platform, the Institute of Microbiology, Chinese Academy of Sciences insists on developing “self-reliant” international cooperation and advocates Global Catalogue of Microorganisms, hereinafter referred to as GCM, which is an international cooperation program for major microbial data resources. GCM program aims to provide a globally uniform data warehouse for valuable microbial resources scattered in various culture collections and in the hands of the scientists around the world. Currently, 127 microbial resource collection agencies in 46 countries and regions have officially participated in the program, providing effective data support for all aspects of physical resources of microorganisms such as gathering, collection, transnational transfer, academic and commercial applications and benefit sharing, and offering the most important support to the implementation and enforcement of the *Convention on Biological Diversity* in the field of microorganisms. On the basis of GCM international cooperation program, WDCM launched GCM2.0 international cooperation program–Global Microbial Type Strain Genome and Microbiome Sequencing Project–for complete coverage of microbial genomes, and established cooperation network for genome sequencing and function exploring of microbial resources covering 30 major culture collections in more than 20 countries, which is expected to complete genome sequencing of more than 10,000 microbial type strains, establish a set of international standard system in microbial resource sharing and exploration, and set up a globally authoritative reference database and data analysis platform of microbiome.

Keywords Microorganisms · Genomes · Sequencing

J. Ma (✉) · L. Wu · J. Zhang
Institute of Microbiology, Chinese Academy of Sciences, Beijing, China
e-mail: ma@im.ac.cn

China’s e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_10

1 The Context of GCM Program

WFCC-MIRCEN World Data Centre for Microorganisms (WDCM Fig. 1), was set up by WFCC in the 1960s, and it is the most important microbial resource data platform in the field of microorganisms throughout the world. In 2010, WDCM got hosted by the Institute of Microbiology, Chinese Academy of Sciences. It is the first world data center in the field of life sciences in China. Ma Juncai, director of The Center for Microbial Resource and Big Data, Institute of Microbiology, Chinese Academy of Sciences, serves as the chairman of the WDCM in China and presides over the work of the center [2]. Taking WDCM as a platform, the Institute of Microbiology, Chinese Academy of Sciences insists on developing "self-reliant" international cooperation and promotes the global informatization construction of microbial resources to a new level by advocating Global Catalogue of Microorganisms (GCM Fig. 3), which is an international cooperation program for major microbial data resources (Fig. 2).

Fig. 1 The website of WDCM

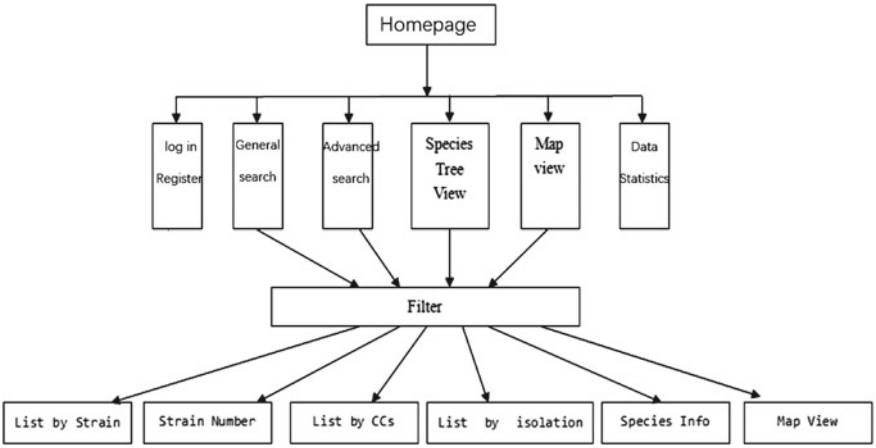


Fig. 2 Data management and service of GCM data platform

Fig. 3 The Website of GCM

2 GCM1.0 Global Catalogue of Microorganisms

The GCM program aims to provide a globally uniform data warehouse for the valuable microbial resources scattered in various culture collections and held by scientists around the world, and to offer information services of microbial stain resources to scientific and industrial circles and industries around the world in the form of a uniform data portal. So far, GCM has integrated detailed information of more than 440,000 microbial physical resources, including microorganisms from special ecological environment with great scientific research and industrial application value. As a big data platform for the integration of microbial digital resources, GCM has also adopted advanced data exploring methods to further extract information on the subsequent research and utilization of microbial resources from more than 6 million published microbial documents and patents worldwide. Therefore, the information platform could provide effective data support for all aspects of physical resources of microorganisms such as gathering, collection, transnational transfer, academic and commercial applications and benefit sharing, and it offers China's most important support to the implementation and enforcement of the *Convention on Biological Diversity* in the field of microorganisms. The integration development framework of the platform adopts java MVC framework and separates the front and back of the integration website to minimize the mutual interference between information storage and information retrieval. MySQL database is used for data storage. In data table design, the basic information and exploring information of strains are differentiated regularly. Currently, the size of the database file is about 20 GB. In the aspect of website monitoring, the monitoring information is updated on a daily basis by using the method of web container access log statistics. Those contents that are monitored are as follows: page view, IP visits, access area, download amount, zone time, etc. GCM relies on strain information offered by culture collections around the world to provide Internet users with query and retrieval, data statistics, literature association, separated sources and indexing of collection sites, etc.

Up till now, a total of 127 microbial resource collection institutions from 48 countries and regions such as the United States, France, Germany, the Netherlands, etc. have officially participated in the program. Meanwhile, it has also established substantive cooperation with regional networks such as ACM, ANRRC, EMbaRC and national networks of Russia, Thailand and Portugal to provide regional data management and sharing with the GCM platform. In order to cater to the implementation of the national strategy of "The Belt and Road Initiative", we have proposed "The Belt and Road Initiative" microbial data resource sharing program on the basis of the GCM platform. The program has been supported by a number of microbial resource institutions. The implementation of the cooperation program could help improve the informatization management level of microbial resources, fulfill the exploration and utilization of the microorganisms with important functions, and promote the development of biotechnology and bio-industry.

3 GCM2.0 Global Microbial Type Strain Genome and Microbiome Sequencing Project

In October 2017, led by the Institute of Microbiology, Chinese Academy of Sciences, GCM2.0 Global Microbial Type Strain Genome and Microbiome Sequencing Project was jointly launched by 12 countries in the world. The launched program is the second phase of the GCM program, including type strain genome sequencing and information analysis, sharing and application of the sequencing data, which is a mighty supplement to the existing GCM1.0 strain data information. The good cooperation between the program and various microbial resource culture collections around the world accumulated in the earlier stage has provided a solid foundation for the resource acquisition of the program, which is also an important condition for us to lead the program.

On October 12, 2017, at the "7th WDCM Academic Seminar", Ma Juncai, the director of WDCM and also the director of the Center for Microbial Resource and Big Data, Institute of Microbiology, Chinese Academy of Sciences, announced the official launch of the Global Microbial Type Strain Genome and Microbiome Sequencing Project led by WDCM and the Institute of Microbiology, Chinese Academy of Sciences and jointly initiated by the culture collections of 12 countries in the world (Fig. 4). At present, 20 culture collections from more than 12 countries such as ATCC of the United States, JCM and NBRC of Japan, KCTC of South Korea, etc.

Fig. 4 The official launch of GCM2.0 global microbial type strain genome and microbiome sequencing project

have joined in the program. The strain resources collected by these culture collections have covered more than 90% of the already known type microbe, of which ATCC and JCM are the largest and most influential type microbe culture collections in the world, which could ensure the availability of resources.

3.1 Whole Genome Sequencing of Type Strains: An Important Breakthrough Point for Decoding the Correlation Between Gene Composition and Function

Type strains are strains that are preserved in pure (reproducible) state as criteria for classification concepts in the process of naming, classifying, recording and publishing microorganisms. At present, there are more than 8,000 sequenced microbial genomes, but the coverage of species is uneven in that a large number of type strains have not been covered, and the data quality is varied, making it impossible for the data to be referred to. It has resulted in a large number of gaps in the systematic classification and genome annotation, etc. of microbial data, which makes the analysis fail to come to an end.

With the diversity of its genes and metabolisms, microorganism is an ideal tool for biotechnology research. In recent years, CRISPR/Cas9 gene editing system from bacterial immune system has rapidly become the most popular technology in life sciences. The whole genome decoding of type strains will make it possible to study the correlation between gene composition and function (such as metabolic activity, virulence, antibiotic production, biomass synthesis, biological fixation of nitrogen, etc.), making great contribution to the research of ecology and biochemistry, and will also further accelerate the discovery of new natural products and drugs. Because it is difficult to be cultured, a large number of valuable microorganisms have not been studied, developed and utilized, and the research methods for the composition and function of environmental and human related microbiome need to be developed urgently. The key of using metagenomic method to analyze microbiome is to obtain high quality genomic reference data. Therefore, the sequencing of type strains will also become an important breakthrough point for microbiome study. Meanwhile, with the reduction of sequencing cost and the improvement of massive data analysis ability, it has become the general trend to launch large-scale sequencing plans and carry out researches based on sequence analysis and function exploration. By taking advantage of our superiority in organization, cost, human resources and technologies, we will seize the highland of international strategic biological resources. Through leading the organization of the program, we will also realize a genuine China-led international cooperation guided by Chinese standards, Chinese databases and Chinese scientists.

3.2 *The Network Establishment Program of the International Microbial Type Strain Genome and Microbiome Sequencing Cooperation*

The program will complete genome sequencing of more than 10,000 bacteria, fungi and archaeobacteria type strains within 5 years, covering all currently known bacteria and archaeobacteria type strains and important fungi type strains, and establish the international microbial type strain genome and microbiome sequencing cooperation network, covering 30 major culture collections in more than 20 countries [1]. It will also select type strains of microorganisms (including bacteria, archaeobacteria and culturable fungi) that have not been sequenced up till now from the microbial resource culture collections throughout the world and complete genome sequencing of more than 90% of the total type strains of microorganisms (Fig. 5).

1. Deposit strains in culture collections in at least two different countries and obtain receiving numbers.
2. Register the receiving number in GCM for the type strain to be a candidate for sequencing, and send the DNA sample to WDCM.
3. The sequencing results are returned.
4. Inform WDCM of the storage number of the sequenced strain confirmed by the culture collection.
5. The sequencing data is associated with the storage number of the culture collection.
6. Authoritative release of original data and annotation data.

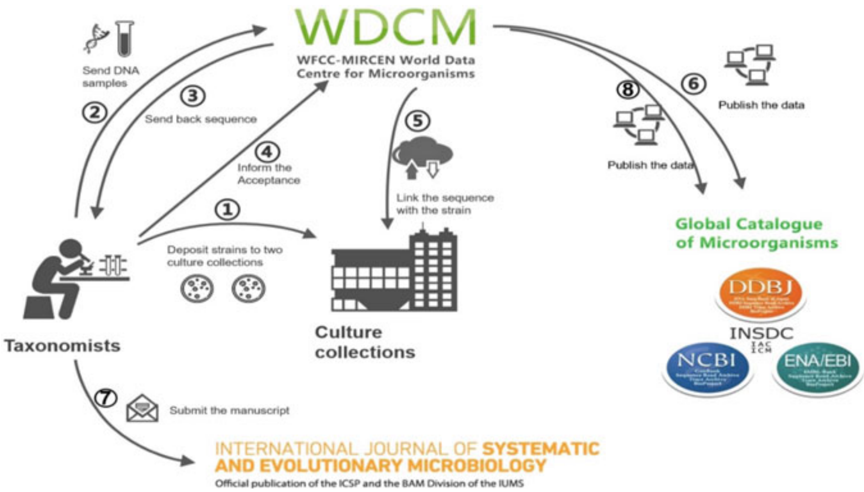


Fig. 5 The flow chart of the international microbial type strain genome and microbiome sequencing cooperation network

7. Submit the paper containing the storage number of the type strain to an authoritative journal for public publication.
8. The sequencing data is published publicly by the authoritative database.

Data standard is the key to successful implementation of a program. The *Data Management and Data Catalog Standard for Microbial Strain Resources*, the establishment of which was led by WDCM, has been officially approved by ISO TC 276 Biotechnology committee, and it is expected to be officially released within two years. It will be the first international data standard in the field of microorganisms. On the basis of data standard, the data generated by this program will also be integrated and shared on the GCM platform. Therefore, we will also form an internationally authoritative microbial data platform.

3.3 GCM Global Cooperation Program Supporting 2019-nCoV Outbreak

On January 24, 2020, Novel Coronavirus National Science and Technology Resource Service System (https://nmdc.cn/nCoV), which was jointly constructed by the Institute of Microbiology, Chinese Academy of Sciences and the Chinese Center for Disease Control and Prevention, was officially launched. The System will timely publish authoritative information on science and technology resources as well as scientific data concerning novel Cov, including the collection of virus strain resources (National Pathogen Microbial Resource Bank), electron micrographs, detection methods, genomes, scientific literature, etc., according to its scientific research progress so as to provide supports for scientific study on 2019-nCoV and special information service of science and technology resources dealing with the current prevention and control of pneumonia caused by 2019-nCoV infection.

On February 18, 2020, Global Coronavirus Data Sharing and Analysis System (https://nmdc.cn/#/coronavirus) was launched by National Microbial Data Center (NMDC). The database includes a total of 3135 coronavirus genomes, including 32,865 nucleic acid sequences from 20,241 strains, separated from 496 different host types and 568 collection sites. The system also provides users with similarity query analysis and phylogenetic analysis based on uploaded genomic sequences, amplified partial sequences or translated protein sequences and data in the database. The system integrates global coronavirus genes and whole genome data, which provides important support and guarantee for Chinese scientists to carry out analysis and research, and promotes the collection and comprehensive analysis and sharing of coronavirus data domestic and abroad. On the other hand, it provides tools such as integrated similarity comparison, phylogenetic analysis, etc., which realizes the integration and standardized analysis and mining process of viromics data, helping scientists to quickly conduct research on virus mutation, traceability, and evolution.

4 Summary and Prospect

At present, the program has already set up the five working groups of bacteria screening, fungi screening, standard operating procedures (SOP), databases, and intellectual property and legal issues, in all of which Chinese scientists have played an important role. The first phase of the program has already started to accept about 800 candidate type strain samples from Belgium, China, Japan, South Korea, the Netherlands, Portugal, Russia, Sweden, Thailand, the United States and Britain. The *Data Management and Data Catalog Standard for Microbial Strain Resources* established by the program has been approved by ISO TC 276 Biotechnology committee as PWI20710. The Chinese Academy of Sciences has already deployed the project of "Research on the Common Technologies of Microbiome in Population and Environmental Health". In 2016, the scientists of the Chinese Academy of Sciences jointly appealed to the state for launching China Microbiome Program, and obtained the instructions of the state leaders. Now the project has been funded, so it is hoped that CAS could take the type microbial genome sequencing program as the starting point, rely on our country's advantages in aspects such as microbial resources research, sequencing technology, and the ability of comprehensive analysis of microbial data, vigorously support the key research and development project of "China Microbiome Program" which covers contents such as human body, agriculture, environment, traditional fermentation, new technologies, etc., and further utilize the international cooperation network established by the program to start the China-led microbiome international cooperation program and capture the strategic commanding point in the field of microorganisms as soon as possible.

Through the implementation of this program, we will take the lead in establishing international standards in the field of microorganisms and set up an internationally authoritative microbial data platform; we will systematically study the physiological functions of microorganisms on a global scale, and establish an integrated research and development application system including biological resource exploration, basic frontier research, technological innovation and industrial development; we will convene a series of brand-name academic conferences and training courses with domain influence to cultivate China's international leading strategic talents and young talents, laying an important foundation for China's talent, resource, technology and industry leadership in the field of microbial resources and even biotechnology.

References

1. Wu L, Ma J (2019) The global catalogue of microorganisms (GCM) 10K type strain sequencing project: providing services to taxonomists for standard genome sequencing and annotation. Int J Syst Evol Microbiol https://doi.org/10.1099/ijsem.0.003276
2. Wu L, Sun Q, Ma J (2017) World data centre for microorganisms: an information infrastructure to explore and utilize preserved microbial strains worldwide. Nucleic Acids Res 45(D1):D611–D618

Juncai Ma, Ph.D., senior engineer, is currently the director of the Center for Microbial Resource and Big Data, Institute of Microbiology, Chinese Academy of Sciences, director of National Microbiology Data Center (NMDC), director of WFCC-MIRCEN World Data Centre for Microorganisms, director of the Biotechnology and Bioindustry Information Center of Chinese Society of Biotechnology, member of Executive Committee of WFCC, chairman of the data management working group of Asian Network of Research Resource Centers (ANRRC) , and Co-chairman of the Working Group of BOLD Mirror, iBOL.

Current Situation and Prospect of EMDB/EMPIAR-China

Tongxin Niu, Yan Zhang, Jun Liu, Bo Zhang, and Fei Sun

Abstract In recent years, cryo-electron microscopy technology has rapidly developed and become a key method in the structural biology field. Moreover, the number of people engaged in cryo-electron microscopy (cryo-EM) is increasing rapidly in China with attendant explosive growth in cryo-EM data. Therefore, to meet the academic demand for access to cryo-EM data and to promote the development of the structural biology field, the European Institute of Bioinformatics established the Electron Microscopy Data Bank (EMDB) in Europe. Furthermore, the Committee for the Study of Cryo-Electron Microscopy of the Chinese Society of Biophysics partnered with the European Bioinformatics Institute to establish a mirror site of the EMDB in China to aid researchers in conveniently uploading and downloading high-resolution electron microscopy data, to promote cryo-EM-related research efforts, and to enhance the international status of China's basic research. This paper introduces aspects of the Chinese mirror website: the construction idea, structure, function realization, visualization tools, and running status. Furthermore, details of an analysis of the total visits to the website from the beginning of the website launch to October 30, 2019, are provided. Finally, future prospects of the Electron Microscope Database are assessed; briefly, we suggest it will play a crucial role in supporting and developing the field of structural biology.

Keywords Cryo-EM · Electron microscopy data bank (EMDB) · Electron microscopy public image archive (EMPIAR) · Mirror Site

1 Introduction

In recent years, cryo-electron microscopy technology has rapidly developed and become a key method in the field of structural biology. Simultaneously, the number

T. Niu · F. Sun (✉)
Institute of Biophysics, Chinese Academy of Sciences, Beijing, China
e-mail: feisun@ibp.ac.cn

Y. Zhang · J. Liu · B. Zhang
Computer Network Information Center, Chinese Academy of Sciences, Beijing, China

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_11

of people engaged in cryo-electron microscopy (cryo-EM) is increasing rapidly in China, yielding an explosive growth in cryo-EM data. Therefore, to meet the academic demand for access to cryo-EM data and to promote the development of the structural biology field, the European Institute of Bioinformatics established the Electron Microscopy Data Bank (EMDB) [1] and Electron Microscopy Public Image Archive (EMPIAR) [2] in Europe; these provide a data sharing service for original cryo-EM data, reconstruction results, and other related information. Because the uplink and downlink data flow are huge when using the website, relying on the European website can be problematic for Asian cryo-EM researchers. Therefore, it was deemed necessary to find a means of expediting the work flow of Asian researchers.

To provide a convenient method for researchers to upload and download high-resolution electron microscopy data, to promote cryo-EM-related research, and to enhance the international status of China's basic research, the Committee for the Study of Cryo-Electron Microscopy of the Chinese Society of Biophysics partnered with the European Bioinformatics Institute (EMBL-EBI) to establish a mirror site of the EMDB in China. The mirror websites (ftp://ftp.emdb-china.org and ftp://ftp.empiar-china.org) and the server (http://emdb-china.org.cn) were established in 2016. The establishment of a mirror site in China can effectively promote extensive exploration of biological information by Chinese scientists to aid in the discovery of solutions to major scientific problems.

EMDB-China and EMPIAR-China provide special access interfaces and faster uploads and downloads of data for cryo-EM researchers in Asia. Simultaneously, the founders developed a 3D interactive visualization tool named VizEMEC to view the density map of the 3D reconstruction. After several optimizations, VizEMEC can now show the internal shape of the density map through an isosurface plus color mapping method. Together with the advanced data reduction transmission strategy, VizEMEC can reduce the network communication and download time greatly. The asymptotic visualization method was adopted to accelerate page response, improve the user interaction experience, and improve the convenience of analyzing and evaluating EMDB data for researchers.

2 Design and Technical Framework

EMDB-China and EMPIAR-China were set up on two separate servers; application and access allocation between the servers is performed by a load-balancing device based on EBI's server configuration in day-to-day normal operations (Fig. 1). During normal operations, the load balancer distributes incoming requests equally to all server sets. However, when a failure or shutdown is detected, the load balancer removes the relevant server sets from the pool serving the EMDB until the set is available again [3]. One or more Nagios servers can monitor the web servers. The Nagios server monitors other relevant servers (MySQL, Solr, and ftp) indirectly by making requests to the web server, which in turn checks whether the other servers are working

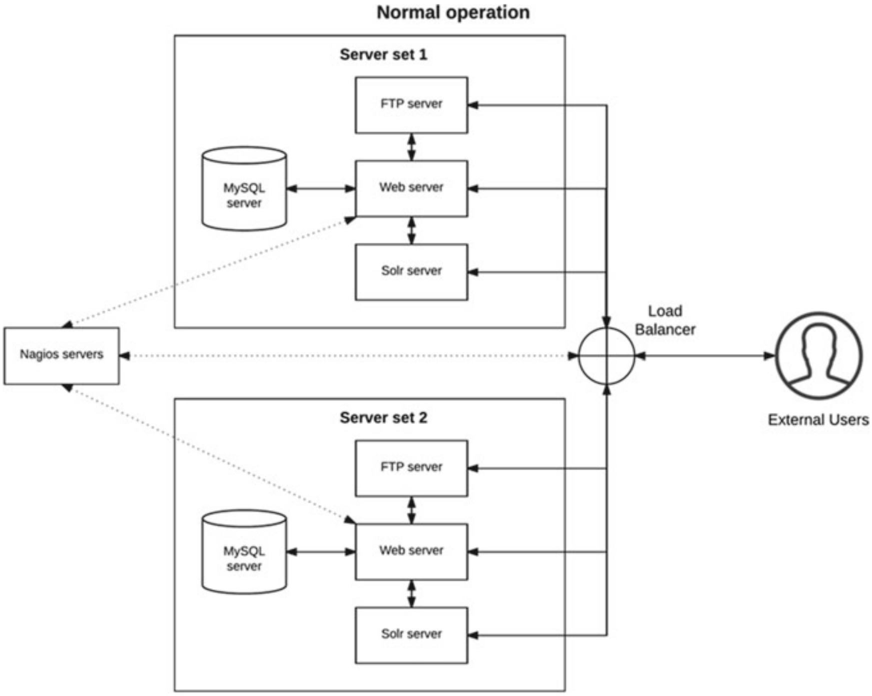


Fig. 1 Server configuration in day-to-day normal operations (from Ardan Patwardhan's document 'EMDB Mirror Recommended Server Configuration')

correctly (from Ardan Patwardhan's document 'EMDB Mirror Recommended Server Configuration').

3 EMDB/EMPIAR-China: Launching the Chinese Mirror Site

3.1 EMDB/EMPIAR-China Data Mirror

The server of the Chinese mirror site automatically synchronizes data from EMBL-EBI at regular times weekly through the China–Europe Terrestrial Cable laid by the China Science and Technology Network. The quantity of EMDB data updated weekly is relatively small (in GB), and thus, we use Rysnc for synchronous update. The quantity of EMPIAR data updated weekly is huge (in TB), and thus, to ensure the stability of data synchronization, we use ASPERA commercial software for synchronous update. Meanwhile, to ensure that Chinese mirror site data are secure, all data are backed up and stored offsite at the Huairou Subcenter of the Computer

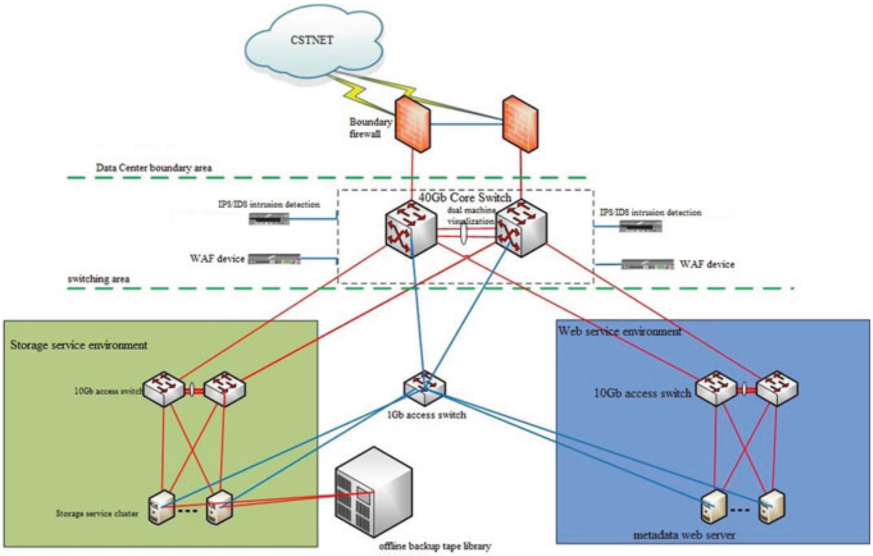


Fig. 2 Topology of data synchronization and offsite backup

Network Information Center, Chinese Academy of Science. All these data are available for download as an FTP site. The topology of data synchronization and offsite backup is shown in Fig. 2.

3.2 EMDB/EMPIAR-China Web Service

3.2.1 Web Service Technology Architecture

Because the data source shown on the EMDB/EMPIAR-China website is XMLle, the website page display is mainly static. The entire system can be divided into four layers, namely data, service, business, and view layers.

The data layer consists of MongoDB, experimental data file system, and Solr index file. The service layer provides basic services such as XML parse server and Solr retrieval server, statistics service, and file service. The business layer provides the main functions of the website, namely EMDB/EMPIAR data retrieval, taxonomy browsing, details browsing, access statistics, XML viewing, download and update time, etc. The view layer uses the Thymeleaf template engine to display the webpage and uses Echarts to display statistical results.

The system was developed using Java programming and the Spring Boot development framework. Spring Boot is a new development framework that can simplify initial setup and development of new spring applications. The special configuration

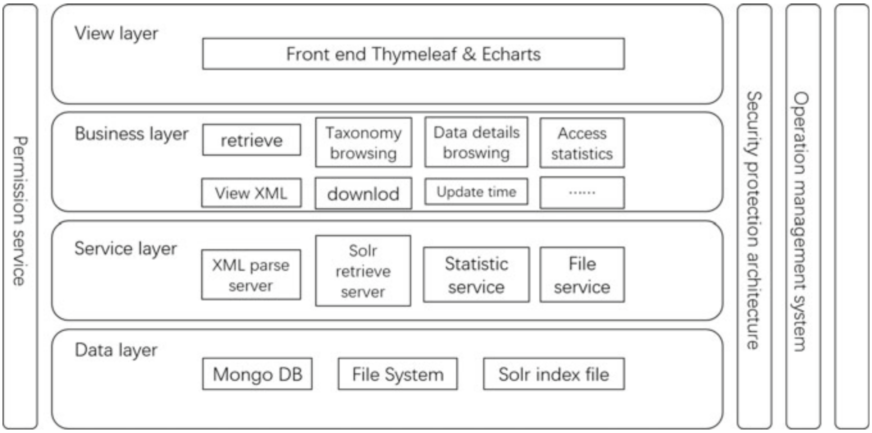


Fig. 3 Architecture of China mirror site implementation technology

of Spring Boot enables developers to work without defining boilerplate configurations. The system uses spring-data-MongoDB to control MongoDB and uses the Solr retrieval engine to provide data index and retrieve function. Solr is an independent enterprise search application server with an API interface, similar to a web service. Solr can generate an index file through submission of a formatted XMLe to a search engine through an HTTP request. Furthermore, it can make a lookup request through the HTTP GET operation and then change the result back into XML format.

First, the system uses JAXB technology to generate java classes according to XML schema. When the data need to be updated, the XML metadata are parsed into java class objects and stored in MongoDB. Subsequently, the system uses the data in MongoDB to generate a Solr index through the interface, complete the website data update, and record the update time. When users access the website, they can search experimental data through the Solr service. The IP and access time are recorded and stored in MongoDB by the system automatically (Fig. 3).

3.2.2 Major Function of EMDB/EMPIAR-China

The main purpose of EMDB/EMPIAR-China is to provide convenient access for users in Asia. The webpage provides EMDB/EMPIAR data, which can be downloaded. EMDB/EMPIAR-China improves the readability of metadata XML files and provides functions such as searching, sorting, and rendering. The homepage of EMDB/EMPIAR-China (Fig. 4) can be used for four main functions: browsing, obtaining statistics, obtaining 3D images through a visualization tool, and transferring data through the FTP site.

Fig. 4 Homepage of EMDB/EMPIAR-China

Data Searching, Filtering, and Sorting Etc.

On the basis of metadata XML files in EMDB/EMPIAR data, full-text retrieval, data browsing and sorting, data filtering, and EMDB/EMPIAR-associated query resolution can be performed.

- Full-text retrieval

Using the provided keywords, XML files are accurately matched and search results yielded. Search results usually contain much more information than is displayed on the webpage, and thus, users can perform secondary retrieval by viewing the metadata XML file online.

- Data browsing and sorting

Both the EMDB and EMPIAR mirror sites support full data browsing and four maximum/minimum sorting categories (Fig. 5): map release time, header release time, density values, and molecular weight. This is helpful for secondary sorting and improves query efficiency.

Because the metadata XML file content is limited, EMPIAR-China can support only released date sorting (Fig. 6).

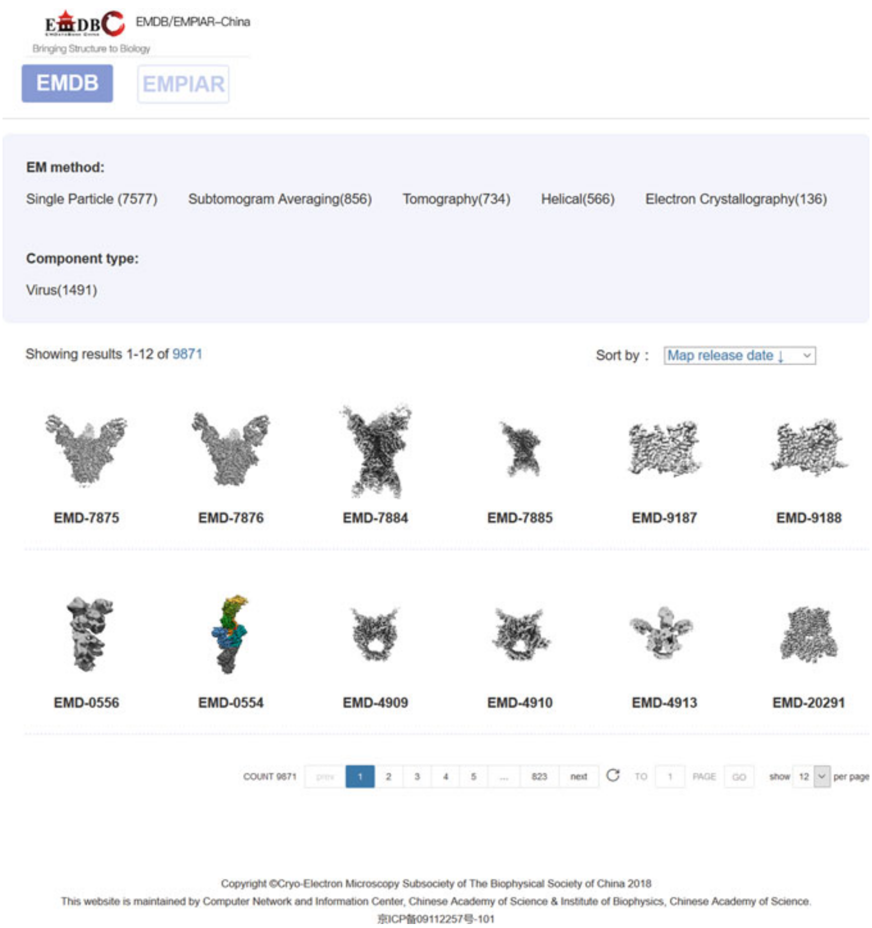


Fig. 5 EMDB-China data browsing and result sorting

- Data filtering

EMDB data can be queried using various filter conditions such as EM method or component type, and the webpage will show related data according to the filtering result.

- EMDB/EMPIAR-associated queries

According to the content of the metadata XML file, the webpage provides a part of EMDB/EMPIAR-associated queries (Fig. 7).

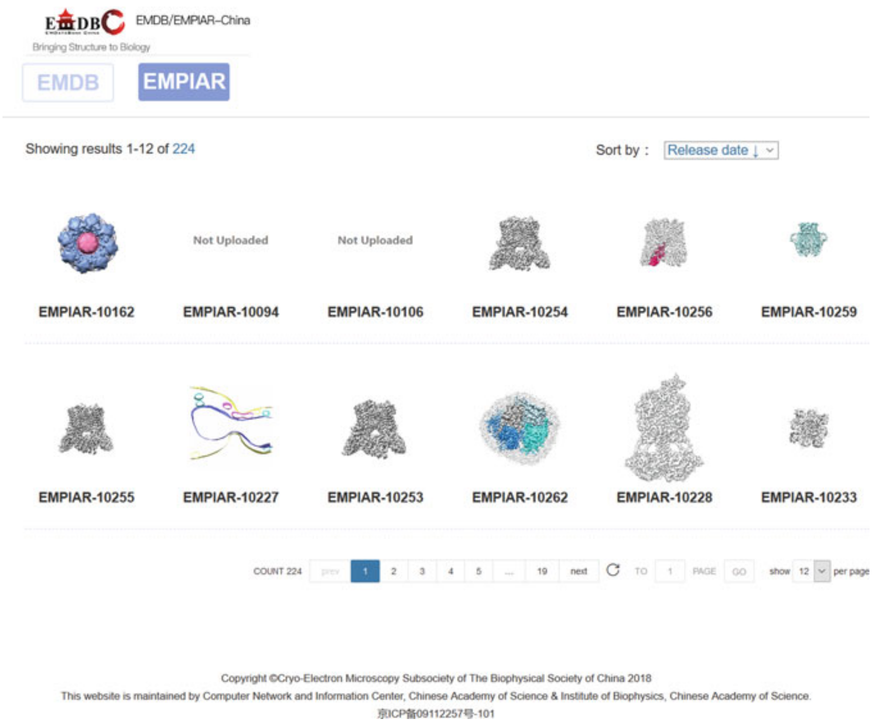


Fig. 6 EMPIAR-China data browsing and result sorting

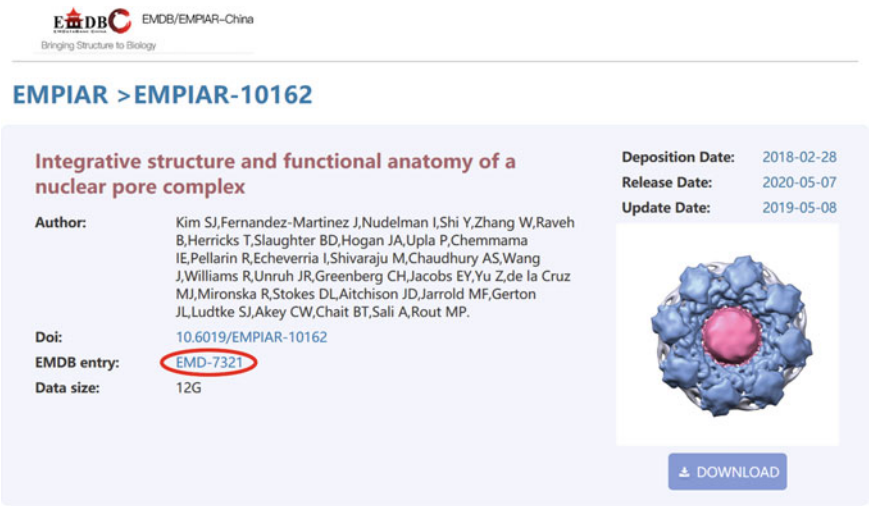


Fig. 7 EMDB/EMPIAR-associated queries

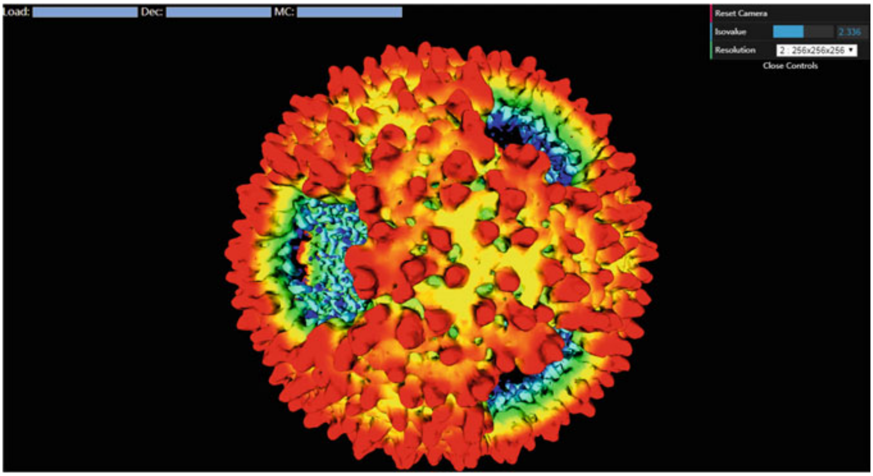


Fig. 8 Screenshot of VizEMEC, a visualization tool for EM density map preview and 3D interaction

Statistics

The statistics function provides statistical results: the annual accumulated EM density map data, highest resolution in a year, and monthly website hits.

Visualization Tools

- EM preview tool

To implement online 3D interaction with the EM density map, researchers must access an interactive preview of the 3D reconstruction data from the perspective of their own research. The project team developed a 3D interactive tool that is based on HTML5 WebGL [4] rendering technology and ZFP [5] volume data compression technology and is named VizEMEC [6]. VizEMEC is suitable for use in a web environment and supports major browsers such as Chrome, Firefox, and Safari. Installation of other plugins or operating systems is not required for this tool. After several optimizations, VizEMEC can show the internal shape of the density map through the isosurface plus color mapping method (Fig. 8). Combined with the advanced data reduction transmission strategy, VizEMEC can reduce the network communication and download time greatly. The asymptotic visualization method is used to accelerate page response, improve the user interaction experience, and improve the convenience of analysis and evaluation of EMDB data for researchers.

VizEMEC contains five modules: a volume-compressed buffer module, volume selected module, volume uncompressed module, isosurface-generated module, and isosurface-rendering module. The volume-compressed buffer module runs on a master server and compresses the original volume of data according to different compression parameters to generate a series of data caches. The other four modules run on a client server's browser. The volume selected module sends a request to the server for the list of available cached data and sorts these data based on specific priorities. Then, the cached data are downloaded according to priority until the usage warning system in the module prevents loading more data. The volume uncompressed module decompresses the downloaded cached data to extract the volume data. The ZFP data decrement algorithm is used for combining the asymptotic online interaction mechanisms such as domain block, interleaved block, precision progression, and residual addition to optimize system loading. The isosurface-generated module acts on the volume data decompressed. It can generate isosurface fatly according to the user's equivalent parameter setting. The isosurface-rendering module uses the WebGL rendering engine matched with a light material setting to render the isosurface.

According to the statistical results, the transmitted data are usually limited to 10% of the original volume of data when interacting remotely if the volume data size is more than 50 MB. The wait time for users is reduced to less than 1% of the original because of the introduction of a multistage loading strategy. The interaction response efficiency is greatly improved.

After a period of internal testing, VizEMEC was officially launched in January 2019. The relevant page is linked to the EMDB/EMPAIR-China data details page, and users can access the visualization tool through the corresponding tool link.

- EM metadata visual analysis tools

To facilitate convenient and comprehensive analysis, evaluation, and query of EMDB data, the project team used an advanced theoretical data analysis model to conduct a requirement analysis, function decomposition, and prototype design for system objectives and tasks based on the structure and features of the metadata. Using the latest visualization tools and techniques, we developed a set of visual exploration and analysis tools named VASEM [6] for use with the overall data in the EMDB.

The user can input information manually while submitting metadata, which may lead to the introduction of errors into the EMDB metadata. Hence, the metadata in the EMDB may be incomplete or inaccurate.

VASEM entails the introduction of a third-party data source for data cross-validation and quality improvement to ensure the accuracy and comprehensiveness of the analysis query process and clean data. VASEM integrates CrossRef, which can be queried through ISSN and DOI, to unify publication names and obtain detailed author information. VASEM integrates NCBI Taxonomy to unify organism names. To achieve the aforementioned design principles and ideas, VASEM integrates a data preprocessing submodule, which runs regularly, synchronizes with the weekly EDMB data update service, and performs the automatic and continuous background data update task. VASEM interface was shown in Fig. 9.

Based on the HTML5 framework and the novel information visualization library echarts, VASEM implements an interactive graphical interface for EMDB metadata

Fig. 9 VASEM interface

through the web. It uses data-driven and informative 2D charts (such as scatter plots, Sankey plots, stacked histograms, and graph plots) and a series of coordinated views to visualize multidimensional EM metadata, including dimensions of time, author, organism, and publication. To compare different data production teams, source organisms, and publications, VASEM provides the function of comparing entities based on data subsets. On the basis of the query results of specific conditions, the EM data produced from different research teams, source organisms, and publications are statistically analyzed. Using Sankey plots, users can easily understand the distribution of research topics and publication venues of research teams. The data in the EMDB can be used to exhibit complex relations through the commonality of each dimension. To make use of the relationship between these data for subset association analysis and query, VASEM provides relation-based interactive force-oriented graph plots that allow users to show or hide certain types of entity relations and to use some specified entity relations to cluster entities to quickly query subsets of data associated with specified data. The EMDB aggregates numerous EM data entities over time. Even after condition filtering, numerous query results may be displayed. To visualize the entire subset of several data entities, VASEM provides a set of subset description and evaluation tools based on an equalized parallel coordinate plot. Users can use this tool to quantitatively evaluate many EM data entities and to compare different authors, papers, organisms, resolutions, and grid sizes. VASEM builds a filter-editing interaction system that combines user query conditions and analysis conclusions into the query calculation process at the same time, allowing researchers to better incorporate unstructured filter conditions into the data query process.

After a period of internal testing, VASEM [6] was officially launched in June 2019. It can be accessed at http://vasem.emdb-china.org.cn. The detailed descriptions of the system design and implementation were summarized in a paper that was accepted by ChinaVis 2019, the top domestic visualization and visual analysis conference, and was recommended for publication in the SCI-indexed international journal *Journal of Visualization*. Simultaneously, a conference report was provided on related work by Jun Liu.

FTP Downloading Service of EMDB/EMPIAR

On the basis of the EMDB/EMPIAR mirror data, the FTP site was set up for users to download EMDB/EMPIAR mirror data. To ensure the security of EMDB/EMPIAR mirror data and the reliability of the FTP service, the data are read only and the FTP service is independently deployed. Users can download the latest EMDB/EMPIAR data from the FTP site.

4 Operational Situation of the EMDB/EMPIAR-China Site

EMDB/EMPIAR-China synchronizes data with EMBL-EBI at 8:00 A.M. every Wednesday; the update time is subject to the FTP site. Approximately 4 h is required to update tens of gigabytes of EMDB data because the amount of data is relatively small. However, approximately 4 days or more are required to update a multiple TBs of EMPIAR data. As of October 30, 2019, approximately 1.6 TB of EMDB data, 140 TB of EEMPIAR data, and more than 9800 sets in total were updated through data mirroring.

Data displayed on the EMDB/EMPIAR-China website (http://www.emdb-china.org.cn) remain unchanged during data synchronization. After data synchronization, the webpage is considered updated. Since EMDB/EMPIAR-China was released in August 2018, it has been revised thrice. As of October 2019, the EMDB/EMPIAR-China homepage has been accessed approximately 20,000 times; the monthly number of visits is shown in Fig. 10. Fifty percent of the visits are from international IP addresses, including those from 113 countries and regions across six continents, except Antarctica; the top three represented countries are Brazil, the United States, and India. Domestic visitors mainly come from 33 provinces and regions, of which the top three are Beijing, Shanghai, and Zhejiang province.

Toward the end of October 2019, the number of times the FTP had been accessed and downloaded exceeded 20,000; associated IP addresses were mainly from domestic regions, the United States, and Canada, and the data were mainly regarding EMD-3228, EMD-10045, and EMD-10160.

5 Summary and Prospects

The amount of cryo-EM data is large and continues to grow as technology develops. This type of data is extremely valuable to academic research. Development of the

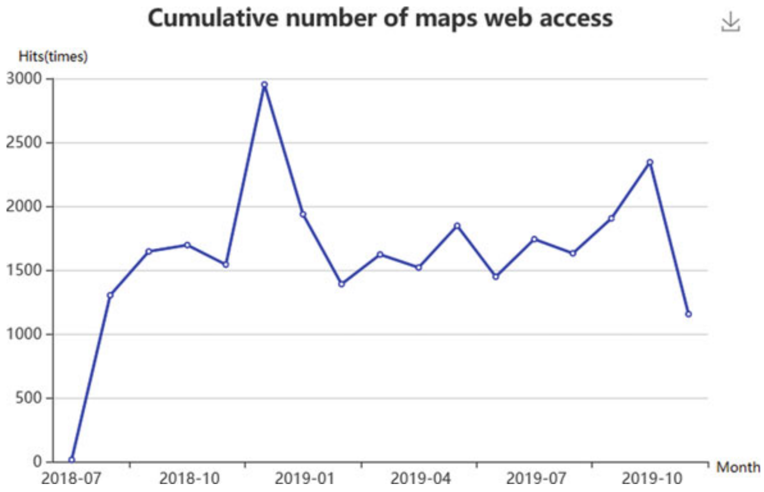


Fig. 10 Monthly visits to the EMDB/EMPIAR-China homepage (until October 2019)

EMDB/EMPIAR-China was thus a critical undertaking to overcome the network bottleneck that prevented efficient access to the original European site.

This paper introduced the Chinese mirror website in terms of the construction idea, structure, function realization, visualization tools, and running status. Moreover, the total visits to the website from the launch of the website to October 30, 2019, were analyzed. Finally, we examined the future prospects of the EMBD and suggested that it will play a crucial role in supporting and developing the field of structural biology.

To provide an improved access experience to scientific research users and improve and add functions, further development of the EMDB/EMPIAR-China is being conducted by the Institute of Biophysics and Computer Network Information Center. The next steps are data validation and intersection.

References

1. Lawson CL, Patwardhan A, Baker ML et al (2016) EMDataBank unified data resource for 3DEM[J]. Nucleic Acids Res 44(D1):D396–D403
2. Iudin A et al (2016) EMPIAR: a public archive for raw electron microscopy image data. Nat Methods 13(5):387–388
3. Young JY et al (2017) OneDep: unified wwPDB system for deposition, biocuration, and validation of macromolecular structures in the PDB archive. Structure 25(3):536–545
4. Opera. WebGL and hardware acceleration. https://dev.opera.com/blog/webgl-and-hardware-acceleration/. Accessed 28 Feb 2011
5. Lindstrom P (2014) Fixed-rate compressed floating-point arrays. IEEE Trans Vis Comput Gr 20(12):2674–2683
6. Jun L et al (2019) VASEM: visual analytics system for electron microscopy data bank. J Vis 22(6):1145–1159

Fei Sun, Institute of Biophysics, Chinese Academy of Sciences; Chief scientist and director, Center for Biological Imaging, Core Facilities for Protein Sciences, Chinese Academy of Sciences. His research interests are mainly related with the structures and functions of biological macromolecules, combining various structural research approaches (majorly crystallography and cryo-electron microscopy (cryoEM)) as well as developing new methodologies to determine the architecture of the biological system, in vitro and in vivo, from nano-scale to meso-scale.

The Ecology of High-Performance Computing

Xuebin Chi

Abstract High performance computing (HPC) embodies a country's comprehensive strength in science and technology, which is an important part of the national innovation system, and is also the strategic commanding point of fierce competition in the world's major developed countries. From the ecological perspective of building high-performance computing, this paper expounds the development trend of this field in detail, and points out the weak links in the application of ecological development, and makes analysis and suggestions.

Keywords High performance computing · Core software · Key applications · Computational ecology

1 Study on the Development of HPC

1.1 Evolution of High Performance Computing in Developed Countries

In modern scientific research, it becomes more difficult to solve problems by theory and experiment, so numerical methods are more favorite used to simulate the physical world in order to solve complex problems. Therefore, computation has become a necessary tool of natural science research. As the scale of problem solving becomes larger and larger, the demand for computing becomes the most direct driving force for the development of high performance computing in developed countries.

1.1.1 Computer Technology Leads the Development of Application

During World War II, it was difficult to manually calculate the trajectory of artilleries, and the need from war to calculate the trajectory of a new type of artilleries prompted

X. Chi (✉)
Computer Network Information Center, Chinese Academy of Sciences, Beijing, China
e-mail: chi@sccas.cn

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_12

the birth of the first computer. On February 14, 1946, the world's first computer ENIAC was born and put into use at the University of Pennsylvania, has had the extremely profound influence on the modern human history. Since then, people kept increasing the level of pursuit of computer technology, especially developed countries headed by the United States have been in leading position in computer technology development. The development of high-performance computers dated back to the 1960s and 1970s, when they were called mainframe.

In general, the development of high performance computers has gone through the following stages (Fig. 1).

Looking back to the history of the computer over the past decades, the application of high performance computing is inseparable from the development of high performance computer technology. The development of high performance computer provides a powerful tool and the material basis for high performance computing applications, and constantly leads the development of domain applications including nuclear weapons research, nuclear material storage simulation, oil exploration, disinformation technology, medical and new drug research, computational chemistry, meteorology, weather and disaster forecasting, industrial process improvement and environmental protection. It has become an important tool for promoting scientific and technological innovation and social progress.

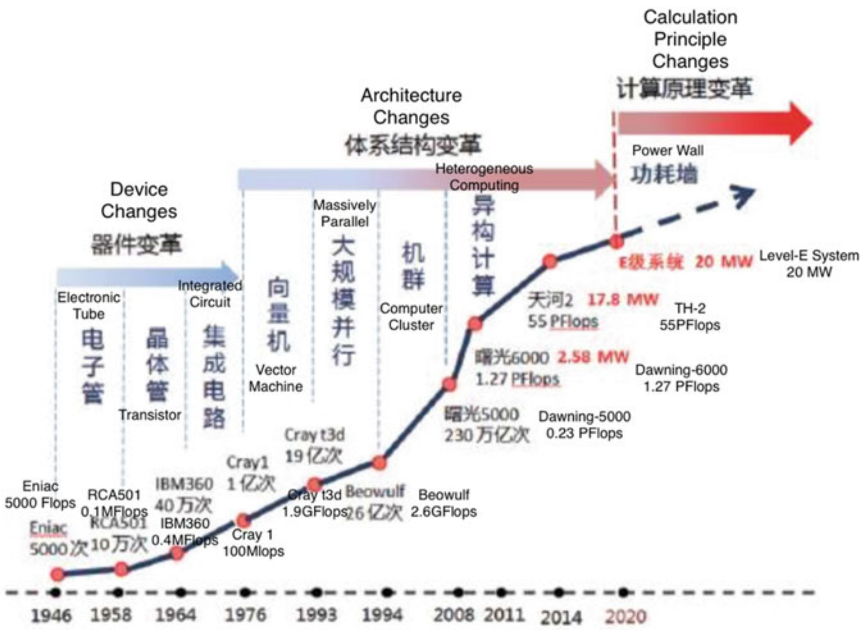


Fig. 1 History of high performance computers. *Source* Zang et al. [1]

Internationally, developed countries, led by the United States, have been pioneers in the development of high performance computing. In 1964, the first high performance computer CDC6600 was born; the 1970s the concept of parallel computing was proposed and soon there came the advent of the iconic parallel computing ILLIAC IV.

In 1982, Cray produced the first parallel vector machine Cray X-MP 2, which uses parallel processing structures such as preemptive control and overlapping operations, computational pipelining, and cross access parallel memory. A variety of applications can be used on CRAY X-MP to solve industrial applications such as the petroleum industry, aerospace, automotive, nuclear research, and chemistry. Scientists and engineers can use CRAY X-MP systems and industry-standard code to solve a wide range of problems. In addition, the software developed for the CRAY-1 system can run on all models of the Cray X-MP series with good compatibility and inheritance, thus greatly protecting the software investment from users.

1.1.2 Application Demands Promote Computer Innovations

(1) HPCC Program Leads the Wave of Integration Between High-performance Computer Production and Application

In the 1990s, the rapid development of supercomputer industry in Japan and other countries posed a series of severe challenges to the United States. Cray, the largest U.S. supercomputer manufacturer, reported annual revenue of less than $1 billion during this period, while other supercomputing companies reported annual revenue of less than $200 million. In the meantime, Japanese supercomputing manufacturers showed more strength. Fujitsu, Hitachi and NEC have annual revenues of $17 billion to $45 billion. Data (see Table 1) show a marked change in the trend of the United States and Japan installing supercomputers as a percentage of the world total in the 1980s and 1990s compared to the same period in the 1990s.

This trend has aroused great concern in the United States. Facing the rapid growth of high performance computing in Japan, HPCC (High Performance Computing and Communications) program, also called grand challenge program attracted considerable attention and support from the U.S. government and Congress. The program was proposed in the context of the coming technological innovation required for the development of national information infrastructure to compare with the challenge of Japan's coming dominance in high performance computing in the 1990s. It is also a

Table 1 United States and Japanese supercomputer installations in percentage of the world (1980–1990)

Year	Percentage	
	USA (%)	Japan (%)
1980	81	8
1990	50	28

Data Source Chongyuan [2]

program for the information superhighway in the United States. The administration's mission is no longer just to build an advanced research infrastructure for the U.S. tech community, but more importantly to strengthen the international competitiveness of the U.S. IT industry, to rival the rapid rise of Japan.

According to the US fiscal report in 1992, the HPCC program has eight government sectors (see Table 2). As can be seen from the budget allocation in the program proposal, four agencies—The U.S. National Science Foundation (NFS), the Defense Advanced Research Projects Agency of the Department of Defense (DAPRA), the Department of Energy (DOE), and the National Aeronautics and Space Administration (NASA)—are key members of HPCC, DARPA is mainly responsible for the research and development of hardware systems, and the NFS is mainly responsible for the software development.

The HPCC program in the United States presented a series of enormous challenges at the time, accelerated the development of high-performance computer systems, and increased the number of developers for algorithms and software. This led the wave of integration between HPC production and application, and hence was a milestone event.

(2) **Co-design Promotes the Construction of Collaborative Design Ecosphere with the Demand of Application**

Co-design means that in the designing process of computer system, the structure design and techniques, algorithms and software are constrained by scientific problems to be solved. The researchers taking computation scientific researches need to participate the production of hardware, software, numerical methods, algorithms and application to ensure the computer architecture is adaptable for those specified programs.

Table 2 Distribution of budget proposals for HPCC program (in million dollars)

Sector name	1991	1992	Responsibility
Defense Advanced Research Projects Agency	183	232.2	Basic technology development
Department of Energy	65	93	Basic technology development
National Aeronautics and Space Administration	54	72.4	Basic technology development
National Science Foundation	169	213	Basic technology development
National Institute of Standards and Technology	2.1	2.9	Development of high speed data communication standard
National Oceanic and Atmospheric Administration	1.4	2.5	Applied research
Environmental Protection Agency	1.4	5.2	Applied research
National Library of Medicine	13.5	17.1	Applied research
Total	489.4	638.3	

Data Source Lan [2]

Fig. 2 Co-design relationships. *Note* APP—Application, SW—Software, HW—Hardware. *Source* https://www.mellanox.com/coe/ [3]

This designing concept combines the expertise of vendors, hardware architects, system software developers, domain scientists, computer scientists, and applied mathematicians. They make joint effort to balance the benefits about hardware, software and bottom-level algorithm design (Fig. 2).

To satisfy the computation demands of energy research, national security and advanced technology, the U.S. Department of Energy defines Co-design as a key factor in achieving the exa-scale computing strategy. They see the application of Co-design with exa-scale computers as both a challenge and an opportunity [4]. At the same time, system software and algorithm innovations are also necessary.

In anticipation of significant changes in architecture and software, many researchers are considering on co-design approaches for embedded computing community development. Therefore, it is necessary to adopt a powerful software and hardware co-design strategy to form a new high-performance computing co-design method. Such a design process does not simply ask "what kind of scientific application can run on exa-scale systems", rather, it calls for "what systems should be built to meet the needs of the most important scientific problems". This leverage provides insight into specific application requirements and broad combinations of computational sciences, and closely focuses on the performance of the scientific applications. This is essential to the general optimization of hardware and applications.

In the 1970s and 1980s, LLNL worked closely with CDC and Cray to design machines using lab applications or agents (such as Livermore Loops) to influence supplier solutions and provide machines with better availability. During the ASCI era (1996-present), LLNL worked closely with IBM, during which IBM delivered five generations of supercomputers (ASCI Blue, ASCI White, ASCI Purple, BlueGene/L and Sequoia), all of which were heavily influenced by the idea of co-design.

In the twenty-first century, the co-designed core hardware/software collaborative simulation (such as structure simulating toolkit) were carried out by the cooperation of Sandia National Laboratory, Oak Ridge National Laboratory, Intel and Cray.

It is clear that the development of co-design promotes the computer innovation, and can break barriers between different subjects and domains. At the moment, a common issue is to found a so-called 'co-design ecology' for HPC production and application development, and to improve coordination is the key factor to break through the bottleneck of software development [5]. As Rob Neely, head of High-performance Computing at LLNL, puts it, "collaborative design is about computing, not simply building machines".

1.2 *Development of High Performance Computing in China*

1.2.1 Rapid Growth of High-Performance Computers, from "Behind" to "Lead"

In 1956, China released the "12 Years' Plan of Science & Technology Development", designating computer science, electronics, semiconductors and automation to be the four prior developing subjects, and formulated the development plan for computer research, production and education. The computer industry in China had begun from this point.

In 1958, machine '103' was handed out. It was the first electron tube based general-purpose minicomputer that was produced in China, hence called 'first-generation computer'. Although it was 12 years later than the first electron tube-based computer ENIAC, it still can be called a miracle in the computer history in China: within 2 years, and without help from western countries.

In the Mid-1960s, China has trial-produced 5 types of transistor computers, and put into small-scale manufacture. This showed the computer industry in China had come into the second generation, and was able to produce computers that fit for conditions of China. Machines of 108-B and 121 were in high standards at that age, and had the output number over 100. Since then, computer industry in China entered a booming stage.

Since 1980s, China has started the high-level development of vector machine, large-scale parallel machine and cluster systems. Series of Galaxy, Sunway, Dawning appeared in succession at that time. In 1991, China released the first microprocessor based tight-coupled parallel computer of BJ-01, which had reached capacity of 150–300 MFLOPS. BJ-01 showed its high operation speed, high efficiency, good performance, reliable system and user-friendly interfaces. It has been an important foundation by providing an excellent platform for researches of parallel computer architecture, parallel operating systems, parallel compiling and parallel algorithms (Fig. 3).

Fig. 3 The first microprocessor based parallel computer in China. *Source* https://www.ccidnet.com/news/newszhuanti/2009/60/pc/

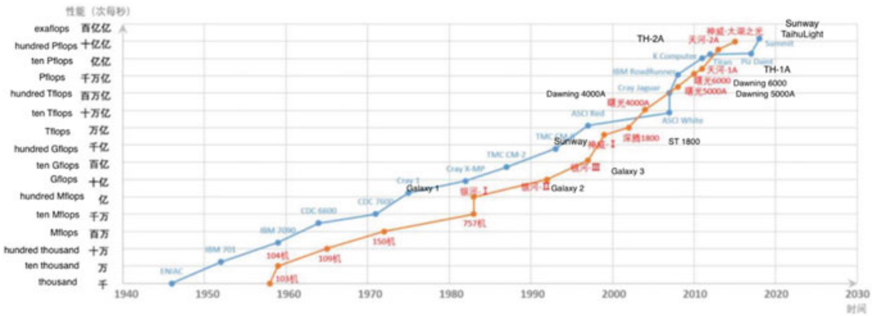


Fig. 4 Comparison of the launch time of HPCs between China and foreign countries

In October of 1993, guided by the 863 Program in China, the first symmetrical multi-processing (SMP) computer Dawning-I was launched. The appearance of Dawning-I successfully broke the embargo of information technology from western countries, and further promoted the independent development of HPC industry in China.

In the past 20 years, independently developed HPCs in China successively reached capacity levels of gigaflops, teraflops and petaflops. Figure 4 shows HPCs produced in China compared with foreign HPCs in the same capacity level. It is clear that the development of HPC industry in China has been in quicken paces. Tianhe-II, Sunway Taihu Light were consecutively ranking the 1st for 5 years in top 500 lists of supercomputers all around the world. Taking Dawning series as an example, from Dawning-I in 1993 to Dawning 6000 in 2010, the Linpack performance has upgraded for about 2,000,000 times, while in the meantime, supercomputers in other countries on the top 500 list only have an average upgrading rate of 32,000 times.

1.2.2 Domestic HPCs Need Supports from Domestic Algorithms and Applications

Domestic HPCs need domestic developed applications. To give full advantages of HPC hardware, it requires proper algorithms and applications to realize large-scale parallel computations. The development of HPC applications calls for deep intersections among mathematics, computational science and other research domains. Chinese scientists have made a great number of contributions to promote the independent development of domestic algorithms and applications.

In 1950s, Prof. Shi Zhongci proposed a new algorithm that integrate variation principle and perturbation theory and calculate the approximate value of helium atom's lowest energy state. Following this, a strong mathematical theoretic support has been provided to subjects including eigenvalue, spline finite element, nonconforming finite element, domain decomposition and multi-grid method.

Since 1990s there appeared a number of self-adaptable finite element software platforms and toolkits in the world, and helped researchers and engineers to make self-adaptable finite element programming. Prof. Zhang Linbo with his team successfully developed the parallel self-adaptable finite element software platform PHG (parallel hierarchical grid). PHG platform supports large-scale parallelization and dynamic load-balancing, and has the capability to support the development of parallel finite element applications on supercomputers from 1 TFlops to 100 PFlops. Parallel finite element programs developed on PHG platform can effectively scale to thousands of processes and tens of thousands of threads to tens of thousands of processes and hundreds of thousands of threads using pure CPU cores, and some programs has realized million-core level on heterogeneous many-core acceleration.

Prof. Sun Jiachang first present the concept of storage complexity in 1996, and suggested that the complexity of algorithms should consider both computational complexity and storage complexity. Being the basic property, computational complexity contains time complexity and space complexity. On the other hand, the storage complexity changes depending on real conditions. Users optimize algorithms for reducing the storage complexity, hence to achieve this, it has to put on new algorithm development. This concept provides firm theoretical basis for the development of domestic independent software and applications.

Besides, with the support from national and regional programs, many domain institutes actively develop applications with independent intellectual property rights, such as numerical simulation on super-large scale virtual drug screening and protein folding and conformation changes, oceanology software, parameter extraction and photolithographic simulation software for integrated circuit designing and producing, magnetic confinement fusion MHD computation, first-principle electronic structure computation, ion channel numerical simulation and structure analyzing software, and so forth. However, from the objective angle, China still falls behind in the area of algorithms and software compared with developed countries such as the U.S. and Japan, thus needs to put on hard effort to catch up with the advanced international standard.

1.3 Deep Integration of Computer Technology and Application Promotes the Development of HPC

In the past 20 years, the development of HPC in the U.S. adopts the strategy of application-driven and integrating the development of technology and application, while China biases to the strategy of technology-driven. China chose to first improve the system performance of HPCs and then stimulate the development of applications. In contrast, the developed countries produce HPC systems according to demands of domain applications, and this strategy effectively avoid wasting on HPC resources. An example that shows how strong domain applications stimulate the development of HPC: the Exascale Computing Project (ECP) in the U.S. is led by the Argonne

National Laboratory, and DOE has founded 25 application developing teams for designing domain applications on exascale supercomputers before the handover of Summit. The target of ECP is not tied to Linpack results, but the geometric mean of the performance of these 25 applications, which means none of them can give a low performance. In the U.S. there are first domain challenges and then produce computers to solve problems, while in China we produce leading machines and then find applications [6]. The result reveals that the priority to hardware computation capability in China leaves a low efficiency of machine utilities, and needs an intermediate period to fully utilize those machines.

To shorten this intermediate period, accelerate the development of HPC and realize the deep integration of software, hardware and applications, Chinese scientists and researchers have made hard efforts, and reached many good achievements.

For example, to solve the bottleneck of high investment, high energy consumption, backwardness of algorithm development, low efficiency in the development of HPC, the Institute of Process Engineering (IPE) of Chinese Academy of Sciences carried systematical researches for 30 years and proposed the unique multi-scale simulation method. In October of 2007 the first single precision CPU + GPU heterogeneous supercomputing system was started building, with the peak performance over 100 TFlops. At that time GPU numerical computation and CUDA were still newly come ideas, and there are barely experiences of such large heterogeneous systems that can be learnt, IPE met plenty unexpected difficulties during this process. However, they consistently stick on the approach of integrating software, hardware and applications. After plenty of experiments and optimizations, they finally got through the process of general design, component selection, integrated installation and application testing in a short period of time, and started serving from February of 2008. Using this system, IPE successfully launched applications including direct value simulation of multiphase flow, micro-simulation of material and nano/micro system, and dynamic behavior modeling of biomacromolecule. All these shows the advantage and bright future of multi-scale discrete parallel computing.

As the development of applications carried on, the system was kept upgrading depending on demands. Meanwhile, IPE cooperated with Lenovo and Sugon to spread 10 heterogeneous supercomputing systems with peak performance around 100–200 TFlops to many institutes in Chinese Academy of Sciences, and hence greatly raised the computation capability. In April of 2010, IPE successfully developed the double-precision supercomputing system with peak performance over 1 PFlops. It was the first system that bases on NVIDIA Fermi GPU in the world, and realized the joint computation across 10 systems, resulting the distributed GPU supercomputing environment that reached 5 PFlops in Chinese Academy of Sciences (Fig. 5).

These show the adoption of the designing ideas of application-driven, efficiency-prior, generalized software, specialized hardware, and the unification of problem, model, algorithm and hardware architecture. This provides an excellent experience for building the ecology of HPC development in China.

Fig. 5 Supercomputing system produced independently by IPE of CAS

1.4 CNGrid Aggregates Resources of HPC Ecology in China

The China National Grid (CNGrid) is continuously supported by 863 Program and National Key R&D Program. CNGrid bases on existing HPC environments, and focuses on researching on key technologies of applications and services, further improving of the resource utilization, developing HPC application services and domain communities that riches in application resources and usefulness, reducing HPC application costs, and promoting the level of HPC application services in all aspects.

Being the northern main node of CNGrid, the Computer Network Information Center of Chinese Academy of Sciences (CNIC) is responsible for operation and management, middleware development and technical support for CNGrid. Compared with other grids in the world, CNGrid has its own advantages. Grids in the U.S. and European countries concentrate on scientific researches, while CNGrid also emphasizes on supporting to domain applications and founding HPC ecology. These domain applications not only require HPC resources, but also the accessing, exchanging and processing of remote and heterogeneous data. CNGrid shows its importance of by supporting these domain applications, and guides the planning, deployment and integration of information application ecology with innovation ideas and methods.

1.5 Study of Future Developing Trend of HPC Ecology

Since the HPC capability reached the level of petaflops in the last decade, major countries and unions has turned to the next target: exa-scale computing (Table 3).

By summarizing the development plans in the U.S., Japan and EU, it is clear that their competition on exa-scale HPC has become ferocious. It can be expected to have exa-scale supercomputing systems no later than 2022. After decades of the development of HPC industry and ecology, they are on their tracks with solid foundation.

Table 3 Exascale HPC development plans in the U.S., Japan and EU

No	Name of the plan/System	Manufacture	Deploying time
1	U.S. ECP Aurora A21	Cray/Intel	2021
2	US ECP Frontier	Cray/AMD	2021
3	U.S. ECP El Capitan	Undecided	2023
4	U.S. ECP NERSC-10	Undecided	2024
5	Japan HPCI Fugaku	Fujitsu	2021–2022
6	EU Mont-Blanc 2020	Atos/Bull	2021–2022

In addition, countries in the world are also speed up to make strategic plan for future HPC ecology. Take US as an example, the Office of Science and Technology Policy (OSTP) released the *National Strategic Computing Initiative Update: Pioneering the Future of Computing* in Nov 14th 2019. Compared with the initiative in 2016, the updated version more focuses on the entirety of hardware, software and infrastructure, and developing innovative practical applications to support the future computing ecology in the U.S. The report suggested to build future computing technology with various software and hardware approaches and use innovative ecosystem to realize the following targets: (i) pioneering computing frontiers, paying more attentions to availability and productivity, lowering barriers in researches and applications, and supporting the integration of edge data resources and traditional computing platforms (including innovative data-driven applications); (ii) providing early accesses to new-type hardware, software and system platforms, recognizing and supporting promising research methods, and reduce system deploying time; (iii) recognizing and lifting priority to software development that satisfying future demands; (iv) encouraging the development, deployment and maintenance of software toolkits, frameworks and systems; (v) encouraging coordinated software development and continuous development of industry, academia and the governmental laboratories in coalitions and other forms, and hence effectively utilizing national computing ecosystems (including edge computing and exa-scale computing).

The development of HPC in China has reached a key point. We must build good ecosystem to realize full-leadership and continuous progressing. The ‘small’ ecosystem means processor-oriented system and application software development, and to make general use of processors. The ‘big’ ecosystem means the coordination among industry, academia and application sectors, and integrating developments of system applications and computing infrastructures to form the competitive scientific industry. It needs better cooperation from education, research and industrial sectors to construct the large innovative ecosystem.

2 Establishing and Developing HPC Ecology in China

2.1 Significance of the Development

HPC ecology is the interaction and interrelationship among HPC systems, models and algorithms of domain applications, application software and developers, and industrialization of achievements. Founding scientific, reasonable, independent and controllable HPC ecology is the urgent demand from the country. It is an essential part to promote social and economic development and effectively to cope with challenges and risks.

2.2 Problems and Challenges

Compared with developed western countries, China still fall behind in depth and range of fundamental researches, independent and controllable key technologies, widespread of applications, and high-quality researchers and developers. There is a long way to found entire HPC ecology in China.

2.2.1 Mismatching Between Software and Hardware: Sharp Internal Contradiction

As HPC industries in China rapidly grows, its internal contradiction becomes sharp. Since the improvement of component integration and operating frequency is increasingly difficult, parallel processing through different levels and scales has been the main approach to improve computation performance, and it brings many problems from parallel programming, data storage and transferring and energy consumption. Therefore, although HPC systems have been applied to many domains in China such as nuclear simulation, complex electromagnetic environment, aircraft design and optimization, complex engineering and large equipment design, global atmosphere changes and weather forecasting, earth environment and resource exploration, life science, material and drug design, the performance of these applications cannot as fast grow as HPC capabilities. This becomes the bottleneck of the development of HPC technology in China.

Although many domestic HPCs have been ranked in the top 500 list, the level of applications does not show advantages. There are some large scale applications developed, but they are more related to data testing and parallel optimization for algorithms and programs, and still lacks of expert agreements. HPC industry in China has always been in the way that dominated by the Ministry of Science and Technology, participated by regional governments, manufactured by companies, and managed by national supercomputing centers. Although we have made magnificent achievements in this way in the past 20 years, it is needed to re-consider the fact that we might be unilaterally chasing the top of lists while ignoring real demands.

2.2.2 Lacking of Domestic Applications and Independent Key Technologies

The development of large-scale parallel software and HPC algorithm is essential to HPC industries in all countries. The weakest link of the HPC ecology in China is software. The real reason of the imbalance in the development of HPC in China is our accumulation of algorithm and software experiences are not enough to catch up with the developed countries.

At the moment, almost all renowned HPC software came from institutes in US and EU (see Table 4). Many key domains in China have to follow the development of HPC applications in developed countries led by the U.S., and adopt foreign standards. We are keen to master independent and controllable key technologies and develop commonly used domestic application software.

At present, most domains in China need to use foreign open-source and commercial software, as independently developed domestic software is still in lacking and immature status. Domains like material science, life science and atmospheric science use foreign open-source software; computer-aided engineering and fluid dynamics use foreign commercial software since domestic software have a narrow scope of

Table 4 List of renowned HPC software

Application domain	Name of software	Developed by	Started year
Atmospheric science	WRF	US	1999
Computational chemistry	Gaussian	US	1970
	ADF	Netherlands	1995
	MOLPRO	UK and Germany	1996
Fluid dynamics	Fluent	US	1983
	LS-Dyna	US	1996
Molecular simulation	GROMACS	Sweden	1991
Materials computation	VASP	Austria	2004
Basic mathematics library	Matlab	US	1984
	BLAS	US	1979
	LAPACK	US	1995
	FFTW	US	1997

applications; quantum dynamics uses self-developed software. Domain-special software usually limits to their domains, and is not very applicable to related subjects. Independent development of open platforms for general-used software resources is also an urgent task. Among the scientific domains, letting domestic scientific users use domestic developed software to realize independent and controllable key technologies is a big challenge to the HPC ecology in China.

2.2.3 Application Development Restricted by Investment and Lack of Talents

The development of large-scale parallel applications heavily depends on the national investment. DOE of the U.S. only spends less than 1/6 of the total investment to hardware, while most of the budget is spent to physical modelling, algorithm design and software development.

China has more biased to HPC hardware productions, and lacked efforts for software and systematic plans. Though there are some large-scale scientific computing programs, the majority of them are just used to overcome certain problems, and are hard to develop to general-used applications. At the moment most of large scientific computing programs comes from foreign countries. The spend to application development only takes up 10% of overall HPC investment in China, while the U.S. spends 6 times more funds than China into application development.

In the report of the 19th national congress of the communist party of China, it is clearly suggested that the training system in higher education in China can't satisfy the talent demand from HPC industry. The scientific evaluation system cannot accurately evaluate the development of HPC applications. The funding management for scientific research cannot well reflect the value difference between software and hardware achievements. The income of software developers is far below to the level in the market, and leads to the heavy outflow of talents [7].

The development of HPC applications has special requirements on related subjects, physical modelling, general algorithm and application development. It is also a big challenge to build talent teams to promote the development of the HPC ecology in China. In general, the lack of cross-subject talents has been a strong restriction to the development of domestic HPC industry [8].

2.3 Suggestions for the Development

2.3.1 Defining the Application-Driven Approach as the Source Force to the Development of HPC Ecology

The HPC industry in China start later than western countries, but in the last 2 decades with efforts paid by scientists and developers, supercomputers of Tianhe-I, Tianhe-II and Sunway TaihuLight consecutively ranked top in the world [9]. However, applications are always the weakness in the development of HPC industry in China. To develop HPC ecology it must target on domain applications. The government should show the leadership, while relevant ministries, sectors and companies could contribute united public fund to satisfy and encourage the development of applications, and promote HPC industrialization and researches [10]. It must be clearly defined that the application-driven approach is the source power to the healthy development of HPC ecology.

2.3.2 Continuous and Stable Support for the Development of HPC Application Software

Domestic subjects including scientific research, economic construction, social development and national defense have great demands on HPC resources, but to turn potential demands to realistic applications it still needs efforts. Most of HPC application developers are scattered in small-scale laboratories and institutes, or in hardware-oriented national key laboratories. There are some application level national key laboratories in domains of nuclear physics, oil production, atmospheric science and geophysics, but they have not formed a united force and lack the long-term support [11]. The development of domestic HPC ecology could go deep into scientific domains, and make the design of software fit for both hardware and applications.

To develop a good application software needs long-term efforts and investments [12]. It is necessary for us to learn advanced experiences and policies from other countries and summarize the weakness and gaps of China, and further continuously and stably supporting the development of HPC applications from aspects of policies, funds and talents to best promote the development of domestic application software.

2.3.3 Recognizing the Importance of Strategic Supports from the Development of HPC Ecology to the Modern Economy System

National and regional governments should continuously invest the HPC applications, and insist on the application-driven approach and further expand application domains. Application domains are not bound to science and engineering, but also include economy, finance, national and social security, etc. [13]. The development of HPC ecology should be unified with national strategic demands, and fully take the advantage of supercomputing to social benefits.

It needs to recognize the importance of the coordinate development of HPC ecology. The demand of nationwide plans of HPC application development is more urgent than the performance competition of supercomputers. We should guide HPC industry to orderly serve regional economy from national strategy level, and incline high quality resources to major scientific challenges and engineering issues [14]. Meanwhile, it is important to found coordination of diverse sectors and institutes, and promote the popularized development of HPC ecology from scientific researches, industrial layouts and application demonstrations [15]. It is also necessary to maintain and improve public computing service platforms such as CNGrid, so that we can effectively lower the barrier of supercomputing, popularize HPC knowledge, improve the development of HPC applications, and provide strategic supporting services to the modern economic system.

3 Summary and Expectation

Independence is the only way to develop the HPC ecology in China, and application-driven is the inexhaustible driving force for the HPC development [16]. The development of HPC algorithms and application software should support domain applications, and guide the design of chips [17]. HPC application software should be able to rapidly adapt to various HPC heterogeneous platforms so that it will no longer become the bottleneck of HPC applications in China. At the same time, the application software needs to show openness, and can be easily integrated into various advanced domain algorithms, so that the application software can be sustainably developed, and finally form representative domain applications and improve the support from the scientific research informatization in China [18]. In addition, in order to realize the ecological development of HPC, one should recognize the importance of the series of coordinated development strategies, including low energy consumption systems, programmable environments and high adaptable chips [19].

Meanwhile, constant improvement of the computing power of HPC systems is a certain result of the development of computer technology [20]. However, whether it is necessary for economic and social development and scientific research will be determined by the trend of development in the world and the characteristics of China in the new age [4]. In the future, HPC systems should be developed according to the application requirements, as only the application requirements are the driving force for the development of computers, and keep the development of HPC ecology in the healthy manner.

Acknowledgements During the process of the construction of HPC ecology environment in our country, the Supercomputing Innovation Alliance actively played the role of demand traction, resource sharing, collaborative innovation and sustainable development. It effectively promotes the innovation of technologies, the popularization of applications and the development of industry, and accelerate the development of HPC ecological integration of production, study, research and use in the HPC industry in China. Through building resource sharing, collaborative work and service mechanism, CNGrid will effectively support the applications in many domains in China, and promote the national HPC ecological environment construction and the development of related industries through technology innovations.

This paper is strongly supported by the above platforms by providing relevant data and use cases. In addition, the experts including Li Guojie, Qian De-pei, Sun Ninghui, Mo Zeyao, Wang Dingsheng, Chen Runsheng, Sun Jiachang, Zhang lin-bo, Ge Wei, and etc. have made long-term efforts to the scientific research work of HPC, and lead to a great number of influential theoretical and practical achievements. All of these help this paper to better explaining and analyzing and the road of HPC ecological environment construction in China.

Thank all of you who work on HPC. Thank my colleagues Zhao Yining and Gu Beibei for providing helps.

This work is supported by the National Key R&D Program of China (2018YFB0204000).

References

1. Zang D, Cao Z, Sun N (2016) Development of high performance computing. Sci Technol Rev
2. Lan C (1992) American high performance computing and communications program. Int Sci Technol Exch (9):10–11
3. https://www.mellanox.com/coe/
4. Achieving exascale computing through hardware/software co-design
5. Li N (2014) To break the bottleneck of supercomputer software development, we must pay attention to the cooperation. Sci Technol Rev (25)
6. Li G (2019) Preface: several strategic issues to consider in developing high performance computing. Bull Chin Acad Sci (6):607
7. https://mil.huanqiu.com/world/2018-07/12393076.html?agt=15438
8. Chi X (2018) National high performance computing environment development report. Science press, p 8
9. Li J (2018) Overview of high performance computing applications. Tsinghua University Press, p 297
10. Yan J (2016) Intelligent computing: leading a new era of supercomputing [EB/OL]. https://news.sciencenet.cn/ht-mlnews/2016/10/357967.shtm
11. Jin Z, Lu Z, Li H, Chi X, Li J (2019) The origin of high performance computing—application status and development of scientific computing. Bull Chin Acad Sci (6):625
12. Ge W, Li J (2011) Multiscale parallel computation of multiphase complex systems. Blue book of China scientific research informatization. Science Press, p 72
13. Chi X, Gu B, Wu H, Wang L, Zhu P (2013) Analysis of development status of high performance computer system and platform. Comput Eng Sci 35(11)
14. Zhou H (2009) China's informatization process. People's Publishing House, p 451
15. Gu B, Wu H, Chi X, Zhao Q (2014) Analysis on the development of high performance computing applications at home and abroad. Res Inf Technol Appl 5(4):82–91
16. Zhiwei X, Chi X, Xiao N (2016) High-performance computing environment: a review of twenty years of experiments in China. Natl Sci Rev 3:36–48. https://doi.org/10.1093/nsr/nww001
17. https://www.top500.org/
18. NSF/DARPA/NASA digital libraries initiative: a program manager's perspective
19. On the role of co-design in high performance computing
20. Inventing new technology with academia and industry

Xuebin Chi, Deputy Director, professor and doctoral supervisor of the Computer Network Information Center of Chinese Academy of Sciences, and director of the Center of Scientific Computing Application and Research of Chinese Academy of Sciences. In 1989, he obtained the doctor degree in Computational Mathematics from the Computing Center of Chinese Academy of Sciences. He is the project leader of "Research on Service Mechanisms and Supporting Systems in the National High Performance Computing Environment (phase I & phase II)" of National Key R & D Program. His main research orientations are parallel computing, software and grid computing technology. He has won 3 second prizes of the National Science and Technology Progress Award, 1 second prize of the Natural Science Award of the Chinese Academy of Sciences.

Advanced Scientific Research Environment Evolution and Cloud Service Architecture Design Integrating 5G Technology

Xu Zhou, Congshan Fan, and Bing Liu

Abstract With the development of 5G network technology and the advancement of cloud computing technology, 5G cloud computing has significant advantages in terms of resource utility ratio, on-demand service, and security. At present, the demand for massive data processing is increasing with "data-intensive scientific research". Compared with the network resources construction of traditional scientific research institutions, 5G scientific research cloud which relies on its excellent network performance, security guarantee, efficient flexible computing resource allocation ability and simple hardware requirements can realize the intelligent and flexible construction of computing resources in the face of different needs, localized specialized services, cloud network collaborative optimization and traffic payment, and fully satisfies the stable and high-speed network usage requirements of researchers and students. This paper introduces the construction requirements of 5G networks and their key technologies and scientific research clouds, and proposes a 5G scientific research cloud architecture. Through the analysis of typical scenarios such as research institutes, large scientific installations, field stations, and university campuses, this paper expounds the practicability and necessity of the intelligent research network constructed by 5G scientific research cloud, and finally summarizes and forecasts it.

Keywords 5G · Cloud computing · Edge computing · Cloud-edge collaboration · Software-defined network

1 Introduction

The 5th generation mobile network is the new generation wireless mobile communication network, which is dedicated to provide the extraordinary user interactive experience. 5G network breaks through the spatial and temporal constraints, greatly shortens the distance between people and things, and realizes the interconnection between people and tremendous devices or objects [1]. In 2012, the International

X. Zhou (✉) · C. Fan · B. Liu
Computer Network Information Center, Chinese Academy of Sciences, Beijing, China
e-mail: zhouxu@cnic.cn

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_13

Telecommunication Union (ITU) organized global industries to carry out relevant 5G research work. The 13th Chinese national "Five-Year Plan" clearly puts forward the requirements and terms of developing and promoting the key technologies of the 5th generation mobile communications and ultra-broadband, and launching 5G commercial applications. Plenty of companies, including Huawei, Zhongxing and Datang, are conducting 5G technology research experiments and have achieved some progress. This procedure accelerates the deployments of 5G network and popularizes and new innovative applications, by making full use of the advantages of political systems and mechanisms, large markets and international cooperation, which is beneficial to boost and improve the China's leading position in international technology competition [2].

The ITU has defined three major application scenarios for 5G: Enhanced Mobile Broadband (eMBB), Massive Machine Type of Connections (mMTC), and Ultra Reliable Low Latency Communication (uRLLC) [3]. Typical applications of eMBB include ultra-high-definition video, virtual reality, and augmented reality. This kind of scenarios not only requires extremely high bandwidth, but also is sensitive to latency given the interactive operations. Typical applications of uRLLC include industrial control, drone control, and intelligent driving control, which focuses on extreme latency sensitive services and has the fundamental requirements on high reliability. The mMTC using scenarios are typically involved in smart cities and smart homes. This kind of applications requires high connection density, meanwhile presents industry diversity and differentiation. 5G industrial application is shown in Fig. 1.

5G combines technologies such as network slicing, edge computing, massive MIMO (Multiple-input multiple-output), high-frequency communications with cloud computing, artificial intelligence (AI), augmented reality (AR), virtual reality (VR), and control technologies. The combination of specific technologies will be

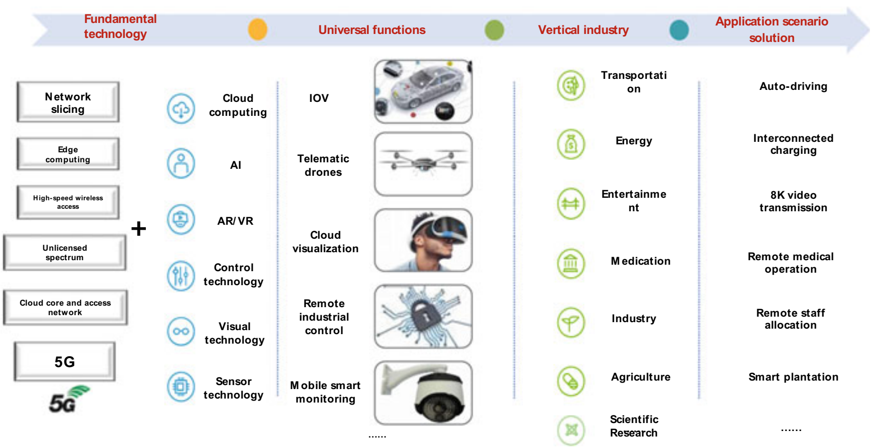


Fig. 1 5G industrial application

deployed in some vertical industries, such as transportation, industry, medical care and agricultural industry, which helps to realize telematics, connected drones, cloud AR or VR, remote industrial control and some other new functions, and finally achieves the goal of Internet of Everything. Integrating 5G into the scientific research cloud can flexibly respond to various scientific research scenarios. Researchers can deploy and install sensors to collect scientific data according to scientific research needs, and the collected data can be transmitted back to the network through high-speed wireless access, while the research procedure can be controlled remotely. The coordination of edge cloud and central scientific research cloud can accomplish efficient data storage, computing and calculation, which supports the visual interactive AR or VR approaches to build multi-dimension research model to provide researchers a more intuitional way to draw their scientific conclusion and acquire their discoveries. This new research paradigm will make contributions to constructing China into the world's major scientific center and innovation highland.

Whereas, the deployment of scientific research cloud and commercial cloud is essentially different due to the different user groups and application scenarios, as shown in Table 1. The scientific research outputs of institutions and universities are based on rigorous scientific experiments. The tremendous data generated by large-scale scientific experiments need to be stored, calculated, and processed timely, which places severe requirements on the latency of cloud platforms. Traditional commercial cloud design paradigm is hard to meet the requirements of some special scientific research scenarios, especially some equipment or installments settled in remote rural areas. As a result, it is also necessary to establish and deploy a proprietary scientific research cloud to tackle with the large-scale data distribution and processing, as well as the resource allocation and scheduling. In addition, the deployment of scientific research cloud in scientific research parks can provide edge cloud services by combining edge computing technology. Reasonably resource schedule, arrangement and management at the edge side closed to users can effectively reduce latency, improve system security, ensure data safety and secure the storage and transmission of data and achievements. At the same time, scientific research cloud can also be jointly deployed with commercial clouds to coordinate network resources, meet the huge computing power requirements, and alleviate network pressure during peak periods.

Table 1 The IT requirements of future research paradigm and scientific installation

Equipment	Network requirement	Storage requirement	Computation requirement
Five-hundred-meter aperture spherical telescope: FAST	10–100 Gbps	1–10 EB	1 EF
Square kilometer array: SKA	10–100 Gbps	1–10 EB	40 PF
High repetition frequency X-ray free electron aurora device	10 Gbps	100 PB	1–10 PF
Major marine technology facilities		0.5 EB	100 PF
Advanced nuclear energy system: ADANS	Quantum encrypted communication	100 PB	600 PF
Shanghai synchrotron radiation facility phase II	1 Gbps	500 PB	20–40 PF
Remote sensing satellite ground station	10 Gbps	100 PB	
Beijing spectrometer experiment III: BESIII	100 Gbps	15 PB	10 PF
JUNO	100 Gbps	30 PB	10 PF
Large high air altitude shower observatory: LHAASO	100 Gbps	30 PB	10 PF
Hashed neutron source	100 Gbps	50 PB	20 PF
High energy photon source	100 Gbps	100 PB	20 PF
Virtual platform for material conversion process	1 Tbps	1 EB	1 ZF

2 5G Key Technologies

With the rapid development of scientific research, the scientific research cloud which utilizes current fixed networks, 4G networks, and WLAN technologies, cannot ensure massive access of tremendous scientific research equipment and cannot provide high data transmission speed for some scientific equipment especially installed in remote areas. At the same time, the lack of spectrum resources further limits the access rate to the scientific research cloud, and the spectrum licensed by telecom operators is not conducive for institutions to establish their private networks. 5G millimeter wave technology has developed new spectrum resources, which can effectively increase the number of connections and promote transmission rate when combined with Massive

MIMO technology. Network in Unlicensed Spectrum is beneficial to deploy institutions' private 5G network flexibly to meet some specific targeted needs. Applying network slicing technology in the scientific research cloud can support different performance indicator requirements in a variety of scientific research scenarios. Adopting edge computing technology can help reduce latency, reduce transmission network congestion, and provide users with flexible and schedulable services through the collaboration between edge and cloud. Cloud Oriented Core Network can overcome the security issue of "Technology Cloud", and meet the requirements of high security, customization, and openness of the network. The key technologies involved in the "5G Scientific research cloud" will be illustrated in detail below.

2.1 Network Slicing

In the development of 5G networks, network slicing technology can provide operators with multiple virtual network services based on a single physical facility. It is a combination of a series of logical network functions which can be used to meet the communication needs in specific application scenarios [4].

Network slicing can implement stream management of network data. Its principle is to divide the actual physical network into several different types of virtual networks at the logical level to meet different scenarios according to targeted performance indicators, such as latency, bandwidth, reliability, so as to support dynamic changeable application scenarios.

Network slicing is one of the distinctive features and advantages of 5G networks. At present, the slice mode generally accepted in the industry is implemented according to three typical application scenarios of 5G [5], as shown in Fig. 2. Network slicing can conduct the network function customization and the arrangement management of the network resources according to the scenario features and

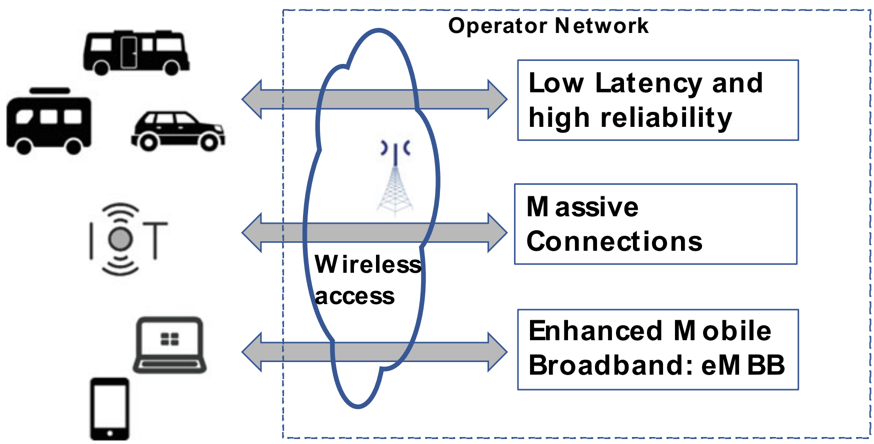


Fig. 2 Three major application scenarios of network slicing

users' requirements for three typical application scenarios containing low latency and high reliability, massive connectivity and enhanced mobile broadband, which realizes the efficient networking, provides personalized services for users and improves the resource utilization and the flexibility of the network services.

2.2 Edge Computing

Edge computing technology is one of the key technologies to achieve the decentralization of 5G networks, which provides some IT capabilities, such as cloud computing, storage, cache and IT service environment at the edge of the network. These capabilities are always deployed generally closed to users or at the wireless access facilities. Edge computing technology is the integration and evolution of mobile base stations and IT technologies [6].

As an open and distributed platform, edge computing provides many network services near the data source at the edge, which can fully meet the needs of industry digitalization as aspect of fast connection, data improvement, intelligent applications and security protection. According to the standards published by ETSI, edge computing is mainly divided into seven application scenarios, as shown in Table 2.

As shown in Fig. 3, edge computing platform is deployed at the edge of the base station, which provide the capability of processing data locally and the ability of intelligent traffic management. This aims to realizes intelligent traffic distribution and supervision, and meanwhile meets superior network service requirements [7]. Edge computing can also provide distributed cache capabilities to accelerate content

Table 2 Seven major MEC application scenarios published by ETSI

Scenario	Features	Issues solved by MEC
Intelligent video acceleration	Large volume	Network congestion optimization
Dense computation assistance	Low latency and massive access	Accuracy and timeliness of information processing, capability of dense computation
Video stream analysis	Large volume	Video stream analysis and intelligent processing
IoV	Low latency and high bandwidth	Timeliness of analysis and decision
IoT gateway	Tremendous traffic data	Capability of local storage and processing
AR/VR	Low latency and high bandwidth	Capability of local processing and requirements of accuracy and timeliness
Network collaboration	Trend of enterprise network service platform	Smart dynamic network choice between operator and enterprise

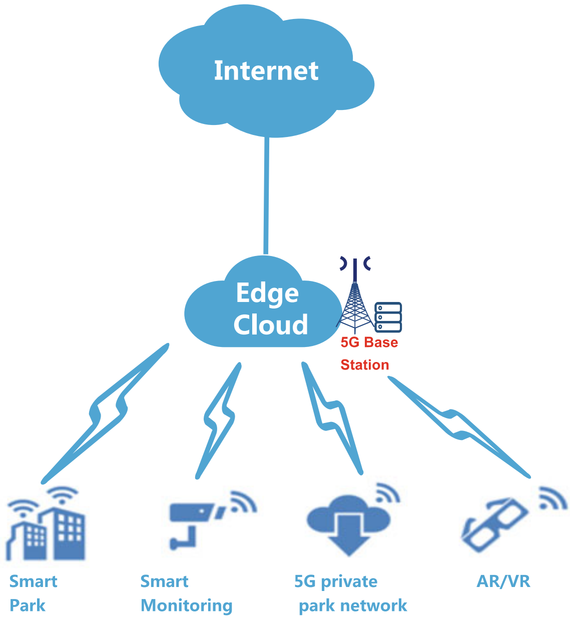


Fig. 3 5G edge cloud deployment architecture

downloads and provide big data collection and localized characteristic services to meet users' various network needs.

2.3 *High Speed Wireless Access*

At present, the spectrum of wireless telecommunication systems is below 6 GHz, and 5G millimeter waves will launch new wireless access technology research in the millimeter wave band (30–60 GHz) [8]. The most prominent advantages of millimeter wave communication are the short wavelength and high bandwidth. The combination of millimeter wave technology and massive MIMO technology can effectively improve antenna gain [9]. The ultra-bandwidth of millimeter waves is almost a thousand times that of LTE, which guarantees the ultra-high speed and massive access in 5G systems.

As shown in Fig. 4, massive MIMO is a large-scale array antenna deployed at the base station side. When the number of antennas is much larger than the number of user terminals, beamforming technology is used to concentrate the antenna energy in a narrower direction for multi-user transmission. When channels tend to be orthogonal, Gauss noise and inter-cell interference will tend to disappear, and spectrum resource can be reused in spatial dimension. This technique can increase cell capacity and spectrum efficiency several times [10], reduce transmission delay and signal attenuation efficiently when complementary with millimeter wave technology.

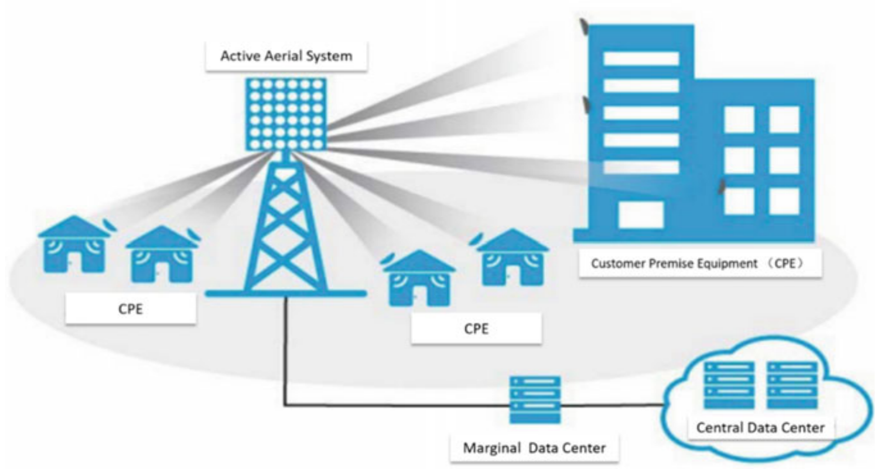


Fig. 4 5G massive MIMO

The integration of millimeter wave technology and massive MIMO can effectively improve the number of wireless access and data transmission rates. The main application scenarios include fixed wireless access and integrated wireless access and backhaul [11]. 5G fixed wireless access replaces the last mile of fiber access and provides gigabit data transmission rates to homes, apartments or businesses at a lower cost. 5G New Radio (NR) integrates wireless access and backhaul to realize wireless backhaul and relay links in 5G NR cells, which can get rid of excessive dependence on wired transmission networks and help deploy and construct 5G cells flexibly and densely. At the 5G scientific research cloud platform, it is more convenient and efficient to use wireless instead of wired network for data transmission, especially for scenarios where some scientific research equipment has been installed and fiber deployment is complicated and time-consuming. In addition, local data processing is no longer limited by wired network, given the ultra-high transmission rate of 5G network.

2.4 New Radio in Unlicensed Spectrum

With the continuous expansion of spectrum deficit in 5G era, networking and deployment of the unlicensed spectrum has become one of the effective solutions. New Radio in Unlicensed Spectrum (NR-U) which based on NR refers to the use of NR protocol to provide access services in the unlicensed spectrum, as the extension and supplement of 5G NR.

5G NR-U includes two modes: licensed assisted access NR-U (LAA NR-U) and stand-alone NR-U (Stand-alone NR-U) [12]. LAA NR-U is an extension of LTE-U/LAA technology in 4G era to 5G NR. Stand-alone NR-U has no longer need licensed spectrum as its anchor. It runs completely independently on the spectrum, which means anyone can deploy his 5G network like Wi-Fi. It can be used to deploy a single access point or its own 5G private network. Stand alone NR-U will share the unlicensed spectrum as fair and open as Wi-Fi [13]. Having stand alone NR-U, 5G and Wi-Fi will move towards fusion.

NR-U contributes to scientific research institutions to flexibly deploy their 5G private network and integrate Wi-Fi network, and provide services for corresponding scientific research.

2.5 Cloud Oriented Core Network and Access Network

With the rapid development of the network, 5G core network will be deployed on the cloud-enabled infrastructures, which replaces proprietary hardware with generic hardware through virtualized cloud computing technology. Clouding of the core network blends various of network devices to realize centralized and unified management and control of multiple software or hardware resources and their related network functions, to improve the capacity of network virtualizing, to provide flexible and dynamic transmission architecture, meeting the coordination and control requirements of virtual machines. 5G core network, which is based on clouding, servicizing and software architecture, enables customization, openness and servitization of network by means of web slices, user face separation control and other technologies, so as to face the Internet of Everything and all walks of life [14]. 5G access network realizes resource sharing and dynamic scheduling, and improving of spectrum efficiency by reducing the number of base station rooms, reducing energy consumption, adopting collaborative and virtualization technology in order to achieve an operation of low-cost, high bandwidth and flexibility. To meet the various challenges brought by the rapid development of mobile Internet to operators (energy consumption, costs of construction, operation and maintenance, spectrum resources) it is necessary to pursue sustainable business and profit growth in the future.

Cloud oriented core network and access network have displayed bright prospects to the operator and the user: cloud oriented core network and access network shows a bright future to operators and users: reducing the costs and energy consumption of operators' purchase/operation and maintenance; rapid business deployment, and reduced innovation period; the network application could achieve multiple version and multiple tenant; different applications, users and tenants are allowed to share a unified platform; different physical areas and user groups can achieve personalized business; promote the network opening, business innovation; incubation of new profit growth point [15].

At present, Amazon deploys the world's first mobile core network on the public cloud. Users can build their own 4G private network with basic charge of $89.99 per month by renting the core network from the cloud platform and simply deploying the edge node as Wi-Fi to connect the core network. The entire mobile private network not only supports high security, but also supports large bandwidth, multiple video streams and low latency application scenarios due to the management and deployment of broader nodes based on SIM card [16].

3 Demands of Scientific Research Cloud Construction

After experiencing experimental science, theoretical science and computational science, scientific research has produced the fourth paradigm represented by "data intensive scientific research", which will promote the qualitative leap of scientific research methods. Scientists need to be strongly capable of calculation while doing the experimental analysis, for instance, seismologists need to sort out the data of sensors after an earthquake; astronomers need to process the observational data of space telescopes. Analogous experiments of large scientific research devices need to store, calculate and process massive data safely and quickly. The experimenters can save a lot of scientific research costs provided relying on 5G scientific research cloud, solving the problem of big data of scientific research through cloud computing, and sharing the configuration of data, software and calculation to make the cooperation between group-members more convenient.

In addition, it is hard to ensure the coverage of signal without dead angle on account of the insufficiency of coverage and access capability of 4G network. The costs of full coverage through Wi-Fi is too high and the disconnection and reconnection while switching the network will affect the browsing experience of user's. We can use specialists of 5G network such as wide coverage, large connection, high bandwidth, low delay to bring, convenient, stable and high-speed browsing experience to users of scientific research institutes and universities at anytime and anywhere through 5G research cloud service. Besides, with the help of research cloud platform, the network traffic costs of users could be greatly reduced by means of unified payment of traffic and distribution control, meanwhile, the computing process of massive data in the edge cloud could save a large number of research hardware costs for universities and research institutions.

It is thus clear that 5G research cloud which have been achieved with the aid of 5G network technology by "edge + net + cloud' framework, can sufficiently meet the various needs of future scientific research for data processing, collection and storage. As shown in Fig. 5, the scientific devices of scientific research institutions need high-speed access capability of the network to meet the rapid processing, computing and storage requirements of scientific data, which need to exploit scientific research cloud's capabilities such as multi-access edge computing and mobile broadband enhancing. Massive scientific data transmission also needs to use network slicing technology to provide a special virtual network to enhance the data transmission capacity in addition to the high performance of 5G network; The multi-access edge computing capability of 5G research cloud is also needed to support massive data

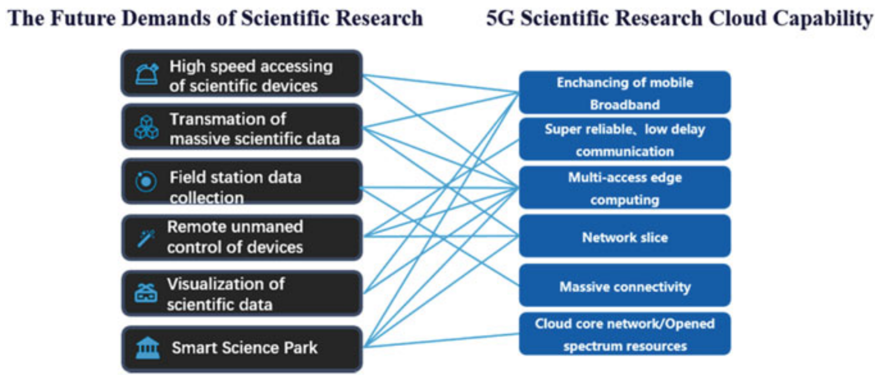


Fig. 5 Demands of 5G scientific research cloud construction

connection, local processing and transmission of data in the field station data collection; In order to provide reliable network communication with low delay and high bandwidth, it is necessary to control the scientific research instruments remotely and unmanned, and to use 5G scientific research cloud's capabilities like network slicing and multi-access edge computing while conducting scientific research experiments.

Scientific data visualization will become increasingly popular in the future scientific research experiments and teaching, and the application of AR/VR and holographic projection technology will be more extensive, which needs the aid of 5G scientific research cloud's multi access edge computing and network access capability of the enhanced mobile bandwidth to ensure the time-validity of data transmission and the efficiency of data processing; In the future, the construction of Smart Science Park will contribute to scientific teaching and personnel training, which requires the capabilities of 5G research cloud such as cloud core network, open spectrum resources and network slicing to conduct intelligent supervision and reasonable scheduling of network resources and ensure the time-validity and security of the network.

4 5G Scientific Research Cloud Architecture

4.1 Design Ideas

The construction of 5G Scientific Research Cloud should follow the principles of manageability, extensibility, advancement, reliability and economy on the whole structure. It has such functions as logical concentration, function distribution, multi-access network management, localization of characteristic services, cloud edge collaborative optimization, and intelligent resource scheduling. 5G Scientific Research Cloud uses 5G key technologies such as mobile edge computing, network slicing, cloud oriented core network and access network, millimeter wave transmission, unlicensed spectrum and SD-WAN and other network optimization means to manage and control network resources through edge network cloud collaboration, so as to realize rapid, intelligent and flexible construction of computing resources, local characteristic services, and cloud network collaborative optimization and network traffic statistics and payments system, it can meet the diversified network use needs of the scene users such as scientific research institutes.

As shown in Fig. 6, in the 5G Scientific Research Cloud, the wireless network side of the edge cloud realizes mass high-speed access to terminal devices by means of 5G Massive MIMO and millimeter-wave transmission. The data traffic is managed, cached, calculated and stored through the edge MEC server, and the network delay is greatly reduced by reducing the return route. Resource management, task coordination and function opening are realized through cloud edge collaboration, and services such as 5G edge cloud transmission optimization, document synchronization, cloud printing, website conversion, media acceleration, wireless roaming, data processing, network traffic statistics and payments are provided. In addition, edge cloud provides customized services for 5G network customers through network function virtualization, network slicing, SDN and distributed machine learning technology, latching on to the vertical industry, providing the service mode of NaaS for 5G network, supporting the rapid deployment of new businesses, providing flexible networking, and supporting the rapid upgrading of network functions. Between 5G Scientific Research Cloud and edge cloud, unified scheduling and optimization of network traffic shall be conducted through SD-WAN and other virtualization network technologies, centralized control of WAN bandwidth and routing, and effective acceleration of network shall be achieved through SD-WAN. On the 5G Scientific Research Cloud, unified resource management and scheduling of data flow shall be carried out by using cloud edge collaboration and cloud network collaboration, and massive

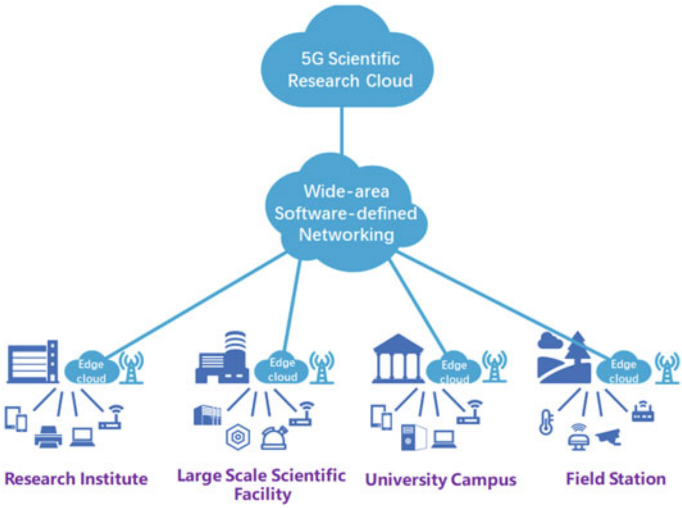


Fig. 6 Schematic diagram of scientific research cloud and edge cloud integration

data shall be processed rapidly, and cloud network function shall be opened and user self-service shall be provided.

4.2 Technology Architecture

The 5G Scientific Research Cloud service architecture is composed of edge, network and cloud, as shown in Fig. 7. Specifically, it refers to edge cloud, transmission network and China Science and Technology Cloud. At the bottom of the edge cloud is the virtualization of computing resources, on which a large number of network functions are virtualized, such as Virtual Customer Premise Equipment (VCPE), Virtual Content Delivery Network (VCDN), Virtual Deep Packet Inspection (VDPI), Virtual Gateway (VGW), which can manage the internal and external network resources of the edge cloud and it can realize more efficient data transmission through network slices. The top layer of network function virtualization is a series of edge cloud services of the China Science and Technology Cloud, including transmission optimization services, cloud print, website conversion, media acceleration, wireless roaming, data processing, flow control and other localized services. The upper layer of edge services of China Science and Technology Cloud needs to realize edge-cloud collaboration, that is, the edge cloud should cooperate with the intermediate transmission network and centralized China Science and Technology Cloud in the whole network, improve the overall efficiency of 5G Scientific Research Cloud, and judge which services need to stay in the centralized cloud and which capabilities need to be transferred to the edge cloud.

The transmission network use CAS's own backbone, the China Science and Technology Network, which includes its own deployed optical fibers and also rents part of the carrier's bandwidth, so the overall deployment and management of the network is crucial to improving transmission efficiency and reducing costs. At the bottom of China Science and Technology network is the virtualization of computing resources. On the basis of this, the virtualized network management is carried out by means of network function virtualization, including virtual switch (vSwitch), virtual router (vRouter), information center network node (ICN node), trial network slice, etc.

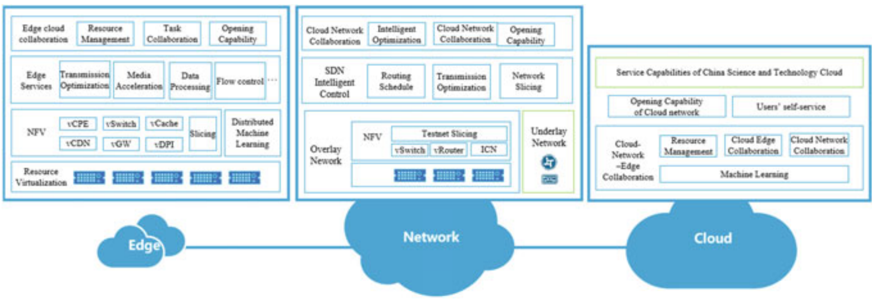


Fig. 7 Schematic diagram of the scientific research cloud service architecture

In the upper layer of network function virtualization, wide area software defined network (SDWAN) is used to search the optimal path for scientific research users to access international scientific research data resources or transmit scientific research data through the functions of routing, transmission optimization and network slicing. SDN intelligent control of the top is the cloud-network collaboration.

China Science and Technology Cloud integrates data and computing tools needed by scientific research users into one website and provides them with specific categories for users to use as needed. In the 5G Scientific Research Cloud service architecture, China Science and Technology Cloud focuses on the underlying cloud-edge collaboration, including resource management, service management, cloud-edge collaboration, cloud-network collaboration, machine learning and other functions.

4.3 Typical Scenarios

4.3.1 Research Institutes

With the rapid development of scientific research and the popularization of big data and artificial intelligence, the current scientific research is the fourth paradigm represented by "data intensive research". Scientists often need to use a large number of sensors and collect a large amount of experimental data for scientific research and experiments, meanwhile, they need to store, calculate and process the large amount of data safely and quickly. However, the current research network or 4G network is not adequate to meet the requirements of research big data on network performance, Wi-Fi network has regional limitations and is not stable enough, and cable optical fiber is limited and not easy to deploy flexibly.

Depending on 5G scientific research cloud, the experimenters can solve the problems of scientific research big data through cloud computing, saving the cost of scientific research, and sharing the configuration of data, software and calculation, so as to the cooperation between group-members more convenient. 5G science and technology cloud service can build a special network of the research institute park, as shown in Fig. 8, through network slice and traffic unified payment, users will make high-speed network access using convenient 5G wireless network. Meanwhile, it provides local computing, caching and network security through edge cloud. In addition, 5G technology cloud can provide IPv4/IPv6 conversion service based on edge cloud, support transparent protocol conversion, and realize IPv6 access and visit on the basis of no modification on system and terminal transformation [17]. When traditional 4G network users access cloud services, their requests need to go through the mobile core network, backbone network, and then through the interconnection exit, which may only return to the local area by detour and multiple hop. The delay is as high as tens of milliseconds, the user experience is poor, and it brings high inter network settlement costs. If 5G technology cloud is adopted, the delay and cost can be greatly reduced, and the online experience of users can be

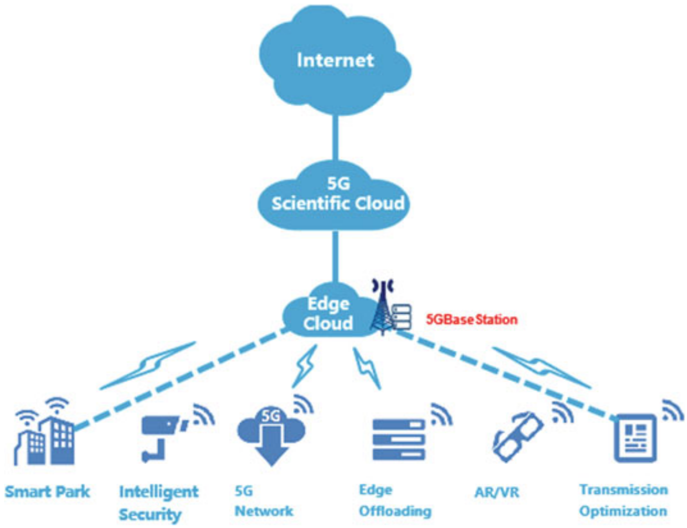


Fig. 8 Schematic diagram of 5G scientific research cloud for integration of research institutes

improved. 5G technology cloud can integrate access management through heterogeneous networks, optimize scientific research traffic management and network transmission, and provide more intelligent park management, security and visualization services through edge computing and 5G network characteristics. The Institute of high energy physics of Chinese Academy of Sciences adopts the intelligent flow identification and control management solution of 5G scientific research cloud service to guarantee the priority transmission of scientific research data. During the epidemic of 2020 novel coronavirus, CNIC urgently built an "International Academic Resource Access Acceleration Service Platform" with the edge computing technology to provide the international resource access acceleration services for 32 frontline scientific research institutions fighting against the novel coronavirus, containing Wuhan Institute of Virology, Institute of Microbiology, Institute of Zoology, and Institut Pasteur of Shanghai Chinese Academy of Sciences, which effectively solved the problem of accessing the international biological scientific research resources sites with poor experience and low download speed, correspondingly provided a strong guarantee for the research of anti-epidemic.

4.3.2 Large-Scale Scientific Facility

Large-scale scientific devices need to carry out real-time remote control and management of equipment and use intelligent park and intelligent security to ensure the normal operation of scientific devices. The data generated are transmitted to the scientific research cloud by high-speed wireless access. New scientific laws are found by analyzing and calculating the data generated by these scientific devices, and the visual display of scientific data is realized.

Once the general large-scale scientific device is opened, it is difficult to monitor its operation state. The transmission of large amount of data needs to occupy a large bandwidth, which will reduce the transmission efficiency, bring a large load to the network, and cause a waste of resources. It is low efficiency and high cost to transfer data to scientific research cloud for processing, and causes great pressure on scientific research cloud, and the traditional visual display method is difficult to realize the visual modeling of massive scientific research data.

In 5G Scientific Research Cloud, for some scenes in which no legal person intervenes during the experiment, remote control and management can be carried out by means of robots [18]. As shown in Fig. 10, data generated by scientific devices can be uploaded to the network for transmission through 5G high-speed wireless access. The transmission of massive data needs to develop a reasonable traffic management scheme. Intelligent traffic management based on edge computing can make use of the processing ability of edge computing. From the beginning of scientific data generation, it can carry out corresponding operations to reduce the load of network transmission. The intelligent traffic recognition based on edge computing and deep learning is used to identify the traffic through deep learning, determine the specific application and traffic type, and lay a solid foundation for the next step of special optimization processing for specific scientific data.

On the basis of traffic identification, the traffic compression based on deep learning is carried out. Because the scientific data has a fixed pattern, using the deep learning model to train the scientific data can obtain more efficient data compression and greatly improve the network transmission efficiency. Edge cloud collaborative data processing helps to respond to user needs quickly, improve processing efficiency, and reduce the load on the cloud. The new visual interactive AR or VR method can help researchers directly observe the simulation and calculation process of scientific data and promote scientific research to find new scientific achievements faster. Based on the construction of China Science and technology network, China VLBI network has realized the networking with the international E-VLBI observation network, which can meet the requirements of high-speed, high-precision and large capacity data transmission (Fig. 9).

4.3.3 Field Stations

In field stations, such as Xishuangbanna Tropical Botanical Garden of Chinese Academy of Sciences, it is necessary to deploy various sensors in a wide range of observation areas or use UAV, ship and vehicle to collect observation data, and remotely control and manage equipment according to different observation needs and data characteristics. The collected data is sent back to the scientific research cloud through 4G network or wireless mesh to complete data storage and processing and realize the visual display of scientific data.

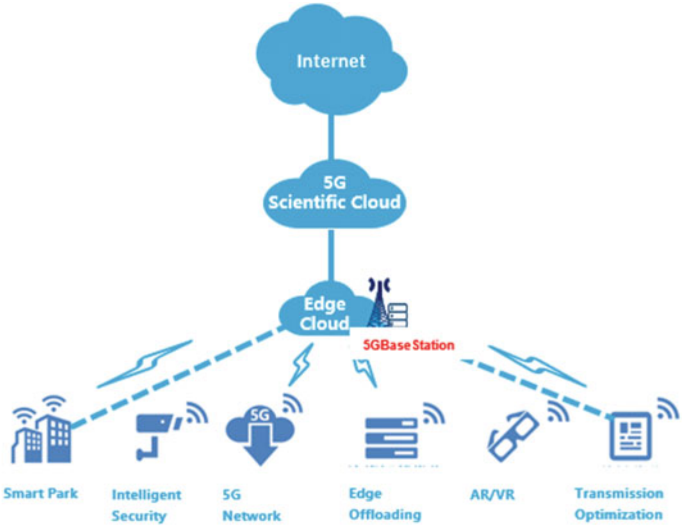


Fig. 9 Schematic diagram of 5G scientific research cloud for integration of large-scale scientific facility

With the expansion of scientific investigation area, the collection of data is increasing. 4G network has limited connection capacity and low transmission rate, which can't guarantee the real-time return of collected data. It is low efficiency and high cost to transfer data to the scientific research cloud for processing, which is easy to cause greater pressure on the research cloud, while the traditional visual display method is lack of three-dimensional and intuitive modeling.

As shown in Fig. 10, in 5G Scientific Research cloud, 5G high-speed wireless access is used to complete the data collection and transmission, with high transmission rate and good real-time performance, and the edge server is deployed at the base station of telecommunications operators by using 5G edge computing technology. MEC has multiple access management capabilities, which can integrate 5G access, Lora, mesh, UAV network and other technologies to realize unified management of virtualization, conduct flexible networking, and complete scientific data return in remote areas. MEC has the function of data diversion, which can directly offload the specified observation data to the local LAN, as avoiding data transmission to the operator's public network [19]. MEC has the function of Lora network server to realize the management of remote Lora gateway. MEC has the function of data preprocessing, which can clean and verify the collected data. MEC has the function of artificial intelligence image recognition, which can carry out intelligent analysis on video and pictures, real-time discovery of key events. MEC builds the big data application platform to realize the storage and processing of the massive data. By using AR or VR technology to realize multi-dimensional visualization model is helpful to support researchers to intuitively discover research results. In Namuco station and Southeast Tibet station, 5G Scientific Research Cloud architecture can establish a

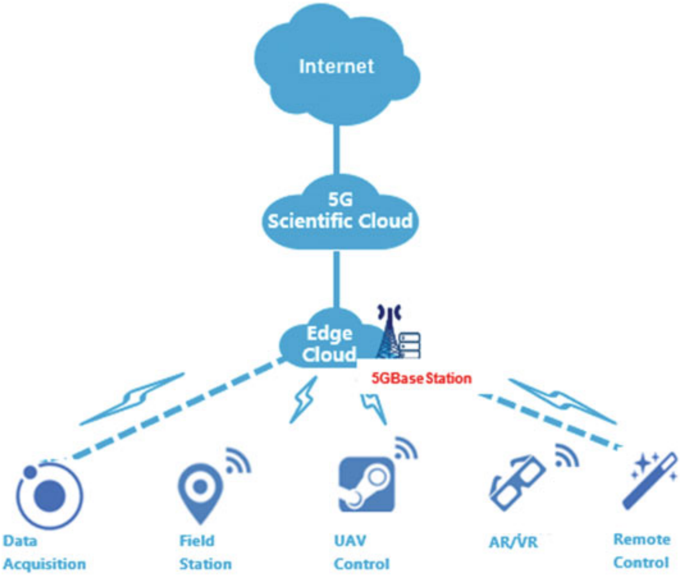


Fig. 10 Schematic diagram of 5G scientific research cloud for integration of field stations

long-term stable operation integration network service platform in extreme environment, and comprehensively improve the automatic data collection and transmission capacity of field stations and the management of field station data.

4.3.4 University Campus

The campus covers a large area, and the dormitory, classroom, library and other areas are crowded with people, while the current 4G or Wi Fi network is difficult to fully meet the Internet needs of users in the campus. The dormitory network has obvious flow tide phenomenon, as the peak service pressure is great during the heavy network use period, and the 4G network congestion problem is serious. The main application of the library network includes many kinds of rich ways, such as the query of the internal network data, the access of the external network educational resources, and the video viewing, which demand a large amount of bandwidth. The classroom is congestion of population and the network congestion is serious. At the same time, smart classroom and interactive teaching are the general trend of future teaching. This interactive teaching will certainly involve mobile interaction, and some on-site teaching will use VR devices, such as applied medicine and foreign languages, which are high bandwidth applications of the network, and are the inevitable development trend [20].

By relying on the characteristics of 5G network, such as continuous wide area coverage, hot spots and high capacity, low delay and high reliability, 5G Scientific

Fig. 11 Schematic diagram of 5G scientific research cloud for integration of campus network

Research Cloud can provide high-speed wireless data collection, scientific data visualization, intelligent navigation and positioning, unmanned machine control, mobile content distribution, intelligent park and security services for campus users. As shown in Fig. 11, 5G Scientific Research Cloud can fully meet the network requirements of smart campus through the integration of edge, network and cloud by aiming at the improvement of teaching quality, high-quality sharing of resources and intelligent management of campus by using 5G cloud technology.

5G smart lecture can build an integrated architecture of edge, cloud, and end through the conference terminal of 5G Scientific Research Cloud. The 4K ultra-high definition teaching live classroom of teachers in schools will be transmitted to more schools, so that more scholars can access the live classroom anytime and anywhere, and be widely used in remote live teaching, teaching and research, ubiquitous mobile learning and other scenes [21]. The immersive interactive learning with 5G cloud AR creates a new learning experience so that students can learn virtual courses by wearing lightweight AR glasses terminals. At the same time, the powerful computing power of the cloud realizes the rendering, presentation and control of AR applications, and transmits AR images to the terminal in real time through 5G network [22].

5G Scientific Research Cloud can realize real-time monitoring management and intelligent analysis of personnel, vehicles and equipment by building intelligent management brain in the campus cloud, taking robots as the carrier and building a safe campus scene based on 5G. Through patrol monitoring, visual recognition analysis, environmental monitoring image recognition and other applications to ensure the safe and efficient operation of the campus, it will be deeply integrated with teaching, research and campus management to build a smart campus.

5 Summary and Outlook

5G scientific research cloud which is built on the bases of 5G network technology and cloud computing technology, can fully integrate network resources through the "edge + net + cloud" architecture. It is a large-scale smart scientific research cloud integrating computing, storage, collaborative distribution, resource sharing and other functions. Its automatic deployment, management and edge data processing services can greatly reduce the cost of scientific research and teaching, and provide convenience for users to carry out diversified scientific research and teaching work and experiments, and support scientific research and talent cultivation preferably. The construction of 5G scientific research cloud will bring researchers and students a large smart scientific research environment with glamorous and broad prospects. In the future, we will continue to conduct innovative exploration on 5G research cloud, and accumulate experimental resources constantly, and enrich experimental contents, optimize service architecture, improve cloud functions, explore new directions and new ideas with more intelligent 5G research cloud to serve scientific research better.

References

1. 5G vision and requirements. IMT-2020 (5G) white paper (2014)
2. Long X, Pan Z, Gao X (2014) The trend of 5G mobile wireless communication development and several key technologies. Sci China Inf Sci 44(05):551–563
3. Li X, Gani A, Yang L, Zakaria O, Anuar B (2008) Mix-bandwidth data path design for 5G real wireless world. World Scientific and Engineering Academy and Society, United States
4. Communications world, the first smart electricity grid service based on network slicing presented by China mobile. Huawei and State grid [EB/OL] [2018.06.28]. https://www.cww.net.cn/%20article?id=435017
5. ETSI white paper no. 11. Mobile edge computing a key technology towards 5G (2015)
6. Xiang H, Xiao Y, Zhang X et al (2017) 5G edge computing and network slicing technology. Telecommun Sci 06:54–63
7. Patel M, Naughton B, Chan C, Sprecher N, Abeta S, Neal A (2014) Mobile-edge computing introductory technical white paper. White paper. Mobile-edge computing (MEC) industry initiative 1(01):1–36
8. Wang X (2007) The development and application of millimeter wave communication technology. Telecommun Inf 10:23–25
9. Zhang C (2016) Research of millimeter wave technology application for 5G. Des Tech Posts Telecommun 06:30–34
10. Zhen Z (2018) Application of large scale antenna in 5G network. Telecommun Inf (09):9–11
11. Bin X (2019) Application of large scale antenna in 5G. Telecom Eng Tech Stand (07):88–92
12. Yu Z, Jian C, Yuehong G (2018) Research on 5G licensed spectrum allocation and unlicensed spectrum utilization technology. Telecom Eng Tech Stand (03):4–9
13. Min X, Nan H, Nan L (2019) Research on 5G unlicensed spectrum networking technology. Telecommun Sci (07):8–16
14. Yi Z, Fang H (2018) Analysis of cloud core network architecture. Technol Innov Appl (08):94–95
15. Jianfeng Z (2018) Core network cloudification deployment recommendations. ZTE Technol 08:1–2

16. Athonet, Athonet BubbleCloud [EB/OL] [2019–02–15]. https://athonet.cloud/
17. Pujol E, Richter P, Feldmann A (2017) Understanding the share of IPv6 traffic in a dual-stack ISP. Springer, Cham, Switzerland
18. Jun W (2016) Research on data acquisition and signal processing systems for large scientific installations. University of Science and Technology of China
19. Wei C, Lun Y (2012) Networking for field stations of Chinese Academy of Sciences. e-Sci Technol Appl (03):46–51
20. Wu YX, Dai J (2017) Research on 5G oriented edge computing platform and interface scheme. Des Tech Posts Telecommun 3:10–14
21. Jinan H, Zhejun Z, Zihua C et al (2018) Towards 5G from key technologies to network deployment. People's Posts and Telecommunications Publishing House, Beijing
22. Zheming Z (2016) The development trend of VR under 5G technology. Mobile Communication Press

Xu Zhou was born in 1976. In 2005, he received his Ph.D from the University of Electronic Science and Technology of China. He is the research fellow in Computer Network Information Center, Chinese Academy of Sciences and the Director of Advanced Network & Technology Development. His research interests include future network architecture, 5G and edge computing. He has published more than 60 papers included in SCI/EI.

Composite Autonomous Anti-disturbance Control for Systems with Multiple Disturbances

Lei Guo and Yukai Zhu

Abstract Complex systems in various fields including robot, aerospace, advanced manufacturing, and others are affected by heterogeneous disturbances and uncertainties resulting from internal disturbances, external disturbances, and model errors. With the increasing of requirements for accuracy, stability, and reliability, enhancing the ability of autonomous anti-disturbance control in complex environment has become the theoretical bottleneck and engineering challenge in the field of automation and artificial intelligence. This paper firstly reviews the traditional control methods that can attenuate or compensate a single kind of disturbance or an "equivalent" disturbance. Then, the paper introduces the theory of "refined" quantification and analysis under the influence of multiple disturbances and uncertainties, and subsequently proposes a composite anti-disturbance control law with hierarchical architecture: Composite Hierarchical Anti-Disturbance Control (CHADC). CHADC includes the disturbance description, disturbance estimation, disturbance compensation in feedforward path and the controller with disturbance attenuation in feedback path. The observability, compensability and repressibility of disturbances are proposed for the external observability and controllability of systems with disturbances. For the attitude control system of spacecraft, the experiment and simulation equipments of whole-loop anti-disturbance control are developed, which includes attitude estimation, controller, and execution parts. This technology can accomplish the whole-loop quantification and analysis of heterogeneous disturbances, as well as simultaneous attenuation and rejection of the heterogeneous disturbances. Consequently, the autonomous anti-disturbance performances of systems can be improved significantly.

Keywords Systems with multiple disturbances · Robust control · Disturbance attenuation and rejection · Disturbance observer · Observability and controllability of systems with disturbances · Autonomous control · Scientific equipments

L. Guo (✉) · Y. Zhu
National Key Laboratory on Aircraft Control Technology, Beihang University, Beijing, China
e-mail: lguo@buaa.edu.cn

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_14

1 Research Background and Problem Requirements for Autonomous Anti-disturbance Control

Complex engineering systems including aerospace and advanced manufacturing generally have complicated dynamics such as nonlinearities, uncertainties, and stochastic characteristics. Meanwhile, the system performances are affected by the external environment disturbances (e.g., aerodynamic, temperature variation, and electromagnetic radiation) and the internal noises (e.g., sensor noises, actuator errors, structure vibrations, and frictions). In modern control theory, the system disturbances generally come from three factors (internal noises, external disturbances, and model errors) and possess multiple sources, types, and channels of physical characteristics. According to the mathematical descriptions of disturbances, the disturbances are pluralistic, which can be described by many types such as step signal, harmonic signal, stochastic variable (Gaussian and Non-Gaussian), uncertain norm-bounded variable, and output variable of neutrally stable system. The multiple disturbances widely exist in the information acquisition, processing, and feedback parts, which affects the stability, accuracy, and reliability of control systems significantly. The ability of control system in adapting to complicated task environment and addressing various disturbances as well as faults needs to be improved urgently. Hence, the autonomous anti-disturbance ability has become the theoretical bottleneck and engineering challenge in the fields of automation and artificial intelligence.

Anti-disturbance has been an eternal topic along with the development of control theory, and the research for anti-disturbance control theory has become a hot and challenging problem in modern control theory and artificial intelligence theory [1–6]. The control systems and environments are becoming more and more complicated, and the requirements for system accuracy, reliability, and rapidity are increasingly demanding. Therefore, disturbance attenuation, compensation, and rejection have become the research hotspot of the field of control in recent years [4–6].

1.1 Development History of Anti-disturbance Control Theory

For the disturbances with different mathematical descriptions, starting with the PID control, a series of anti-disturbance control methods have been proposed by scholars. According to the anti-disturbance capability, these control methods can be classified into two kinds: disturbance rejection (compensation) methods and disturbance attenuation methods.

Disturbance rejection methods are proposed based on the principle of invariance [7–9]. PID control has the ability to compensate step disturbances and still is the commonly used control method in industry process at present. Internal model control and output regulation theory developed from 1960s and 1970s can compensate neutrally stable disturbances. Developed from 1990s and proposed by Chinese scholar, the active disturbance rejection control (ADRC) [9–11] has been developed greatly in recent years. In the framework of ADRC, the internal and external disturbances are regarded as the "total disturbance" which is subsequently estimated and rejected with the aid of extended state observer. At present, ADRC has been successfully applied in many fields including missiles, chemical industries, and unmanned aerial vehicles.

Another popular disturbance compensation method—disturbance observer-based control (DOBC) appears in 1980 s. The basic idea of DOBC is to estimate the influence of external disturbances by using observer and then compensate this influence in the feedforward channel. DOBC has been widely applied in robotics, electromechanical systems, platform drive systems, and hard disk drive systems [12, 13]. In the early stage of DOBC, single-input single-output linear systems in the frequency domain are mainly studied. In [14], the DOBC method is studied for some nonlinear systems at a relatively earlier stage, and an important reduced-order structure of DOBC is proposed. However, the systems are required to have fixed relative orders, and the disturbances are limited to be harmonic and load signals. So far, the DOBC method has been extended to general theoretic control systems and various application fields (e.g., aerospace) [15–17]. The above-mentioned disturbance rejection or compensation methods can achieve less conservative control performances in the presence of disturbances. However, specific theoretical restrictions are needed for the system or disturbance models.

Another important anti-disturbance control technique is the disturbance attenuation method, which can reduce the influence of disturbance to output by controlling the gains of transfer channels. When the disturbance is described by the Gaussian noise, the performance optimization can be achieved by controlling the output variance in the framework of the Gaussian stochastic control theory (e.g., Kalman filter and minimum variance control) developed in the late 1960 s [1, 18]. However, the statistical characteristics of disturbances are needed to be known for the stochastic control. In order to overcome this restriction, robust control methods including H_∞ and H_2 control are proposed by scholars in the early 1980 s. In these disturbance attenuation methods, the disturbance influence to system performance is minimum in the norm meaning via the optimization control technique [19–21].

1.2 Existing Problems

(1) **Conservativeness exists for the control methods that only attenuate or compensate the single disturbance**. Although great progresses have been achieved for the traditional anti-disturbance control methods, the fragility and conservativeness still exist. For instant, the disturbance compensation control theories based on the principle of invariance are only suitable for specific systems and disturbances. The stochastic control theory and robust control theory study the disturbance attenuation problem of stochastic noises and norm-bounded disturbances, respectively. A typical problem is that these methods only consider the single equivalent disturbance and achieve the disturbance compensation or

attenuation by grasping the “principal contradiction”. However, the “principal contradiction” may be hard to be determined sometimes. In other words, the “secondary contradiction” may also induce bad results [22, 23]. When the hardware condition is fixed, many disturbances cannot be compensated and attenuated by control algorithms alone, which is related to the compensability and repressibility of disturbances [24].

(2) **Multiple disturbances exist inevitably**. With the development of information acquisition, computer, and data technologies, the understanding for multiple disturbances (internal, external, and model) is ever-growing, and the ability of disturbance description and modeling is improved greatly. On the other hand, the requirements for the accuracy and reliability of control systems in practical engineering are increasing demanding. Hence, the autonomous anti-disturbance control theory for systems with multiple disturbances which contain unknown dynamics (e.g., uncertain dynamics, stochastic dynamics, and unmodeled dynamics), internal disturbances, and external disturbances has become a difficulty in international control science and artificial intelligence fields. It is of theory and engineering significance to extend from single-disturbance systems to multiple-disturbance ones and accomplish the goal of simultaneous disturbance rejection and attenuation [23].

(3) **Multiple disturbances are involved for whole-loop system design**. For a century, great achievements have been obtained for modern control theory. However, suitable tools for solving practical engineering problems in complex disturbance environment are still lacking. The reason is that the control algorithms, however accurate they may be, could no longer be applicable in the presence of sensor noises or actuator deviations because the control systems contain the whole control loop including information acquisition (sensing), processing (modeling and control), and execution. The existence of heterogeneous disturbances is inevitable due to the consideration of whole control loop. It is essential to extend the control algorithms design to control system design (sensing and actuation, environment and task) from the viewpoint of control discipline development.

Disturbance for control system is like disease for human body. Anti-disturbance control can be regarded as a signal processing procedure including “disease prevention, disease diagnosis, disease remedy, and disease recuperation”. The “disease diagnosis” can be seen as the refined description, modeling, and understanding of disturbances, which is vitally important. “Knowing one’s self and knowing the enemy” requires that not only the control model, sensor together with actuator, and environment together with task, but also the multiple disturbances should be understood as well as possible. Here, the disturbance estimation, observation, and identification are the core problems. The process of “disease remedy” requires the ability of “finding a right antidote” based on the specific characteristics of disturbances. For the complicated control systems subject to heterogeneous disturbances, it is essential to combine the disturbance estimation, feedforward compensation, and feedback attenuation, such that the multiple disturbances can be “crushed one by one”.

Supported by the projects (e.g., National Natural Science Foundation of China), our team is devoted to the research of anti-disturbance control theory for systems with multiple disturbances in the recent decade. The framework of refined anti-disturbance control with "knowing one's self and knowing the enemy, finding a right antidote, and crushing one by one" is proposed. A series of pioneering and innovative research results have been achieved. Now, we will introduce them from the aspects of theory study and engineering application, respectively [22–34].

2 Refined Anti-disturbance Control Theory and Methods for Systems with Multiple Disturbances

The unknown disturbances and uncertainties of complex systems come from different sources including model unknown dynamics, internal noises, and external disturbances. According to their stochastic and dynamic characteristics, different mathematical descriptions including stochastic (Gaussian and Non-Gaussian) signal, step signal, harmonic signal, derivative-bounded signal, neutrally uncertain variable, norm-bounded variable, and neutrally stable exo-system are adopted. The traditional control methods only consider the single disturbance from the aspect of control systems and consider either the compensation or the attenuation purpose in the control performances. Our team proposes the framework of refined anti-disturbance control which includes the quantification, description, and modeling of multiple disturbances, the estimation and learning method of unknown disturbances, and the external performance analysis methods (e.g., the robustness, sensitivity, observability, and anti-disturbance controllability for multiple disturbances). The time-domain hierarchical framework of disturbance observation, feedforward compensation, and decoupled stabilization of systems with multiple disturbances is established. Moreover, the CHADC and estimation theory are established for systems with multiple disturbances [22–24]. The detailed research achievements are shown as follows.

2.1 Unknown Disturbance Estimation and Performance Analysis Methods for Systems with Multiple Disturbances

In the traditional anti-disturbance control methods, may it be the disturbance compensation or the disturbance attenuation approach, multiple disturbances are lumped as one "equivalent disturbance". On one hand, important functions and great achievements have been obtained in practical engineering by only considering the "principal contradiction". On the other hand, sometimes this thought is a kind of method that "trying to attend to big and small matters all at once". It should be noted that the "secondary contradiction" may also induce bad results. For the complicated systems that are “full of diseases”, the research for multiple disturbances is essential. In practical engineering, some disturbances may be obtained by sensor measurement directly, but most of the disturbances, errors, and uncertainties can only be obtained through the dynamic estimation and calculation from the measurement information. Moreover, mechanism analyses for disturbance influence and transfer are made possible as a result of the effective descriptions for multiple disturbances. This is the reason why the control theories for systems with multiple disturbances that contain feedback loops are more advanced than uncertainty quantification theories.

The concepts including disturbance estimation, unknown input estimation, uncertainty estimation, and disturbance observation are all processes to estimate the disturbance dynamics based on the measurement dynamics. The traditional disturbance estimation methods rely on the extended state observer, adaptive estimation, etc. However, the disturbance estimation errors cannot be quantified in some of the above-mentioned estimation methods. As a result, the disturbance estimation accuracy cannot be ensured. What is more, some methods require certain parameters such as the frequency of disturbance.

To solve the above-mentioned problems, our team proposes the time-varying disturbance estimation methods which can ensure the performances of disturbance estimation errors. The adaptive, self-learning, and robust disturbance estimation methods including robust disturbance observer [25], fuzzy disturbance observer [26], neural network disturbance observer [27], and adaptive variable structure disturbance observer [28] are proposed for unmatched, uncertain, and unknown disturbances. The framework of modeling, description, and analysis for the systems with multiple disturbances is established. Furthermore, the stability analysis and the convex optimization methods are proposed based on the novel generalized potential functions with the adaptive parameter bases. Consequently, the conservativeness of the traditional robust analysis and control methods has been reduced, the system adaptation ability to uncertain parameters has been improved, the multi-objective optimization based compatibility design and separation control problems have been solved under the algebraic and differential constraints, and the multi-objective optimization based learning and estimation methods under different types of disturbances are obtained. These disturbance estimations methods will be effective tools for signal recognition in future.

2.2 Composite Hierarchical Anti-disturbance Control and Filter for Systems with Multiple Disturbances

The traditional disturbance compensation methods are based on the principle of invariance, and many restrictions of systems and disturbances essentially exist in these methods. Hence, the system observability and controllability should be analyzed by fully considering the system input-output structure, the actuator amplitude constraint, the input constraint of control channel, and the under-actuation [29].

For the disturbance attenuation methods, the conservativeness and controller high-gain phenomenon are easily induced due to the insufficient utilization of disturbance characteristics.

The ability to understand disturbances has been improved greatly due to the development of information technology. Up to now, how to propose the control methods that can simultaneously compensate and attenuate the disturbances according to the characteristics of systems and disturbances is still an open problem. The traditional methods that are only limited to feedback control are no longer applicable due to the channel restrictions in the control methods with multiple disturbances.

In our team, by combining the disturbance estimation, feedforward compensation, and feedback attenuation for the first time, the CHADC and estimation methods with "two layers and three parts" are proposed, and a practical theory framework of refined anti-disturbance control for systems with multiple disturbances is proposed. In our methods, the challenge that is the simutaneous presence of uncertain dynamics and unmatched channel in the system and disturbance models has been overcome, and the disturbance models have been extended to many types including the harmonic signals with unknown frequencies, the derivative-bounded variables, and the output variables of neutrally stable exo-systems. A theory framework of "DOBC + X" has been formed in the CHADC framework, which includes DOB + PID control, DOB + H_∞ control, DOB + adaptive control, DOB + variable structure control, etc. For instant, in [30], for a nonlinear system, the disturbances are divided into two types, including the signal with exo-system model and the signal with unknown parameter function, then a composite DOBC and adaptive control method is proposed. For the inertia variation and the center-of-mass deviation of the space manipulator during unknown target capturing, the actuator error is considered simultaneously, a neutrally uncertain dynamic model with less conservativeness is established in [31], and a novel composite DOBC + H_∞ anti-disturbance control scheme is proposed, where the disturbance compensation and attenuation difficulty in the presence of neutrally uncertainty is overcome. For the uncertain nonlinear systems with sensor fault, actuator fault, and norm-bounded disturbance, a series of anti-disturbance fault detection methods are proposed, where the sensitivity of fault detection has been improved and the false alarm rate has been reduced [32].

In order to further enhance the autonomous anti-disturbance control ability of systems in the complex environment with multiple disturbances, by combing two commonly used disturbance rejection methods (DOBC and ADRC) for the first time, an enhanced anti-disturbance control (EADC) is proposed in the framework of CHADC recently, where the DOBC and the ADRC are in the feedforward and feedback loops, respectively [24, 33]. In the EADC method, the DOBC is used to estimate and compensate the disturbances with partially known information, while the ADRC is used to address the other disturbances with uncertain dynamics. Compared to the single DOBC or ADRC, better anti-disturbance capability can be achieved by the EADC. Meanwhile, the EADC can reduce to the traditional DOBC or ADRC conveniently according to the disturbance characteristics. An example of attitude control system of flexible spacecraft is given in [33] to show the application of EADC, where the separation, description, and decoupled estimation of the composite disturbances including flexible vibration, inertia uncertainty, and environment disturbance in the input channel of attitude control system have been solved. Furthermore, the EADC has been further extended to the systems with unmatched disturbances.

Most of the traditional stochastic control methods are only applicable for the Gaussian stochastic systems. To overcome this restriction, the purpose of non-Gaussian stochastic control system is to control the probability density function of system output, which is an infinite dimensional distributed parameter system with integral and positive constraints. Our team has overcome the difficulty in transforming the infinite dimensional distributed parameter systems to the finite dimensional systems with ellipsoidal constraints. The non-Gaussian filtering, control, fault detection and isolation model is established based on the statistical information of non-Gaussian variables and functional distance optimization. The adaptive self-learning filters and control methods of stochastic distribution systems are proposed based on the convex optimization design. Moreover, the general entropy optimization criterions of multi-dimensional and multi-channel residual systems and the iterative calculation methods of output entropy and higher moment are proposed, where the attenuation of non-Gaussian disturbance and the fault detection have been achieved simultaneously [34].

Compared to the traditional anti-disturbance control methods, the composite autonomous anti-disturbance control has the following characteristics. ① The systems with multiple disturbances are considered. The research scope of anti-disturbance control theory has been extended to the systems with multiple disturbances from the systems with single disturbance. ② Tailorability: the system anti-disturbance capability has been enhanced significantly by combing DOBC with other control methods such as PID control, robust control, variable structure control, adaptive control, ADRC, etc. ③ Fineness: the disturbance information and the other dynamic characteristics of uncertainties have been fully utilized in the controller design. As a result, different types of disturbances can be rejected and attenuated simultaneously, and the refined disturbance compensation of uncertain systems can be accomplished. Moreover, the attenuation ability of stochastic disturbances has been improved essentially by utilizing the non-Gaussian characteristics of stochastic signals. ④ Non-fragility: the proposed methods are applicable to the control inputs, plants, and disturbance models with multiple uncertainties.

A practical research direction of anti-disturbance control has been formed by these theoretical achievements, the research scope of the traditional disturbance compensation and attenuation theory has been extended, and successful applications have been achieved in many fields such as robots, unmanned aerial vehicles, traffic, and energy systems. Now, we will primarily introduce the applications in the field of spacecraft attitude control.

3 Composite Autonomous Anti-disturbance Control System Technology of Spacecraft and Its Application

Characteristics such as strong coupling, high dynamics, and multiple constraints exist in spacecrafts. With the development of space technology and the urgent requirements of some major projects including high-resolution earth observation, deep space exploration, and precision strikes, the accuracy, autonomy, and reliability of the attitude determination and control systems on spacecrafts are faced with increasingly challenging demands. Meanwhile, the environments of the new generation of spacecrafts are becoming more and more complicated, and the uncertainties and unknown factors are increasing. The environmental disturbances, the performance degradation of the device components, and the significant increase of the failure and risk caused thereby will seriously restrict the accuracy and reliability of the attitude determination and control systems. Take the Mars exploration mission for example. In addition to the external environment influences (e.g., the long-term high and low temperature, electromagnetic disturbance, solar radiation, high-energy particle radiation, and solar wind), the uncertainties of Mars atmospheric environment, the failure of components, and the lack of measurement and control techniques are present. The technology of spacecraft anti-disturbance attitude estimation and control under multiple composite disturbances and uncertainties has become a bottleneck technology which needs to be tackled.

3.1 Composite Disturbances Filter and Attitude Determination Method of Spacecraft

The attitude measurement system of spacecraft is a multi-sensor information fusion system. The difficulties include three aspects. That is, the disturbances have non-Gaussian stochastic characteristics, the multi-sources and multi-types of disturbances exist, and the disturbances possess composite characteristics. In our team, the refined composite disturbance filter and attitude determination technology is proposed, by which the accuracy, autonomy, and reliability of the attitude determination systems can be improved significantly in the presence of multiple composite disturbances.

In most of the existing attitude estimation methods of spacecrafts, all the disturbances are assumed to take a single form (Gaussian white noise). However, the disturbances with multiple sources and multiple types are present in practical space tasks. The complex non-Gaussian and uncertain characteristics exist in these disturbances, which can be described by the unknown neutral disturbance, norm-bounded disturbance, and stochastic noise. The model of multiple disturbances of gravity gradient measuring satellite is given in Fig. 1.

For the combined attitude determination system with multiple disturbances including Gaussian stochastic variable, uncertain norm-bounded variable, and derivative-bounded variable, the composite anti-disturbance estimation method is

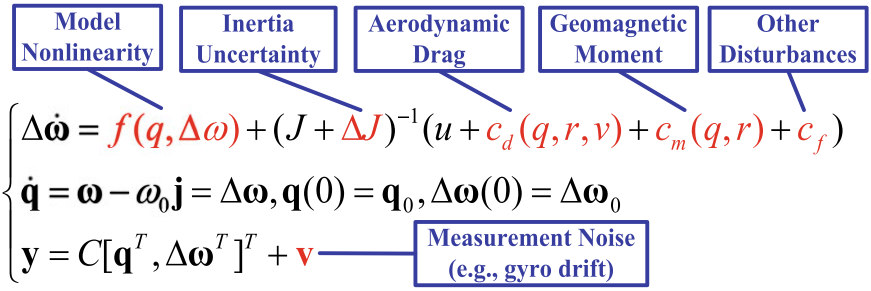


Fig. 1 The model of multiple disturbances of gravity gradient measuring satellite

proposed based on the disturbance observer (DOB) and Kalman filter. The DOB is used to estimate and compensate the harmonic and non-neutral disturbance (e.g., Markov process) in real time, and the Kalman filter is used to attenuate the Gaussian stochastic noise. As a result, the accuracy of attitude determination has been improved greatly. For the multi-sensor information fusion system with the unknown modeled disturbance, the Gaussian noise, and the norm-bounded disturbance, the composite anti-disturbance estimation method is proposed by combing the DOB with the H_∞ (H_∞/H_2) optimization methods [23]. The block diagram of the composite disturbance filtering method based on the DOB + Kalman is given in Fig. 2.

Furthermore, the software and instruments of anti-disturbance test and analysis are developed for the satellite attitude estimation systems. Compared to the latest methods of GOCE satellite of European Space Agency, the attitude estimation accuracy has been improved greatly.

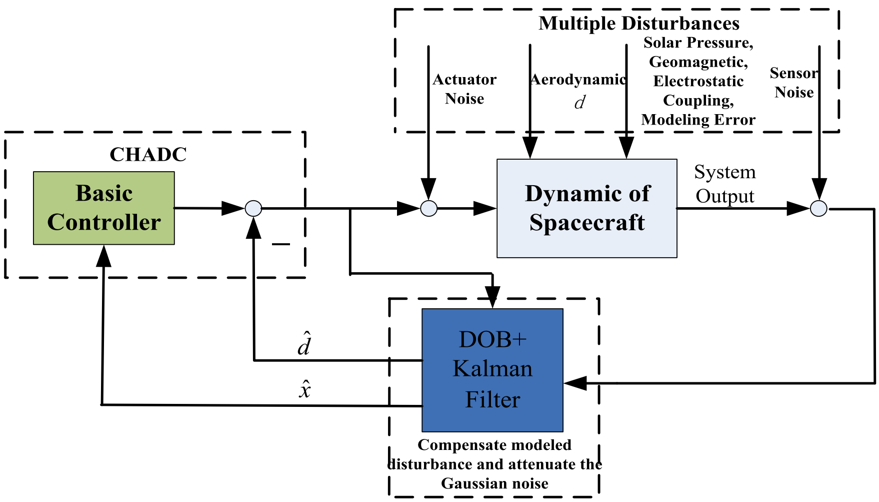


Fig. 2 The block diagram of composite filter method based on DOB + Kalman

3.2 Composite Hierarchical Anti-disturbance Attitude Control Technology of Spacecraft

The attitude control system of spacecraft is subject to the influence of many factors including the external environment disturbances, the actuator errors, the structure vibrations, and the unmodeled dynamics. The structure of the closed-loop control system is becoming complicated due to the coupling of different disturbance attenuator and compensators. The stability and disturbance attenuation performance cannot be guaranteed by the traditional single controller in the uncertain situation, which is a technical bottleneck that needs to be solved. Moreover, in the traditional fault detection and fault-tolerant control methods, only the systems with faults are considered. However, the persistent disturbances in the system tend to contaminate the properties of residual signals, reducing the sensitivity of fault detection and leading to the increase of false alarm rate.

(1) **Composite Hierarchical Anti-Disturbance Attitude Controller Design with Disturbance Estimation and Compensation Channels**
In the spacecraft dynamics systems, the models of multiple disturbances have the unique characteristics of unmatched channel, nonlinearity and multiple constraints. In our team, by using the non-smooth estimation and compensation in the inner loop and the predictive feedback in the outer loop, the hierarchical structure is proposed. The composite hierarchical anti-disturbance attitude control method of spacecraft is proposed in the presence of multiple constraints and unmatched channels, where the effective compensation of uncertain disturbances has been achieved. Furthermore, for the rapid attitude maneuver of spacecraft, a composite anti-disturbance nonlinear predictive guidance method is proposed, where the rapidity of spacecraft response to attitude control commands and the autonomy of spacecraft attitude control system have been improved. For the control requirements of deep space detectors, multiple disturbances have been attenuated and compensated effectively in the presence of input and measurement time-delays, the system accuracy and reliability have been improved.

(2) **Anti-Disturbance Fault Detection and Self-Learning Fault-Tolerant Control Method of Spacecraft**
For the attitude control system subject to external disturbance torque, structure vibration, and actuator fault, a kind of fault detection and diagnosis methods are proposed based on the disturbance compensation of additive and multiplicative composite disturbances, where the conservativeness of fault detection has been reduced. A kind of design methods of self-learning rapid attitude reconstruction and fault-tolerant controllers are proposed, where the online real-time learning and autonomous compensation of faults can be achieved. As a result, the shortcomings of passive strategies (e.g., look-up table) can be overcome. The fault diagnosis and fault-tolerant control model of nonlinear system with non-Gaussian disturbance, norm-bounded disturbance, and time-varying fault is established, and the self-learning fault-tolerant control method which can

estimate the states, detect the fault, and attenuate various types of disturbances is proposed.

On the basis of the innovative theories, our team developed an integrated testing and analysis platform of anti-disturbance attitude control based on the real-time simulator, the disturbance simulator, the attitude control module, and the real reaction wheel. The "disturbances–methods–performances" quantitative analysis of multiple disturbances and uncertainties of aircrafts, the anti-disturbance control method design, and the system performance evaluation can be accomplished. Specifically, the influence mechanism description and the cause-and-effect analysis of multiple disturbances in extreme and subtle conditions can be achieved. This test platform can provide technical support for the development of a series of tasks and equipment projects.

4 Summary and Prospect

From the viewpoint of anti-disturbance control, the understanding of disturbances may be the primary tasks, which can be regarded as a sensing process without sensors. On this basis, different anti-disturbance methods have different advantages. However, with the development of data technology and the enrichment of knowledge on the disturbances, the control methods which can "find a right antidote" according to the disturbance characteristics have better control performances. Compensating the disturbances with partially known information and attenuating the disturbances that cannot be compensated are the main characteristics of CHADC theory which contains disturbance estimation, feedforward compensation, and feedback attenuation. It needs to be pointed out that no matter how advanced the methods are, they always rely on the constraints of overall system and hardware conditions. Not all disturbances can be addressed to a satisfactory performance by algorithms alone. For different disturbances, it is a research priority to analyze the disturbance observability and anti-disturbance controllability (e.g., compensability and repressibility of disturbances) in the presence of system constraints. On this basis, it is an important research direction to redesign the system parameters and task purposes and to conduct the reconstruction and optimization works.

The following steps are needed for the design of control systems with CHADC. Starting from the control tasks and environments, the descriptions and representations of the systems and disturbances are given. On this basis, the mechanisms of disturbance influence and disturbance transfer are analyzed, which can be seen as a process of uncertainty quantization. Furthermore, the disturbance observability and controllability are analyzed, and the disturbance estimation methods are developed. The third step is to "find a right antidote" and "crush one by one", and the anti-disturbance control methods with composite structures which can compensate and attenuate the disturbances simultaneously are proposed. The first and second steps above can be regarded as parts of "disease prevention" and "disease diagnosis". The third step is a process of "disease remedy". In the fourth step, the task reconstruction and system redesign are considered in the presence of disturbances and faults, which can be seen as a process of “disease recuperation”.

In the presence of complex disturbance environment, the following problems need to be further considered for the composite autonomous control technology in the future. ① From the control point of view, the analysis of observability, compensability, and repressibility of multiple disturbances should be emphasized. The studies of disturbance and fault should be integrated. The optimization of the “disease prevention, disease diagnosis, disease remedy, and disease recuperation” process should be achieved, and the research of control system design, rather than control method research, should be considered. ② From the system point of view, the whole loop optimization of information acquisition, fusion, control, and execution, the disturbance estimation, attenuation, and compensation methods based on small loops, and the transition from algorithms to software, chips, systems, and even products should be considered. ③ From the simulation point of view, the refined description and modeling methods of multiple disturbances and faults based on the multidisciplinary intersection, the closed-loop causal analysis, traceability analysis, refined qualification, and the refined performance evaluation under extreme and subtle modes should be accomplished. ④ From the intelligence point of view, the active and passive immunity and adaptation ability of control algorithms in the presence of various of disturbances should be improved, and the combination between anti-disturbance algorithms (e.g., learning, predictive) and anti-disturbance mechanisms should be achieved. Moreover, the extension from objective disturbances to active confrontations should be considered.

Acknowledgements This paper is supported by the National Natural Science Foundation of China “the integrated test and analysis platform of anti-disturbance attitude control” (grant 61127007) and “spatial motion information bionic intelligence sensing flight instruments” (grant 61627810).

References

1. Guo L, Cheng DZ (2005) Introduction to control theory: from basic concepts to research frontiers [M]. Science Press, Beijing (in Chinese)
2. Wu HX, Hu J, Xie YC (2009) Characteristic model-based intelligent adaptive control [M]. China Science and Technology Press, Beijing (in Chinese)
3. Chai TY (2016) Industrial process control systems: research status and development direction [J]. SCIENTIA SINICA Inf 46(8):1003–1015 (in Chinese)
4. Bao WM (2013) Present situation and development tendency of aerospace control techniques [J]. ACTA AUTOMATICA SINICA 39(6):697–702 (in Chinese)
5. Feng CB, Zhang KJ (2004) Robust control of nonlinear systems [M]. Science Press, Beijing (in Chinese)
6. Guo L, Fang JC (2017) Recent prospects on some problems of navigation guidance and sensing technology [J]. SCIENTIA SINICA Info 47(9):1198–1208 (in Chinese)
7. Davison EJ (1976) The robust control of a servo mechanism problem for linear time-invariable multivariable systems [J]. IEEE Trans Autom Control 21(1):25–34
8. Byrnes CI, Francesco DP, Isidori A (1997) Output regulation of uncertain nonlinear systems [M]. Springer, Berlin

9. Han JQ (2009) From PID to active disturbance rejection control [J]. IEEE Trans Ind Electron 56(3):900–906
10. Gao ZQ (2014) On the centrality of disturbance rejection in automatic control [J]. ISA Trans 53:850–857
11. Huang Y, Xue WC (2012) ADRC: thought, application and theory analysis [J]. J Syst Sci Math Sci 32(10):1287–1307 (in Chinese)
12. Nakao M, Ohnishi K, Miyachi K (1987) A robust decentralized joint control based on interference estimation [C]. In: IEEE International conference on robotics and automation, pp 326–331
13. Kim KH, Baik IC, Moon GW et al (1999) A current control for a permanent magnet synchronous motor with a simple disturbance estimation scheme [J]. IEEE Trans Control Syst Technol 7(5):630–633
14. Chen WH, Ballance DJ, Gawthrop PJ et al (2000) A nonlinear disturbance observer for robotic manipulators [J]. IEEE Trans Indus Electron 47(4):932–938
15. Guo L, Feng CB, Chen WH (2006) A survey of disturbance observer-based control for dynamic nonlinear system [J]. Dyn Contin Discrete Impuls Syst Series B: Appl Algorithm 13E:79–84
16. Chen WH, Yang J, Guo L et al (2016) Disturbance-observer-based control and related methods-an overview [J]. IEEE Trans Ind Electron 63(2):1083–1095
17. Li SH, Yang J, Chen WH et al (2014) Disturbance observer-based control: methods and applications [M]. CRC Press, Boca Raton
18. Astrom KJ, Wittenmark B (1989) Adaptive control [M]. Addison-Wesley, MA
19. Zhou KM, Doyle JC (1998) Essentials of robust control [M]. Prentice Hall, New Jersey
20. Shen TL (1996) H_∞ control theory and application [M]. Tsinghua University Press, Beijing (in Chinese)
21. Huang L (2003) Fundamental theory on stability and robustness[M]. Science Press, Beijing (in Chinese)
22. Guo L, Chen WH (2005) Disturbance attenuation and rejection for systems with nonlinearity via DOBC approach [J]. Int J Robust Nonlinear Control 15(3):109–125
23. Guo L, Cao SY (2013) Anti-disturbance control for systems with multiple disturbances [M]. CRC Press, Boca Raton
24. Guo L, Zhu YK, Li WS (2018) An enhanced anti-disturbance control approach for systems subject to multiple disturbances [C]. In: 2018 Chinese automation congress, Xi'an, pp 2785–2790
25. Wei XJ, Guo L (2010) Composite disturbance-observer-based control and $H\infty$ control for complex continuous models [J]. Int J Robust Nonlinear Control 20(1):106–118
26. Wu HN, Liu ZY, Guo L (2014) Robust L_∞-gain fuzzy disturbance observer-based control design with adaptive bounding for a hypersonic vehicle [J]. IEEE Trans Fuzzy Syst 22(6):1401–1402
27. Sun HB, Guo L (2017) Neural network-based DOBC for a class of nonlinear systems with unmatched disturbances [J]. IEEE Trans Neural Netw Learn Syst 28(2):482–489
28. Zhu YK, Qiao JZ, Guo L (2019) Adaptive sliding mode disturbance observer-based composite control with prescribed performance of space manipulators [J]. IEEE Trans Ind Electron 66(3):1973–1983
29. Guo L, Wang CH, Wang Y (2018) Preliminary study on causality and cause-effect control [J]. Control Decis 33(5):835–840 (in Chinese)
30. Guo L, Wen XY (2011) Hierarchical anti-disturbance adaptive control for nonlinear systems with composite disturbances and applications to missile systems [J]. Trans Inst Meas Control 33(8):942–956
31. Zhu YK, Qiao JZ, Zhang YM et al (2019) High-precision trajectory tracking control for space manipulator with neutral uncertainty and deadzone nonlinearity [J]. IEEE Trans Control Syst Technol 27(5):2254–2262
32. Cao SY, Guo L, Wen XY (2011) Fault tolerant control with disturbance rejection and attenuation performance for systems with multiple disturbances [J]. Asian J Control 13(6):1056–1064

33. Zhu YK, Guo L, Qiao JZ et al (2019) An enhanced anti-disturbance attitude control law for flexible spacecrafts subject to multiple disturbances [J]. Control Eng Pract 84:274–283
34. Guo L, Wang H (2010) Stochastic distribution control system design: a convex optimization approach [M]. Springer, London

Lei Guo was born in 1966. He established and developed the composite hierarchical anti-disturbance control theory and non-Gaussian stochastic control theory. In recent five years, he hosted and participated more than 10 national key projects. He had more than 150 SCI papers, 2 English monographs, and 50 national patents. He obtained the State Technological Invention Award of China in 2018, the State Natural Science Award of China in 2013, and the National Defense Technological Invention Award of China in 2017 (all ranked first). His research interests are the anti-disturbance control theory and application, stochastic control theory, and high-accuracy navigation and control system technology.

Evolution Towards Quantum Metrology and the Action Agenda for Developing IT-Enabled Metrology

Xiang Fang

Abstract Quantum metrology and informatization of metrology are two major directions for metrology innovation. A new generation of information technologies, represented by the Internet, big data and artificial intelligence, are booming and has brought a far-reaching impact on scientific research around the world. In 2018, the General Conference on Weights and Measures (CGPM) at its 26th meeting adopted a Resolution on the revision of the International System of Units (SI) to redefine all SI base units in terms of physical constants. This revision terminated the history of using artefacts as measurement standards, and ushered in a new era of "modern metrology" where measurements are largely based on quantum technologies. China has played a vigorous part in the global research cooperation enabling the SI revision. Furthermore, China has proposed to establish a quantum-based national advanced measurement system where quantum measurement standards and a "flattened" traceability chain will be the core features. This action will help incorporate quantum metrology into informatization programmes of China, accelerating the integration between quantum technology and information technology.

Keywords International system of unit (SI) · quantum metrology · Informatization · National advanced measurement system

- Quantum metrology and IT-enabled metrology are two developing trend of measurement science. The driving forces behind are the fast advances of emerging information technologies, such as big data and artificial intelligence, which have significantly shaped the science innovation around the world. In fact, metrology always evolves with science and technology, taking the advantages of new science and technology to pursueever-higher measurement accuracy and ever-broader measurement range. In November 2018, the CGPM at its 26th meeting adopted the Resolution to revise the SI from 20 May 2019. This decision marks the arrival

X. Fang (✉)
National Institute of Metrology (NIM), Beijing, China
e-mail: fangxiang@nim.ac.cn

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_15

of a quantum metrology era, opening the opportunity for the use of new technologies, especially quantum technologies, to develop measurement solutions in a long future. It provides China with a historic opportunity to establish a quantum-based national advanced measurement system, which could underpin the whole science innovation in China.

1 Evolution of Metrology Over the Past 150 years

Metrology is the science of measurement and its application. The objective of metrology is to ensure accuracy, stability, comparability, traceability and consistency of measurements. It involves almost all aspects of human society from science, legislation to industry, and supports more than 60% of global economy.

The Metre Convention signed in 1875 established an international metrology framework that later established the SI, an internationally recognized system of units with seven base units at the core: the unit of time, length, mass, electric current, thermodynamic temperature, amount of substance, and luminous intensity. The treaty also created the International Bureau of Weights and Measures (BIPM) as an international organization to conserve international prototypes. The objective of world-wide uniformity of measurement is then achieved through the provision of a framework for metrologically "traceable" measurements, which contains the definition of internationally accepted units of measurement, the realization of units of measurement by measurement standards, and the establishment of a metrological traceability chain that transfers quantity values of measurement standards to measuring instruments.

Quantum metrology has developed for about half of century since 1960s, driven by the development of quantum theory and quantum technology. In 1960s, the unit of time, the second, was the first to be redefined on the frequency of the ground state hyperfine transition in the caesium 133 atom, which has significantly improved the measurement accuracy of time frequency. Then in the 1980s the meter, the unit of length, was redefined based on the fixed numerical value of the speed of light. In 2005, the International Committee for Weights and Measures (CIPM) unanimously agreed on the preparation of new definition of the SI units. After decades of joint efforts by metrological scientists around the world, in November 2018, the General Conference on Weights and Measures (CGPM) at its 26th meeting adopted a Resolution to revise the SI, changing the world's definition of the kilogram, the ampere, the kelvin and the mole. From 20 May 2019, all SI units are defined in terms of constants that describe the natural world. This will assure the future stability of the SI and open the opportunity for the use of new technologies, including quantum technologies, to implement the definitions.

This revolution has two main characteristics: one is the quantization of SI units. Realization of SI units using quantum technologies has greatly improved the measurement accuracy and expand the measurement range. For example, defining the meter on speed of light and realizing this definition by laser waver length has improved the measurement accuracy of length by nearly 3,000 times to nanometer scale and the measurement range from planet scale to universe scale. The other characteristics is the "zero chain" of traceability. Quantum metrology is combined with information technology to enable shorter traceability chain, more rapid metrological calibration and more accurate measurement results. This will completely change the previous dissemination pattern based on physical artefacts and disseminate primary standard directly to measuring instruments, saving costs and improving product quality.

2 China's Efforts and Achievements in Developing Quantum Metrology

2.1 Initiatives of World Major Science Leaders

Most developed countries in the world have recognized that, the innovation towards quantum metrology, featured by quantum measurement standards and a much flattened traceability chain, perfectly fits the fourth industrial revolution that is based on Cyber-Physical Systems.

In U.S., the National Institute of Standards and Technology (NIST) has been giving high priority to quantum metrology in its investment plan over the past twenty years. After playing a leading role in the revision of the SI, NIST launched in 2017 a NIST-on-a-Chip programme to develop intrinsically accurate, quantum-based measurement technologies that can bring measurement services directly to the user.

In Europe, a European Metrology Research Programme (EMRP) was launched in 2007 to support six priority metrology areas including the fundamental SI. Now the programme was completed and succeeded by its second phase, the European Metrology Programme for Innovation and Research (EMPIR), where quantum metrology remains as a funding priority. The EMPIR is an integrated part of Horizon 2020, the EU Framework Programme for Research and Innovation.

2.2 Contributions of China to the SI Revision

Thanks to the wide recognition of quantum metrology importance by Chinese science community and authorizes, the planning on research and development activities has been made at state level. In the 13th five-year plan period from 2015 to 2020, a national research programme called NQI Programme was funded by the Chinese Ministry of Science and Technology to support, among others, new-general measurement standards based on quantum effects, advanced calibration technologies, and determination of some fundamental constants in relevance to the SI revision. This programme has enabled a series of research activities towards SI revision carried out at NIM:

1. NIM determined the Boltzmann constant by two primary thermometers, the cylindrical acoustic gas thermometer (c-AGT) with a measurement uncertainty of 2.0×10^{-6}, and the Johnson noise thermometer (JNT) with a measurement uncertainty of 2.7×10^{-6}. Results made by both of the two methods have contributed to the CODATA 2014 and 2017 adjustments of the constant, and the CODATA 2017 recommended value of the Boltzmann constant was used for the redefinition of the kelvin. Besides, NIM Vice Director Dr. Duan Yuning contributed to the top coordination of worldwide research efforts for the kelvin redefinition in his capacity as the president of the CIPM Consultative Committee for Thermometry (CCT).
2. NIM has proposed the unique Joule Balance solution to measure the Planck constant and to realize the new definition of the kilogram. The NIM-1 J Balance designed for determining the Planck constant was established in 2007, followed by the current NIM-2 J balance developed since 2013. In May of 2017, the relative uncertainty of NIM-2 is 2.4×10^{-7} ($k = 1$). A series of improvements are still in progress.
3. New definition of the mole is based on the Avogadro constant, which can be determined by measuring the number of atoms in 1 kg single-crystal spheres that are highly enriched with the ^{28}Si isotope. In this experiment, the molar mass of enriched silicon is one key quantity to be measured. NIM has used two independent methods for this measurement, including a self-developed NR-ICP-MS method. Result of the CCQM-P160 "Isotope ratios and molar mass of highly enriched silicon" have shown the excellent measurement results made by NIM for both Si isotopic abundances and molar mass, as the only one NMI using two different methods.
4. Research for the possible redefinition of the second is underway. Time is the most important base unit in the new SI. Metrological scientists around the world are working on a more accurate optical clock. Since 2006, NIM started the research on the next generation of time and frequency standard, the strontium optical lattice clock, with uncertainty of 2.3×10^{-16} on its first round evaluation in 2017. NIM measured the absolute frequency of strontium optical lattice clock with respect to cesium fountain clock NIM5 with uncertainty of 3.4×10^{-15} and the result of which was adopted by the Consultative Committee for Time and Frequency (CCTF) in calculating the recommended value of frequency of the new ^{87}Sr optical clock. Currently, NIM is working on its second strontium optical clock, with uncertainty expected to reach a few parts in 10^{-18}, to support the redefinition of the second that may take place in 2026.

3 An Outlook to Informatization of Metrology

Recognizing the role of quantum metrology in underpinning science innovation and enhancing industrial competitiveness, many developed countries has turned their investment focus to quantum metrology. Under the framework of Horizon 2020, the Quantum Technologies Flagship of the European Union aims to place Europe at the forefront of the second quantum revolution and has identified four funding priority areas in its ramp-up phase: quantum communication, quantum simulation, quantum computing, quantum metrology and sensing. The United Kingdom government launched in 2013 a National Quantum Technologies Programme to accelerate the translation of quantum technologies into the marketplace. A Quantum Metrology Institute at the National Physical Laboratory (NPL) will form part of the programme. Similar programmes also includes the above-mentioned NIST-on-a-Chip programme where chip-loaded and quantum based measurement standards can be deployed nearly anywhere and anytime. China has made exciting achievements in the search for SI revision, but still lags behind developed countries in the field of quantum metrology due to shortage of strategic planning, scientists, research facilities, and interactions with industries.

3.1 Quantum Metrology Supports Management of Scientific Research Data

Robust supply of high quality data is a basic task of informatization. Quantum metrology will help build a comprehensive information network and improve data quality. It will support data aggregation, information integration and correlation, which can realize automatic management and quality control of scientific research data.

Based on the new SI, four base units, the kilogram, the ampere, the kelvin, and the mole, are defined by the Planck constant, the elementary charge, the Boltzmann constant, and the Avogadro constant, respectively. Further, the definitions of all seven base units of the SI are now uniformly expressed using the explicit-constant formulation. To most people, the SI didn't change, because a so-called "continuity conditions" were established before the transition to help ensure that there would be no change in magnitude of any of the SI base units, and hence no change in any units derived from the base units. In fact, the change in the numerical value of the voltage unit is only of one millionth, and the change in that of the electrical resistance unit is even smaller. The change to the SI will only affect top-level metrology institutes and calibration laboratories that require the smallest measurement uncertainty. For ordinary users, workers and most researchers, the new definitions will not affect them and their measurement results are still valid. You may take the SI revision for granted, but in fact it happened as a result of the persevering research and collaboration of measurement scientists around the world over decades, so that all constants used in the redefinition are accurately determined and the SI was kept stable when the revision came into force on 20 May 2019.

From another perspective, however, the change is "huge". Firstly, artefacts are replaced by invariable "constants" in nature, which will ensure the long-term stability of the SI. Secondly, the defining constants are not affected by space, time or human factors, so that the new definitions are very objective. Thirdly, new definitions can be realized for any range of measurement, enabling traceable measurement of extremely small or large quantities. Fourthly, a new definition can be realized by more than one method, which opens the opportunity for the use of potential technologies to implement the definitions.

3.2 Quantum Metrology Supports Sensor and Intelligent Instrument Industry

The sensor and intelligent instrument industry is an underpinning sector of economy, and is in the forefront of promoting IT and industry integration. The accuracy of measuring instruments is ensured thought traceability to highly accurate measurement standards. Quantum mechanics and Micro-Optical-Electro-Mechanical System (MOEMS) have provided opportunities to develop quantum-based measurement standards and sensors for multi physical quantities and multi parameters, providing high accuracy and in-time self-calibration to intelligent instruments in their life cycle. For example, chip-level quantum measurement standards using effects of atoms, electrons and photons are able to realize the calibration of magnetic field, acceleration, gravity, inertia, pressure, temperature, time and other physical quantities, and is less affected by noise. Now quantum magnetic sensors and Hall effect sensors have been commercially available, quantum sensing technology has been applied to material research such as the coefficient test of stainless steel lattice fracture magnet. It will help with the technical upgrading of our high-end instrumentation industry.

3.3 Quantum Metrology Supports Other Key Technologies

Quantum metrology may provide solutions to some bottleneck problems in developing key technologies. Quantum metrology will shorten the traceability chain, enabling faster, more accurate and embedded calibration. By putting chip-level quantum measurement standards into measuring instruments, high-accuracy measurement can be disseminated directly to measuring instruments, providing vital support to manufacturing and engineering, improving the product quality of many industries. This will be needed by industries manufacturing aeroengines, high-end digital machine tools, and nuclear power equipment, etc.

Defining the second based on atom properties has greatly improved the accuracy of time frequency measurement, which ensures the accuracy of satellite navigation and positioning systems including the Beidou Satellite Navigation System of China, making potential positioning precision even close to millimeters. Thanks to the time synchronization technology based on atomic clocks, the current time synchronization level of China's power grid operation has reached 1 μs, and the time synchronization between 5G communication base stations has reached 100 ns.

Temperature is a most difficult physical quantity to measure directly. The accuracy of temperature measurement depends on stability of electrical or mechanical properties of temperature sensing elements, which may restrict temperature measurement range and even cause failure of measurement in some extreme conditions. Redefinition of the kelvin based on the Boltzmann constant has fundamentally solved the defects of the above-mentioned temperature scale and problems in practical temperature measurements, so that in principle accurate measurements from extremely low temperatures to extremely high temperatures can be realized. In addition, the new temperature measurement method is based on quantum physical phenomena, so it can realize self-calibration.

Defining the kilogram in terms of fundamental physical constants ensures its long-term stability, and hence its reliability, as we do not need to worry about the loss or damage of the International Prototype of the Kilogram. Redefinition of the kilogram means that advances in science and technology can make mass measurements more accurate. Mass measurement plays a crucial role in a large variety of fields such as aerospace, intelligent transportation, high-speed railway, auto manufacturing, Biopharmaceutical, semiconductor industry, etc.

In the new definition of the mole, the amount of substance unit is expressed with the Avogadro constant. The mole no longer exists on the mass unit of "kilogram". The concept of the amount of substance is clearer and is easier to understand. Nevertheless, redefinition of the mole has brought a new challenge to metrology, because providing traceability to the mole will go beyond physical measurements and require accurate measurements in chemical fields, involving preparation and analysis of ultra-high purity materials, precise measurements of isotopic compositions and measurements of atomic weights of elements. Redefinition of the mole will strongly promotes the development of metrology for micro-particles, promoting the development of relevant theories and cultivate new disciplines, and enhance people's understanding of the micro world. It will also facilitate science innovation in other fields, e.g. measurements in microbiology and biochemistry fields can be traceable to the SI.

The rapid development of quantum physics has greatly enhanced people's understanding of light and the capability to use light. In recent years, exploration of entangled photons has provided new potentials of using quantum information technology. The rapid development of such technologies as quantum optics, single-photon detector, single-photon source, resolvable photon number detector and photon detection, has greatly promoted the development of metrology in photometry and radiometry, but on the other hand has brought new challenges to single-phone metrology. With the continuing development of photon detection technology, the quantum nature of electromagnetic radiation has led people to consider using counting photon numbers as a way to trace the optical radiation measurement standard to the SI. If the frequency of these photons is known, the number of photons obtained by photon counting can be known as the energy of spectral radiation.

3.4 Establish a Quantum Metrology Research Base in Shenzhen

The Shenzhen Special Economic Zone (SEZ) plays a strategic role in supporting the informatization of scientific research of China. It is also a test bed of the integration of quantum metrology and information technology. Shenzhen has been given a special status with preferential policies for developing 5G, artificial intelligence, life science, and biopharmaceutical, etc. It also stays at the heart of the Guangdong-Hongkong-Macao Greater Bay Area.

In order to explore a new research and collaboration mode for developing quantum metrology, NIM has proposed to establish an Institute of Technical Innovation (ITI) in Shenzhen, which has been strongly supported by the Shenzhen government. This initiative was endorsed by Shenzhen government its annual government work report delivered on January 18, 2019. It has been identified by Shenzhen as one priority task for R&D as to establish a national advanced measurement system featured by quantum standards and short traceability chain.

Building the ITI in Shenzhen will not only help NIM to take the opportunity of quantum metrology revolution to develop a national advanced measurement system. It will also add value to Shenzhen itself, enhancing Shen Zhen's strategic position as high-tech center of China and the heart city of the Greater Bay Area. ITI will both provide well-conditioned metrology laboratories and create a platform for open and inclusive collaborations that may lead to theoretical and technical innovation in quantum metrology. The Shenzhen Metrology Institute will also join its part bringing its special experience in marketing and industrial service.

The ITI is expected to support Shenzhen in its technical innovation and developing customer-oriented manufacturing industry. At present, Shenzhen is making great efforts in improving research facilities. Representative programmes include the Daya Bay Neutrino Laboratory, the China Environment for Network Innovations (CENI), the China National Gene Bank (CNGB), and the proposed Gravitational Wave Detection facility. All of these will need high-accuracy and high-stability measurement standards to assure measurement accuracy and data quality. Advanced measurement technology will also help Shenzhen itself to solve technical challenges arising from key industries. The ITI can build a supporting system comprising quantum measurement standards, quantum chips, advanced measuring instruments, and advanced measuring technology, and to create a full-chain eco-system for innovation that consists of basic research, applied technology, commercialization, and Techfin.

References

Xiang F (2019) To work faster on national advanced measurement system for a high-quality greater bay area economy. Shenzhen Special Zone Daily, 03 April 2019. edition A10

Fang F et al (2019) Time—from astronomical second to atomic second, metrologytechnology (5):07–10

Feng X et al (2019) Evolution of the kelvin, the SI unit of thermodynamic temperature, measurement technology, (5):52–54

Haiyong G et al (2019) Development and prospect of photonic measurement technology, metrology technology (5):64–67

Gao W et al (2019) International measurement system and post new SI era, measurement technology (5):72–80

Li Z et al (2019) Contribution of national institute of metrology, China to the redefinition of the kilogram, metrology technology (5):28–33

Liu H et al (2019) From candle light to the candela—the unit of luminous intensity, metrology (5):68–71

Ren T et al (2019) China's contribution to accurate determination of Avogadro constant, metrology technology (5):60–63

Shen P et al (2019) Evolution of the unit of electromagnetic measurement, measurement technology (5):36–42

State Administration for Market Regulation, International System of Units (SI) Fundamentally better [Z] (2019)

Sun S et al (2019) China's maintenance and contribution to the unit of length, 1 (19-si 07), measurement technology (5):14–17

Wang Q (2019) Optical clock—towards a new definition of the second, metrology technology (5):11–13

Zhang J et al (2019) Redefinition of the SI unit of thermodynamic temperature in China, measurement technology (5):43–46

Zhou K et al (2019) On determination of Boltzmann constant by Josephson noise thermometer, measurement technology (5):47–51

Mr. Xiang Fang is Member of the 13thNational Committee of Chinese People's Political Consultative Conference (CPPCC), Director of National Institute of Metrology (NIM), China, and a research fellow of NIM. Prior to his current position, he had served successively as Vice Director General of the former National Research Center for Certified Reference Materials (NRCCRM), Vice Director General of NIM, Vice Director General of China National Institute of Standardization (CNIS), Chief Engineer and then Vice Administrator of Standardization Administration of the People's Republic of China (SAC).

Mr. Xiang Fang has a long-term involvement in standardization, metrology and testing, and analytical instrument development and application.He has lead the implementation of a number of state-level research projects funded by national research programmes such as the National Key Technologies R&D Programme, the Scientific Instrument Research and Development Programme, and the Natural Science Foundation of China. With many unique innovations made to mass spectrometry and mass spectrometer development and use, he now holds many innovation patents and is the winner of twosecond prize of the National Science and Technology Progress Reward (as group leader) and a few first prizes of the provincial and ministerial science awards. Mr. Xiang Fang has published over 30 scientific papers. He has also cultivated a high-quality research team in mass spectrometry.

From 2007 to 2014 when he worked at SAC, he has overseen the standardization work at local levels and at service and high-tech industries in particular. Under his leadership, a series of new polices and mechanisms have been established, concerning the reform of Chinese standardization system, the interaction between science and standards, and the disposal of patents in standards, etc.

He has also participated in the development of a number of state-level strategic plans in scientific and standard fields, including the National Long-term Science and Technology Development Plan(2006-2020), the top science plan in China, the 11th and the 12th Five-year Plans for Science and Technology Development of China, the 12th Five-Year Plan for Technical Standards of China and the National Standardization System Construction and Development Plan (2016-2020).

Development and Services of Resource Discovery System of National Science and Technology Library

Yiqi Peng, Boer Wu, and Zhongqi Shen

Abstract In order to solve the problems of serious shortage of scientific and technical literature resources and extremely low document guarantee rate in China, the National Science and Technology Library (NSTL) was established in June 2000 by the Ministry of Science and Technology and several other departments, which aims to build a national scientific and technical literature guarantee system and provide public welfare and basic information services to the whole country. After nearly 20 years of development, NSTL has become the strategic guarantee base and a national sharing platform for scientific and technical literature resources. However, the problem of unbalanced and insufficient supply of scientific and technical literature is still severe in China, and there is an urgent demand for the next generation of knowledge service systems, such as big data and cloud computing based ones. To this end, NSTL is making every efforts to develop the national scientific and technical literature resource discovery system, seeking to play its role as a national core platform to provide informational strategic support for China's scientific and technological innovation and development. Its main achievements including promoting the establishment of scientific and technical literature's metadata resource integration and fusion system, strengthening semantic knowledge organization, building up a series of thematic portals and an international scientific and technological citation system based on the discovery system, and forming a big data discovery system of scientific and technical literature.

Keywords Scientific and technical literature · Information guarantee · Resource discovery system · Information sharing

Y. Peng (✉) · B. Wu · Z. Shen
National Science and Technology Library, Beijing, China
e-mail: pengyq@nstl.gov.cn

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_16

Scientific and technical literature is the carrier of scientific and technological information, achievements and knowledge recorded in the form of text, symbols, graphics, audio and video. It's an important manifestation of scientific research activities. At the same time, as the significant strategic resource of the country, the platform of scientific and technical literature is the vital foundation and supporting system for scientific and technological work and innovative development [1].

1 The Development of the National Scientific and Technical Literature's Information Guarantee System

1.1 Original Intention and Mission

In the middle and late 1990s, the purchase volume of China's scientific and technical literature, especially the external one, has been hovering at a low level, and the basic satisfaction rate of foreign scientific and technical literature is less than 1/3, which seriously restricts the scientific and technological innovation activities. In order to solve the problem of severe shortage of scientific and technical literature resources and extremely low document guarantee rate, the National Science and Technology Library(NSTL) was established in June 2000 [2].

NSTL's mission is to centrally reflect the will of the country, and according to the principle of "centralized purchasing, separate processing, synthetical online publishing, and resource sharing", to collect and develop scientific and technical literature information resources in the field of science, engineering, agriculture, medicine, etc. It also aims to make full use of the modern advanced technologies to develop the national scientific and technical literature information guarantee system, and provide document information services to the whole country [3]. Besides, NSTL is organized according to the four pillar fields of science, engineering, agriculture, and medicine. The participating institutions are National Science Library, Chinese Academy of Sciences, Institute of Scientific and Technical Information of China, China Machinery Industry Information Institute, China Metallurgical Information and Standardization Institute, China National Chemical Information Center, Library of the Chinese Academy of Agricultural Science, Library of the Chinese Academy of Medical Sciences and other literature and information organizations. NSTL is committed to become Chinese authoritative collection and service center of scientific and technical literature resources, a modern information technology demonstration and application center, and a hub of scientific and technical literature information resource system.

1.2 Organizational System

NSTL established a brand-new organizational operation and management mechanism at the beginning of its establishment. To be more concrete, as a virtual institute of scientific and technological information resources, NSTL implements the director responsible system under the leadership of the Council. The director is responsible to the council and in charge of the daily work, the council members are composed of representatives from six departments such as the Ministry of Science and Technology and the Ministry of Finance, representatives of the four major information institutions of science, technology, agriculture and medicine, and famous scientists.

1.3 Function and Influence

After nearly 20 years of development, NSTL has been built into the strategic support base for national scientific and technological information resources and the indispensable foundation of the national innovation system, which have played an important role in the guarantee of national scientific and technical literature information resources.

(1) The continuous and stable development of core resources has laid the foundation for China's guarantee of scientific and technical literature. The category of high-quality printed literature ordered by NSTL has remained stable at about 24,000, which is the largest in China. There are more than 140 online foreign language scientific and technical literature resource databases (covering more than 20,000 online foreign language periodicals of key disciplines) ordered by a variety of ways, including national license, joint purchasing and funding support. Besides, NSTL has continuously collected 6,000 open access (OA) literature resources, and provided 3,018 important foreign language scientific and technical periodicals retrospective services for public service based institutions in a state-authorized manner, which effectively makes up for the structural deficiency of foreign language literature in China, forming a resource guarantee situation of the collaborative development of printed literature and electronic literature, ordered literature and open access resources [4].

(2) The nationwide service system has improved the public welfare guarantee ability and narrowed the information gap between departments and regions. In addition to the main site, NSTL has also established 40 service sites across the country, forming a nationwide scientific and technical literature service system. Each year, the usage amount of NSTL's ordered foreign language electronic resources is more than 60 million. And more than 1.2 million printed-documents-based original text services are provided. The whole-process information tracking and thematic service targeted at the teams of national major special projects for science and technology, and the information support services

for the national major strategies, such as the ones for “The Belt and Road Initiative”, “Yangtze River Economic Zone”, and “Beijing-Tianjin-Hebei Coordinated Development”, have been well received by the project team and relevant government departments. Furthermore, NSTL also provides knowledge-based solutions to key enterprises, supporting a large number of enterprises’ information tracking of thematic topic, knowledge organization management and other important tasks. The thematic literature support activities for the western region where the document supply is inadequate, carried out by NSTL have bridged the information gap between the eastern and western regions of China and promoted the balance of scientific and technical literature services.

(3) NSTL has established a resource library for foreign language scientific and technical literature, and its collection is close to the international mainstream bibliographic index database’s counterparts. Specifically, making full use of the advantage of member organizations’ rich database development experience as professional information agencies, NSTL has established the foreign language scientific and technical papers and citations database, forming the largest foreign language abstracts and citations database in China, and this provides strong support for information search, literature discovery, and intelligence analysis.

(4) The strategic guarantee for scientific and technical literature based on printed copy has relieved university libraries’ worries about the comprehensive digital transformation. Adopting the resources development method of centering on printed copy, NSTL has realized a comprehensive collection of core scientific and technical literature abroad, which can be stored and used for a long time. For this reason, university libraries no longer need to consider the collection problem of foreign language literature, but only to order digital documents in bulk from the perspective of utilization. There are more than 6,000 categories of printed periodicals ordered by NSTL are unique in China. Meanwhile, a number of high-quality scientific and technical literature of small disciplines and minor languages are also subscribed to meet the literature needs of these niche scientific research groups.

(5) Leading the infrastructure construction of scientific and technological information to support the transformation from document services to big data mining services in this filed. The development of knowledge organization infrastructures require the accumulation of relevant data and technologies, and huge investment in human and material resources. Relying on member organizations’ decades of accumulation, NSTL has successively built a large number of infrastructure tools, such as “Chinese Digital Library Standard Specification”, “Chinese Thesaurus”, “English Super-Thesaurus”, “Science and Technology Field Classification List”, “Institutional Specification Documents”, “The Document Name Specification Table”. And these tools have been applied in member organizations, key libraries, the Project of National Digital Publishing, and related projects of the Chinese Academy of Engineering, which have improved the level of knowledge-based and semantic organization of the scientific and technical literature in the industry.

2 The Situation Faced by the National Scientific and Technical Literature's Information Guarantee Work

Facing the increasingly diversified knowledge needs, the rapid development of the Internet and information technology, and the progress of the open and integrated new media, the scientific and technological information guarantee environment is undergoing revolutionary changes, and new forms of knowledge creation, organization, dissemination and utilization are taking shape. Under the circumstances, the national scientific and technical literature information guarantee work are facing new challenges.

2.1 The Problem of Unbalanced and Insufficient Supply of Scientific and Technical Literature is Still Severe

With the continuous enhancement of China's scientific and technological strength, researchers have also put forward higher requirements for the scientific and technical literature information guarantee work. On the one hand, scientific and technological personnel's using behaviors and habits of literature resources have undergone fundamental changes, and the guarantee work of scientific and technical literature information needs to adapt to new scientific research paradigms and knowledge production processes. On the other hand, it needs to be based on in-depth processing and organization of massive literature data, which make new requirements and pose challenges to the supply methods and system tools of scientific and technological information services. In addition, the unbalance between demand and supply capacity of scientific and technical literature in different regions and systems requires the government's support and guarantee of targeted literature. Due to insufficient funds, some regions in China purchase less foreign language scientific and technical literature resources, resulting in an extremely imbalanced overall literature guarantee level. From a regional perspective, there is a large gap between economically developed regions' literature guarantee level and economically underdeveloped regions' counterparts. Although China's scientific and technical literature guarantee capabilities have steadily increased in recent years, the scientific and technological information needs of different regions, industries and innovators vary. The problem of unbalanced and insufficient supply of scientific and technical literature is still severe.

2.2 *The Opening and Integration Need to Innovate the Mode of Information Guarantee Services*

In the past 20 years, scientific and technical literature have completed the development process from printed publications, coexistence of printed and electronic resources, to pure electronic publishing, and even semantic publishing. New publishing models and publication forms have continued to emerge, making digital resources become the mainstream of publication and utilization of scientific and technical literature information. The printed copy centered resource development mode is changing to the printed copy based digital resource guarantee system [5]. The diversification of publishing, dissemination, display, and utilization brought by digital publishing not only improves the availability and convenience of scientific and technical literature resources, but also increases the limitation and use-cost of literature services. In addition, facing the development trends of open and fusion, such as digital publication of scientific and technical literature, open access, artificial intelligence processing, and big data analysis, the development of a national scientific and technical literature resources guarantee system needs to adapt to the demands of researchers for “one-stop” acquisition and “knowledge-based” discovery of scientific and technical literature in the context of big data; it also needs to build a new digital business management system that is oriented to digital resources and takes into account the needs of printed copy resource management [6].

2.3 *Big Data and Cloud Computing Have Generated the Demand for the Next Generation of Knowledge Service System*

With the advancement of big data, cloud computing, artificial intelligence and other technologies, the correlation of information content and the fine-grained organization of knowledge have become important trends. Also, the organization and disclosure of information resources are developing towards the fine-grained, structured, semantic, and correlative directions. The basic unit of information resources organization used to be “paper”, but now it deeps into the knowledge units such as sections, chapters, diagrams, formulas, citations, and thematic objects, and extend to knowledge objects such as authors, institutions, projects, data sets, tools, and topics. This contribute to form the knowledge network that can be excavate and extend, conduct in-depth knowledge services of users’ research and innovation process, expand resource integration and services based on the research process and knowledge life cycle, and fully support users to make scientific decisions in R&D and market exploitation [7]. Therefore, the design of the next-generation knowledge service system has become the primary task of the building of China’s scientific and technical literature platform.

3 Building of the National Scientific and Technical Literature's Resources Discovery Platform

Under the new situation, information guarantee work of scientific and technical literature in china is facing new challenges. In the context of growing development of the digital environment and increasing demand for innovation, to solve the unbalanced and insufficient problem existing between the demands and services of scientific and technical literature, and achieve the established goals of national guarantees, the development of the information guarantee system for national scientific and technical literature needs to continuously adjust its resource development and service goals according to changing information environment and user demands. In addition, on the basis of guaranteeing external printed scientific and technological information resources, it is necessary to strengthen the supply of digital scientific and technological information resources, deepen the integration, long-term preservation and knowledge-based organization of various resources, build a national scientific and technical literature resource discovery platform, and develop a big data discovery system for scientific and technical literature.

3.1 Promoting the Establishment of Scientific and Technical Literature's Metadata Resources Integration and Fusion System

To establish the digital scientific and technical literature guarantee system based on printed documents, it is necessary to promote the sustainable development of metadata resources. On the one hand, the development of the "big" metadata system of China's scientific and technological information resources and the formation of an integral national metadata database of scientific and technical literature need to: ① expand the multi-channel acquisition method of scientific and technical literature metadata resources, from purely processing to multi-channel of processing, collection, donation, legal deposit, purchasing, etc.; ② conduct real-time discovery, collection, standardization and preservation of metadata such as network resources and open access resources; ③ uniformly process the metadata of different sources, types, carriers and formats through the cooperation with domestic and overseas publishers, relevant information agencies, etc.; ④ strengthen the integrity and update timeliness of metadata. On the other hand, in order to realize functions such as structured storage of national metadata, resource and service scheduling, custom reorganization of metadata, resource positioning and discovery, it is necessary to ① integrates the internal business systems, such as document subscription and selection, joint cataloging, resource processing and data warehousing, to form a unified digital business management system; ② formulate the functional requirements and system framework of national metadata database, so as to establish a metadata integration and fusion system; ③ implement the mapping registration of multi-source metadata's format, for integrating multi-source metadata at the three levels of data, information, and semantics; ④ design a functional model covering the entire process of metadata resource acquisition, integration, and information services [7].

3.2 Carrying Out the Development of Knowledge Organization System and Its Application Demonstration Intensively

The building a knowledge organization system with China's independent intellectual property rights, the conducting of application demonstrations and key technology research based on the scientific and technical knowledge organization system, can promote the development and utilization of NSTL's scientific and technical literature resources, and give full play to there source benefits. Therefore, it is necessary to build an organization system of foreign scientific and technical knowledge targeted at domain ontology, develop a system that can effectively serve the organization of scientific and technical literature, and focus on building a unified super scientific and technical thesaurus with a certain scale. Establishing a sustainable development mechanism for the national scientific and technical knowledge organization system, and supporting the continuous maintenance and update of Simple Knowledge Organization System (STKOS) to ensure its sustainable development. The promoting of the open service for the scientific and technical knowledge organization system, the support for large-scale application of NSTL's scientific and technical literature and open application service for the national scientific and technical information service institutions, will make the scientific and technical knowledge organization system become the information infrastructure which support various information institutions and research institutions in China. At present, the system has accumulated 530 million pieces of various types of data. NSTL will continuously strengthen the development of data and knowledge resources, reinforce the description and disclosure of resources, and gradually improve the data processing capabilities and the efforts to integrated disclosure.

3.3 Strengthening the Application of Semantic Knowledge Organization of Scientific and Technical Literature and Correlation Technology and Methods

The development of scientific and technological innovation requires deepening the knowledge organization and service capabilities of scientific and technical literature information. On the one hand, the massive information resources should be organized in a knowledge-based way, and the intelligence research based on big data analysis should be strengthen, to support the personalized knowledge services based on user issues and in-depth analysis. On the other hand, it is also necessary to rely on massive information to provide analysis and estimation on technological and industrial development strategies, so as to support decision-making [8]. Therefore, it is also essential to carry out in-depth organization and knowledge disclosure of multi-granular scientific and technical literature information, and adopt a scientific and technological knowledge organization system to support the development of application services and third-party services based on knowledge organization tools. Besides, the establishment of content disclosure and indexing mechanism of metadata, the gradually realization of the automatic semantic indexing and semantic association of the full content set, and the support for the discovery, linking, and reorganization of knowledge objects and relationships across the content types is also needed. Exploring revealing methods by using bibliographic titles, chapter knowledge points, informative abstract, entity descriptions, etc., and achieving multi-dimensional links to concepts, institutions, funds, authors, citation relationships, similar literature, bibliography, chapters, etc., to satisfy multi-level and all-dimensional information retrieval and scientific evaluation needs such as scientific research, technological innovation, performance management and the assessment of resource development.

3.4 Building a New Generation of Scientific and Technical Literature Resource Discovery Platform

As a brand new intelligent search and one-stop acquisition tool for academic information, the resource discovery system can comprehensively reveal and integrate various resources. Through ways such as metadata fusion, federal retrieval, resources linking, resources reorganization, data mapping, and data association, massive heterogeneous document information resources are gathered, which enables users to retrieve, display, sort, and acquire all the library resources in one interface, realizing the integration of multiple service methods and the selection of various access methods, which meets the needs of users in the Internet era. Accordingly, NSTL has built a new generation of scientific and technical literature resource discovery platform, deeply integrating the network service system, the international scientific citation retrieval system, the retrospective literature service system, the open resource integration service platform, the information portal from key fields and enterprise, and so on, which were successively built in different developments periods, from three aspects of data, service, and system. It also integrates the scientific and technical literature's abstracts information (secondary literature) retrieval services and its full-text providing services to realize functions such as resource integration, data organization, integrated retrieval, resource positioning and scheduling, and intelligence discovery. Furthermore, NSTL expand sits services to aspects such as semantic retrieval, knowledge navigation, bibliometric analysis, and the construction of scientific research network, and establishes domestic scientific and technical literature resource discovery centers, resource allocation and scheduling centers, and knowledge service centers, to provide one-stop public services from discovery to acquisition, and from literature retrieval to knowledge discovery.

At the same time, the discovery system expands the discovery scope of scientific and technical literature resources. Users can retrieve third-party resources in addition to the ordered literature, including OA resources, nationwide access resources, and retrospective resources. It also provides multi-channel resource acquisition methods, including local acquisition, full text delivery, interlending, OA downloading, retrospective downloading, DOI link, single article order, etc. And the interface and the tool system that supports open use is established, providing services such as professional vocabulary navigation for all disciplines, intelligent expansion of concepts, and homonymy discrimination. Still, cross-language retrieval based on knowledge organization tools is realized; and based on third-party tools, the bilingual translation (English-Chinese, Japanese-Chinese) and cross-language retrieval of the search terms, and the translation of search results is provided. Moreover, multi-dimensional association and disclosure of literature resources is completed; then based on normative data, correlation technologies, and visualization technologies, exploratory user experiences is provided, such as multidimensional facets, knowledge navigation, and content association.

3.5 Developing Thematic Portals Based on the Discovery System and International Scientific and Technological Citation System

The development of thematic information portals in key areas is an important means for NSTL to provide professional and knowledge-based support services, and a significant part of the discovery system. Based on the mode of “expert + platform + data”, the thematic portals develop a tool platform that support the new data management services and analysis services, providing users with thematic intelligence data management, data analysis services and data product release services [9]. The development will focus on national key areas such as energy, environment, transportation, and artificial intelligence, as well as key technology areas such as information technology, space technology, energy and resource technology. Tracking and monitoring related fields and technologies by collecting, analyzing, selecting, reorganizing and in-depth processing the network information. Providing information translation reports, hot topic recommendation, the matic intelligence news and other services, by means of establishing a service mechanism for data exchange and scenario fusion with the NSTL’s next-generation knowledge discovery platform. Making full use of knowledge computing models such as statistical analysis, correlation analysis, cluster analysis, co-occurrence analysis, etc., to realize the service organization of innovative talents’ distribution maps, accurate discovery of experts, scientific and technological development trends, hot issue’s research and estimation, research topic analysis, research output analysis, research institution analysis, technical route analysis, research project analysis and so on, which then provide users with comprehensive support about science and technology innovation.

Based on massive citation data, the international scientific citation system can reveal the progress of the world scientific research, and display the relationship of the cooperation, crossing, reference and utilization between different discipline among scientific research [10]. NSTL's international scientific and technological citation system sets document discovery, citation link and citation analysis as a whole, and preferentially selects excellent foreign language periodicals in science, engineering, agriculture, and medicine to process citation data, reveal and calculate the correlation and association strength between the documents, which provides an analytical tool for researchers to understand the context of world scientific research and development. And the system becomes one of the important tools for knowledge discovery services.

4 To Promote and Deepen the National Scientific and Technical Literature's Resource Discovery Service

The development of the national scientific and technical literature resource discovery platform is based on NSTL's original network service system, reforming its front desk, retrieval and personalized services, as well as knowledge services, including function optimization of retrieval, results presentation and delivery, designing of personalized service and user community function, the adjustment and designing of the WeChat public account and APP function. After the completion of the system, the number of user registrations, visits, literature retrieval, secondary document views increased substantially compared with the past. In the promotion and deepening of the national scientific and technical literature resource discovery service, the following aspects should be strengthened.

4.1 To Develop a User Management System of Unified Authentication

Establishing a user management system of unified authentication is the primary prerequisite and basic means to achieve the interconnection, fusion and correlation between scientific and technical literature information resources and service platforms, which can greatly facilitate users to switch between multiple systems, and utilize system resources and functions more simply and efficiently. The unified authentication of the user identity needs to determine whether a user is a legitimate user. Once the user's identity is authenticated, which resources the user can access and how they can be accessed are determined. The establishment of a user

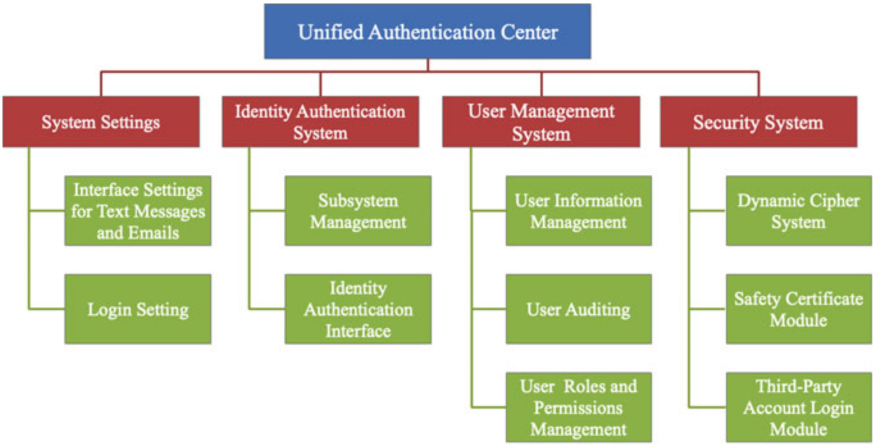


Fig. 1 The functional framework of the unified authentication system

management system for the unified authentication of the national scientific and technical literature resource discovery service is to establish a unified user authentication center that integrates the users from the national scientific and technical literature resource discovery platform, online bibliography system, international scientific citation system, thematic portals for various disciplines, and the introduced third-party information services system into unified management, to realize the unified identity information storage and ensure the consistency of user information (Fig. 1).

4.2 *To Innovate the Scientific and Technical Literature's Public Service System Covering the Whole Country*

Based on the national scientific and technical literature resource discovery platform, NSTL has established a nationwide scientific and technical literature public service system [11]. In order to continuously optimize the service sites system and improve the service quality and efficiency, NSTL has established a system of unified service sites' cloud platform, enabling 40 service sites to develop, manage and serve users on one system, achieving the unified management, training, consulting of users across the country, and forming its reliable service coverage in areas where the scientific and technological information is in short supply, and the high-tech parks, major enterprises and industries located. Besides, supplementing information pushing services, one-on-one consulting services, characteristic tool services, etc., promoting embedded resource utilization tools, providing resource integration tools that integrate national platforms, regional platforms, third-party agencies, local agencies, etc., and improving personalized information resource guarantee and service capabilities, can allow users to fully experience the great convenience provided by the new information environment for information discovery.

4.3 To Deepen the Intelligence Service System for Major Tasks and Important Decisions

Facing the different needs of various types of innovative subjects for scientific and technological information, NSTL relies on a new generation of scientific and technical literature resource discovery platform to establish a variety of online scientific and technological information automatic monitoring and decision supporting information service platforms. On the one hand, it deeply develops decision-making information support services for government departments, provides regular services such as briefings, news and trends, special reports, research and investigation reports to government departments, supplies high-level scientists with cutting-edge thematic analysis information on science and technology, and think tank supporting information services. On the other hand, developing the matic intelligence services for national key research and development programs, through various serving way such as sci-tech novelty retrieval, customizable subject retrieval, institutional knowledge base building, development of scientific and technological information automatic monitoring service platform, special topic dynamic newsletters, patent analysis and intelligence research reports, to conduct support and guarantee service of scientific and technological information and provide intelligence services for the front-line scientific research arrangement, key R&D problem tackling, disruptive technology priority selecting, and industrial transfer and transformation.

4.4 To Promote the Competitive Intelligence Services for Regional Economic Development

The guarantee service of national scientific and technical literature also need to oriented to the main battle field of national economic, to promote regional economic innovation and technical progress of enterprises. Therefore, on the one hand, NSTL carried out the building of a scientific and technological information service system oriented to the three major national strategies. Specifically, guided by major strategies such as the development of 'the Belt and Road Initiative', the coordinated development of Beijing-Tianjin-Hebei region, and the development of the Yangtze River Economic Belt, to build 'NSTL's information service platform for major national strategies'. Furthermore, relying on the platform to carry out services such as the matic literature, briefings for special topics, monitoring reports, scientific and technical reference briefings, as well as analysis of the technological competitiveness of countries along the Belt and Road, and information research and services about the coordinated development of the Beijing-Tianjin-Hebei industry and ecological environmental protection in the Yangtze River Economic Belt, to form planning and customized intelligence research reports, and vigorously develop information services to support the country's three major strategies. On the other hand, actively develop scientific and technical literature services for enterprise innovation. To be more specific, based on the building of the integrated industrial information service platform, enterprise competitive intelligence system, and scientific and technical achievements transformation platform, NSTL will further deepen the professional information retrieval, industry development situation analysis, competitive intelligence landscape service, patent novelty searching and technology opportunity analysis, patent portfolio analysis, etc. Moreover, strengthening the thematic service capabilities for key industries, high-tech enterprises, small and micro innovative enterprises, innovation incubation parks, etc., to provide knowledge-based information products and expand the information services that support enterprise product R&D and independent technological innovation.

5 Conclusion

Nowadays, with the rapid development of science and technology, a new type of collaborative scientific research and learning environment is taking shape. Along with the rapid development in science and technology in China,, the variety and quantity of information demand for scientific research innovation are also increasing, which urgently needs to build a large-scale, converged, networked, knowledge-based, and connected resource discovery platform for scientific and technical literature resources. Currently, China's information guarantee system of scientific and technical literature has gradually entered the stage of digitization, networking and intelligentization. NSTL will continue to play its role as the national core platform, providing strategic support for the implementation of the innovation-driven development strategy of china's scientific and technological information, lead and promote the development of science and technology information undertakings in China. Future efforts include ① gradually alleviating the unbalance and insufficient problems in the utilization of scientific and technological information resources in different regions, industries, and fields; ② adapting to the profound changes in the research environment, information environment, and user needs, to improve its knowledge service level in the new era and meet the requirements for knowledge services in the big data environment; ③ developing into the core knowledge infrastructure of China.

References

1. Yuan H, Meng L (2002) The practice of information resources co-construction and sharing in the network environment—including the discussion of the construction and development of National Science and Technology Library. J China Soc Sci Tech Inf (1):57–62
2. Wu B (2010) Establishing NSTL—an important chapter in the development history of Chinese scientific and technical literature. Dig Lib Forum (10):4–6
3. Articles of Association of National Science and Technology Library. Lib Inf Ser (7):90–91 (2000)

4. Zeng J, Deng S (2011) The analysis of resource development and service development of National Science and Technology Library. J Lib Sci China 37(02):30–35
5. Wu B, Zhang J, Jie Y, etc (2015) The hotspot in international digital resource development and its effect to NSTL. Dig Lib Forum (4):2–6
6. Zhang J, Yu Q, Huang Y, etc (2016) Metadata standard design of NSTL unified literature. Dig Lib Forum (2):33–38
7. Peng Y, Wu B, Shen Z (2016) The 13th five-year plan for the development of National Science and Technology Library. Dig Lib Forum (11):12–20
8. Ding J (2015) Research on the development of scientific and technical literature resource sharing platform of NSTL. Res Lib Sci (20):39–41, 51
9. Cheng B, Jin Q, Zhang J, etc (2017) Construction and prospect of NSTL information portal in library and information science. Dig Lib Forum (11):54–58
10. Meng L, Zhang J, Liu X, etc (2010) Building database of international science citation and expanding information service of NSTL. Dig Lib Forum (10):58–61
11. Qiao X, Liang B (2008) The change for service mode of National Science and Technology Library in the new era. Digital Lib Forum (12):50–52, 69

Yiqi Peng, is the director of the National Science and Technology Library(NSTL), and the former deputy director and inspector of the Basic Research Department of the Ministry of Science and Technology(MOST). He has devoted himself in national fundamental research management for a long time in MOST. For example, he has been engaged in the management, basic work and construction of stateskey laboratories of the National Basic Research Program of China, participated in the ITER Program's special project of national nuclear fusion research in China and basic research international cooperation program, and taken part in formulating related plans and policies of national basic research.

The Development of Chinese Sociological Knowledge Space Based on Linguistic Representation

Huashan Chen

Abstract The traditional subject analysis is mainly based on the qualitative analysis of the text, supplemented in part by the method of literature measurement and analysis such as word frequency and bibliographic citations. Based on the full text of the papers published in the top two journals of Chinese sociology, Journal of Sociological Studies (1986–2015) and Chinese Journal of Sociology (2006–2015), this paper studies the changes, development characteristics and sub-discipline characteristics of sociology in the past 30 years by means of modern text semantic analysis. This paper also shows the feasibility of applying modern natural language analysis techniques as measurement on the subject of knowledge space in social sciences.

Keywords Word embedding · Knowledge space · Social network analysis

1 Knowledge Graph Mapping Based on Natural Language

Sociology of knowledge is the study of the relationship between human thoughts and its social context and of the effects that prevailing ideas have on societies. The development of contemporary sociology of knowledge, which focuses on the study of the production, storage, dissemination and application of knowledge, has begun to move increasingly away from insight-based qualitative analysis towards empirical studies that are primarily quantitative. Sociology increasingly attaches importance to the role of knowledge in social development and change, and addresses a range of issues or ideas in the formulation and implementation of social policies.

From the perspective of knowledge graph, the knowledge space or idea space can be seen as a high-dimensional vector space made up of a large number of vocabularies. Each published academic paper or material can be seen as a scholarly sample, and the overall sample is drawn from the latent knowledge space within discipline. Therefore, the knowledge space can be measured based on the frequency of academic vocabulary (nouns, terms, theoretical concepts and scholarly names) and

H. Chen (✉)
Institute of Social Development, Chinese Academy of Social Sciences, Beijing, China
e-mail: chenhs@cass.org.cn

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_17

their co-occurrence observed in the academic papers. Mathematical models based on natural language semantics have been developed and evolved quickly in computer science with the proposed the Distributed Representations and Word Embeddings models [1, 2]. The theoretical basis of this language model is that words with similar contexts have similar semantics [3], or rather, the semantics of words are determined by their context. The word vector/embedding model has made notable success on semantic analogy tasks [1, 4]. With word vector model, word in natural language are transformed into low-dimensional dense vectors (Dense Vector) that can be manipulated by computer and algorithms, so that character encoding that would not otherwise be directly computable becomes a computable series of vectors, based on which word-to-word relationships can be extracted, such as finding similar words (synonyms, etc.), semantic associations (China—Beijing = UK—London), etc. More importantly, the multidimensional character representation of word vectors applied to the knowledge space gives it the meaning of geometric representation, which is richer in information than the traditional symbol-based and rule-based methods of knowledge space theory, and can be computed based on observed data to construct the knowledge space while the traditional methods of construction must be based on expert judgment.

1.1 Data Sources and Word Vector Modeling

With the goal of constructing the intellectual space of sociology in Chinese, this study systematically collects 3,674 papers with full-text published from 1985 to 2015, using two professional sociology journals, Journal of Sociological Studies and Chinese Journal of Sociology, which are also currently the only two sociological journal in China, to present the discipline development since the reconstruction of sociology after the economic reform and opening up policy.

Since the 1980s, Chinese sociology has undergone a process of recovery and reconstruction, starting with remedial courses, expanding the discipline and growing rapidly. During the thirty years of disciplinary recovery and reconstruction, sociology has generally undergone a systematic process of introducing Western social theories and research methods to emphasize localization and indigenization, from simple transplantation to the pursuit of subjectivity and self-awareness. In terms of essay writing style, from early theoretical thinking and prose writing to the beginning of emphasizing the chapter structure of the essay and research norms ("foreign stereotyped writing"), the question, theory (literature), hypothesis, data, measurement, methods, findings and conclusions [5]. In terms of paper format, it also began to become more and more standardized, and until the late 1990s, most articles published in the Journal of Sociological Studies had no references and only a few articles listed the main references. Since its revision in 1999, the editorial style of articles in Journal of Sociological Studies has become standardized with a formatted editorial style.

On modeling the word vector from the article full text, we are concerned not only with the knowledge space constituted by sociological conceptual vocabulary, but also with the knowledge space constituted by sociological figures. The use of conceptual vocabulary varies not only due to its rapid migration but also tranlation by personal preferences, thus making it difficult to accurately track conceptual semantics. The vocabulary of sociological figures itself does not change, but is only reconstructed and represented through scholarly texts by posthumous understanding of their relevant theories and studies. Therefore, it is more helpful to trace the temporal changes of the sociological discipline and to grasp the currents of the disciplinary by constructing an intellectual space for sociological figures.

The editorial style of both Journal of Sociological Studies and Chinese Journal of Sociology uses Chicago Style Citation format, an Author-Date System using parenthetical citations in the text to reference the source's author's last name and the year of publication. Each parenthetical citation corresponds to an entry on a References page that concludes the document. Below are simple example for CSC:

(Goman 1989, 59), or (Fairbairn and Fairbairn 2001), or (MHRA 2004).

Thus, as opposed to numbered citation, the Author-Date citation format allows the cited author to have a very close textual proximity to the disciplinary concepts in question, and the use of word vector model helps to discover the connections between names and concepts.

This paper uses the GloVe word vector model proposed by Pennington et al. (2014) and it is generally accepted that the GloVe model using word covariance matrix helps to exploit global statistical information and works better on small corpus.

1.2 Measurement and Model Fitting

Word embedding models fall into the category of unsupervised learning, so the vast majority of efforts to improve word vector models in current natural language processing practices rely on word similarity datasets such as WordSim-353 for metrics and use them as criteria for evaluating model fit quality. However, this evaluation is based on a common vocabulary and is no longer applicable when undertaking a similarity measure of a professional disciplinary text. Realistically, there is no professional vocabulary dataset for sociological texts, let alone a professional vocabulary similarity dataset.

Therefore, in this modeling task, the evaluation of the fit to the word vector model was carried out in the form of expert evaluation. In GloVe modeling, the two core hyperparameters are the word window of word co-occurrences and the vector dimension, so a set of parameter combinations (window step values: 30, 50 and 100; dimension: 50, 100, 200) are applied to fit the model resulting nine models, respectively, and five experts are asked to evaluate the fit of the nine models and at last an optimal model was selected.

In order to reflect the process of reconstruction that China's sociology has undergone over the past three decades, this study divides the data into six epochs with 5 year period by the year of publication of the articles. By screening the names of the authors of the papers and the names of the scholars cited in each epoch (mentioned more than three times), an undirected network with the names of the scholars as the nodes to represent knowledge associations are established by computing the Cosine correlation for each epoch. The final results are six epochs of knowledge space networks with nodes of scholar names representing the temporal of development of the Chinese sociology discipline.

2 Evolution of the Chinese Sociological Knowledge Space

2.1 Rapid Expansion of Knowledge Space and the Process of Reconstruction of Chinese Sociology

As of 2018, sociological rehabilitation and reconstruction has been carried out for forty years, and many scholars have classified sociological rehabilitation and reconstruction according to different criteria at different stages. For example, Li Wei divides the process of sociology restoration and reconstruction into three stages: the social survey practice stage (1979–1989); the growth stage (1990–1999); and the prosperity stage (2000-) [6]. Feng Xiaotian divided the development of the domestic sociological community in the field of research methods in 1979–1999 into three stages: 1979–1985, 1986–1992 and 1993–1999, and then characterized the three stages of development in terms of "learning"–"practice"–"improvement" [7]. Fang Ming and Wangjie, on the other hand, divided the first decade of sociology (1979–1989) into three phases: the nascent phase (1979–1982), the second phase (1983–1985), and the third phase (1986–1990) [8]. However, all the above studies lack an overarching perspective.

In our study based on the six epochs of knowledge space networks, it's now possible to show the growth and development of Chinese sociology rehabilitation. The network size has gradually increased from 621 nodes in 1986–1990 to 2199 nodes in 2011–2015 and the network density has also increased in parallel. This shows that since the restoration and reconstruction of sociology, the number of Western scholars introduced into Chinese has been increasing, the number of domestic scholars engaged in sociological research has been increasing, and the sociological intellectual output and research frontiers have been expanding simultaneously at any time. In terms of growth rate (see Fig. 1), the sociological knowledge space demonstrates a typical S-shaped growth curve (see Fig. 2), which is also a classic law of exponential growth of knowledge, a process from the infancy (1986–1990), development (1991–2005), expansion (2006–2010) and stabilization (2011–2015) of a discipline (Table 1).

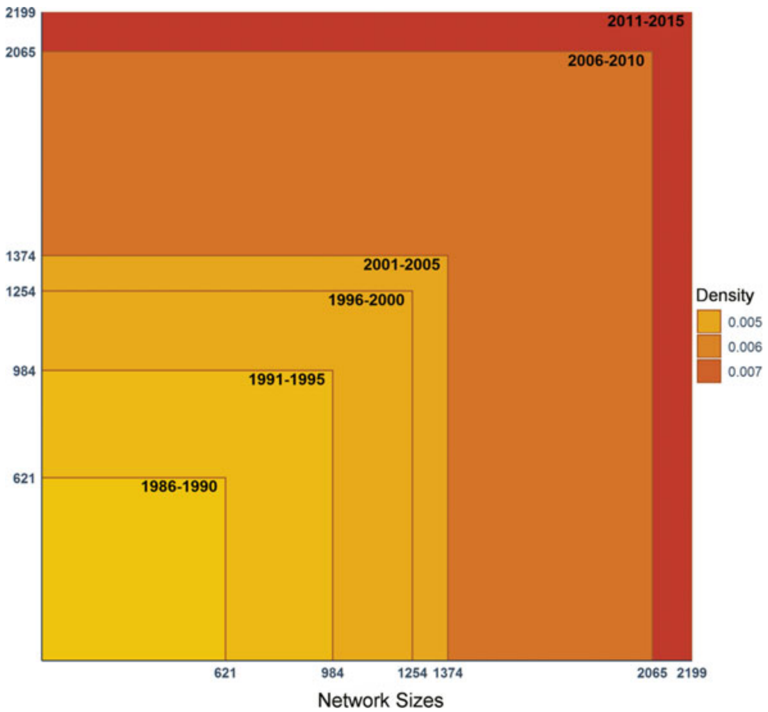


Fig. 1 Growth of knowledge space (6 epochs)

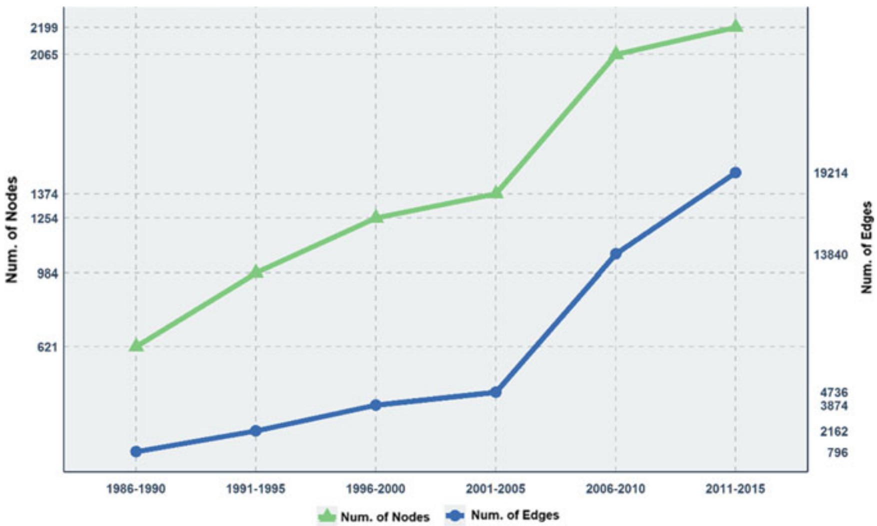


Fig. 2 Growth rate of knowledge space (6 epochs)

Table 1 Features of 6 epochs of knowledge space network

	No. of nodes	No. of edges	Graph density	Modularity	No. subgroups	No. of cut points	Degree centrality (mean)	Degree centrality (max)	Degree centrality (variance)
1986–1990	621	796	0.00414	0.7211	21	148	2.564	15	4.46
1991–1995	984	2162	0.00447	0.5305	16	97	4.394	26	9.86
1996–2000	1254	3874	0.004931	0.4362	15	51	6.179	30	13.97
2001–2005	1374	4736	0.005021	0.4078	12	32	6.894	34	15.20
2006–2010	2065	13,840	0.006494	0.3354	9	6	13.40	68	62.09
2011–2015	2199	19,214	0.007951	0.3196	5	0	17.480	105	150.60

2.2 Network Dynamics and Evolution of Knowledge Space

A dynamic holistic network perspective helps researchers examine the dynamics of network development, its latent structure and how it evolves. Small World phenomenon is often used to explore the formation of networks as a starting point. A small-world network is a network with short average path length and large clustering coefficient relative to a random network of equal size and density. In network systems with small-world features, local behavior leads to global results, and the relationship between local and global dynamic features is often dependent on the structure of the network [9–12]. For example, there may be a few star scholars in the knowledge network, surrounded by a large number of collaborators or knowledge associates. In knowledge sociology point of view, star scholars usually have more academic publications, more knowledge bearers, and more academic resources, star scholars tend to receive the majority of attention and social recognition, thus supporting the main network structure. In addition, from the point of view of the development of discipline, the star effect in the knowledge space network is an indicator of whether a discipline is conservative or innovative, that is, whether the discipline is more focused on classic or front line.

We use the Scale Free Graph to detect the presence of "star scholar" effects in knowledge space networks. In network analysis, this test also determines whether there are power law in the distribution of network node centrality [13, 14]. In this study, since we are only concerned with whether the network node centrality of each epoch is more consistent with the idempotent distribution feature, linear regression is used to fit the node centrality values to the centrality frequency and to compare the fitted regression linear coefficients between the six epochs. In Fig. 3, the fitted R^2 of linear regression of the power distribution for the six networks of sociological knowledge space gradually decreases from 0.8 in the 1986–1990 to 0.615 in the 2011–2015, indicating that there is a strong academic star network effect in the knowledge network in the early days of sociology reconstruction, while the academic star effect subsequently decreases while the discipline becomes more and more specialized in the branch disciplines.

2.3 Trends in Indigenization in Chinese Knowledge Space

By extracting the names of the top twenty scholars with the highest degree of network centrality in the six epochs, we can see that from the overwhelming majority of classical philosophers and sociologists such as early Marx, Engels, Lenin, Weber, Hegel, and Durkheim in the 1986–1990 epoch, through the intermediate transition where native sociologists began to occupy a certain percentage, to the current middle-aged native sociologists occupying the majority in the 2011–2015 epoch. Thus, according to Kuhn's theory of paradigm shift, the continuous efforts of Chinese scholars since the restoration and reconstruction of Chinese sociology have not only expanded the

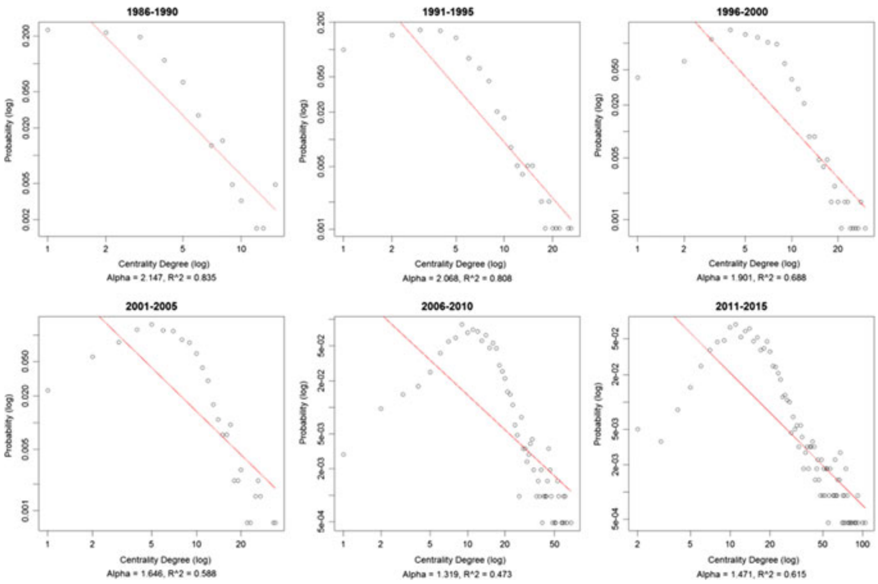


Fig. 3 Distribution of network centrality and fitted linear regression

field and space of sociological research, but also contributed to the continuous expansion of the intellectual space of Chinese sociology, through the study and introduction of Western philosophy and sociological theory, and their digestion and application to Chinese research.

If one compares scholars from China and abroad, one can see that the proportion of domestic scholars has increased year by year, from 25% (1986–1996) to 65% (2011–2015) (Figs. 4 and 5). There is no doubt that this change has a very strong Chinese character. The disruption of the academic bloodline and the disruption of the talent pool caused by the disappearance of sociology for more than two decades in the early stage of recovery and reconstruction [15] not only resulted in fewer corresponding research results, but also the research results of existing Chinese scholars on Chinese society were not effectively inherited, and thus a large number of foreign sociological theories were cited and borrowed to conduct research on Chinese issues. In addition, the mainstream theories and methods of sociology are almost always based on the study of European and American societies by Western scholars and most sociology practitioners have to devote a lot of time and energy to study Western sociology and imitate the work of Western scholars [16]. The growth in the number of foreign scholars has slowed down significantly, reflecting a slowdown in the introduction of foreign sociological works.[1]

[1]In extracting the names of scholars, only scholars with Chinese translations of their names are retained for foreign scholars, that is, scholars whose works have been translated into Chinese and published. The translation of the work into Chinese implies formal introduction of the work of

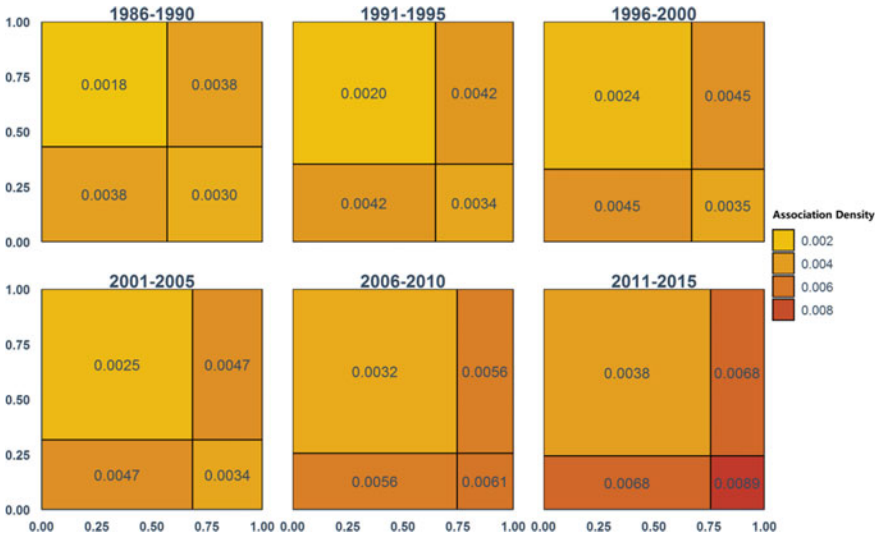


Fig. 4 Size and association density of Chinese and foreign scholars

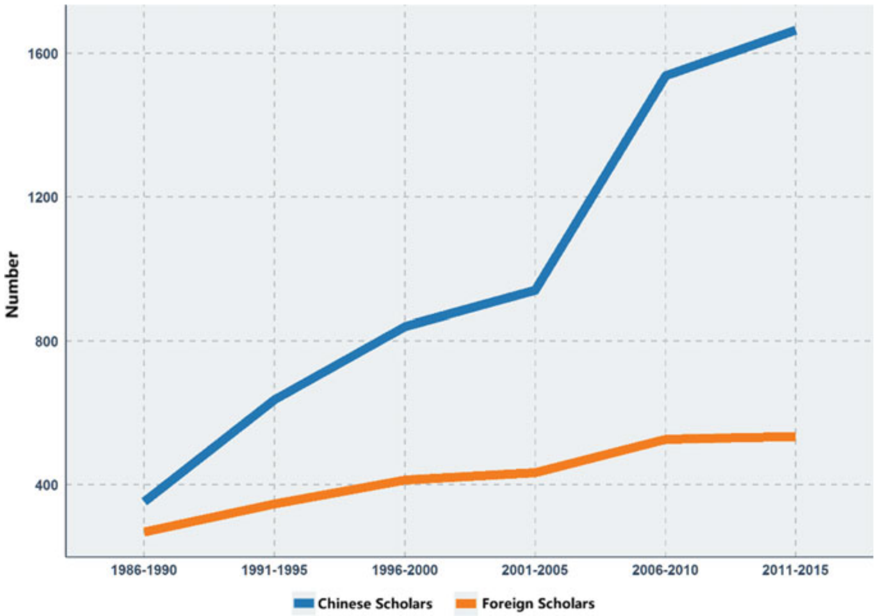


Fig. 5 Growth in the number of Chinese and foreign scholars

3 Characteristics of Knowledge Space Network and Discipline

There are two important theoretical ideas in the field of scientific knowledge research. One is Lakatos's theory of the scientific research program, which holds that a family of theories is composed of branch theories and has a "hard core" that constitutes the development of the program, with protective bands around the "hard core" of theories, such as various auxiliary hypotheses, but also the assumptions needed to describe the initial conditions [17]. The second is Thomas Kuhn's scientific paradigm theory. Kuhn argues that paradigms arise from scientific discoveries. Specifically, a paradigm is a common belief, a universally recognized scientific achievement that includes scientific concepts, laws, theories, problem-solving models, paradigms, applications and tools, etc., and a paradigm is also a theoretical perspective or research perspective that is recognized and shared by the scientific community [18]. In the definition of the concept of paradigm, it is clear that there is a close relationship between the paradigm and the scientific community, i.e. the scientific community is the realistic basis for the existence of the paradigm and the paradigm is the link that sustains its existence, development and growth. On one hand the paradigm exists on the basis of scientific communities, which can be either a large discipline [19] or a community of people that are differentiated and reorganized from the large community [20–22]. On the other hand, the paradigm is also the foundation and the link that sustains the scientific community, after all, the paradigm provides a unified perspective and insight to the members within the community, making professional communications within the community. However, when the paradigm is in crisis, the scientific community itself will be in chaos, gradually disintegrating and subsequently reorganizing under a new paradigm.

How to measure the "hard core" of knowledge space is a difficult issue, and for a long time there have been few relevant metrological studies applying the concept of the "hard core circle". In the 1960s, Price's statistical analysis of papers revealed the existence within each field of a "community of a hundred" representing the frontiers of research, who communicated and interacted with each other through e-mail, unpublished manuscripts and scientific commuting mechanisms [23]. However, this actually measures the behaviour of academic cooperation rather than the content of the domain of knowledge, and is therefore still some way from the concept of "hardcore" knowledge. The evolution and formation of different paradigm groups within economics is more observable and structural than in other social sciences. Therefore, economics has more research on paradigms, but generally takes the character or research platform as the core to show the development of different stages and paradigms of economics through specific discourse [24, 25]. With the development of statistical analysis and visualization techniques, some researchers have conducted network analyses of cross-citation relationships and co-authors of papers from a bibliometric perspective, attempting to describe and analyze the evolution and

---

foreign scholars. There are large numbers of foreign scholars whose work are directly quoted and referenced.

development of network structures in different disciplines [26], but fewer studies have directly addressed the "hardcore" analysis of knowledge.

In this study, the concept of structural embeddedness used by Moody et al. [10, 27] is used to measure the "hard core" of knowledge. The so-called structural embeddedness refers to whether a network subgroup can easily be cut into multiple discrete subgroups. If a subgroup needs to delete multiple nodes to achieve the purpose of being broken up, it indicates that the cohesion of the network subgroup is strong, and the opposite is weak. In this study, as the knowledge mapping is constructed through the measurement of full-text articles, the association of scholar names reflects the knowledge associations inherent in the knowledge domain, rather than mere scholarly collaboration and communication as Price measured. Therefore, if a subgroup of scholars is more densely and closely associated, it reflects more intersection of their research content, fields of study and knowledge link. Therefore, the subgroups with the highest network embeddedness can be extracted as knowledge "hard-core circles" based on the six temporal networks used in this study.

In order to better reflect the changes in the hardcore circles of knowledge networks, this study distinguishes two sets of indicators for the strength of network relationships, i.e., one is the structural embeddedness indicator computed based on the overall knowledge map structure and the other is the structural embeddedness indicator only for networks with strong relationships, that is, cosine similarity coefficient greater than or equal to 0.4.

By calculating network embeddedness indicators for the six epochs of the strong association knowledge network (see Table 2), it can be seen that the "hardcore circle" in the strong association knowledge network is almost entirely composed of classical Western social theorists, and extends from the early sociological founders Marx, Condé, and Weber to the major sociological theorists. Among all the six time periods, only Mr. Fei Xiaotong was ranked in the subgroup with the highest network embeddedness in the 2006–2010 peoch.

Table 2 Hardcore circles in strong association networks (six epochs)

Epoch	Scholars in hard core circle
'1986–1990'	Comte, Engels, Hegel, Marx, Weber
'1991–1995'	Comte, Durkheim, Engels, Hegel, Marx, Merton, Parsons, Spencer, Weber
'1996–2000'	Comte, Durkheim, Engels, Hegel, Giddens, Marx, Merton, Parsons, Spencer, Weber
'2001–2005'	Comte, Boudieu, Durkheim, Engels, Habermas, Hegel, Giddens, Marx, Mead, Merton, Parsons, Spencer, Weber, Zimmer
'2006–2010'	Bauman, Boudieu, Brown, Comte, Durkheim, Engels, Fei Xiaotong, Freud, Foucault, Habermas, Hegel, Giddens, Kant, Levy, Luhmann, Mannheim, Marx, Mead, Mills, Moss, Merton, Parsons, Spencer, Smith, Turner, Tocqueville, Weber, Zimmer
'2011–2015'	Boudieu, Comte, Durkheim, Engels, Freud, Foucault, Habermas, Hegel, Hobbes, Hume, Giddens, Kant, Locke, Marx, Moss, Parsons, Rousseau, Smith, Tocqueville, Weber, Zimmer

Table 3 Chinese scholars in hardcore circles of '2011–2015' epoch weak association knowledge network

Bian, Yanjie	Cai, He	Cao, Zhenghan	Fei, Xiaotong	Feng, Shizheng	Li, Youmei
Guo, Yuhua	Li, Lulu	Li, Meng	Li, Peilin	Li, Qiang	Su, Guoxun
Liu, Shiding	Liu, Xin	Lu, Xueyi	Qu, Jingdong	Shen, Yuan	Zhao, Dingxin
Sun, Liping	Wang, Mingming	Yang, Shanhua	Ying, Xing	Zhang, Jing	Wu, Wenzao
Zheng, Hangsheng	Zhou, Feizhou	Zhou, Xueguang	Lin, Nan	Liang, Shuming	

In the overall knowledge network, where weak associations are included, there is a rather different result. Among them, the highest network embeddedness coefficient is 20 in the 2011–2015 epoch, with a total of 84 figures, including 29 Chinese scholars (see Table 3) and the 55 foreign theorists. The difference between these two kinds of strong and weak networks indicates that, due to the disciplinary inheritance, Western theorists and their knowledge systems are still the academic sources and theoretical hard cores (strong networks) of contemporary Chinese sociology, however, with the development of Chinese sociology and the deepening of sociological practice, Chinese sociology has more and more native elements in research topics and research cores. In terms of the paradigm shift, it can be argued that the paradigm shift in Chinese sociology will emerge from the weakly associated networks and will have more Chinese characteristics.

4 Conclusions

This study measures the knowledge space of Chinese sociology through the introduction of word embedding techniques in the natural language processing field, and provides a systematic and holistic review of the more than 30 years of recovery and reconstruction in Chinese sociology based on the constructed six-epoch knowledge space networks.

Sociology has been facing a problem since the day it was introduced to China: the indigenization of the discipline. This issue has also been one of the hot topics in the academic community. Through a quantitative analysis of text in Chinese sociology over the past three decades (1986–2015), this paper summarizes the growth of the knowledge space since the reconstruction of sociology in general: sociological knowledge has been increasing, the rate of knowledge growth began to stabilize after a period of explosive growth, and the proportion of domestic scholars increased year by year. As the paradigm transplanted from the West, sociology is able to describe and explain Chinese society to some extent. Chinese sociology is beginning to produce preliminary results at the level of research theory, research concepts and research methods, and based on this, more in-depth exploration and study of Chinese issues.

However, sociology is a discipline that studies social processes and social phenomena in specific social contexts, and research outside of social contexts is of little significance [28], coupled with the increasing number of scholars engaged in sociological research, many scholars have gradually found that the established Western research theories, research concepts and research methods cannot well describe and explain Chinese society, and there is a very obvious tension that the Western sociological paradigm is in crisis in Chinese research. Through the measurement and analysis of the "hardcore circle" of sociological knowledge space, this study demonstrates the difference between the classical "hardcore" and indigenous "hardcore" of sociological knowledge, and from the data analysis results, it can be seen that the trend of indigenization of sociological paradigms is becoming more and more obvious and strong.

This study also demonstrates a new way of systematically exploring sociological research of knowledge from a holistic perspective, in addition to traditional qualitative analysis and bibliometric methods. The semantic-based knowledge mapping approach is more capable of exploring the evolution of academic concepts and connotations than the bibliometric approach. Unlike the natural sciences, which are well-defined and quantifiable, there is a certain ambiguity and extensiveness in the definition of concepts in the humanities and social sciences. Even for the same or similar concepts, different scholars often use different terms or conceptual systems to express them. The same vocabulary of terms is often recreated or applied to new fields of study by later scholars, resulting in conceptual migration. Traditional measures that simply identify disciplinary hotspots and potential directions of development by keywords are therefore often less accurate. Through the construction of dynamic semantic knowledge mapping, it is possible to identify historical changes and migrations of concepts to reflect the development of the discipline on the one hand, and the latest semantic migrations to judge potential research trends on the other hand, so as to make forward-looking predictions for the development of the discipline.

Scientific research activities and outcomes are dependent on specific institutional arrangements and social processes. The dissemination of academic knowledge, organizational mobilization and academic collaboration are important impact variables that affect academic output. In the field of humanities and social sciences, the mapping of knowledge based on unstructured textual data with auxiliary data from relevant research institutions would help to incorporate the above-mentioned factors into the academic study process, thereby better examining the level of intellectual output of research institutions and moving away from the purely quantitative number counting approach for evaluation.

References

1. Mikolov T, Chen K, Corrado G et al (2013) Efficient estimation of word representations in vector space. arXiv preprint arXiv:1301.3781
2. Hinton GE (1984) Distributed representations. Carnegie-Mellon University Pittsburgh, PA

3. Harris ZS (1954) Distributional structure. Word. 10(2–3):146–162
4. Pennington J, Socher R, Manning CD (2014) GloVe: Global vectors for word representation. In: Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP), pp 1532–1543
5. Peng Y (2010) Foreign stereotyped writing and social science norms. J Sociol Stud. 02:180–210
6. Li W (2016a) Keeping up with the times: the development of research since sociological recovery and reconstruction. J Sociol Stud 6:73–94
7. Feng X (2000) Twenty years of sociological methodology: applications and research. J Sociol Stud. 1:3–13
8. Fang M, Wang Z (1989) Review and prospects: charting a new course in the development of sociology in China. J Sociol Stud 1989(1):5–14
9. Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. Nature 393(6684):440–442
10. Moody J, White DR (2003) Structural cohesion and embeddedness: A hierarchical concept of social groups. Am Sociol Rev 68(1):103–127
11. Newman ME, Barabási A, Watts DJ (2006) The structure and dynamics of networks . Princeton University Press, Princeton; Oxford
12. Gulati R, Sytch M, Tatarynowicz A (2012) The rise and fall of small worlds: Exploring the dynamics of social structure. Organ Sci 23(2):449–471
13. Barabási A, Albert R (1999) Emergence of scaling in random networks. Science 286(5439):509–512
14. Clauset A, Shalizi CR, Newman MEJ (2007) Power-law distributions in empirical data. SIAM Rev 51(4):661–703
15. Li Y (2016b) Reflections on the construction of the academic discourse in sociology with Chinese characteristics. J Sociol Stud 5:27–37
16. Xie Y (2018) Getting Out of the myth of the localization discussion in Chinese sociology. J Sociol Stud 2
17. Lakatos I (1976) Falsification and the methodology of scientific research programmes. In: Can theories be refuted? Springer, pp 205–259
18. Merton R (1967) On theoretical sociology: five essays, old and new. Simon and Schuster
19. Kuhn TS (1986) The structure of scientific revolutions, 2nd edn. University of Chicago, Chicago
20. Price DJDS, Tukey JW (1963) Little science, big science. Columbia Univ. Press, New York
21. Gaston J (1978) The reward system in British and American science. Wiley
22. Crane D (1988) Invisible colleges: diffusion of knowledge in scientific communities. University of Chicago Press
23. Price DJDS (1963) Little science, big science. Columbia University Press, New York
24. Ye H (2015) Beyond the neoclassical: the fourth revolution in economics and the fourth Synthesis. South Econ V33(8):1–31
25. Ma T (2014) The paradigm structure of Western economics and its evolution. Chin Soc Sci 10:41–61
26. White HD, Wellman B, Nazer N (2004) Does citation reflect social structure?: longitudinal evidence from the Globenet interdisciplinary research group. J Am Soc Inform Sci Technol 55(2):111–126
27. White DR, Harary F (2001) The cohesiveness of blocks in social networks: Node connectivity and conditional density. Sociol Methodol 31(1):305–359
28. Xie Y, Fan Z, Lu, Z et al (2012) The realistic approach to social science. Yunmeng Study 33(3): 5–10.

Huashan Chen is an associate researcher of the Institute of Social Development of the Chinese Academy of Social Sciences and Director of the Center for Social Climate Research. He has been a visiting scholar at the University of Michigan and Stanford University. He is currently engaged in empirical sociological research based on big data, mainly in the fields of organizational sociology, Internet and society, quantitative research methods, and social network analysis, and is chairing a number of empirical research projects related to big data, some of which were funded by the National Social Science Foundation.

Management and Service System of CAS Key Laboratory

Hongfei Hou, Xiaoning Li, Xuerui Bai, Jue Wang, Fan Yang, and Ying Wang

Abstract Key laboratory system is an important part of the scientific research and technological innovation system, the core strength of basic research, applied research and high-tech frontier exploration in Chinese Academy of Sciences (CAS). The purpose of developing the management service platform of Key Laboratory is to realize the informatization of the management service of CAS Key Laboratory (CAS-KLMSS), so the managers can conveniently understand and master the research and operation of the Key Laboratory, and more support will be provided for the implementation of innovation-driven development strategy. This paper designs a platform architecture system based on data flow and approval process for different users and work scenarios according to the actual needs of laboratory management. It introduces the idea of generating system "Database" through filling of laboratory annual reports, to make all data captured, accounted and analyzed. The security mechanism of the system is developed to ensure the encryption of all data transmission and reading. Finally, it outlooks the future development trend of the informatization of the key laboratory management in CAS.

Keyword Key laboratory · Informatization · Database · Security mechanism

1 Background

The world is currently making much account of the construction of scientific research informatization with the development of data-intensive scientific research paradigms. The United States, the European Union and other developed countries and regions take the promotion of scientific research informatization as a strategic measure

H. Hou (✉) · X. Li · X. Bai · Y. Wang
Bureau of Frontier Sciences and Education, Chinese Academy of Sciences, Beijing, China
e-mail: hfhou@cashq.ac.cn

J. Wang
Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing, China

F. Yang
CAS Xunlian Intelligence Supply Chain Network Technology (Beijing) Co., Ltd, Beijing, China

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_18

to enhance innovation ability and international competitiveness, and invest huge amounts of funds to improve the level of scientific research application [1]. The development of laboratory management in foreign countries was earlier. In the 1960s, western countries put forward the concept of Laboratory Information Management System. With the rapid development of information technology, laboratory information management has evolved from simple operations such as simple data collection, storage, modification and querying in the early years to a new management system for overall management of laboratory. At present, the network-based laboratory management system has been widely used in the field of education and scientific research area, medical environmental protection, chemical manufacturing, food safety and other industries in developed countries or regions such as in Japan, Australia, Europe, and the United States. These systems can mostly be set up and operated by users independently, which can be well adapted to laboratory business development, to improve efficiency of laboratory management [2, 3].

In recent years, China has also clearly put forward a series of requirements for accelerating the advancement of scientific research informatization in government's reports. In 2014, the State Council issued several opinions on improving and strengthening the management of scientific research projects and funds of the central government (GF [2014] No. 11), which proposed to strengthen management innovation and overall coordination, strengthen the disclosure of scientific research projects and fund management information, optimize management processes, and improve the scientific, standardized, and refined level of management [4]. It was clearly required to improve the national science and technology management information system, strengthen trace management and information disclosure throughout the project implementation process in the plan for deepening the management reform of the central financial science and technology (GF [2014] No. 64) [5]. As an important carrier for discipline construction and development, the key laboratory system of CAS has become an important base for undertaking major tasks of the State and the CAS. It is the core strength of basic research, applied basic research and high-tech frontier exploration, which has been attracting, cohesive and cultivating a large number of high-level innovative talents [6]. There are 4 national research centers, 82 national key laboratories, and 217 academy key laboratories, which is covering 103 professional research institutes, 1 joint research institute, and 2 universities in CAS. The original offline management model has been unable to meet the needs of scientific researchers and the departments of science and technology management with the rapid development of information technology. The Chinese Academy of Sciences has launched the construction of key laboratory management service platform (CAS-KLMSS) in order to speed up the informatization of scientific research since 2017.

In two years, an information management system has built, which integrates basic laboratory information maintenance, notification announcements, annual report filling, statistical analysis, and online expert evaluation.

2 Features

CAS-KLMSS strives to provide comprehensive information management services for key laboratories, research institutes of CAS and Universities. It has now opened several functional modules such as Laboratory Management, Achievement Management, Annual Report, Data Analysis, Online Assessment and Notification.

2.1 Lab Management Module

The main functions of Lab Management Module include basic laboratory maintenance, laboratory personnel management, role rights assignment, etc. Users can independently maintain complete laboratory information, including laboratory name, code, research nature, focal area, subject point, relying unit, director/deputy director, contact information, and website link; carry out the management of laboratory personnel, such as the names, ages, titles and research directions of fixed, floating and historical personnel in the laboratory; lab administrators can enable different permissions for internal members of the laboratory according to different needs, such as query management permissions for laboratory directors, and reporting permissions for laboratory members.

2.2 Achievement Management Module

The main task of Achievement Management Module is to collects the projects hosted or participated by laboratory members and the output of scientific research achievements, including the publication of monographs, patent applications, and awards. This module has the function of open and dynamic maintenance of project information in order to ensure flexible management and accurate. The platform will automatically record the operation logs during the maintenance process to ensure that the data can be audited and traced back.

2.3 Annual Report Module

The goal of Annual Report Module is to achieve the function of annual report filling, data summary, workflow review, preview printing. Users can directly make supplementary modifications on the basis of the information of the previous year.

2.4 Data Analysis Module

The powerful function of Data Analysis Module is to capture and analyze all kinds of statistical data. The system provides multi-dimensional analysis for managers' final decision based on projects, funds, personnel, and results, for example, the return on investment of fund achievements, investment of project personnel, personnel funds in related fields, etc. All analysis results can be displayed visually through pie charts, line charts, bar charts, and scatter charts.

2.5 Online Assessment Module

Online Assessment Module is relatively independent. It is specially used for the evaluation of key laboratories of the Institute, including the functions of expert management, evaluation material query, expert online scoring and evaluation, real-time summary of scoring and evaluation information, etc. In order to ensure the accuracy and accessibility of the evaluation report, it can automatically summarize the contents of the annual report of the participating laboratories in the past five years and generate the evaluation report according to the specified format. The function of online scoring and real-time statistics of scoring results are realized through distributed data cache middleware.

2.6 Notification Module

Homepage will automatically update and push the latest notification announcement content after the user logs in.

3 Key Technologies

3.1 System Structure

There are four layers of architecture in CAS-KLMSS platform, which are composed of the basic layer, platform layer, business layer, and access layer. The base layer provides computing power and storage resources for the system, and on-demand allocation and rapid deployment of resources are achieved by the cloud platform. The platform layer plays a linking role, providing a series of cluster services, such as resource management, middleware, service integration, distributed scheduling centers, message middleware, workflow engines, development frameworks, ETL data tools, relational databases, distributed databases, distributed file systems, in-memory database, distributed cache, distributed NoSQL, big data analysis engine, etc. The business layer directly provides customers with a friendly interactive interface, and designs and develops multiple modular component applications based on human–computer interaction usability specifications (see Fig. 1).

The system adopts the design concept of matrix organization structure to realize the different requirements of vertical and horizontal leadership management relationship in the organization structure. It improves the problem of insufficient horizontal management ability in traditional linear management mode and quickly configures the horizontal management department. It does not affect the original organizational structure and personnel ownership relationship, and can adjust the appointment of horizontal organizational structure personnel at any time, and configure the

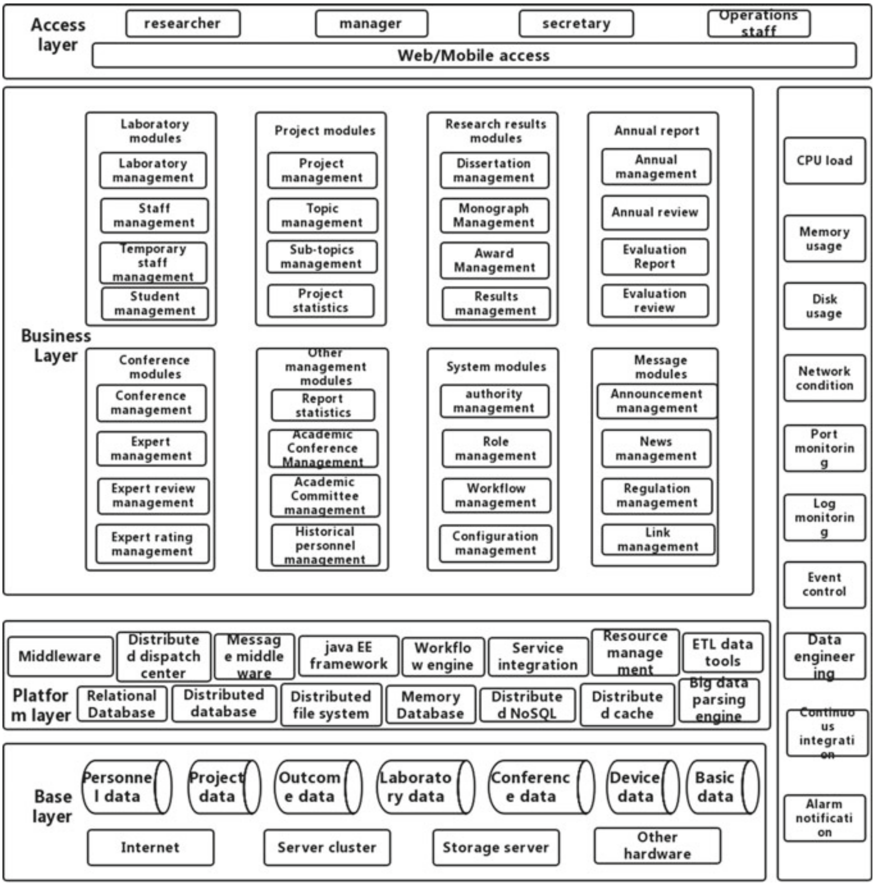


Fig. 1 System architecture

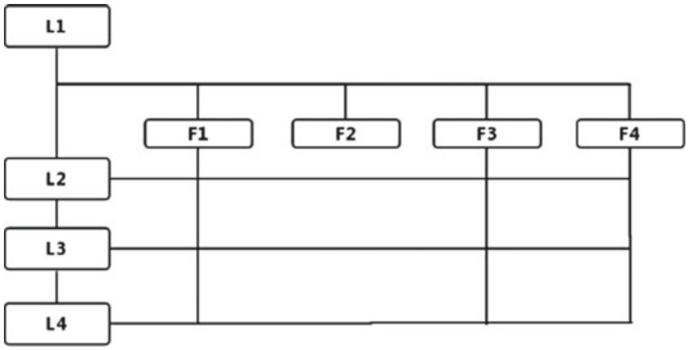


Fig. 2 Matrix organization structure

system role and authority, so that the platform manager's personnel management efficiency can be comprehensively improved. The design diagram of matrix organization structure is shown in Fig. 2.

3.2 Collaborative Management

The CAS-KLMSS platform aims to break the relatively independent management processes and modes, and solve the problem of inefficient exchange and communication of information between laboratories. The analysis data of the laboratory related business would be reorganized and divided according to the theme, to abstract the application-oriented data in the traditional data, and classify the data at a higher level to form a logical relationship between the abstract concept object and the real data object. The object container pool would be established, in which the workflow is used as the engine to realize the information flow and collaboration of abstract objects. When business is processing, specific data objects are obtained from the container pool for business operations. It would get through the information of each organizational structure and daily business processing, reduce intermediate links, and form a unified, standardized, concise, and standard business process. For example, the system can record related logs of personnel information changes and form related archive analysis data; the system also encapsulates various types of data into API services to provide services for annual report module. The annual report module automatically fills related data of the laboratory and personnel through open data, and realizes data sharing and linkage between modules through loosely coupled services.

3.3 Data Warehouse

The database is a topic-oriented, integrated, non-volatile, and time-varying data set that supporting management decisions. The data warehouse is a different type of database whose concept changes the definition of time in a database (not time-varying). Data warehouse is no longer only for storing and retrieving data, nor only for application data, but for both the operating system and the decision support system. Data warehouse make the firm foundation for trusted data and it can bring many benefits. For example, the date can be gotten quickly because the data has been in a waiting state in the data warehouse. For analysis and decision-making, the integration of data is consistent. If a new analysis method is needed, the data warehouse can provide a data foundation; If it is necessary to conduct compliance inspections or audits, there will be credible data foundations to support analysis and so on. The construction of the data warehouse is mainly designed to meet multiple future management goals. SAP's BO analysis tool is used to establish a unified data management and application publishing platform, which can realize thematic databases and shared databases for different businesses of key laboratories. It provides decision-making data support services for management departments, shares data resource of the whole key laboratory system and meets the demand of fast and efficient data sources. The main goal is to transform data into information or form data standards.

Building data warehouse requires a complete set of infrastructures, including extraction/transformation/loading (ETL) technology, data mart, and operational data storage (ODS). It adopts ETL technology to create a data warehouse, integrates many different forms of application data into a single form, and realizes unified definition of data. The system uses data mart to meet the different needs for aggregated data, each different organization in the data mart has its own data perspective, and all data is derived from granular data in the data warehouse. Based on such granular data, different departments can establish different interpretations of the data and maintain coordination with the public data warehouse [7].

The main data content of a data warehouse includes metadata, data organization, data integration and data storage. The platform also established a set of data analysis system in order to better provide decision analysis. As the data of personnel, projects, achievements and awards are multi-source heterogeneous, it is necessary to be cleaned, transformed and aggregated. It takes corresponding measures for missing values, outliers, duplicate values, and useless values of different data objects to obtain the final aggregated data. For example, formulating a strategy based on field importance for missing personal data, the gender, birthday and city are supplemented automatically through the ID card information. In order to ensure the accuracy of the final aggregate data, constraint and association rules are formulated for the abnormal data in the achievement data. Setting classification rules according to priorities for project sources and types, the data would be repeatedly clean to meet the requirements for data aggregation.

3.4 Distributed Database

The supporting units of key laboratories of CAS are distributed all over the country, and the data between laboratories are usually scattered. Each laboratory will naturally maintain data related to its own work, so that the entire information assets of the key laboratory will be divided into information islands. On other hand, the information between laboratories should urgently be flexibly exchanged, shared, managed and used in a unified way with the expansion of the application of the information system. The system has established a distributed database which provides a bridge for linking these small information islands.

The distributed architecture system has high reliability and availability. The reliability is further improved through appropriate redundancy. The failure of one node will not cause the system crash or data loss. However, there are many inherent technical difficulties due to the distributed environment of data. Based on the consideration of some aspects of the real system, 12 rules of the distributed database system need to be weighed and selected [8].

What the model of distributed database management be used in this platform (shown in Fig. 3) is Mycat, an open-source middleware based on Mycat of Alibaba Cobar, to realize distributed management [9]. It is managed through the master–slave switchover of zookeeper and the cluster deployment of Mycat, while the dynamic management of cluster nodes is realized through Mycat balance. The construction of MySQL database is based on binlog master–slave replication, read–write separation, and multi-master simultaneous write operations for key tables in the system. The system reads common data by performing query operations from the database, and some data maintain the database binlog data into MongoDB through the mq message queue. The front-end service implements efficient and fast queries by querying MongoDB.

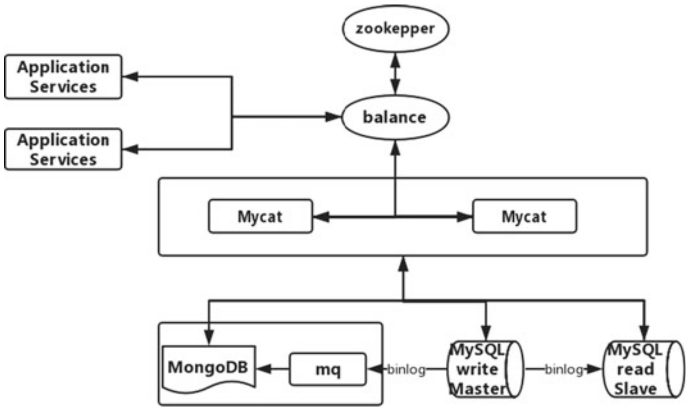


Fig. 3 Distributed database architecture diagram

3.5 Security Mechanism

Data security is very significant in the process of key laboratory information management. Data leakage, tampering and identity legitimacy are the crucial issues to be solved in the process of CAS-KLMSS platform construction. The system adopts SSL encryption transmission protocol https to realize the security of information transmission between the web browser and the server, to make sure all communication between client and server is encrypted. Certificate public key issued by the server through CA certification is used to achieve the transmission and handshake of the symmetric key with the client. Then, the client communicates through the cipher text channel encrypted by the generated symmetric key, so as to solve the problem of trusting the host and the problem of data leakage and tamper resistance during transmission.

4 Application Effectiveness

CAS-KLMSS is committed to serving and promoting the development of the key laboratory system in CAS. It provides information support guarantee for all kinds of key laboratories to explore scientific frontier and further solve major basic and critical scientific problems.

From the perspective of function realization, the current progress and effectiveness of the system are as follows. ① The main functions cover the operation management of key laboratories, the collection and statistics of project, achievement and award, the expert evaluation and review, notification and other modules, which provide a full range of information services for key laboratory system operation. ② The system has achieved unified management and decentralization of authority. The management department is responsible for allocating the account, and opening up the internal use authority based on the overall responsibility of the director of the key laboratory. The secretary of the key laboratory is responsible for maintaining basic information, uploading annual reports, and reviewing data. Laboratory members submit basic data of projects, results, and personnel. ③ The annual report of key laboratories is submitted online, and all relevant laboratory information is stored in the cloud server according to the "field", which can be easily captured, statistic, and analyzable. The main users cover more than 300 key laboratories, and a total of nearly 25,000 new accounts have been created. ④ A complete online evaluation process is established. The system realizes multiple functions such as paperless evaluation, online evaluation and real-time scoring.

From the perspective of system operation, the current progresses are as follows. ① The system strengthens management collaboration, eliminates Information Island, optimizes and improves laboratory information management. It improves the daily work efficiency, and realizes the safe, efficient and controllable sharing of information and data through online collaborative management. Relevant logs are recorded throughout the process, enabling querying and tracing of data. Information sharing association mechanism is formed. ② The system contains a data verification and feedback mechanism during the importing of Excel template, abnormal data would be marked. ③ The platform can switch and analyze different word files through the word file upload function, transform structured data into template data, it is greatly simplified the complexity of filling in and submitting by automatically generating summarized word and PDF files. ④ It adopts a distributed design, which eliminates the bottleneck of the stand-alone mode and forms a read–write separation. Database combining file and memory enhances the efficiency of users' frequent access to a large number of data in the process of evaluation, and avoids the service disruption in the expansion, so that users can get a good experience without perception.

5 Summary

It is an important starting point to construct and optimize the CAS-KLMSS for strengthening the systematic management and coordination of the key laboratory in CAS. Next, the system would closely focus on the overall national layout and the 14th Five-Year Plan of CAS, optimize the operating system and strengthen vigorously R&D on functions such as intelligent matching of technical requirements and achievements to ensure sustainable stability. More novel functions would be developed such as Flexible expansion, smart matching, docking with ARP system of CAS and so on. We are striving to play a demonstration role in the field of construction of scientific research informatization infrastructure, resource sharing and scientific research informatization.

References

1. China's e-Science Blue Book (2017) China Industrial Credit Publishing Group
2. Fan D (2017) Design and implementation of web-based university laboratory management system . Qingdao University, Shandong
3. Liu W (2011) Design and implementation of laboratory information management system. University of Electronic Science and Technology of China, Sichuan
4. Guowuyuan (2020) Opinions on improving and strengthening central financial research projects and fund management [EB/OL]. https://www.gov.cn/zhengce/content/2014-03/12/content_8711.htm
5. Guowuyuan (2020) Plan for deepening the management reform of the central financial science and technology plan [EB/OL]. https://www.gov.cn/zhengce/content/2015-01/12/content_9383.htm
6. Chinese Acadamy of Sciences (2020) The 14th Five-Year Plan of CAS key laboratory system [EB/OL]. https://wenku.baidu.com/view/1dc52249900ef12d2af90242a8956bec0975a5ae.html.

7. Inmon WH (2014) Data architecture: a primer for the data scientist. Morgan Kaufmann, San Francisco
8. Shao P (2005) Distributed database system and its application. Science Press, Beijing
9. Zhou J, Feng Z, Chen S, Zuo Y (2016) Distributed database architecture and enterprise practice—based on Mycat middleware. Publishing house of electronics industry, Beijing

Hongfei Hou, director of Key Laboratory Division, Bureau of Frontier Sciences and Education Chinese Academy of Sciences, senior engineer. He graduated from Department of Geology, Central South University in 1988, and engaged in scientific research at the Institute of Remote Sensing Applications Chinese Academy of Sciences between 1988-1994. Since 1994, he has been engaged in the management of key laboratories at Chinese Academy of Sciences. He has published more than 10 research and management papers.

Achievements of Informatization in Interdisciplinary

Current Situation and Prospect of CERNET

Ying Liu

Abstract China Education and Research Network-CERNET was built in 1994. The Ministry of education is responsible for the management. The Tsinghua University and other top universities are responsible for the construction and operation of the national backbone. In the past 25 years, with the care and support of national ministries and commissions, domestic cooperation units, all sectors of the society and with the efforts of universities and all scientific and technological personnel who undertake the construction and operation tasks, CERNET has been gone through from scratch, developed from small to large and become the largest national academic Internet in the world. It is also an important infrastructure of educational informatization in China and an integral part of national informatization infrastructure. This paper focuses on the main progress of CERNET since 2018 and puts forward the development imaginations for future.

Keywords China education and research network · CERNET · CERNET2

1 Overview

China Education and Research Network-CERNET was built in 1994. It is a national academic computer network supported and constructed by Chinese government; and the Ministry of education is responsible for the management. The Tsinghua University and other universities are responsible for the construction and operation of the national backbone. CERNET is one of the major Internet backbone networks with international export rights in China. After 25 years of efforts, CERNET backbone network connects 41 core nodes in 36 cities of China with the bandwidth of 100G/10G, and 1800 universities in China are connected into CERNET, the number of users reached 20 million. CERNET supports lots of education informatization projects such as online enrollment of college entrance examination, digital library, China national grid, modern distance education, etc., which has developed into

Y. Liu (✉)
Institute for Network Sciences and Cyberspace, Tsinghua University, Beijing, China
e-mail: liuying@cernet.edu.cn

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_19

the largest national academic Internet in the world, an important infrastructure of educational informatization in China and an integral part of national informatization infrastructure.

Since 2003, CERNET has united more than 100 universities to participate in the China Next Generation Internet Demonstration Project-CNGI which has been approved by the State Council and organized jointly by the State Development and Reform Commission and other eight ministries and commissions, building the largest core network of CNGI-CNGI-CERNET2/6IX, making important breakthrough in the key technologies of next generation Internet. CNGI-CERNET2/6IX has become the important infrastructure supporting next generation Internet technology researches, application development and industry evolution. In 2016, the State Development and Reform Commission approved the Internet + major project support project "The IPv6 demonstration network for Education" (CERNET2). On the basis of CNGI-CERNET2/6IX, which has been built and put into operation for more than ten years, a large-scale and high-performance IPv6 next generation Internet demonstration network for education is built. The backbone network has 41 core nodes with 100 Gbps bandwidth, and the number of IPv6 users is more than 10 million. The "Internet +" technology test and application demonstration have been carried out. IPv6 transition technology, real source address verification technology and other tests are in progress to provide experimental verification platform for the national "Internet +" plan, promote China's accelerated development of the IPv6 next generation Internet, enhance the national network space security capability, and plays an exemplary role in supporting the "Internet +" plan and advanced layout of the next generation Internet.

The following parts introduce the development of CERNET from CERNET backbone, CERNET2 backbone, CERNET/CNGI interconnection center.

2 Development Status of CERNET Backbone

Based on the "211 Project" third phrase CERNET construction project, the coverage of CERNET optical fiber transmission network has been extended to 29 provinces/autonomous regions/municipalities. The backbone network bandwidth has been upgraded to 10–100 Gbps, connecting 38 core nodes in 36 cities, of which the interconnection bandwidth of 23 core nodes in 21 cities has reached 100 Gbps, and the total bandwidth has reached more than 3 Tbps. As of December 2019, CERNET has access to 1873 universities with more than 20 million users, and the total traffic of CERNET backbone network reached 1091 Gbps (CERNET backbone topology is shown in Fig. 1).

According to the requirements of users, CERNET has further completed the upgrading of 38 core nodes access equipment and capacity expansion of CERNET bandwidth. Carry out the user experience measurement of CERNET backbone, improve the network access experience of CERNET backbone users, and provide

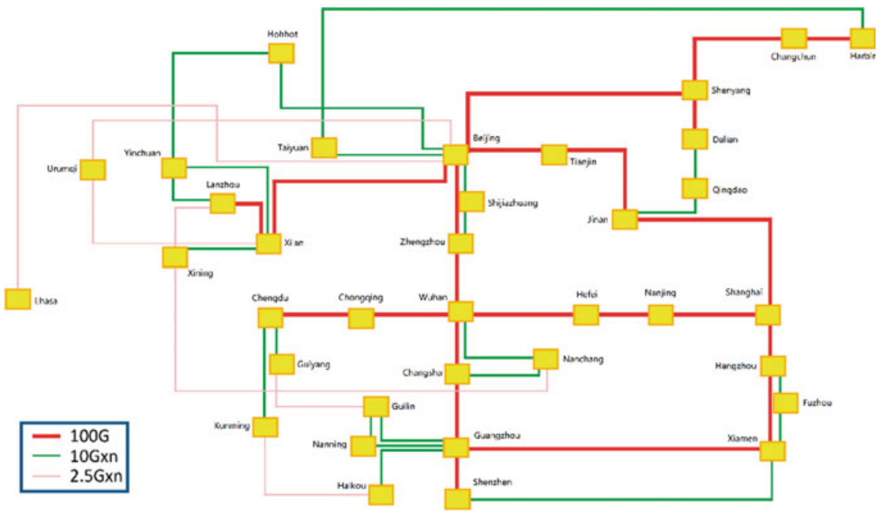


Fig. 1 CERNET backbone topology diagram

testing services for Qingdao National Laboratory for Marine Science and Technology, Wuxi National Supercomputing Center and other major national scientific research facilities. At the same time, CERNET further improves the construction of high-performance network management and security system, improves the security of CERNET users' routing information, and ensures the safe, stable and reliable operation of CERNET backbone network. CERNET has accomplished lots of important network security guarantees for national key network infrastructure.

Since 1997, EDU.CN domain name has been operated and managed by CERNET network center, providing authoritative resolution service of secondary domain name EDU.CN in China. At present, the total number of edu.cn domain names is 6324, and serving root domain names at all levels for more than 400 billion times for one year.

3 Development Status of CERNET2 Backbone

In 2016, the National Development and Reform Commission approved the second batch of "Internet +" major project support project "IPv6 demonstration network for education field" (CERNET2 phase II). The goal of the project is to build a large-scale IPv6 next generation Internet demonstration network in the field of education. Based on the infrastructure of CERNET2, develop "Technology test and application demonstration" to provide experimental verification platform for the implementation of the "Internet +" action plan.

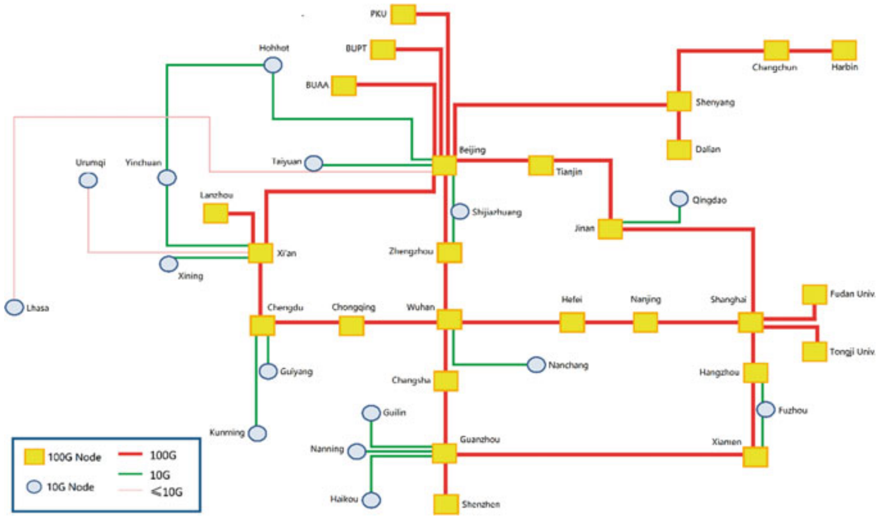


Fig. 2 CERNET2 backbone topology diagram

In 2019, CERNET network center unites 41 universities completed CERNET2 backbone construction by implementing "Internet +" project—"IPv6 demonstration network for education field". The core nodes of CERNET2 backbone network have increased from 25 to 41, covering 36 provinces and cities nationwide. The bandwidth of backbone network reached 100 Gbps, the total bandwidth reaches 3 Tbps, By implementing CERNET/CERNET2 unified access system, all CERNET users can access the CERNET2 backbone network through IPv6.

At the same time, CERNET has accomplished the IPv6 upgrade of EDU.CN domain name resolution system, forming IPv6 support capability of domain name registration, resolution and management. By the end of 2019, EDU.CN has supported 1417 IPv6 domain names in total, with an increase of 663 in 2019.

CERNET2 has been steadily developed and innovated in 2018–2019. With the growing up and rapid development of large-scale online education and big data of scientific research in China, as the solid foundation CERNET will be more effective on the development of China education and research.

The CERNET2 backbone topology diagram is shown in Fig. 2. The traffic changes of CERNET and CERNET2 backbone in 1994–2019 are shown in Fig. 3.

4 Development Status of CERNET Internet Exchange Point

CERNET is responsible for the construction and operation of CERNET Beijing Exchange Point (CERNET-IX), CNGI Beijing Exchange Point (CNGI-6IX) and CERNET Hong Kong Exchange Point (CERNET-HKIX), located in Beijing and

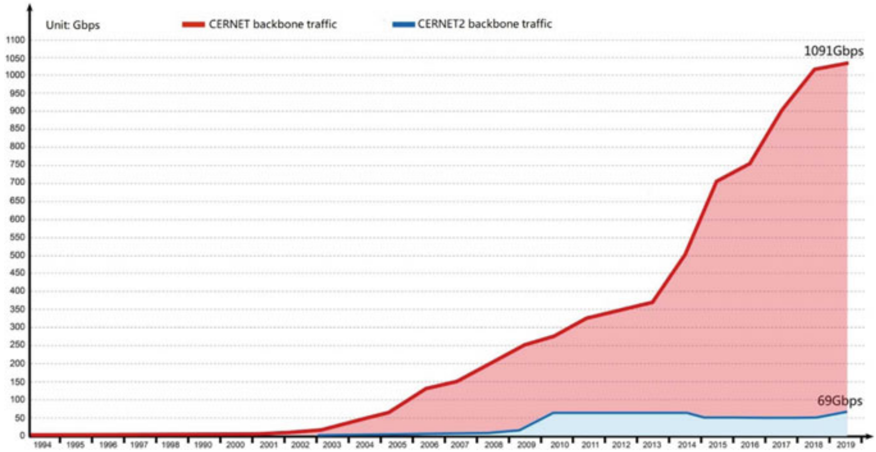


Fig. 3 Traffic changes of CERNET and CERNET2 backbone network from 1994 to 2019

Hong Kong, respectively. High-speed links are made between China Science and Technology Network, China Telecom, China Unicom, China Mobile, other domestic Internet and next generation Internet test networks. High-speed links are also achieved with the international next generation Internet academic networks, including Internet2 in the United States, GEANT2 in Europe and APAN in the Asia–Pacific region.

At present, the Ministry of Industry and Information Technology has set up internet backbone direct connection points in 13 cities, and CERNET has deployed in 11 Internet direct connection points. As of December 2019, the total domestic interconnection bandwidth of CERNET is 331 Gbps.

Since 1995, CERNET has set up its only international export in Beijing. It has been directly connected with Internet2 in the United States, GEANT2 in Europe, APAN and other academic networks in the Asia–Pacific region, providing a dedicated international channel for education and scientific research. As of December 2019, the total international interconnection bandwidth of CERNET is 350 Gbps.

Since 2004, Tsinghua University has won the bid to undertake the work of TEIN Internet backbone management and operation. It provides inter-continental network connection and operation services for the interconnection of national academic networks in Southeast Asia, South Asia and European academic networks, and supports international collaboration in education and scientific research among Eurasian countries. In 2016, Tsinghua University continued to be designated as the direct participant of TEIN project and TEIN's follow-up project—Asi@connect. Tsinghua University is responsible for the operation of TEIN NOC, providing transcontinental network services for the interconnection of more than 20 Asia–Pacific countries NRENs and European NRENs, supporting international collaboration in education and research between Eurasian countries, and organizing the international collaboration project Asi@connect.

5 Support Major Application of Education and Scientific Research

The successful construction of CERNET demonstration project took the lead in providing advanced Internet services for teachers and students of universities in China, cultivating the first group of Internet users in China, supporting the first batch of Internet applications in China, and making significant contributions to the development of higher education and scientific research in China.

Over the past more than twenty years, CERNET has supported and promoted a large number of network application innovation services, completed the basic support system for public network application, including network security service system and video service system; provided digital certificate services for more than 100 application systems for education system; established video service center, HD video conference service platform distributed in 38 core nodes and management system; provided a convenient environment for domestic and international academic exchanges between universities, and become a critical support platform for international collaboration and exchanges. Meanwhile, CERNET has completed the construction and promotion of key discipline information service system, established distributed information service nodes in CERNET Network Center, Beijing, Shanghai and Guangzhou, completed the construction of 54 key discipline information resource systems, and formed a distributed and large-capacity university key discipline information service system covering 11 discipline categories. CERNET has supported a number of national education information projects, including online distance enrollment of college entrance examination, digital library, education and research grid, modern distance education, etc. CERNET has built 10 information resource image systems including top universities, famous international academic organizations around the world and 12 information resource image systems of key disciplines, as well as lots of well-known academic websites in China.

CERNET provides special services for national scientific research. On the basis of existing services, CERNET featured services focus on university informatization construction in 2018, mainly including: ① campus network infrastructure construction services: University website IPv6 upgrading and transformation technical services; ② international bandwidth support services: international dedicated video support services, international reservation bandwidth services,, support for Sino-foreign cooperation in running schools, overseas literature publishers support services; ③ network security service: network security detection service, network security cloud service; ④ characteristic application service: Eduroam international academic network roaming service, CARSI resource sharing service of China University Identity alliance, high-speed data sharing network service; ⑤ IPv6 characteristic service: launch next generation Internet technology innovation project, next generation Internet technology innovation competition, etc.

CERNET actively provides platform and support for national education. CERNET provides an important support for the video conference system of the Ministry of education. In 2018, CERNET cooperated with the Ministry of education to upgrade the video conference terminal system and add 14 new university branch venues. At present, the video conference system of the Ministry of Education is the largest video conference department of government and enterprise organs in China. In the past seven years, CERNET has guaranteed hundreds of video conferences. In 2017, CERNET began to provide college entrance examination information service website security inspection. In 2018, CERNET cooperated with the Ministry of Education to complete the vulnerability scanning of enrollment websites for 1890 universities, notified users to deal with emergency and high-risk vulnerabilities, providing a safe network environment for college entrance examination enrollment.

6 Conclusion and Expectation

2019 is the 25th anniversary of CERNET construction. In the 1990s, CERNET built China's first national Internet backbone. At the beginning of this century, CERNET built the world's largest pure IPv6 next generation Internet backbone. Looking back on CERNET's 25 years of history, the construction of CERNET originates from the needs of national strategy; the development of CERNET grasps the pulse of the times. CERNET has been gone through from scratch, developed from small to large and become an important infrastructure of educational informatization in China and an integral part of national informatization infrastructure.

At the same time, with the overall goal of "national urgent need and first class in the world", CERNET has developed from the Internet technology innovation group of 10 universities to hundreds of universities by undertaking a series of national projects. Over the years, CERNET has established a wide range of communication and collaboration relations with the Industry-University-Research circles at home and abroad, formed a collaborative innovation platform in the field of Internet technology in China, become an important experimental infrastructure for the research of key technologies of Internet and the next generation of Internet in China, as well as an important training base for Internet innovative talents in China.

At present, a new round of scientific and technological revolution and industrial transformation are accelerating the evolution. Artificial intelligence, big data, Internet of things and other new technologies, applications and formats are in the ascendant. The Internet has ushered in more powerful development momentum and broader development space.

Based on the above important background, CERNET will focus on the following aspects:

First, we will continue to play an exemplary role in the national scale deployment of IPv6, actively develop IPv6 users, push all access units to complete IPv6 access, support the "Internet +" action plan and the advanced layout of the next generation Internet, and strive to the development of the IPv6 and the network security of China.

Second, Based on innovation, we will strengthen joint research on core technologies, build future network science and technology infrastructure, strive for breakthroughs in core technologies and standards of the Internet, and make contributions to deepen China's diplomatic discourse in the international competition in cyberspace.

Third, relying on the advantages of universities, strengthen personnel training, vigorously cultivate the needed innovation talents in the field of network and IT, and further improve the level of network space security protection.

Forth, serving the construction and development of "Double First-Class" universities, and constantly improve the quality and ability of network services. Supporting discipline construction and scientific research, and further supporting major national research infrastructure, research facilities, and collaboration in major international research projects.

Ying Liu, Associate Research Fellow in Institute for Network Sciences and Cyberspace, Tsinghua University. She served as Secretary General China Computer Federation Technical Committee. The main research directions are next generation interconnection development planning, network architecture, multicast routing algorithm research, multicast routing protocol design, high performance router architecture. As the project and project leader, she undertook and participated in many key scientific research projects at the provincial and ministerial levels, including 973 projects, 863 projects, projects of the National Natural Science Foundation, projects of the National Science and Technology Basic Conditions Platform, and projects of the National Science and Technology Support Plan, etc.

Implementation of the Informatization Application Scenario for Prevention and Control of Desertification in the China-Mongolia-Russia Economic Corridor of the Belt and Road Initiative—Taking the China-Mongolia Railway (Mongolia Section) as an Example

Juanle Wang, Haishuo Wei, Jia Song, Hanlin Wang, and Kun Bu

Abstract The areas of the China-Mongolia-Russia economic corridor in the Belt and Road Initiative are characterized by a complex natural geography, a fragile ecological environment, and serious desertification. Its impact on the main traffic trunk lines between China, Mongolia, and Russia is not clear, which brings risks and challenges to infrastructure construction. The completion of dynamic monitoring and risk assessment of desertification in areas along the China-Mongolia-Russia economic corridor is urgently required. Faced with the above problems, this study, based on research information and GIS technology, carried out a desertification remote sensing inversion algorithm, a big data application platform, and multi-source data fusion and integration, and established application scenarios for desertification risk control. In this application, based on the extraction effect comparison of different characteristic space models, a desertification information extraction model algorithm suitable for the China-Mongolia-Russia economic corridor is constructed. Using modes of big data batch processing and real-time processing, the desertification information along the corridor was extracted, analyzed, and dynamically monitored. Combined with historical data, the diagnosis and testing of desertification patterns and changes within 200 km of both sides of the China-Mongolia Railway (Mongolian section) from 1990 to 2015 have been completed. This application scenario is oriented to multi-source, multi-scale, and long-time series earth observation satellite data. Based on the big data batch processing technology driven by cloud computing, it realizes high-throughput processing, rapid analysis, and the visual display of long-time series desertification

J. Wang (✉) · H. Wei · J. Song · H. Wang
Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China
e-mail: wangjl@igsnrr.ac.cn

K. Bu
Northeast Institute of Geography and Agroecology, Chinese Academy of Sciences, Changchun, Jilin, China

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_20

dynamic information. It has the ability to monitor desertification dynamically and can provide information and decision-making support for the prevention and control of desertification in key areas of the China-Mongolia-Russia economic corridor.

Keywords The belt and road initiative · China-mongolia-russia economic corridor · Desertification · Risk prevention and control · Informatization · Application scenarios

1 Introduction

1.1 Research Significance

The China-Mongolia Railway is the main cross-border traffic line in the China-Mongolia-Russia region and the core foundation for the construction of the China-Mongolia-Russia economic corridor in the Belt and Road Initiative. The natural geography of the China-Mongolia-Russia economic corridor is complex and diverse, with distinct latitudinal and zonal differentiation and a fragile and sensitive ecological environment. Desertification is particularly serious in the region. In 2017, according to data from the Ministry of Natural Environment and Tourism of Mongolia, more than 76.8% of the land suffered from varying degrees of desertification [1]. With the serious desertification problem increasing in Mongolia, the environmental changes will inevitably affect the China-Mongolia Railway (hereinafter "the Railway"), bringing risks to the construction of transportation infrastructure in the China-Mongolia-Russia economic corridor. Therefore, in order to ensure the security of the cross-border regional ecological channel and promote the construction of an economic corridor, there is an urgent need to build a refined method system for extracting desertification information and accurately grasp the desertification pattern along the Railway. This will provide important support for regional desertification risk control, ecological security, and sustainable social development.

1.2 Research Status

Research into desertification monitoring by remote sensing began in the 1970s [2]. The lack of precipitation and strong evaporation in the area of the railway located in arid and semi-arid regions leads to sparse vegetation growth and a simple group structure. Desertification information is easily confused with other information about weak vegetation cover [3]. It does not have the typical characteristics of healthy vegetation in the spectral line, and it has no obvious strong absorption valley or reflection peak, which makes the vegetation spectral information obtained from remote sensing images extremely weak and even difficult to detect [4]. However, the vegetation index is widely used in the early stage to describe land degradation due to its simplicity

[5–7], but the vegetation coverage in arid and semi-arid areas is relatively low, which means that the soil spectrum interferes with the vegetation index [8, 9]. Therefore, the application of a single vegetation information remote sensing method in arid and semi-arid areas is not ideal [10].

Since the 1990s, remote sensing data sources and index products have become more abundant [11]. Many large-scale desertification monitoring studies have emerged to reflect the characteristics of the long-time series in the Mongolian Plateau and surrounding areas. Liu et al. determined the fractional vegetation cover (FVC), modified soil adjusted vegetation index (MSAVI), albedo, land surface temperature (LST), and temperature vegetation dryness index (TVDI) through inversion of national oceanic and atmospheric administration (NOAA) data and moderate-resolution imaging spectroradiometer (MODIS) data. Furthermore, they provided data on the distribution of desertification land in China and Central Asia from 2001 to 2010 at a spatial resolution of 1 km [12]. Through inversion of MODIS data, Zhuo determined the normalized difference vegetation index (NDVI), MSAVI, FVC, LST, and drought index to build a quantitative desertification remote-sensing monitoring index system and generate a status map of desertification in the Mongolian Plateau for 2006 at a spatial resolution of 1 km [13]. Unulbart et al. extracted information about desertification on the Mongolian Plateau from MODIS data and analyzed the spatial distribution pattern of desertification and its variation trend from 2001 to 2010 [14]. These studies generally rely on the long-time series characteristics of the satellite product data itself and form large-scale data products with coarse resolution from one or more indexes. Most of them seek to master the spatial and temporal characteristics and changing trends of desertification on a macroscopic scale. However, it is difficult to reveal the status and dynamics of desertification to a detailed degree and provide direct and accurate data support for desertification control and regional risk prevention and control.

Since the beginning of the twenty-first century, in order to explore the fine methods of extracting desertification information, some scholars have tried to build a desertification extraction model using multidimensional remote sensing information [15–17]. Zeng et al. used the albedo and NDVI to build the Albedo-NDVI feature space to conduct a study on desertification. They found that multi-dimensional remote sensing data had clear biophysical significance and could reflect the surface coverage, hydrothermal combination, and changes in land desertification [18]. However, due to the considerable influence of the soil background on the NDVI, the vegetation condition cannot be well expressed in areas with sparse vegetation. More and more studies began to try to use different types of vegetation index to extract desertification information. Qi et al. found that MSAVI can eliminate or reduce the impact of soil and vegetation canopy background and is more sensitive to vegetation status [19]. Feng et al. proposed building an Albedo-MSAVI feature space model by replacing NDVI with MSAVI and applying it to the study of soil salinization [20]. In addition, due to the different degrees of desertification, different topsoil textures are produced. More serious desertification corresponds to rougher surface soil particle composition [21]. Therefore, the topsoil grain size index (TGSI) can be used to represent soil surface particle size and as an assessment indicator of land degradation [22]. Through experiments, Lamchin et al. found that the correlation between NDVI and TGSI was weak, while those between NDVI and Albedo and between TGSI and Albedo were strong [23]. This provides the foundation for constructing the Albedo-TGSI feature space model. Based on Landsat 8 data, Wei et al. constructed three feature space models of Albedo-NDVI, Albedo-MSAVI, and Albedo-TGSI in northwest Mongolia, and analyzed their respective mechanism characteristics and applicable conditions [3]. The above studies all reflect the advantages of multi-source feature space modeling in revealing fine desertification information. However, all are only able to obtain the inversion results in the local area and lack the inversion ability to extend to a larger geographical area. One reason lies in the failure to recognize the applicability of the model, and another reason lies in the failure to combine the geographical differentiation of the research area with the applicability of the model.

At the same time, remote sensing data, following the arrival of the era of remote sensing big data, presents the "three high" characteristics (high temporal resolution, high spatial resolution, and high spectral resolution). Desertification information extraction usually involves the processing of remote sensing data of large spatial scale and long-time series, which requires the support of a large amount of remote sensing data, and it is typically a data-intensive task. The precise inversion of desertification based on remote sensing data is currently a frontier research field. It improves the problem of low precision in extracting desertification information seen in past research, which was caused by using a single data source. However, it also inevitably raises new questions. Detailed inversion of desertification requires multiple models and multiple data sources. In the past, single-machine serial computing methods consumed a lot of time for manual intervention. Big data technology can solve the problem of low computing efficiency using data and computing intensive methods.

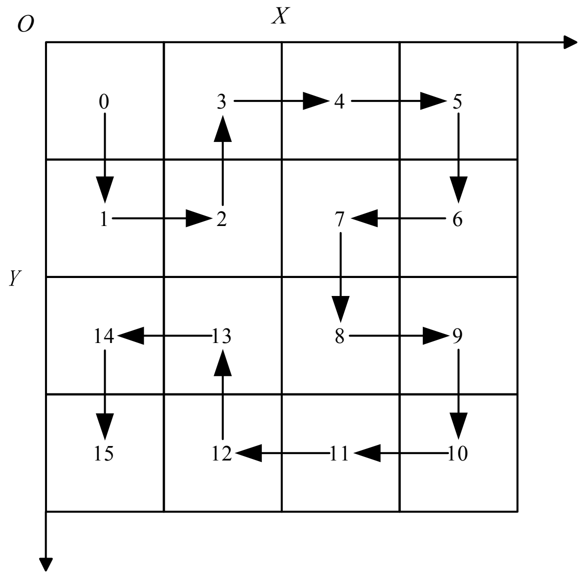
2 Study Area and Data Sources

2.1 Study Area

The Railway starts in Beijing and ends in Ulaanbaatar, extending to the Ulan-Ude Station in Russia, which belongs to Russia's Siberian Railway. The Railway stretches from the south to the north of the Mongolian territory and passes through the following major cities: Zamyn Uud (a frontier port between China and Mongolia), Sainshand (the fourth largest city in Mongolia), Bor Undur (a world-class fluorite production area) [24], Ulaanbaatar (Mongolia's capital and the largest city in Mongolia), Zonhala and Darhan (emerging industrial cities), Erdenet (the center of the copper and molybdenum industries), and Suhbaatar (a frontier port between Mongolia and Russia).

The study area encompasses land within an empirical distance of 200 km on either side of the Railway, as shown in Fig. 1. The area covered 13 provinces, including

Fig. 1 Hilbert space filling curve



Suhbaatar, Dornogovi, Dundgovi, Govisumber, Hentiy, Tov, Ulaanbaatar, Orhon, Bulgan, Arhangay, Selenge, Darhan, and Hovsgol. The land is predominantly high-lying and dominated by a temperate continental climate. Spring and autumn are short, and there is minimal precipitation. Strong winds and rapid weather changes are the biggest features of the region's climate. The southern part of the study area was covered by desert steppe, semi-desert land, and desert land, and the northern part was covered mainly by real steppe and forest. The main plants found in the region are *Achnatherum splendens*, *Populus euphratica*, Mongolia *Imperata cylindrica*, and other similar plants [25]. Along the China-Mongolia Railway was densely populated and economically dominated by the animal husbandry and mining industries.

2.2 Data Sources

The remote sensing data source used in this study is the 30-m spatial resolution Landsat series data downloaded from the United States Geological Survey (USGS) website (source: United States Geological Survey, https://earthexplorer.usgs.gov/). The auxiliary data used include the following: (1) vector map of Mongolian administrative divisions in 2013; (2) vector data of buffer zones within a distance of 200 km on either side of the Railway (source: Human-earth thematic database of the Chinese Academy of Sciences, https://www.data.ac.cn); (3) land cover classification data along the Railway in 2015 [26], Mongolia geographical zoning map, and the online map from Google Earth.

3 Application Scenario Implementation Scheme

3.1 Distributed Remote-Sensing Data Storage

With the continuous explosive growth of remote sensing image data, the traditional centralized remote sensing data management system cannot effectively support the storage and management of remote sensing big data owing to its poor scalability, low fault tolerance rate, and high storage cost. However, the development of big data technology provides an effective solution for distributed storage, sharing, and retrieval of remote sensing data.

The Hadoop Distributed File System (HDFS) is a distributed file storage system that uses a streaming data access mode to store massive data and runs securely and efficiently on a computer cluster [27]. The advantage of the HDFS is that it can shield the differences in computer hardware in the underlying cluster and present the storage capacity of the whole cluster in the form of the whole. Data redundancy backup can be completed automatically, and the operation is simple, efficient, safe, and reliable.

The HDFS uses master/slave architecture. An HDFS cluster is composed of one "Namenode" and a certain number of "Datanodes." As a central server, the "Namenode" manages the namespace of the file system, as well as client access to the files. There is typically one "Datanode" per node in a cluster, and it is responsible for managing the data stored on its node. Within the file system, a file is divided into one or more blocks that are stored on a set of "Datanodes." The "Namenode" performs namespace operations on the file system, such as moving, deleting, or renaming files or directories. It is also responsible for recording the mapping of data blocks to specific storage "Datanode." The "Datanode" is responsible for handling the actual read and write requests from the file system client. Data blocks are created, deleted, and replicated by the "Namenode" in a unified schedule.

The HDFS can reliably store very large files across nodes in a large cluster. For all except the last one, it divides large files into a series of blocks of the same size (128 MB by default). For file system fault tolerance, all data blocks have copies (by default, each block is stored in triplicate), and the block size and copy coefficient of the file can be manually configured. Files in the HDFS are written once, and there is a strict requirement that only one writer can write at any one time. The HDFS can directly store massive remote sensing data sets, and the system can partition and back up remote sensing data without converting the format of the data.

3.2 Remote Sensing Data Preprocessing

In the preprocessing stage of desertification information extraction, radiometric calibration, atmospheric correction, tile cutting, and a cloud mask are needed. A distributed data processing scheme based on the image level is designed for radiometric calibration and atmospheric correction in the preprocessing stage. In order to provide the public with Landsat 8 data products that support the study of surface changes, the USGS launched a set of atmospheric correction software LaSRC (Landsat8 Surface Reflectance Code) for Landsat 8 images in 2016, and its cloud detection algorithm has a high accuracy when examining thicker clouds [28]. However, the complexity of the LaSRC installation environment (which requires more than a dozen dependent packages to be installed in a Linux environment, with dependencies also existing between these packages) makes it difficult to export applications and provide a public service. In this study, a popular Docker container technology is proposed to solve the problems of the difficult construction and poor portability of the LaSRC software environment. Containers and virtual machines have similar resource isolation and allocation advantages, but the architectural approach is different. Container technology is more portable and efficient than virtual machines [29]. Then, based on container technology, the LaSRC program can be deployed at one time and transplanted in many places [30]. Meanwhile, container technology can be combined with cloud computing to realize multi-container parallel computing on a cloud platform.

Apache Spark is currently the most widely used big data computing engine in the market, and it can be applied to massive offline data batch processing, real-time data flow processing, machine learning, graph computing, and other scenarios [31]. Spark's core data abstraction is a resilient distributed database (RDD), a distributed data set that can operate in parallel and has fault tolerance. The calculated aggregation of data is realized through a series of RDD transform and action operations. Its advantage compared with Hadoop MapReduce is that the intermediate data can be stored in memory instead of on disk, thus avoiding multiple data reads and writes, and significantly improving the computing efficiency. Using a distributed computing architecture means breaking the data into pieces, the premise of distributed computing is to slice the data. After atmospheric correction, the remote sensing data are cut into smaller and more structured tile data sets to realize distributed parallel computing of the remote sensing data based on tile.

The cloud mask of Landsat series remote sensing data can be used to judge the cloud using a QA (Quality Assessment) band, and then the image values contained in RDD can be converted to achieve the removal of the cloud information of other bands. Cloud mask and band information extraction are essentially two or more band calculations, and the mathematical operation of pixel values at the same position of the same image can be realized through the transformation operation of RDD.

3.3 Correlation Analysis of Geographical Differentiation Law and Feature Space Modeling

According to the natural and human factors, such as the terrain, climate, hydrology, and population resources of the main provinces along the China-Mongolia Railway

(Mongolia section), combined with the results of research into Mongolian animal husbandry pasture zoning, and taking that 200 mm is a coprecipitation line in Mongolia as the dividing line between arid and semi-arid areas, the main provinces along the China-Mongolia Railway (Mongolia section) are divided into two parts, the south and the north [32–34]. Taking into account the effects of topography and river runoff on local climate, the arid and semi-arid areas in the south and north are further subdivided, and the southern subregion is classified as the southern Gobi region. The northern subregion is classified as the Tov and its northern region and the eastern Mongolian Plateau in the east. The eastern Mongolian Plateau mainly includes the Hentiy and Suhbaatar provinces, with an average elevation of approximately 1000 m, and the topographic fluctuations are the smallest of the three subregions, with annual precipitation of about 300–400 mm. Tov and its northern districts mainly include seven provinces and one city, which are Hovsgol, Arhangay, Bulgan, Selenge, Orhon, Darhan, Tov, and Ulaanbaatar. The population accounts for more than half of the country, with an average altitude of approximately 1500 m and rainfall of approximately 300–400 mm. The southern Gobi area mainly includes five provinces: Dornogovi, Dundgovi, Omnogovi, Ovorhangay, and Govisumber. The average altitude is <1500 m, the terrain is relatively flat, and the annual precipitation in more than half of the areas is below 100 mm.

Combined with the background data of land cover in Mongolia and the vegetation coverage data along the China-Mongolia Railway (Mongolian section), the distribution pattern along the China Mongolia Railway (Mongolia Section) as a whole can be found, from northwest to southeast, to change from forest to real grassland and desert steppe, and then to barren, and has obvious latitudinal progressive change, with the vegetation coverage gradually decreasing. The geographical divisions also show the pattern distribution, which is suitable for the vegetation change and terrain characteristics of the study area. That is to say, the land cover of Tov and its northern region is mainly forest and real grassland, with the highest vegetation coverage. The land cover of the eastern Mongolia Plateau is mainly real grassland and desert steppe, followed by vegetation coverage. The land cover of the southern Gobi region is mainly barren, with the lowest vegetation coverage.

This research team completed the applicability analysis of three feature space models of Albedo-NDVI, Albedo-MSAVI, and Albedo-TGSI in northwestern Mongolia as a test area in 2018. The study found that the Albedo-NDVI model is suitable for areas with high vegetation coverage and large forest ratios. The Albedo-MSAVI model is suitable for areas with relatively low vegetation coverage. The Albedo-TGSI model is suitable for areas with extremely low vegetation coverage and the widely distributed Gobi and barren areas. Therefore, to extract desertification information, the Albedo-NDVI feature space model was used in Tov and its northern region. The Albedo-MSAVI feature space model was used in the eastern Mongolian Plateau, and the Albedo-TGSI feature space model was used in the southern Gobi region.

3.4 Principle of the Distributed Remote Sensing Data Model

The analysis problems of remote sensing big data can be summarized in the following steps. Suppose X is the input remote sensing data, $F(X)$ is the method applied to x, and the result is Y. Ordinary data processing tasks can be expressed as $Y = \mathrm{F}(X)$. Remote sensing big data calculations require data segmentation based on the remote sensing data segmentation algorithm to divide the original data X into N smaller datasets [35], such as $X = \{X_1, X_2, \ldots., X_N\}$. The sliced dataset and calculation program are sent to each computing node through the network, and then each node starts the computing resources to perform the data processing task $F(X_N)$, and, finally, summarizes the calculation results of each node through the network.

Optical remote sensing data are mainly used in desertification research. Their essence can be understood as a multi-dimensional matrix with metadata. The multi-dimensional matrix is sliced and cut into tile data, and the tile data is coded so that remote sensing data can be sliced and merged based on the coding. The big data application program transmits the sliced data and the model algorithm to each computing node through the network, and then each computing node performs the calculation of the model. Finally, the result data are output to the distributed file system or database by reading the encoding.

If the front-end visualization is not involved after slicing the remote sensing data, then there is no need to build an image pyramid, but only the segmentation based on the original resolution image data. The basic idea of image segmentation is to divide the target image data into non-overlapping and equal-sized image blocks from top to bottom and from left to right. The shape of the image block is generally square, and the size of the image block is generally set to 256 pixel $\times$ 256 pixel or 128 pixel $\times$ 128 pixel. As the original images are mostly not regular squares, this results in the inevitable redundancy of some boundary data when using square-cut images. In this experiment, in order to avoid losing information, the boundary tiles are filled by means of completion. The raster is set to no data and filtered directly during the calculation.

The Hilbert curve is a spatial index curve that can fill each slice unit in turn and pass only once, as shown in Fig. 1. Hilbert curve index coding is used to encode the slice unit that runs through, and each unit code is used as the unique index of the unit. Through this encoding method, adjacent slice data can be stored next to each other physically, so that spatially adjacent objects will also be stored adjacent to each other, thereby reducing the input and output time of the image blocks. After an emergency event, remote sensing images of the area where the event is located are usually loaded, and Hilbert curve coding is easier for loading images of adjacent areas. Therefore, the use of Hilbert curve coding can achieve a uniform distribution of massive spatial data, reduce the time of inbound and outbound images, and improve the efficiency of data retrieval.

3.5 Construction of a Feature Space Model and Fine Extraction of Desertification Information

Based on the preprocessed Landsat 8 remote sensing image data inversion, various surface reference variables, such as NDVI, MSAVI, TGSI, and Albedo, along the China-Mongolia Railway (Mongolia section) were obtained.

The NDVI calculation formula is as follows [36]:

$$NDVI = (NIR - RED)/(NIR + RED) \tag{1}$$

In the formula, NIR is the near-infrared band (band 5), and RED is the red band (band 4).

The calculation formula for MSAVI is as follows [19]:

$$MSAVI = (2NIR + 1 - \sqrt{(2NIR + 1)^2 - 8(NIR - RED)})/2 \tag{2}$$

The TGSI calculation is as follows [37]:

$$TGSI = (RED - BLUE)/(RED + BLUE + GREEN) \tag{3}$$

In the formula, BLUE is the blue band (band 2), and GREEN is the green band (band 3).

Albedo' s formula is as follows [38]:

$$\begin{aligned} Albedo = {} & 0.356BLUE + 0.13RED + 0.373NIR \\ & + 0.085SWIR1 + 0.072SWIR2 - 0.0018 \end{aligned} \tag{4}$$

In the formula: SWIR1 is short-wave infrared band 1 (band 6), and SWIR2 is short-wave infrared band 2 (band 7).

As NDVI, MSAVI, and Albedo have a strong negative correlation, TGSI and Albedo have a strong positive correlation. Therefore, by dividing the feature space in the vertical direction that represents the trend of desertification, different desertification areas can be effectively separated. The calculation formula is as follows:

$$DDI_{MAX} = K_{MAX} * NDVI - Albedo \tag{5}$$

$$DDI_{MID} = K_{MID} * MSAVI - Albedo \tag{6}$$

$$DDI_{MIN} = K_{MIN} * TGSI + Albedo \tag{7}$$

Among them, DDI_{MAX} is the desertification grading index of the Albedo-NDVI feature space model of Tov and its northern region; DDI_{MID} is the desertification grading index of the Albedo-MSAVI feature space model of the eastern Mongolian Plateau, and DDI_{MIN} is the Albedo-TGSI feature space model of the southern Gobi region; K_{MAX}, K_{MID}, and K_{MIN} are determined by the slope of the straight line fitted in the feature spaces of the Albedo-NDVI, Albedo-MSAVI, and Albedo-TGSI, respectively.

The feature space is constructed mainly by calculating the linear relationship of the relevant bands, and the desertification level is divided using a clustering algorithm. Spark MLlib machine learning components have provided corresponding interface implementations and only need to implement the corresponding data frame data structure for remote sensing data to directly call methods. The Spark data frame is a distributed data collection, such as RDD. Each piece of data consists of several named fields. Conceptually, it is equivalent to tables in relational databases or data frames in R Language and Python, but at the bottom, the data frame uses more data structure optimization. Compared to RDD, the data frame is more suitable for the analysis of structured data. At the same time, Spark provides many machine learning calculation algorithms for the data frame data model. The data frame can be regarded as a higher-level abstract data structure than RDD, and it is also more suitable for machine learning and other operations on structured data. Tile images are obtained from the remote sensing image data slices and then packaged into a data frame data structure, where: the tile_index field represents the Hilbert index address of the tile; cell_num represents the index position of the pixel in the tile; the band field is each image's meta value. The Hilbert space-filling curve is then used to index the image, and the tile length and width are each set to two pixels to obtain the data frame data of the entire image.

The feature space construction and clustering algorithm based on the data frame must first obtain the index data required by the feature space, such as NDVI, Albedo, MSAVI, and TGSI. These indices need to be normalized. Then, through the regression class provided by org.apache.spark.ml, the slope a of the linear regression of Albedo-NDVI, Albedo-MSAVI, and Albedo-TGSI, respectively, is determined. According to the formula $a \times k = -1$, k can be calculated. Bringing the k value into the exponential expression of the desertification difference can calculate the desertification DDI. Then, based on statistical principles, the DDI is called the K-means (K-means) algorithm based on the pixel value through the clustering algorithm, and the degree of desertification is divided into five categories (non-desertification, low desertification, medium desertification, high desertification, and severe desertification). The K-means (K-means) algorithm is based on the inherent natural grouping in the data. It sets its boundaries at relatively large differences in data values, calculates each classification situation, and automatically selects the classification situation with the smallest variance value. Thus, the difference between the categories is the smallest and the difference between the classes is the largest, and the optimal classification result is obtained.

In order to obtain more accurate desertification data, it is necessary to separate sand information from other desertification information in advance and divide it into a separate category. In general, the reflectivity of sandy land in various bands (except the thermal infrared band) is high, so after adding the reflectance data of multiple bands, the value of the sandy land must be the highest. Therefore, based on this characteristic, the sandy land information can be extracted before the desertification information is extracted. First, the reflectance of the BLUE, GREEN, RED, NIR, SWIR1, and SWIR2 bands were added and summed. Then, the natural discontinuity method is used to divide it into six categories, with the highest grade being sandy land. Finally, the discontinued point classification method is used to classify the synthesized images into six categories, and the highest value category is sand.

In order to verify the accuracy of the desertification data products along the Railway obtained in this study, verification points are uniformly distributed across the study area. Based on high-resolution Google Earth data, true-color Landsat 8 remote sensing image data, and other texts and pictures related to desertification in Mongolia, the verification points are interpreted, and Mongolia's desertification is collected during the same period. The relevant data are compared with the results obtained in this study to complete the accuracy evaluation.

The research is based on the desertification distribution data obtained along the Railway, and it objectively analyzes the pattern of desertification along the Railway, in order to further understand the overall spatial zonal distribution of desertification areas along the Railway, discover the distribution law of desertification in different regions along the Railway, and provide detailed data and method support for desertification prevention and control in the region.

The general technical route of thisstudy is shown in Fig. 2.

4 Applied Scenario Result Analysis

4.1 Desertification Distribution Pattern Along Railways in 2015

Table 1 shows the desertification area and proportion. The non-desertification areas are mainly distributed across large blocks in the northern part of the China-Mongolia Railway. There is specific distribution in the east of Hovsgol, the northeast of Arhangay, most of Bulgan, Orhon, and Selenge, the east of Darhan, northern and central parts of Tov, Ulaanbaatar, the northwest area of Govisumber, and the north and west of Hentiy. The area of the non-desertification regions covers about 189,010.03 km^2, accounting for 45.42% of the total study area. Desertification areas are mainly distributed across the south, central, eastern, and northern regions along the Railway. The area of desertification is about 223,604.92 km^2, accounting for 53.49% of the total study area. Among them, the low desertification areas are mainly distributed across the northern and central parts of the Railway. There is specific distribution in the western border and the southeastern part of Bulgan, the central northern part of

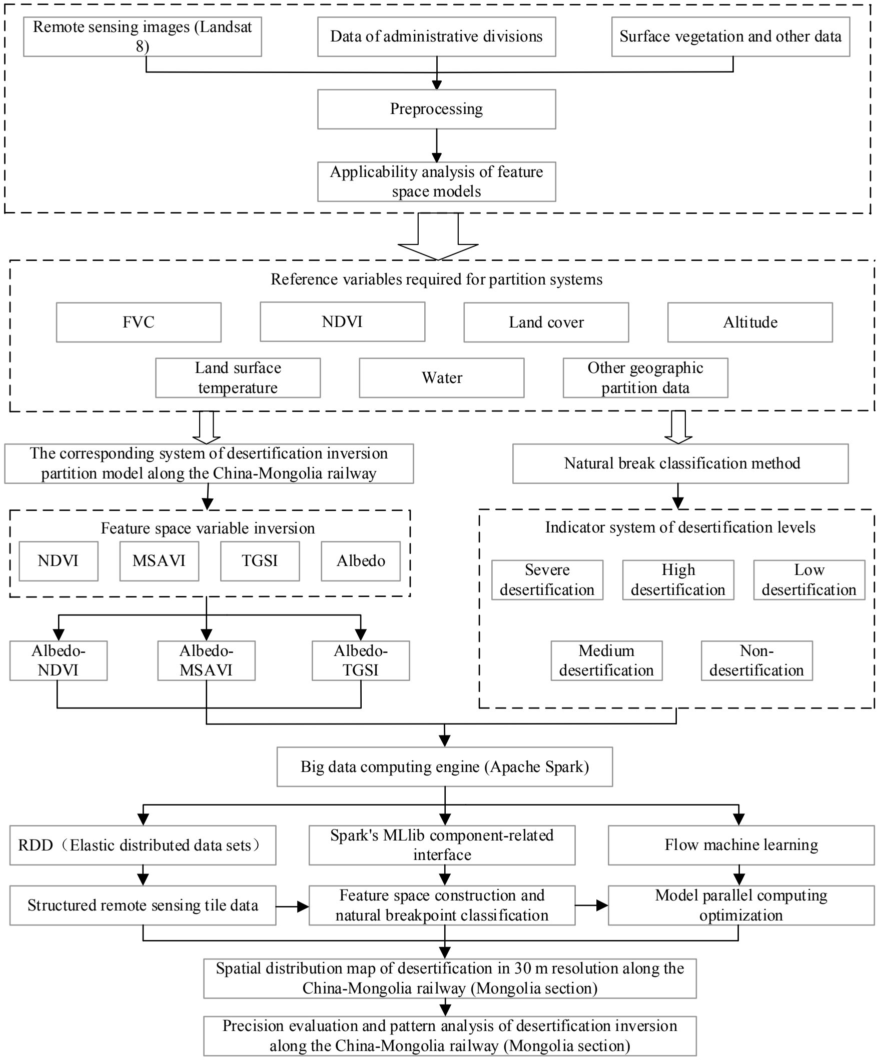


Fig. 2 General Technical Roadmap (NDVI: Normalized Vegetation Index; MSAVI: Improved Soil Regulated Vegetation Index; TGSI: Topsoil Granularity Index; Albedo: Surface Albedo; Jenks: Natural Discontinuity Point Classification; Apache Spark: A fast and general-purpose computing engine designed for large-scale data processing; MLlib: a machine learning algorithm)

Selenge, the western part of Darhan, the central and southern parts of Tov, the southwest and east of Hentiy, the northwest of Suhbaatar, and the northeast of Dundgovi. The low desertification area covers about 50,928.10 km^2, accounting for 12.24% of the total study area. The medium desertification areas are mainly distributed across the central and eastern parts along the Railway. There is specific distribution along

Table 1 Desertification area and proportion along the Railway in 2015

Level	Area (km^2)	Proportion (%)
Non-desertification	189,010.03	45.42
Low desertification	50,928.10	12.24
Medium desertification	58,978.85	14.17
High desertification	106,216.68	25.52
Severe desertification	6481.29	1.56
Sand	4545.05	1.09
Total	416,160.00	100

the southern border of Tov, the northern and central parts of Dundgovi, the south-eastern part of Govisumber, the southeastern part of Hentiy, the southwestern part of Suhbaatar, and scattered areas in the northern and eastern parts of Dornogovi. The area of medium desertification is about 58,978.85 km^2, accounting for 14.17% of the total study area. Areas with high desertification are mainly distributed across south of the Railway. They are specifically distributed across the central and southern parts of Dundgovi, the south border area of Govisumber, most of Dornogovi, and the southwest of Suhbaatar. The high desertification area covers about 106,216.68 km^2, accounting for 25.52% of the total study area. The severe desertification areas are mainly distributed across the central part of Dornogovi in strip shapes, and there is sporadic distribution in the central part of Dundgovi in small blocks. The severe desertification area covers about 6481.29 km^2, accounting for 1.56% of the total study area. Sand is concentrated in the central part of Dornogovi and sporadically distributed across the central part of Dundgovi. Its distribution is concomitant with areas of severe desertification. The area of sand is about 4,545.05 km^2, accounting for 1.09% of the total study area.

4.2 Distribution Pattern of Land Degradation Along the Railway in the Periods 1990–2010 and 1990–2015

Table 2 shows the distribution, areas, and proportions of the newly increased land degradation regions along the Railway. The main distribution is across the southeast of Dornogovi, west of Suhbaatar, south of Hentiy, north of Govisumber, north of Dundgovi, the southern border of Tov, and the western border of Bulgan. Land degradation areas covered 43,970.60 km^2, accounting for 10.57% of the total land area within a distance of 200 km on either side of the Railway. There were three main types of land degradation: non-degraded land degraded into desert steppe, desert steppe degraded into barren, and non-degraded land degraded into barren. Of these degraded land types, non-degraded land degraded into desert steppe covered the largest area of about 20,091.10 km^2, accounting for 45.69% of the total degraded land area. It was mainly distributed across strips in the western border region of

Table 2 The area and proportion of newly increased land degradation areas along the Railway

Time	Degradation	Area (km^2)	Proportion %	Restoration	Area (km^2)	Proportion %
1990–2010	Non → desert steppe	20,091.10	45.69	Desert steppe → non	11,010.90	47.53
	Non → barren	5,304.79	12.06	Barren → non	1,745.03	7.53
	Non → sand	0.06	0.00	Barren → desert steppe	10,232.80	44.17
	Non → desert	18.62	0.04	Sand → desert steppe	105.51	0.46
	Desert steppe → barren	17,771.30	40.42	Sand → non	4.29	0.02
	Desert steppe → desert	49.28	0.11	Desert → barren	54.48	0.24
	Desert steppe → sand	7.85	0.02	Desert → desert steppe	4.28	0.02
	Barren → sand	474.25	1.08	Desert → non	7.27	0.03
	Barren → desert	253.35	0.58			
	Total	43,970.60	1	Total	23,164.56	1
1990–2015	Non → desert steppe	32,574.35	55.84	Desert steppe → non	11,330.70	33.26
	Non → barren	9,541.76	16.36	Barren → non	3,586.15	10.53
	Non → sand	15.37	0.03	Barren → desert steppe	18,710.99	54.92
	Non → desert	10.80	0.01	Sand → desert steppe	355.93	1.04
	Desert steppe → barren	13,862.02	23.76	Sand → non	4.62	0.01
	Desert steppe → desert	38.45	0.07	Desert → barren	12.62	0.04
	Desert steppe → sand	66.12	0.11	Desert → desert steppe	16.12	0.05
	Barren → sand	2,105.81	3.61	Desert → non	52.25	0.15
	Barren → desert	122.58	0.21			
	Total	58,337.26	1	Total	34,069.38	1

Bulgan, the southern border region of Tov, the northern border region of Dundgovi, and the southern part of Hentiy. The desert steppe land shifted to barren and covered the second largest area of about 17,771.30 km^2, accounting for 40.42% of the total degraded land area. It was mainly distributed across the middle part of Dundgovi, the northern part of Govisumber, the southeastern part of Hentiy, the western part of Suhbaatar, and the northeastern and southeastern parts of Dornogovi. Areas of non-degraded land that shifted into barren covered about 5,304.79 km^2, accounting for 12.06% of the total degraded land area. It was sporadically distributed across southeastern Dornogovi, northern Dundgovi, northern Tov, and northern Hentiy. However, an area of land of approximately 23,164.56 km^2 was restored to varying degrees in regions along the Railway and accounted for approximately 5.57% of the total land area. There were three main types of land restoration: desert steppe to non-degraded land, restoration of barren to desert steppe, and restoration of barren to non-degraded land. The restoration from desert steppe to non-degraded land has the largest area, about 11,010.90 km^2, accounting for about 47.53% of the total restored land. It is mainly distributed across the northwest of Suhbaatar, the southeast of Hentiy, and the sporadic areas of western Tov. The area of restoration from barren to non-degraded land is about 1,745.03 km^2, accounting for about 7.53% of the total restored land. It is mainly scattered across central Hentiy and northeast Bulgan.

From 1990 to 2015, the newly degraded land areas were mainly distributed across the southeastern part of Dornogovi, the western part of Suhbaatar, the southern part of Hentiy, and the eastern part of Hentiy, the northern part of Dundgovi, the southern part of Tov, and the western part of Bulgan. Land degradation areas covered 58,337.26 km^2, representing 14.02% of the total land area within a distance of 200 km on either side of the Railway. Of these degraded land types, the largest area was represented by non-degraded land that changed into desert steppe, covering about 58,337.26 km^2 and accounting for 55.84% of the total degraded land area. It was mainly distributed across the southern part of Hentiy, the southern borders of Tov, sporadic regions of northern Tov, and in the western borders of Bulgan. The second largest area was desert steppe degraded into barren, covering about 13,862.02 km^2 and accounting for 23.76% of the total degraded land area. It was mainly distributed across the southeastern and northeastern parts of Dornogovi, the western part of Suhbaatar, and the middle part of Dundgovi. Areas of non-degraded land that changed to barren covered about 9,541.76 km^2 and accounted for approximately 16.36% of the total area. They were sporadically distributed across southern Hentiy, northern Dundgovi, and on the southern borders of Tov. Approximately 34,069.38 km^2 of land was restored to varying degrees, accounting for approximately 8.19% of the total land area. There were three main types of land restoration: restoration of barren to non-degraded land, restoration of desert steppe to non-degraded land, and restoration of barren to desert steppe. The restoration from barren to desert steppe is the largest at about 18,710.99 km^2, accounting for about 54.92% of the total restored land. It is mainly distributed across southwest Suhbaatar, north Dornogovi, and southeast Dundgovi. The area of the restoration from the desert steppe to the non-degraded land was next at about 11,330.70 km^2, accounting for around 33.26% of the total restored land. It is mainly distributed across scattered areas of northwestern Suhbaatar, southeastern Hentiy, and northern Dundgovi, and southwestern Tov. The area of restoration from barren to non-degraded land is approximately 3,586.15km^2, accounting for about 10.53% of the total restored land. It is mainly distributed across northeast Dundgovi and northeast Bulgan.

4.3 Analysis of the Driving Forces Behind Desertification and Land Degradation Along the Railway

(1) Mongolia is located in the hinterland of the Mongolian Plateau and the central region of the Eurasian continent. It has a temperate continental climate with obvious seasonal changes that include great differences in average temperature and less precipitation throughout the year [39]. According to a statistical analysis of meteorological data acquired from 12 meteorological stations along the China-Mongolia Railway [40], air temperature showed a significant fluctuation and a slow rising trend during 2000–2015, and the difference between the maximum and minimum annual average temperature was 3.14 °C (as shown in Fig. 3). Temperature fluctuations create an adverse effect on the normal growth and succession of vegetation, and they can cause a reduction in vegetation coverage and productivity and severe grassland degradation, thereby accelerating the land degradation process. Overall, the area and degree of land degradation on the east side of the Railway are weaker than those on the west side. The reason for this phenomenon is that the east Asian monsoon can only reach the eastern edge of Mongolia, bringing sufficient rain and a relatively appropriate temperature to the eastern region, which is conducive to the growth of vegetation and, to a certain extent, prevents land degradation. Global warming not only causes higher temperatures in the Gobi region in summer, but also warmer and increased rainfall in spring and autumn. In the Gobi region, many areas of dry vegetation in summer can grow in spring and autumn, and seasonal land recovery occurs.

(2) As Mongolia is located in arid and semiarid regions, precipitation has a significant influence on vegetation growth in this region. According to a statistical analysis of meteorological data acquired from 12 meteorological stations along the Railway [40], the average annual precipitation from 2000 to 2010 was 2,494.20 mm, and from 2011 to 2015 was 2,361.42 mm, thereby declining

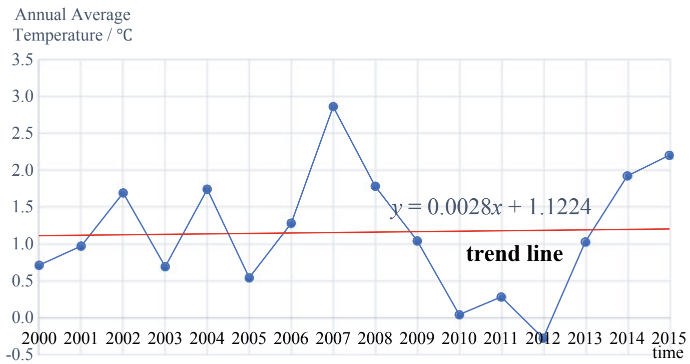


Fig. 3 The annual average temperature along the Railway

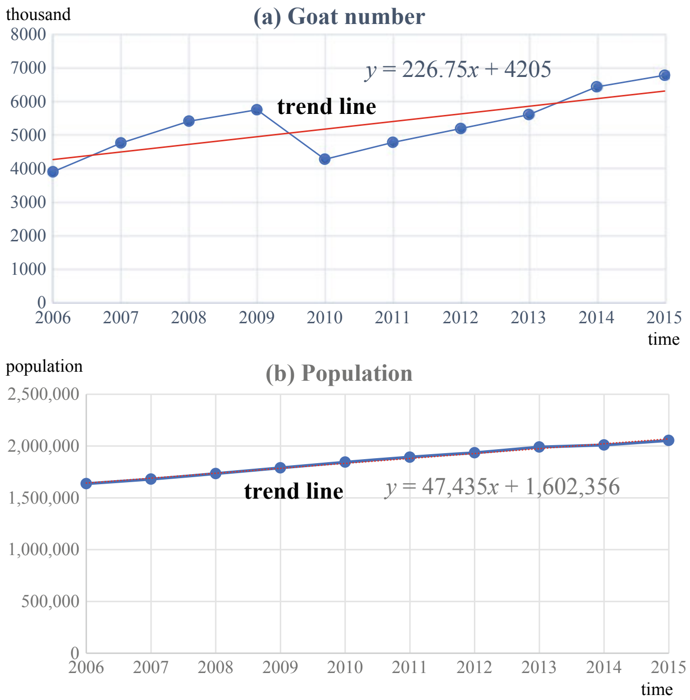


Fig. 4 Annual population and annual number of goats along the Railway

overall. Vegetation growth was suppressed with decreased precipitation. Since the precipitation along the Railway mainly occurs from June to September, the heavy precipitation in summer and autumn will aggravate the soil erosion, and precipitation is scarce in spring and winter, which will further aggravate the drought in spring and winter, inhibit the growth of vegetation, and lead to the acceleration of land degradation. Uneven rainfall also significantly reduced the storage capacity of the southern Gobi region, increasing the risk of flooding during the rainy season.

(3) Overgrazing and population migration have aggravated desertification and land degradation along the Railway. The animal husbandry and mining industries are Mongolia's two mainstay industries. With the increased demand for cashmere at home and abroad of Mongolia, the number of goats raised in Mongolia increased by about 3.3857 million in the 10 years from 2006 to 2015, with an increase rate of 75.70% [40] (as shown in Fig. 4a). Due to the goats' great foraging capacity, they will chew the herbivore root directly during a bad year for the pasture, which causes serious damage to the pasture and directly promotes grassland degradation. At the same time, in order to develop the mining industry, mining enterprises in southwestern Mongolia have also encroached on pasture grounds, resulting in a lack of sufficient grazing space for herders who are forced to

move from the southwest of Mongolia to the northeast through the Railway. As a result, the population in eastern Mongolia has increased. Located along the Railway, Ulaanbaatar, Darhan, and Tov have convenient transportation and rich resources. They are the major political and economic centers of Mongolia, as well as being important destinations or transit stations for the migration of herders. The region's population increased by 397,600, representing an increase of 33.41% in the decade from 2006 to 2015 [40] (as shown in Fig. 4b). Affected by the mining area, the pastures migrating to the east and north also aggravate overgrazing in the areas adjacent to the Railway [41].

(4) Infrastructure has accelerated land degradation. To promote economic development, Mongolia has actively constructed railway and road infrastructure, and there has been an increasing demand for engineering construction soil. To reduce construction costs, many construction engineering companies often take soil directly from adjacent infrastructure, and do not adopt any reclamation measures after the soil is taken. Consequently, the ground surface is bare and severely crushed in these areas, which accelerates the land degradation process. At the same time, the country's roads were mostly built alongside railways. To ensure safe operation of the Railway, the Mongolian government has built a network of culverts every 30 km to 40 km or so, in order to connect the east and west sides of the Railway. As a result, the Railway has become a "wall" dividing the east and west, increasing the cost and difficulty of crossing the Railway to the west, leading to increasing numbers of human activities concentrated on the side closer to the road and convenient for transportation.
(5) Irrational mining and local over-cultivation of farmland have accelerated the process of land degradation. The south of Mongolia is rich in mineral resources and has large reserves of coal, fluorite, tungsten, gold, iron, and tin. There are also several strategic mineral bases, such as Tavan Tolgoi (the world's largest untapped coal mine), Narinsuhai coalfield, and Oyu Tolgoi gold and copper mine (the largest gold and copper deposit in the world) [42]. With the development of the mining industry, the random abandonment of mineral deposits and accumulation of mining abandoned soil increases the source of sand and dust, increases sand and dust flow, and aggravates land degradation. North of Darkhan and Selenge Province are two of Mongolia's most agricultural areas. From 2006 to 2015, the area of agricultural land in these provinces has grown rapidly, with an annual growth rate of approximately 280.12% and an average annual growth rate of 28.01% [40]. The increase in agricultural land area is bound to destroy a large amount of grassland at the cost of accelerating the land degradation process.
(6) Rapid urbanization and existing land use systems increase the risk of land degradation. In the 1990s, Mongolia enacted a "*Civil Liberties*" policy with respect to citizens' residential location choices [39]. It has prompted many herders to move to resource-intensive regions and greatly accelerated the urbanization process along the Railway. However, in the process of urban construction, the intensive use of construction land is insufficient, resulting in the waste of many

land resources, which reduces the local vegetation coverage and increases the risk of land degradation.

In summary, in view of the main driving factors of land degradation along the Railway, it is suggested that Darkhan and Selenge Province make reasonable plans for farmland reclamation, control the growth rate of agricultural land, and return farmland to grassland. The example of Ulaanbaatar and Tov points to reasonable planning and urban construction plans, strengthening the construction of urban land intensive utilization, and giving full play to the guiding role of the government, corresponding preferential policies to encourage people to settle on the west side and accelerate the construction of the west side's railway infrastructure, in order to ease the problems of the eastern population and urban concentration. It is suggested that the mining technology of many mining enterprises in the southern region should be improved, as well as seeking a scientific solution to mining abandoned soil and enhancing awareness of environmental protection. The rational planning of the structure of livestock breeding is suggested to reasonably match the species of livestock breeding and control the growth rate of goat breeding. A further suggestion is that the Mongolian government should rationally plan the land use system and reasonably guide urban residents and rural herders to choose the way the land is used. In the future, it is suggested that the monitoring and prevention of land degradation should focus on regions with a light degree of land degradation and strong resilience, in order to improve the regional capacity of preventing and controlling climate and environmental change and ecological risks, and promote the sustainable development of the China-Mongolia-Russia economic corridor.

5 Conclusions

In this study, based on the Mongolian geographical division, land cover data, and vegetation coverage index, the area along the Railway is divided into three geographical divisions: Tov and its northern region, the eastern Mongolian Plateau region, and the southern Gobi region. According to the applicability relationship between various feature space models and different geographical regions, three geographical divisions of Tov and its northern region, the eastern Mongolian Plateau region, and the southern Gobi region were constructed to build the Albedo-NDVI, Albedo-MSAVI, and Albedo-TGSI feature space model wasteland algorithm. Comprehensive large batch and real-time data processing two kinds of processing mode realized the China-Mongolia-Russia economic corridor along the Railway, communication analysis and dynamic monitoring of desertification information extraction, the completion of the diagnostic test of desertification pattern and change along the Railway during 1990–2015, and then the provision of information support and decision support for the desertification risk prevention and control in the China-Mongolia-Russia economic corridor and the Belt and Road initiative. The practical application shows that, in 2015, about 53.59% of the land along the Railway was affected by desertification to different degrees. The desertification area was mainly distributed across the south, center, the east, and sporadic areas in the north of the region, and the degree of desertification gradually increased from northwest to southeast. In the past 25 years, the newly added land degradation areas along the Railway are mainly in the central part of the country. While desertification is taking place, part of the land is being restored, but the land recovery capacity lags far behind the land degradation process. According to the analysis, natural factors and social and economic factors contribute to the process of desertification and land degradation in the region. Among them, temperature fluctuation and precipitation reduction are the inducing factors. Population migration and overgrazing, infrastructure construction, and rapid urbanization aggravate desertification and land degradation.

References

1. Yueliu C, Hong C (2019) Nearly 80% of land in Mongolia suffers from desertification in different degrees [EB/OL]. https://world.people.com.cn/n1/2017/0617/c1002-29345905.html.
2. Basso F, Bove E, Dumontet S et al (2000) Evaluating environmental sensitivity at the basin scale through the use of geographic information systems and remotely sensed data: An example covering the Agri basin (Southern Italy) . CATENA 40:19–35
3. Haishuo W, Juanle W, Kai C et al (2018) Desertification information extraction based on feature space combinations on the Mongolian Plateau. Remote Sensing 10(10):1614
4. Guli J, Xi C, Bao AM (2009) Coverage extraction and up-scaling of sparse desert vegetation in arid area. Chin J Appl Ecol 20(12):2925–2934
5. Holm A, Cridland S, Roderick M (2003) The use of time-integrated NOAA NDVI data and rainfall to assess landscape degradation in the arid shrubland of Western Australia. Remote Sens Environ 85:145–158
6. Geerken R, Ilaiwi M (2004) Assessment of rangeland degradation and development of a strategy for rehabilitation. Remote Sens Environ 90:490–504
7. Wessels K, Bergh F, Scholes R (2012) Limits to detectability of land degradation by trend analysis of vegetation index data. Remote Sens Environ 125:10–22
8. Kremer R, Running S (1993) Community type differentiation using NOAA/AVHRR data within a sagebrush-steppe ecosystem. Remote Sens Environ 46:311–318
9. Xiaohui Y, Junlong C (2006) Progress on remote sensing-based desertification assessment. World For Res 19(6):11–17
10. Xiangting L, Jie B, Guanglu. L, et al. (2013) Comparison of methods based on MODIS for estimating sparse vegetation fraction across desert in Xinjiang. Arid Land Geography 36(3):502–511
11. Rasmussen K, Fog B, Madsen JE (2001) Desertification in reverse Observations from northern Burkina Faso. Global Environ Change 11:271–282
12. Aixia L, Changyao W, Jing W et al (2007) Method for remote sensing monitoring of desertification based on MODIS and NOAA/AVHRR data. Trans CSAE 23:145–150
13. Zhuo Y (2007) The ration remote sensing method study of desertification of Mongolia plateau based on MODIS data. Master's Thesis, Inner Mongolia Normal University, Huhehaote, China
14. Unurbaatar B, Caoligeer YH (2014) The spatial and temporal changes of desertification in the Mongolian plateau from 2000–2010. In: Proceedings of risk analysis and crisis response in information technology-china disaster prevention association risk analysis professional committee annual meeting, china disaster prevention association risk analysis professional committee, Beijing, China, 1–2

15. Yayun L, Xiuchun Y, Xiaohua Z et al (2009) The application of remote sensing technology to land desertification monitoring. Progress Geography 28(1):55–62
16. Albalawi EK, Kumar L (2013) Using remote sensing technology to detect, model and map desertification: a review. J Food Agric Environ 11:791–797
17. Guo Q, Fu B, Shi P et al (2017) Satellite monitoring the spatial-temporal dynamics of desertification in response to climate change and human activities across the Ordos Plateau, China. Remote Sensing 9:525
18. Yongnian Z, Nanping X, Zhaodong F et al (2006) Albedo-NDVI space and remote sensing synthesis index models for desertification monitoring. Sci Geogr Sin 26:75–81
19. Qi J, Chehbouni A, Huete AR et al (2015) A modified soil adjusted vegetation index. Remote Sens Environ 48:119–126
20. Juan F, Jianli D, Wenyu W (2018) Research on soil salinization in Weikui Oasis based on the characteristic space of Albedo-MSAVI. China Rural Water Hydropower
21. Zhenda Z, Tao W (1990) An analysis on the trend of land dessertification in northern China during the last decade based on examples from some typical areas. Acta Geogr, Sin
22. Liu Q, Liu G, Huang C (2018) Monitoring desertification processes in Mongolian Plateau using MODIS tasseled cap transformation and TGSI time series. J Arid Land 10(1):12–26
23. Lamchin M, Lee WK, Jeon S et al (2017) Correlation between desertification and environmental variables using remote sensing techniques in Hogno Khaan, Mongolia. Sustainability 9:581
24. Yunjie W, Lin Z, Xuelin L et al (2008) Land use change and its driving factors in Mongolia from 1992 to 2005. Chin J Appl Ecol 19:1995–2002
25. Xiuxian Y (2011) Study on the flora of seed plants in the Mongolian Plateau. Ph.D. Thesis, Inner Mongolia Agricultural University, Huhehaote, China
26. Juanle W, Haishuo W, Kai C et al (2019) Spatio-temporal pattern of land degradation along the China-Mongolia railway (Mongolia). Sustainability 11(9):2705
27. Feng W, Baohua L (2010) Modeling and analysis of hadoop distributed file system. Telecommun Sci 26(12):95–99
28. Skakun S, Vermote EF, Roger JC, et al. (2019) Validation of the LaSRC cloud detection algorithm for Landsat 8 images. IEEE J Selected Top Appl Earth Observ Remote Sensing 1–8
29. Xuan F (2018) Research on Hadoop performance optimization based on docker technology. Ph.D. Thesis, Nanjing University of Posts and Telecommunications, Namjing, China
30. Lirao M, Chen L (2017) A cluster scaling method based on Docker container. Comput Appl Softw 34(1):34–39
31. Zaharia M, Xin RS, Wendell P et al (2016) Apache spark: a unified engine for big data processing. Commun ACM 59(11):56–65
32. Qingshan Y, Lisha G, Xiumin L (1994) Mongolian Geography. Northeast Normal University Press, Changchun, pp 17–32
33. Fuqiang W (2010) Study on sustainable development of Mongolian grassland animal husbandry. Inner Mongolia University, Huhehaote, China
34. Yifan L, Juanle W, Junxiang Z (2016) Landscape pattern analysis of Mongolia based on the geographical partitions. Arid Land Geography 39(4)
35. Chi M, Plaza A, Benediktsson JA et al (2016) Big data for remote sensing: Challenges and opportunities. Proc IEEE 104(11):2207–2219
36. Carlson TN, Ripley DA (1997) On the relation between NDVI, fractional vegetation cover, and leaf area index. Remote Sens Environ 62:241–252
37. Xiao J, Shen Y, Tateishi R et al (2006) Development of topsoil grain size index for monitoring desertification in arid land using remote sensing. Remote Sens. 27:2411–2422
38. Liang S, Shuey C J, Russ A L, et al. (2003) Narrowband to broadband conversions of land surface albedo: II. Validation. Remote Sens Environ 84:25–41
39. Buren GW (2011) Research on current status, causes and prospect of desertification in Mongolia. Master's Thesis, Inner Mongolia University, Huhehaote, China
40. National Statistics Office of Mongolia (2018) Mongolian statistical information service [EB/OL]. 29 Oct 2018. www.1212.mn

41. Bo CG (1996) NDWI-A Normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens Environ 58:257–266
42. Tobu NC, David A, Riziley (1997) On the relation between NDVI, fractional vegetation cover, and leaf area index. Remote Sens Environ 62:241–252

Juanle Wang is a professor, doctoral supervisor, and the deputy director of the earth data science and sharing laboratory at the Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences. He is a member of Scientific Committee of World Data System of International Science Council, and also the director of the Renewable Resources and Environment Data Center in WDS, deputy director of the natural resources information system committee of the Chinese Society of Natural Resources, deputy director of the informatization branch of the Chinese Society of Environmental Science, and editorial board member of Data Science Journal and Chinese Science Data. He has presided over projects of the national natural science foundation, special research projects in environmental protection, public welfare, industry, and key projects around national science and technology basic work, etc. He has participated in the construction of the national science and technology basic condition platform, the information science database of the Chinese Academy of Sciences and the big data application demonstration platform for a long time. He has published more than 100 academic papers, five monographs, and two atlases. He won the second prize in the national science and technology progress award in 2014 and the Peixiu excellent map award in 2018. His main research interests include resource and environmental science data sharing, spatial information systems in the Belt and Road Initiative, and disaster risk reduction knowledge service systems.

Establishment and Operational Application of China National Air Quality Forecast and Early Warning Device

Zifa Wang and Jianjun Li

Abstract Air pollution episodes frequently occurs in China, and accurate forecast and early warning of air pollution episodes is crucial for the air pollution control in China. Around this demand, this paper introduces the design and construction of the China National Air Quality Forecast and Early Warning Device. The device was established based on an integrated design of hardware and software. The key technologies of big data fusion, fast emission inversion, multi-model ensemble forecasts, regional heavy pollution episode identification, and an efficient-safe-stable intelligent management platform are constructed. The interactive and rapid display technology of big data based on Web-GIS is developed. An information technology system has been established for the exchange and release of national air quality forecast information and early warning of heavy pollution episodes. The device can provide high-precision forecast for city-scale air quality in 7 days and can forecast the trends of regional air quality in the next 7 to 10 days over Beijing-Tianjin-Hebei region, China and East Asia. Such advanced technologies have greatly improved the performance and accuracy of the air quality routine forecast in China and strongly supported the achievement of the five-year goal of Air Pollution Control Action Plan and environmental diplomacy in China.

Keywords Air quality forecast · Early warning of heavy air pollution · Software and hardware integration · Multi-model ensemble · Big data fusion

1 Introduction

With the acceleration of urbanization process, regional heavy air pollution episodes frequently occur in China and present the characteristics of large-scale and long-lasting. In 2013, the State Council issued the "Air Pollution Prevention and Control

Z. Wang (✉)
LAPC, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China
e-mail: zifawang@mail.iap.ac.cn

J. Li
China National Environmental Monitoring Centre, Beijing, China

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_21

Action Plan", which requires the establishment of an emergency monitoring, forecasting and early warning system for new air quality standards to properly deal with heavy pollution episode. Early warning and emergency control of regional heavy pollution episode need high-precision forecast for the future 3–7 days, trend forecast for the future 7–15 days, and puts forward higher requirements for technical support such as refined simulation of pollution process evolution and rapid response for emission control. At the same time, the main pollutants $PM_{2.5}$ and O_3 are caused by many complex factors with difficulty in forecasting. The establishment of the National Air Quality Forecast and Early Warning Device is crucial for meet the national demand for pollution prevention and control. It involves multi-disciplinary integration and multi-disciplinary technology research and development of atmospheric pollution chemistry and physical mechanisms, pollution emissions and environmental quality monitoring data applications, efficient simulation calculations, etc. Moreover, China faces the technology demand of heavy pollution process forecast and early warning under complex changing conditions and international environmental protection cooperation such as East Asia. The construction of the device needs to constantly adapt to new changes and needs a strong openness and international characteristics.

However, at that time, the large-scale forecasting of provinces and regions in China had not yet been carried out, and there was no National Air Quality Forecast and Early Warning Device. Beijing, Shanghai and Guangzhou which have provided important technical support for major activities such as the Beijing Olympic Games, Shanghai World Expo and Guangzhou Asian Games, have set up urban-scale air quality forecasting systems for the three traditional pollutants of PM_{10}, SO_2 and NO_2. The forecast time limit is only 1–3 days, and it cannot deal with the forecast and early warning of the new standard compound pollutants $PM_{2.5}$ and O_3. In the past decade, while the technology of air pollution emission inventory has become more advanced, the national operational monitoring network implementing the new air quality standards has developed steadily. The progress has been made in advanced air pollution super-research stations and large-scale comprehensive observation experiments for mechanism research. The high-performance computing technology in China, which affects the timeliness of forecast, has also made great progress. The basic conditions of multi-discipline integration and Multi-technology Cooperative Engineering R&D are formed to support the forecast and early warning of regional complex atmospheric heavy pollution process. However, the application of the above-mentioned technologies and interdisciplinary subjects is basically limited to laboratory research.

In order to meet the urgent demand of national air pollution prevention and control, organized and implemented in accordance with the system engineering theory, the National Air Quality Forecast and Early Warning Device is designed and established. It is established bases on an integrated design of hardware and software, multi-model ensemble forecasts and multi-technology integration. Based on this device, this project takes Beijing-Tianjin-Hebei and its surrounding areas as a demonstration, develops the key technologies to early warning of heavy pollution episodes, and constructs a national-regional-provincial-key city early warning system that covers the whole country. It is applied to the national regional early warning and emergency response of heavy pollution episodes, and provides guarantee for major national activities and key operational scientific and technological support for the government's national-level air pollution prevention and control.

2 Main Ideas of This Project

The air pollution prevention and control in China is very important, and it is necessary to build the National Air Quality Forecast and Early Warning Device. In design, it is necessary to develop a large number of interdisciplinary applied technologies, which have the dual nature of engineering and research. After completion, it requires long-term stable operation and can continuously adapt to new changes. At the same time, it needs strong openness and international characteristics. With this as the starting point, the National Air Quality Forecast and Early Warning Device is applied to national air pollution prevention and control management, regional early warning and emergency response of heavy pollution and air quality guarantee of national major activities, providing strategic, basic and forward-looking key operational support for national air pollution prevention and control. Figure 1 shows the overall technical route for device construction and application.

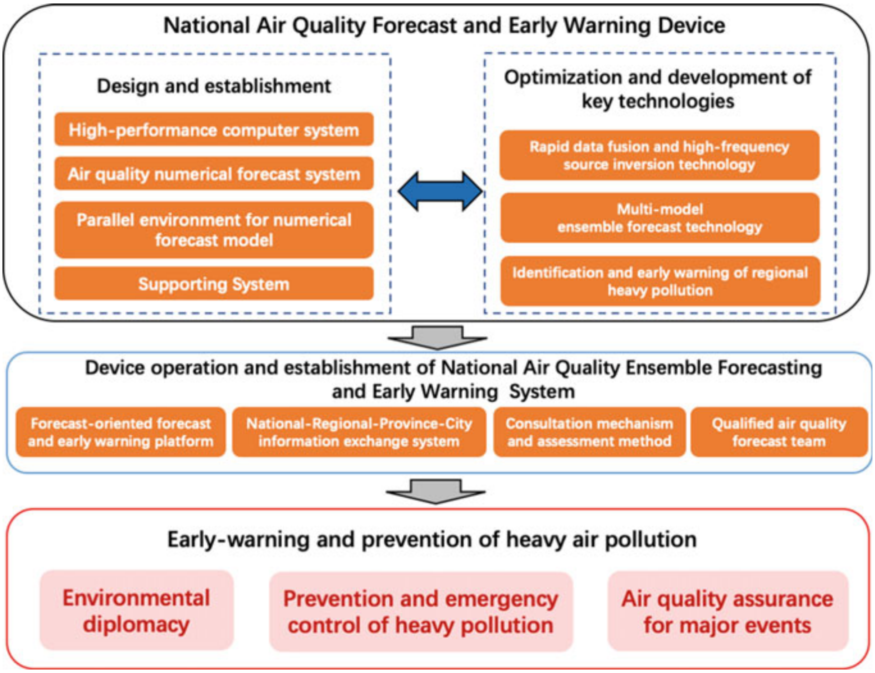


Fig. 1 Overall technical route for device construction and application

The core of the construction of the National Air Quality Forecast and Early Warning Device is to develop advanced forecasting technology, and to integrate the establishment of the operational support system of the national forecasting products. It has the ability to provide high-precision forecast for the future 7-day air quality and trend forecast of the future 10-day air quality, so as to support the emergency response to heavy air pollution episodes in China. As an operational support infrastructure, when designing the National Air Quality Forecast and Early Warning Device, it is necessary to not only effectively support national air pollution prevention and regional heavy pollution episode warning, but also ensure its long-term stable operation with relatively low costs.

(1) Establishing a forecast and early warning device to meet the above requirements.

Under the demands of rapid installation and operation of a single facility, the application of massive observation and emission data as well as forecast and early warning information requires high-performance computer to integrate multi-technology, cross-system platforms, cross-application interfaces, and cross-database. It faces rapid storage and efficient stable computing problems. The design needs to support multi-copy real-time synchronization, distributed lock, cache update, concurrent read–write technology for big-data rapid storage. Adopting the patent technology of high performance computing blade system and high efficiency water cooling system, developing the integrated management technology of the whole system, cooperating to integrate high density computing blade system and high-efficiency water cooling system, the key point is to develop a highly intensive, highly reliable, highly maintainable and highly expandable computing platform.

The forecast, early warning and emergency response for heavy pollution process require the ability with quick response or fast computation of this system. In order to solve the problem and extracting effective product information for completing the forecast and early warning process, the integrated multi-model air quality prediction technology is developed combining with high performance computing and a prediction process-oriented and efficient operating system. The interactive source-acceptor correlation fast visualization technology is developed, which achieves the multi-view stereoscopic display of air quality data. It focuses on the design and development of an intelligent forecast and early warning platform based on an integrated design of hardware and software.

(2) Optimizing the device

Under the changing meteorological conditions and pollutant emission, it is a bottleneck and a key problem to stabilize and improve the forecast accuracy of regional heavy pollution process. The effective optimization method is adopted to screen out the key factors from many factors influencing the simulation. The techniques such as the adaptive fusion application of massive multi-source data, multi-method optimization ensemble forecast, and the early warning and identification of regional heavy pollution episodes are developed. At the same time, the multi-level optimization and high-performance algorithms such as full-module calculation optimization and hybrid parallel calculation of the chemical solver are developed. In order to cope with the changing meteorological conditions and pollutant emission, the forecast accuracy of heavy air pollution process has been continuously improved with new technologies.

(3) Developing key products for routine forecast and early warning

Under the new standard, the routine forecast starts on a zero basis, and the China National Air Quality Forecast and Early Warning Device is used as the core to carry out the forecast and early warning application of the heavy pollution process in Beijing-Tianjin-Hebei. Based on the application experiences of Beijing-Tianjin-Hebei, a technical guide for air quality forecast and early warning has been issued. The training of key national forecasters at provincial and key cities nationwide has focused on the establishment of a nationwide forecasting and early warning technology system and the establishment of a professional forecaster technical team. To develop a national forecast information exchange system, a national forecast information release system and a national regional early warning consultation system based on the devices are established with the products distributed to the air quality forecast departments of key regions and cities. It is a very important issue to solve the establishment of a national routine forecast system that supports regional air pollution forecasting and early warning, and guarantees national major activities. At the same time, scientific research, environmental diplomacy and social services will be established to support the government's efforts in the prevention and control of air pollution and environmental protection.

The core is to achieve an integrated design of hardware and software, multi-mode integration and multi-technology integration, bases on "scientific research leads routine forecast, the needs from routine forecast promotes scientific research", through the "optimization with operation" of the device, and effectively adapt to the new changes in external conditions.

3 Key Components of This Project

3.1 Design and Establishment of National Air Quality Forecast and Early Warning Device

The National Air Quality Forecast and Early Warning Device is designed, including high-performance computer system, ambient air quality numerical forecasting model system, ambient air quality numerical forecasting model system parallel environment system, and support system. The high-performance computer system includes a computing subsystem, a management service subsystem, a storage subsystem, a network subsystem, and a cabinet subsystem. Ambient air quality numerical forecasting model system comprises regional air quality integrated forecasting module, real-time data fusion module, regional pollution source analysis and source-tagged

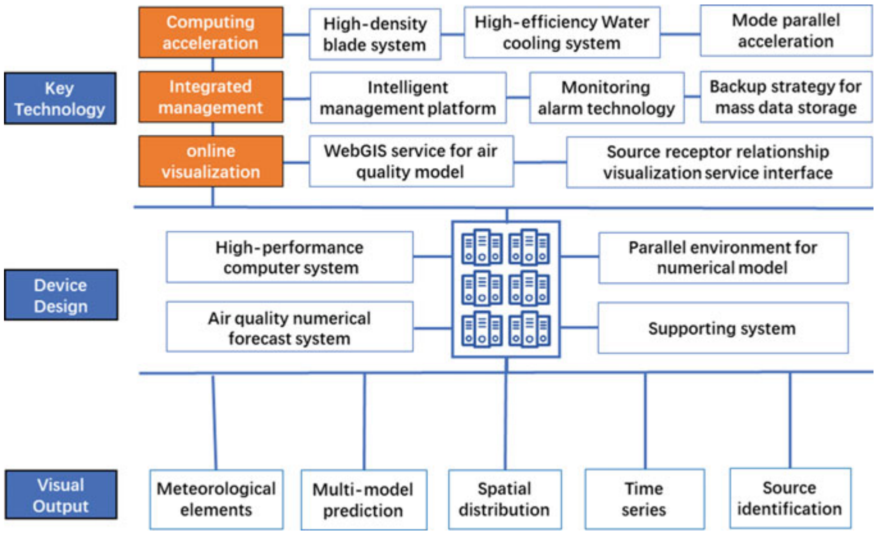


Fig. 2 Key contents of the China national air quality integrated forecast and warning device

module, forecasting and warning software platform. Ambient air quality numerical forecasting model system parallel environment system includes Linux operating system, cluster monitoring software, cluster scheduling software and parallel environment. Supporting system involves ArcGIS platforms, map data, antivirus systems, Web security gateways, and databases. Figure 2 is the project team's design of the key content of the National Air Quality Integrated Forecast and Warning Device.

According to the device design, around the key challenges faced by traditional numerical forecasting, such as mode parallel acceleration, rapid online visualization of massive data, and efficient and stable operation, the following key technical breakthroughs have been achieved.

3.1.1 Advanced Computing Acceleration Technology

In the computer room with extremely limited space, a high-performance computer system based on a high-density, high-performance computing blade system and a high-efficiency water-cooling system is designed to achieve a high-performance computing cluster system for air quality operational forecasting in Beijing, Tianjin and Hebei and its surrounding areas. Its peak computation performance is up to 130 trillion. For the reliability and timeliness of routine forecasting, technologies such as routine forecast and scientific research parallelism, multiple copies of stored data, real-time synchronization, global non-blocking full-line-speed communication, and efficient unified scheduling of software-hardware resources are used. On the premise of ensuring the daily safety of the system, the overall running time of the daily numerical forecasting system has been reduced by one third. It developed a new parallel acceleration technology of air quality model under the new processor architecture. For the gas chemistry module that takes a long time to calculate in the air quality mode, it designs a new framework for chemical kinetics simulation to adapt to the use of single instruction multiple data (SIMD) technology in new processors [1] to achieve fine-grained parallelization through vectorization, construct loop and integral main branches to simultaneously operate multiple spatial points in the model on the vector processing unit [2, 3]. This technology enables the air quality forecasting model to achieve more than two times faster calculations, and realizes a multi-model and multi-nesting rolling forecast across the country (the forecast is updated every 6 h during heavy pollution).

After the integration of the above technologies, the hardware computing efficiency has been improved by more than 30%, and the parallel computing efficiency of the air quality forecasting model has increased by more than 2 times, which has solved the contradiction of limited resources and huge simulation calculations.

3.1.2 Efficient-Safe-Stable Integrated Management System

The construction of a multi-model air quality forecasting system is a complex project, involving dozens of functional modules, thousands of parameters related to physical and chemical processes, and terabytes of massive data. The hardware environment of the entire system is complex, multiple sets of network interaction, frequent reading of large amounts of data, rapid storage of massive data, and high failure rate. The previous manual service with inefficient operation and maintenance management is far from meeting the stability requirements under the forecast and early warning requirements. Based on this, a set of system-wide, workflow-type intelligent management platform has been developed to ensure that daily results are provided on time to provide forecast service products. The technology adopts multi-task concurrent processing, state-flow management, and breakpoint recalculation to ensure the platform's fault-tolerant mechanism, which greatly improves the stability and timeliness of the business. It develops operational forecast monitoring and warning technology, which has the functions of multi-mode parallel monitoring, multi-forecast area monitoring, forecast mode full process sub-module monitoring, forecast mode operation process abnormal alarm. It realizes the real-time automatic monitoring of warning mode and timely handling of the forecast mode. It develops a storage and backup strategy for mass data and products in forecast mode. It exploits automatic storage and backup tools and implement classified archiving and cleaning of historical forecast data and products. According to the characteristics of software and hardware environment, it develops software and hardware collaborative optimization technology, solves the problems of high memory consumption of mode calculation, unsuitability of parallel file system and unreasonable allocation of computing resources, etc., which significantly improve the speed and stability calculation mode. After integrating the above technologies, the failure rate of the mode system is reduced to below 1%.

3.1.3 Fast Online Visualization of Massive Simulation Prediction Data

The air quality model can provide massive three-dimensional gridded concentration and data analysis of pollution sources, but the visualization of massive data has always been a key technical problem in the construction of air quality forecasting systems. In the past, the urban air quality forecast used static customized pictures, which could not provide a full range of multi-view visual analysis. Moreover, because the structure of the source analysis simulation data is very complicated, it is stored according to the standard paradigm of traditional relational databases. The daily storage records are in the millions, which makes it difficult to query and commercialize. It develops a proprietary WebGIS service for air quality model forecast data, realize the direct generation of geographic data that conforms to the network map service specification from the original model data, which can display pollutant data for different models, regions, times, and layer heights. The technology achieves model data parallel visualization and divide data processing into multiple task processes to greatly increase the speed of online forecast data visualization, making the online visualization time of a single pollutant distribution map less than 0.15 s and solve the rapid visualization of forecast data analysis and data storage challenges. It exploits a visual service interface for source-recipient relationship tracking. In the data layer, multi-threaded data processing and storage of the source analysis mode are used. The horizontal and vertical partitions of the database table structure are designed, and the indexing and caching mechanisms are established to improve data retrieval efficiency. At the service layer, it constructs a network service interface that represents a state transition application framework, operate on data resources through network service addresses, and use dynamic compression and serialization technologies to improve access efficiency. In terms of performance, the innovative introduction of a map animation form to show the source-recipient relationship is clear at a glance, which realizes the efficient processing and display of source analysis model prediction big data.

The integration of the above technologies greatly improves the dynamic interaction ability between forecasters and model forecast data and helps forecasters quickly interpret the pollution transportation of the surrounding cities to the research area and provides decision-making basis for targeted haze control.

3.2 Optimizing the Device and Developing Key Technologies for Forecasting and Early Warning of Heavy Pollution Episodes

High-precision air pollution is a worldwide problem, and the sources of uncertainty in forecasting are very complicated. Regional and cross-regional air pollution episodes frequent occur in China with large pollution range, high concentration of fine particles, long duration, long transmission across regions, and complicated influencing factors. The prediction of the occurrence, process and peak concentration level of

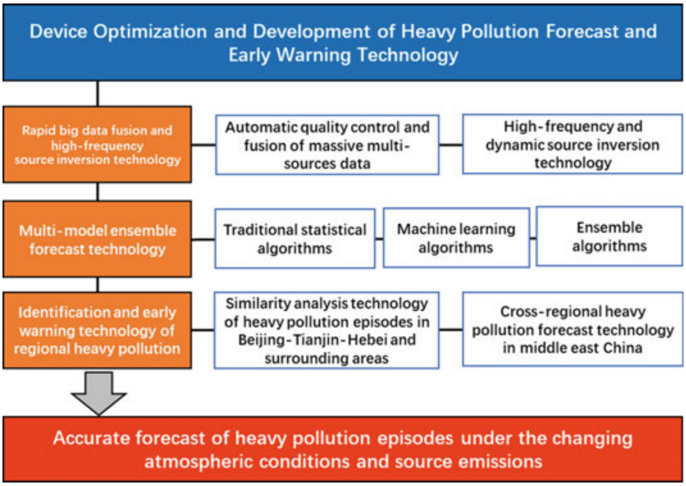


Fig. 3 Technical route of key technology development for device optimization and heavy pollution prediction and early warning

heavy pollution needs to be fast, detailed and accurate, without missing any heavy pollution process while reducing empty reports and false reports, which makes the early warning very difficulty. In addition, due to the urbanization process and various pollution control measures, air pollution emissions have been in a process of rapid change, which has brought serious challenges to the prediction and early warning of air pollution. In response to this problem, research and development of an atmospheric composite pollution prediction and early warning optimization technology system such as rapid large data fusion, high time–frequency sources inversion, multi-model ensemble, and regional heavy pollution episode identification have been developed to capture the changing of meteorological conditions and emissions. It solves the problem of continuously improving the accuracy of atmospheric heavy pollution forecasting under the rapid changing weather conditions and emissions. Figure 3 is the technical route of key technology development for device optimization and heavy pollution prediction and early warning. The key technical breakthroughs achieved by the project team through research and development are as follows.

3.2.1 Rapid Integration of Environmental Big Data and High Time–frequency Emission Inversion Technology

This project developed environmental big data rapid automated quality control and fusion technology to solve the problem of the explosive growth of air pollution monitoring data in China and the difficulty for application [4, 5]. The technology can complete the real-time quality control and rapid assimilation of 6 pollutant observation data from more than 1500 sites of the National Atmospheric Environmental

Monitoring Network within 10 min. The re-analysis data set of the hourly resolution of major pollutants in the country since 2013 is constructed, which significantly improves the accuracy of the initial field concentration of air pollution prediction and early warning, and makes the $PM_{2.5}$ prediction accuracy during heavy pollution improved by more than 20% [6]. According to the problem of rapid changes in atmospheric pollution emissions, to overcome the inversion bottleneck of the current inversion method on the changes of endogenous emissions at the weekly scale, we develop source inversion algorithms under biased assumptions. It breaks through the problem of optimal estimation of high-frequency change mode errors, research and develop high-time frequency emission (1-day resolution) inversion method for heavy pollution forecasting and early warning [7, 8]. It is applied to the inverse estimation of emission reductions of regional pollution source control measures such as emergency emission reductions and reduces the forecast error of major pollutants such as NO_2 during emission reduction by about 40%.

3.2.2 Integrated Forecasting Technology Combining Machine Learning and Traditional Statistical Algorithms

By combining the machine learning algorithm with the traditional statistical algorithm, the adaptive ensemble forecasts function of different time periods, different regions and different pollutants is realized. The prediction results are dynamically optimized. The optimization integration algorithm is based on a multi-model air quality forecast system, which fully integrates forecast data of different air quality models, different nesting areas, and different timeliness, and expands the representativeness of the members of the ensemble, and taps the information advantages of different data [9, 10]. The test results show that the optimization integration algorithm is computationally efficient, and the time for the single-core operation of the ensemble forecast module is less than 20 min, which satisfies the operation requirements of forecast and early warning. Using the optimal integration algorithm, the prediction accuracy is 25% higher than that of the single model.

3.2.3 Early Warning Comprehensive Identification Technology of Regional Heavy Pollution Episodes

In the process of device construction and development, the automatic search and prompt technology of regional heavy pollution episode information was proposed and implemented [11, 12]. This technology uses numerical simulation of pollutant concentration and distribution to realize automatic screening, judgment and search of forecast information. According the classification of regional early warning to search and judge contiguous cities and contiguous ranges, it gives prediction information such as the extent, degree, time and population of the heavy pollution episode. It develops similarity analysis technology of regional heavy pollution episode, identify key factors of air quality forecast and heavy pollution episode forecast. It compares the trigger sensitivity of different models to the heavy pollution episode and the trigger rule of atmospheric diffusion conditions, obtain regional air pollution characteristic rule by correlation analysis and the main influencing factors, and classification and screening to obtain the key guidance products for the prediction of the atmospheric heavy pollution episode required for different scales. It develops expanded forecast technology for the mid-east cross-regional heavy pollution episode [13]. Air pollution in the central and eastern part of China has the characteristics of long-distance transport across regions. For large-scale heavy pollution episode with wide coverage and long duration, on the one hand, effective information for cross-regional forecast is extracted from conventional regional air quality forecast products. On the other hand, cross-regional heavy pollution episode ensemble analysis is carried out from the aspects of heavy pollution characteristics of adjacent regions, heavy pollution transportation, accumulation in the region, and probability of occurrence of cross-regional heavy pollution episode to provide cross-regional forecast products in the central and eastern regions. This provides scientific basis and technical support for cross-regional multi-city collaborative heavy pollution emergency management and control, minimizing the impact of large-scale cross-region heavy pollution.

Integrating the above-mentioned key technologies, a key technology system for regional heavy pollution forecast and early warning in China has been established, and solve the problem of accurate forecast of regional atmospheric heavy pollution episode under the background of constantly changing weather conditions and pollution source emissions.

3.3 Establishment of National Routine Air Quality Ensemble Forecast and Early Warning System

In order to cope with the large-scale regional pollution in China and carry out heavy pollution emergency response timely, it is necessary for the country and the localities to have strong heavy pollution forecast and early warning capabilities. However, the previous work foundation is very weak, and there is a lack of high-level forecasters and standard procedures for forecasting and warning. National and local forecast and early warning information cannot be shared in time. To solve this problem, based on this device, a standardized process and method for forecast and early warning operation was created, and data exchange technology was developed. Breakthroughs in consultation information screening and numerical forecast evaluation technology, and multi-level, multi-sector and multi-purpose consultation system were designed. Figure 4 is the technical route established for the operation of the device and the National Routine Air Quality Integrated Forecast and Early Warning system. The key platforms, systems, consultation mechanisms and evaluation methods are as follows.

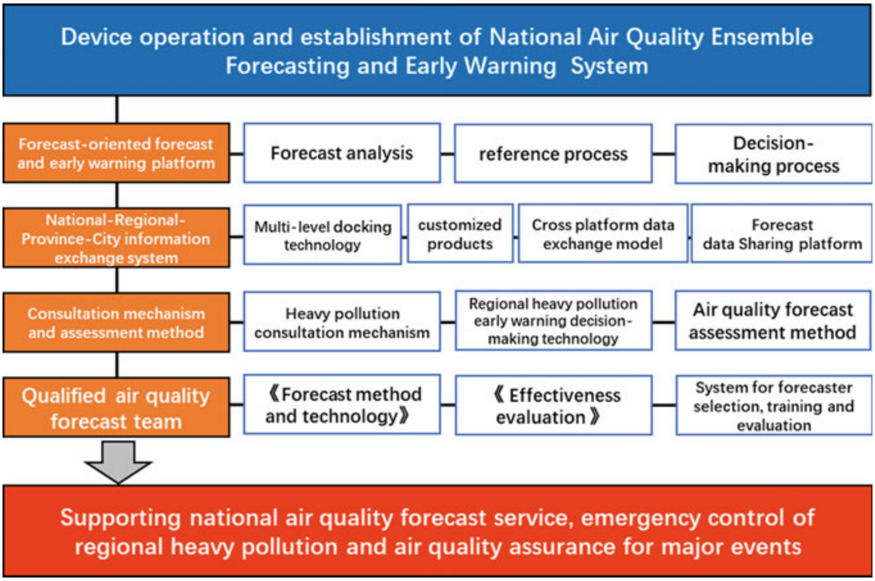


Fig. 4 Technical route established for the operation of the device and the national routine air quality integrated forecast and early warning system

3.3.1 Operational Forecasting Platform Based on Forecasting Process

In the past, the design of the operation platform was mostly from the perspective of software programming, and it was rarely designed according to the forecast method or forecast process, which often caused blind, invalid and repetitive operations. The design of forecast process-oriented, ergonomic and systematic regional ambient air quality forecast operation platform is proposed [14]. The logical distribution of platform functions is completely sorted according to the forecast workflow, through the summary of a large amount of practical experience, and the operation and application of inductive analysis. On the basis of the routine regional forecast work in the existing massive data analysis and processing and the heavy regional air pollution prevention and control technology support needs, a platform functional logic design technology that emphasizes clarity, fluency and meets the requirements of efficient work support has been formed. This guides the forecaster to quickly and orderly complete the forecast analysis process part (model analysis, pollution source analysis, atmospheric condition analysis, real-time air quality analysis and objective correction), reference process part (similar case analysis, AQI probability statistical analysis, magnitude change statistics Analysis, forecast error analysis and ensemble/statistic forecast analysis, etc.) and decision-making process (forecast live record, forecast result preservation, historical review, early warning management, forecast release management, etc.). It avoids disorder, inefficiency and repetitive work caused by massive simulation and data applications.

3.3.2 Nation-Region-Province-City Forecast and Warning Information Interaction System

Taking the air quality forecast technology and the operational application of product exchange as the main objective, we develop multi-scale forecast numerical product applications, network transmission, information security, operational demonstration and assessment analysis, and design an Nation-Region-Province-City integrated system for operational forecast processes. It builds a nationally ensemble air quality forecast data sharing service platform and forms national standards and specifications for information exchange and network transmission information security for forecast products [12]. According to the technical support requirements of forecast operations and environmental quality management, we analyze and customize key parameters such as categories, areas, range of time and space, spans, and resolutions of different levels of forecast guidance products. We develop cross-platform data sharing and exchange models for the individualized needs of regional, provincial and urban forecast platforms, establish information transmission specifications for data at different levels, different sources, different dimensions, and different attributes, and form an integrated solution for the distribution and exchange of forecast product information in a unified national, regional, provincial, and city forecast technology system. Figure 5 is the overall design of the Nation-Region-Province-City forecast and warning information interaction system.

3.3.3 Joint Consultation Mechanism and Effectiveness Evaluation Method for Early Warning of Heavy Pollution Episodes and Guarantee of Major Activities

For heavy pollution episode, it develops joint consultation with the technical research and development area with China's environmental monitoring station and monitoring station at all levels for doing operation forecast warning system construction and management, and set ups industry standards and specifications related to air quality forecast warning [15–17]. It establishes the national and regional air quality forecast operation consultation system including internal consultation, consultation, expert consultation and departments set up heavy pollution forecast and early warning information release and management mechanism, and heavy pollution weather forecast emergency work mechanism, etc. [13]. It contains integrated regional early warning and decision support technology of heavy pollution episode, to cope with regional heavy pollution weather and regional air pollution joint prevention and control requirements, comprehensively uses various forecast methods and products to indicate the source of atmospheric pollution and simulate regional pollutant transportation and the contribution of pollutant emissions from different regions and cities, providing scientific decision-making basis for the early warning, emergency management and control response of the heavy pollution episode of the management department [14]. In the process of device development and application, through the summary of the air quality prediction and early warning practice in Beijing,

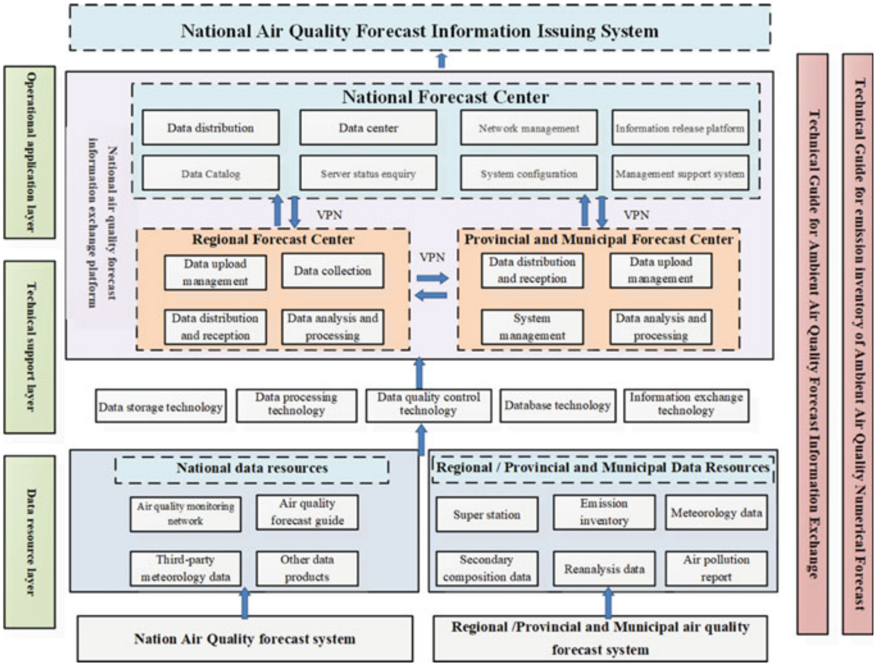


Fig. 5 Overall design of the Nation-Region-Province-City forecast and warning information interaction system

Tianjin, Hebei and surrounding areas, we explored and proposed technical methods for regional air quality and atmospheric heavy pollution episode forecast effectiveness evaluation, and established a rapid assessment method for regional forecast effectiveness based on representative cities and a method for evaluating the effectiveness of air heavy pollution episode forecast based on multi-factor deviation analysis. When conducting routine regional air quality forecast, the target area is first divided into several regions by means of cluster analysis, and then air quality forecast and effectiveness evaluation are carried out separately for each region, focusing on the three major processes of atmospheric heavy pollution evolution assessment of the forecast effect of key elements, including duration, scope of impact and degree of impact.

4 Application of This Device

The device supports routine and special tasks such as national air quality forecast, central and eastern regional and cross-regional atmospheric heavy pollution episode forecast and early warning, international-level major activity guarantee, and East Asia international environmental cooperation [18]. The application are as follows:

4.1 Supporting the Information Release of the National Air Quality Forecast

The construction of the device forms the core foundation of the national operational system, and the construction experience is promoted in provinces and cities nationwide. This forms the relevant technical standard such as regional forecast and early warning methods and technologies, and has constructed the core of national forecast technologies and methods, driving the establishment of a national operational system that includes key cities, provincial and region. The products directly support regional and provincial and key cities to carry out forecast, and provide strong technical support for forecast and early warning and air pollution control for provinces and cities nationwide. This achieved the trend forecast of air quality for the future 7–10 days in 6 key regions (Beijing-Tianjin-Hebei region and surrounding areas, Yangtze River Delta, South China, Southwest, Northwest, Northeast), 31 provincial forecast of the next 5 days and Information about AQI forecast of cities in the next 5 days in municipalities, provincial capital cities, and cities with separate plans. The major products of nationwide forecast information services and technical support include.

The air quality prediction products from the model prediction system for the future 7–15 days, including the distributions of hourly and daily average concentrations of major pollutants, AQI, aerosol extinction coefficient (AOD), and visibility forecast; The time-varying graphs of $PM_{2.5}$ components and their vertical concentrations such as nitrate, sulfate, ammonium, black carbon, primary organic matter, and secondary organic matter at urban air quality monitoring stations; The guidance products for forecasting the various pollutant concentrations, AQI, primary pollutants, and air quality levels at various monitoring stations and cities; The backward trajectory analysis chart of 100, 500, 1000 m relative to the ground height of urban air quality monitoring stations and other designated points. According to early warning rules and conditions, based on Web-GIS technology, the regional pollution episodes over Beijing-Tianjin-Hebei region are dynamically displayed and the relevant early warning information reports and charts are generated including early warning levels, affected cities, the time of occurrence, impacted area, influenced population, and the probability of heavy pollution occurrence. Forecast and early warning operational module includes automatic warning, forecast correction, forecast and early warning consultation, forecast and early warning signing, forecast and early warning information production, etc. National air quality assimilation products from the model assimilation system include hourly and daily average concentration charts of major pollutants which are generated from the reanalysis data set, historical monitoring data, graphic products of predicted concentrations of major pollutants, AQI forecasting effect assessment function, comparison between the optional time-period forecast results of cities and points and the monitoring live, error test, objective scoring, etc. Product charts for the analysis of the main pollutant concentration sources in the future 7 days over Beijing-Tianjin-Hebei and its surrounding monitoring stations contain three categories, (1) the source analysis products of regional sources and industry sources, (2) the destination tracking products such as relevant analysis charts of the impact extent and impact scope of the primary pollution concentration in the future 7 days by the pollution emission sources of different regions and industries over Beijing-Tianjin-Hebei prefecture-level cities and its surrounding provinces, and (3) the mutual transfer and contribution analysis product chart of the primary pollution concentration over Beijing-Tianjin-Hebei prefecture-level cities and the surrounding provinces.

4.2 Supporting Early Warning of Heavy Pollution Episodes and Emergency Response with Emission Control

The release of public service information on early warning of heavy pollution has significantly improved the influence and credibility of national environmental protection decisions. Information disclosure is gradually increasing. For example, 69 warning messages of heavy pollution episode were issued from 2015 to 2018, which provided important guidance for public travel and protection, and won the trust and praise of the public. "Breath and work together" has gradually become the code of conduct for the whole society. Giving interviews with CCTV, Xinhua News Agency, People's Daily, Central People's Broadcasting Station and other media effectively guides the public to pay attention to and participate in the prevention and control of air pollution. Meanwhile, it increased environmental protection publicity and public confidence in the country's effective efforts to curb pollution. The device continues to provide effective heavy pollution forecast and early warning products. It has successfully predicted 69 times of heavy pollution episode of particulate matter over Beijing-Tianjin-Hebei and its surrounding areas between 2015 and 2018. The accuracy of the prediction of the impacting degree of heavy pollution is close to 80%, which provides a cracking tool for the frequent atmospheric pollution problems in autumn and winter. Beijing-Tianjin-Hebei and its surrounding areas have launched heavy-pollution weather warnings. Due to the timely implementation of emission reduction measures, joint actions, and precise response, the impact of the heavy pollution episodes has been greatly reduced, the peak concentration of pollution has dropped, and the duration has been shorten, which strongly supports the "Ambient Air Pollution Prevention Action Plan" requiring that the average annual $PM_{2.5}$ concentration over Beijing-Tianjin-Hebei region must drop by 25%.

Taking the regional heavy pollution process on December 16–22, 2016 as an example, the model system predicted 3–4 days in advance that a large-scale extreme heavy pollution process would occur in Beijing-Tianjin-Hebei and its surrounding areas and the red warning of the entire Beijing-Tianjin-Hebei area were needed to activate in advance to respond effectively. The red warning is generally few activated and has a large impact. And the air quality was still excellent at that time. Given the above, the model prediction result was inevitably questioned by the public and public opinion. But in view of the excellent and stable forecast performance of the model system in the past, the Ministry of Environmental Protection was full of confidence in the forecast results and withstood the pressure of public opinion and insisted on launching the "Red Warning under the Blue Sky" in the entire Beijing-Tianjin-Hebei region. Facts had proved that the model system successfully predicted the accumulation and removal trend of this heavy pollution episode, and the prediction results of the pollution range and degree were in good agreement with the actual measurement, which directly supported the early and effective response of this large-scale extreme heavy pollution episode. In addition, it greatly improved the credibility of the Ministry of Environmental Protection's emergency control of heavy pollution and air pollution treatment among the public.

4.3 Supporting Air Quality Assurance for Major National Activities, Promoting International Cooperation, Economic Development and Environmental Protection

This device provides crucial technical support for successful completion of environmental quality assurance for major events such as the Parade on Anti-Japanese War Victory Memorial Day in 2015, Zhengzhou SCO Summit, Hangzhou G20 Summit in 2016, the Belt and Road Forum in 2017, Xiamen BRICS Summit, Qingdao SCO Summit in 2018, and Shanghai Import Expo. The achievements of this device have played a significant role in the consultation of air quality assurance forecasting in many large international conferences, major events and important commemorative events in recent years. Through on-site participation or remote service, it provides large-scale air quality forecast products of the central and eastern regions for undertaking city forecast consultations, supplies crucial forecast information references from multiple models and multiple angles, offers scientific basis for potential pollution control, and effectively promotes the successful accomplishment of the air quality assurance goals of various major events.

The device has strong openness and international characteristics, and was selected as the "Best Practice" award of the first Digital China Construction Summit. It was also selected for the large-scale achievement exhibition of "Five Years of Endeavor" as a major achievement in environmental protection. And it has achieved the release of a national-level air quality forecast information platform and provided timely and efficient regional forecast and early warning information services for the heavy pollution process through CCTV and other mainstream media and network channels, effectively enhancing the national environmental impact and credibility. As a technical exchange material, it has conducted exchange discussions with relevant national institutions in the United States, Britain, Belgium, France, Germany, Japan, South Korea, Australia and other countries, and has been widely recognized by foreign parties; It has played a considerable role in receiving visits of the environmental ministers or experts from Israel, South Korea, the United States, the European Union, etc., as well as in the "Belt and Road" and the training of South-South cooperation countries; Based on the device products, it has provided analysis reports many times for cross-border pollution transmission between China and South Korea, and offered technical basis for relevant news and public opinion. The achievements and applications of this device have effectively promoted more extensive and deeper technical exchanges and cooperation between China and other countries in the field of air quality forecasting, and continuously improved China's international image and status.

The device has developed a number of crucial technologies with international advanced levels, such as optimization of core algorithms at the internal module level of the air quality model, visualization of multiple pollutant source-receptor relationships, the operational assimilation of air pollution data, adaptive inversion technology of air pollution sources and the commercialized air quality integrated forecasting technology. All of them has reached the international advanced level, which has promoted the progress of international air quality operational forecasting technology. It has supported scientific research and results application of the environmental protection public welfare projects, national science and technology supporting projects, and strategic leading science and technology projects of the Chinese Academy of Sciences such as "Multi-scale air quality forecaster docking technology and operational demonstration research", "Research on the Influence of the Atmospheric Boundary Layer Process of Beijing-Tianjin-Hebei City on the Formation of Heavy Pollution", "Research on the Emergency Plan for the Heavy Air Pollution Process in Beijing-Tianjin-Hebei Region", "Beijing-Tianjin-Hebei Environmental Air Quality Monitoring Forecasting and Prevention and Control Technology Research and Demonstration", "Special Numerical Model and Collaborative Control Scheme for Atmospheric Haze Tracking and Control", etc.

5 Summary and Outlook

At present, the problems of poor atmospheric quality, heavy ecological damage, and many hidden environmental hazards in China are still very serious, affecting and damaging the health of the people, which is not conducive to sustained economic and social development. Through the continuous development and improvement of the simulation, prediction and early warning device platform, the results and products of the device can be applied to the prediction and assessment of the impact of air pollutants on environmental risks, health, ecology, biodiversity, surface water, and oceans. By utilizing the products of this device such as the source emissions and the generation, transmission and deposition of the atmospheric pollutants, this device can be extended to the environmental risk and environmental health simulation.

In future, it is necessary for this device to further integrate technologies such as big data assimilation, heterogeneous accelerated computing and artificial intelligence learning. And it is needed to couple models, such as the environmental risk assessment model of the diffusion, transmission, and deposition of pollutants for water bodies, plants and soil, and the assessment model of the pollution control technology. Also it should develop a new generation adaptive intelligent system for the atmospheric environment modeling, prediction and control, which integrates advanced information technology and numerical modeling theory. Eventually, it will need to realize high-precison full-process simulation, prediction and assessment of regional atmospheric components and toxic and hazardous pollutants.

References

1. Feng F, Wang Z, Li J et al (2015) A nonnegativity preserved efficient algorithm for atmospheric chemical kinetic equations. Appl Math Comput 271:519–531
2. Wang H, Chen H, Wu Q, et al (2017) GNAQPMS v1.1: accelerating the global nested air quality prediction modeling system (GNAQPMS) on Intel Xeon Phi processors. Geosci Model Dev 10:2891–2904
3. Wang H, Lin J, Wu Q, et al (2019) MP CBM-Z V1. 0: design for a new carbon bond mechanism Z (CBM-Z) gas-phase chemical mechanism architecture for next-generation processors. Geosci Model Dev 12(2):749–764
4. Wu H, Tang X, Wang Z et al (2018) Probabilistic automatic outlier detection for surface air quality measurements from the China national environmental monitoring network. Adv Atmos Sci 35:1522–1532
5. Huang S, Tang X, Wang Z, et al, Evaluating the PM2.5 pollution over Beijing−Hebei−Tianjin region based on observations, simulations, and data assimilation results. Climat Environ Res (in Chinese) 21(6):700−710
6. Zheng H, Liu J, Tang X, et al (2018) Improvement of the real-time $PM_{2.5}$ forecast over the Beijing-Tianjin-Hebei region using an optimal interpolation data assimilation method. Aerosol Air Quality Res 8:1305–1316
7. Feng F, Wang Z, Tang X (2016) An adaptive algorithm for inversion of atmospheric pollution sources based on shooting method. Atmos Sci (in Chinese) 40(4):719−729
8. Kong L, Tang X, Zhu J, et al, Improved inversion of monthly ammonia emissions in China based on the Chinese ammonia monitoring network and ensemble Kalman filter. Environ Sci Technol 53:12529−12538
9. Wu J, Xiao L, Yan P et al (2017) Application of optimal consensus forecast in urban ozone prediction. J Environ Sci (in Chinese) 33(4):213–220
10. Huang S, Tang X, Xu W, et al, Improving Beijing PM_{10} forecast using multi-model ensemble and multiple linear regression. J Environ Sci (in Chinese) 35(1):56–64
11. Wang X, Liu B, Li J, et al (2015) General methods and basic principles of regional environmental air quality forecasting. China Environ Monitor (in Chinese) 31(1):134–138
12. Wang X, Chen J, LI J, et al (2015) A preliminary study on the forecast method of the range of urban environmental air quality index. China Environ Monitor (in Chinese) 31(6):139–142
13. Wang X, Zhao Y, Huo X, et al (2016) Affecting factors and improvement methods of ambient air quality forecast based on numerical model. China Environ Monitor (in Chinese) 32(5):1–7
14. Liu B, Wang X, Li J, et al (2016) Design of a forecasting method-oriented system operation platform for systematic air quality forecasting. China Environ Monitor (in Chinese) 32(2):1–10
15. China National Environmental Monitoring Centre (2014) Technical guidelines for forecasting and warning methods of ambient air quality, 1st edn. China Environmental Science Press, Beijing
16. China National Environmental Monitoring Centre (2017) Technical guidelines for forecasting and warning methods of ambient air quality, 2nd edn. China Environmental Science Press, Beijing
17. China National Environmental Monitoring Centre (2018) Technical guideline for the assessment method of environmental air quality forecast effectiveness. China Environmental Publishing Group, Beijing
18. Bai Q, Li J (2017) The role and enlightenment of environmental monitoring and early warning in the response to heavy pollution weather. Environ Protect (in Chinese) 8:45–48

Zifa Wang is the Director of State Key Laboratory of Atmospheric Boundary Layer Physics and Atmospheric Chemistry (LAPC), Institute of Atmospheric Physics, Chinese Academy of Sciences. For over 20 years, He has been focusing on the study of atmospheric environment and modeling. He developed a nested air quality prediction modeling system (NAQPMS) for air pollution study, such as Asian dust storms and air quality forecast. This model has been widely used in several cities including Beijing, Shanghai, Zhengzhou, Shenzhen, Shenyang, and etc., for real time forecasting of the air quality. This model was also involved in a multiple model inter-comparison project (MICS-Asia) to investigate the transport and deposition mechanisms of atmospheric pollutants including ozone, and the Asian dust particles. He have operated air-quality modeling projects across China and my group has conducted air quality forecasting in many international field campaigns, such as CAREBeijing, PRD, LTP, etc. He was in charge of the air quality modeling and forecast for the Beijing Olympics in 2008. He developed the Ensemble air quality forecast Modeling System (EMS), which has been successfully applied to air quality forecast during the Beijing Olympics 2008, the Shanghai World Expo 2010, and the Asian Games 2010, the Beijing-APEC CHINA 2014, the commemoration of the 70th anniversary of the end of World War II 2015. He is author and co-author of 300+ SCI papers with a total citation of 9300+. My H-index is 52. He has been honored by "2011 China CAS Outstanding Science and Technology Achievement Prize of the Chinese Academy of Sciences" (Principal Contributor) and "Chinese National Prize for Progress in Science and Technology" in 2017.

Practice on Big Data Platform for Subject Knowledge Discovery in Stem Cell

Zhiqiang Zhang, Zhengyin Hu, Ning Yang, Yi Wen, Xiaochu Qin, Yibing Song, and Guangjin Pan

Abstract The project of "E-Science Application for Knowledge Discovery in Stem Cells" is an important part of the 13th Five-Year Plan e-Science Programme of Chinese Academy of Sciences. Focusing on the critical needs of science and technology (S&T) innovation and subject knowledge discovery in stem cell research for e-Science, the project makes comprehensive use of big data and new generation of artificial intelligence technologies to develop a platform of "Stem Cell Subject Knowledge Discovery (SCKD)", which can provide professional, efficient and accurate subject knowledge discovery services for the research and S&T management activities in stem cell. This paper introduces the project's objectives, tasks, key technologies, achievements and service effect, and discusses the future work of SCKD.

Keywords e-Science · Subject knowledge discovery · Big data platform · Stem cell

1 Introduction

Stem cells technology is a focus and frontier of life science at present. Regenerative medicine taking stem cell technology as the core is expected to become the third disease treatment approach after medicine therapy and surgical therapy, which is developing the major scientific breakthroughs and huge industrial drive [1]. In stem cells, scientific research big data including S&T literatures, scientific data, clinical trials, drugs and S&T service resources,etc. is increasing rapidly, and the scientific research is becoming a data-driven knowledge discovery activities accordingly. Big data-driven knowledge discovery and technology breakthroughs are turning into new engines of S&T innovation, thus subject knowledge discovery platform which

Z. Zhang (✉) · Z. Hu · N. Yang · Y. Wen
Chengdu Library and Information Center, Chinese Academy of Sciences, Chengdu, Sichuan, China
e-mail: zhangzq@clas.ac.cn

X. Qin · Y. Song · G. Pan
Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, Guangdong, China

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_22

integrate subject scientific research big data and knowledge computing has become an essential tool for scientific research and plays an important role in e-Science. Scientific research big data in stem cell has the characteristics of huge quantity, various types, complex relationships and wide variety of sources, etc. The technologies of subject knowledge discovery include automated knowledge extraction, structured knowledge organization, semantic linkage and knowledge computing, etc. It is key how to efficiently and accurately mine valuable knowledge from the massive multi-source heterogeneous scientific research big data. At present, there are some similar subject knowledge discovery platform such as Stem Cell Commons[1] which is powered by Harvard Stem Cell Institute, Galaxy[2] which is an open source, web-based platform for data intensive biomedical research, etc., which mainly provide subject research data management, computing and analysis services. They are designed to only handle the subject scientific data, which are unfit for the subject knowledge discovery scenarios such as research focus, frontier and trend analysis, as well as key technology mining and technology foresight based on big data knowledge computing.

Aimed at the new situation and challenges of e-Science application in the era of big data and oriented to the demand of subject knowledge discovery in stem cell, the SCKD was developed by combining some text mining and new generation artificial intelligence technologies such as information extraction, natural language processing, ontology, knowledge fusion, machine learning, knowledge graph, knowledge computing and visual analysis, etc., which are based the S&T literature and data resources of Chengdu Library and Information Center, Chinese Academy of Sciences (CLIC) and stem cell research resources of Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences (GIBH). The SCKD implements the stem cell multi-source heterogeneous data fusion and knowledge linkage, effectively breaks down data silos and supports professional knowledge computing services, which can provide "comprehensive, professional, accurate and efficient" subject knowledge discovery services for the research and S&T management activities in stem cell of GIBH and other relevant research institutions of Chinese Academy of Sciences. The SCKD will advance the application of data-driven subject knowledge discovery, improve the fusion of scientific research activities and e-Science, and promote the application level of e-Science in stem cell.

[1] https://hsci.stemcellcommons.org/.

[2] https://usegalaxy.org/.

2 Background of SCKD

2.1 Developmental Needs of Stem Cell and Regenerative Medicine

The latest "2016 National Economic and Social Development Statistics Bulletin" released by the National Bureau of Statistics of China in 2017 shows that the current population of China aged 60 and above is about 230 million, accounting for 16.7% of the total population. By 2050, the total elderly population in China is estimated to exceed 400 million and the aging level will reach more than 30%. This severe aging trend has brought about great changes in the pedigree of major diseases that endanger the health of our people. Tissue and organ damage, failure of cardiovascular and other functional organs, degenerative diseases, cancer and other major disease types closely related to aging have become major disease types. However, for these diseases, traditional treatment methods such as medicine therapy and surgical treatment often have little effect, which is difficult to meet the growing medical needs of the people at this stage. At the same time, with the implementation of the national marine strategy and space strategy, human beings will also face great health challenges such as hypoxia, high pressure and radiation in the continuous development of deep sea, space, plateau and other fields. This series of difficult problems cannot be completely solved only through conventional medical methods. Regenerative medicine research with stem cell technology as the core would bring revolutionary changes to the prevention and treatment of diseases closely related to aging, such as cardiovascular and cerebrovascular diseases, cancer, diabetes, Parkinson's disease, Alzheimer's disease, etc., which lays a foundation for the birth of new subversive medical technologies.

At present, regenerative medicine has become an important research field that governments, science and technology and business circles of various countries pay close attention to and invest heavily in, and has become a strategic must-compete field that represents the strength of national science and technology. Since 1999, the research findings in stem cells and regenerative medicine have been selected as the ten most significant discoveries in science community in the year 11 times by Science Magazine. In 2012, the Nobel Prize in Physiology or Medicine was awarded to two scientists in the field of nuclear reprogramming and induction of multifunctional stem cells, which shows that the science community attaches great importance to stem cells and regenerative medicine. In China, the importance of regenerative medicine has attracted great attention from relevant decision-making departments and scientific and technological circles. In 2015, the Ministry of Science and Technology of China launched the National Key Research and Development Project of "Stem Cells and Transformational Medicine", aiming at promoting research on stem cells and regenerative medicine at the national level and enhancing China's core competitiveness in this field as a whole. The Chinese Academy of Sciences began to implement the Strategic Pioneer Research & Development Programme of "Stem

Cell and Regenerative Medicine Research" in 2011 and achieved a number of innovative research results. Therefore, the Chinese Academy of Sciences is preparing to set up a new institute of "Innovative Research Institute of Stem Cells and Regenerative Medicine", which will promote the third medical health revolution around stem cells and regenerative medicine.

2.2 e-Science Requirements for Subject Knowledge Discovery in Stem Cell

The essence of e-Science is the informatization of scientific research activities themselves. The characteristic is to make full use of network information infrastructure and information technology, promote the exchange, collection and sharing of S&T resources, and change the scientific research organization and activity mode. With the development of e-Science, a series of scientific research activities such as scientific instruments, S&T experiments and academic exchanges produce massive, heterogeneous and diversified data and information every moment. Scientific research is increasingly becoming a data-driven knowledge discovery activity. The era of d-Science (Data-Driven Science) is coming, and it has entered the fourth paradigm of data-centered thinking, design and implementation of scientific research activities [2]. The fourth paradigm—"Data-Intensive Scientific Discovery" is coming, which make scientific discoveries through the processing and analysis of massive data. The research objects in various disciplines or subjects are no longer a single isolated system, but a complex innovation system covering a wider range and involving multiple disciplines or subjects. The large amount and rapid growth of data in various disciplines or subjects have led to the dual development of each discipline or subject—"X-Informatics" [2]. Taking the discipline of biomedicine as an example, Bio-informatics [3] used to understand the biological significance contained in a large number of data and Medical Informatics [4] used to analyze patient health and social health data and information have emerged.

The fourth paradigm has been widely used in the discipline of life and general health. Its core is the integration of multi-source heterogeneous data and information and the analysis of the massive data. In terms of data and information integration, more and more relevant biological data have been obtained from biological macromolecule sequencing and "Human Genome Project", which has rapidly promoted the development of Bio-informatics. The development of a variety of biomedical databases and platforms, such as Swissprot, GenBank, PharmGKB, IPA, etc., provides rich resources for the development of a variety of knowledge mining technologies and tools. In terms of data analysis, many big data analysis software and platforms have also emerged. For example, IBM has launched Watson[3] Medical System based on big data analysis and artificial intelligence. Open Phacts [5], a biomedical knowledge discovery system supported by 7th Framework Programme

[3]https://www.ibm.com/watson.

of European Union, helps small molecule screening and new medicine design. Open Phacts uses knowledge graph technology to link various data sets from molecules to genomes and then to patients, and uses depth learning algorithms to discover potential knowledge and implicit knowledge linkages [5]. Benevolent Bio, a British biotechnology company, used JACS (Judgment Augmented Cognition System) [6] which is a subject knowledge discovery platform based on big data to extract useful information from a large number of academic papers, patents, clinical trials, patient records and other data worldwide and discover clues of new medicine research and development. With the help of JACS, Benevolent Bio has labeled many potential compounds that can be used to treat amyotrophic lateral sclerosis, which greatly improved the efficiency of drug research and development [6].

According to the actual scientific research activities, the application requirements of e-Science for knowledge discovery in stem cell are as follows.

(1) Analysis of research focus, research frontiers and development trends in stem cell. Science and research administrative staffs hope e-Science can help dynamically discover the research focus, research frontiers and development trends in stem cell by making statistics, analysis and mining of a large number of S&T information and data such as S&T policies, scientific research fundings, scientific papers, patents, standards, clinical trial, new medicine instructions, etc. base on bibliometrics, scientometrics and knowmetrics. These will help provide scientific basis and reference for "Benchmarking and Catching up".
(2) Key technologies mining and technology foresight in stem cell. Researchers hope e-Science can help develop some visual cell tissue prediction models and other knowledge discovery tools through data mining and knowledge computing of multi-source heterogeneous stem cell research data, which will help to predict cell layout changes of cancer and other diseases by combing with knowledge graph, machine learning, deep learning and other information technologies. What's more, it can solve some complex key technical problems of stem cells and reveal a large amount of rich basic information of cell biology, accelerate the progress of stem cell research, cancer research and medicine development, and provide support for more accurate technical foresight.
(3) Integrated management of stem cell scientific research data and information. In addition to S&T literatures and information, it is often necessary to utilize and analyze a large number of fine-grained data objects, including knowledge entities, scientific instruments, experimental animals, experimental data, fundings, etc. during stem cell research activities. Through the development of an integrated stem cell scientific research data and information management platform, the isolated island of data from different sources can be opened up, and the research efficiency and information service level can be effectively improved.

3 Technical Framework of SCKD

3.1 General Objectives

Focusing on the e-Science requirements for subject knowledge discovery in stem cell, SCKD aims to implement the integration and deep processing of multi-source heterogeneous data, S&T information management, professional knowledge computing and subject knowledge discovery services in stem cell by comprehensively utilizing big data and a new generation of artificial intelligence technologies. SCKD provides a "one-stop" information retrieval and knowledge discovery services with accurate knowledge expression and linkages and rich visual interfaces, which helps researchers to carry out subject knowledge discovery research, and promotes the national technological foresight and strategic layout in the stem cell.

3.2 Technical Framework

Around the above general objectives, the SCKD is developed from three aspects: stem cell knowledge graph, knowledge computing environment and knowledge discovery services. Among them, the knowledge graph focus on data integration and fusion, the knowledge computing environment is the analysis tool, and the knowledge discovery services are the service interfaces. The technical framework of SCKD is shown in Fig. 1.

3.2.1 Construct Stem Cell Knowledge Graph

Stem cell knowledge graph timely, accurately and standardizedly integrate S&T literatures, S&T information, research data and S&T service resources in stem cell by comprehensively using knowledge extraction, knowledge mining, knowledge fusion and visualization technologies. The knowledge entities contained in it are automatically extracted, structured and linked to form a knowledge graph, which realizes the effective integration of "multi-form, multi-granularity, multi-dimension" data, information and knowledge and provides high-quality data support for subject knowledge discovery. The construction of stem cell knowledge graph involves three parts: basic data aggregation, knowledge entities mining and linking. The technical flowchart of stem cell knowledge graph is shown in Fig. 2.

(1) Basic data aggregation. Stem cell knowledge graph aggregates multi-source heterogeneous scientific research data such as papers, patents, fundings, projects, clinical trials, drugs, regulations, experts, institutions, etc. in stem cell, and a long-term update mechanism is established. Relational database, graph database and Solr index are respectively used to effectively store and manage the

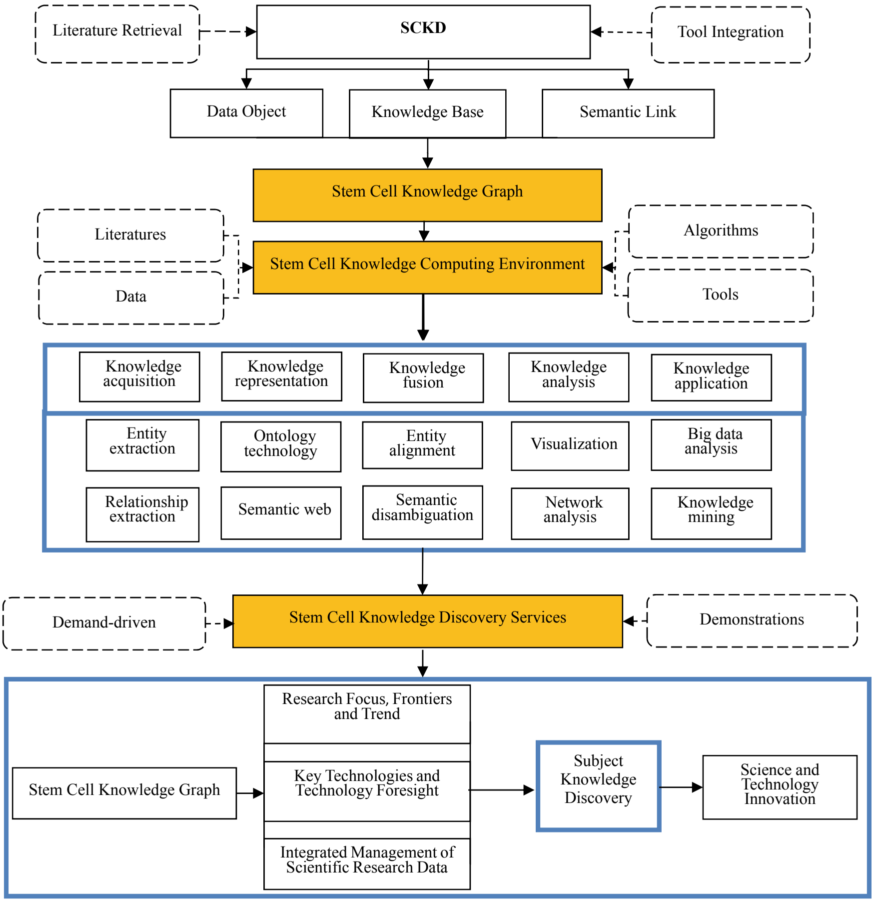


Fig. 1 Technical framework of SCKD

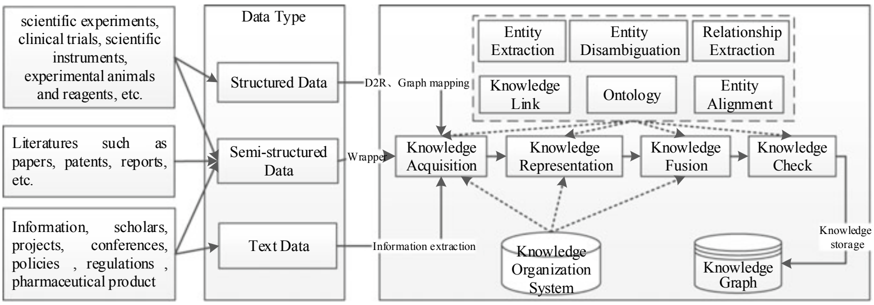


Fig. 2 Flowchart of stem cell knowledge graph construction

raw data of scientific research data, RDF triples and the comprehensive service data oriented to knowledge discovery services.

(2) Mining knowledge entities. Knowledge entities are the keys to construct a knowledge graph. The research adopts ontology-based knowledge extraction technologies to mine domain knowledge entities from the content of some stem cell basic data. Firstly, referring to some domain ontologies and knowledge organization systems such as UMLS, DisGeNET, and StemCell Commons and so on, an upper structure of stem cell knowledge organization system which includes these perspectives of scientific instruments, experimental animals, experimental technologies, cells, organs, diseases, genes, etc. is set up. Then a mature biomedical domain knowledge extraction tool SemRep [7] is used to extract knowledge entities and their relationships from some core S&T literatures in stem cell, so as to enrich the knowledge organization system of stem cell. After that, according to the knowledge organization system, other S&T information and data are tagged with these semantically labels to form the instances in the knowledge graph.

(3) Knowledge linking. Based on some scientometrics indicators and text mining technologies, semantic relationships between various S&T information and knowledge entities in the stem cell knowledge graph are linked based on citation, acknowledgement, cooperative network and co-occurrence of knowledge entities, etc. At present, there are 58 kinds of semantic relationships in 3 categories as "literature-literature, literature-knowledge entity, knowledge entity-knowledge entity" in the stem cell knowledge graph. The core relationships are those between "knowledge entity-knowledge entity" in the form of Subject-Predicate-Object (SPO) semantic network.

3.2.2 Develop Stem Cell Knowledge Computing Environment

The stem cell knowledge computing environment integrates algorithms, models, software and tools in stem cells and related subjects. A workflow-based knowledge computing environment is developed to realize the functions of data management, data cleaning, data mining and report output. The knowledge computing services such as visual analysis and knowledge reasoning are provided for subject knowledge discovery. It is functionally divided into three parts as data management, data processing and data analysis.

(1) Data management. Data management module is the foundation of knowledge computing environment, which mainly provides data management functions necessary for intelligence analysis and knowledge discovery such as data connection, integration and storage. In addition to directly utilizing stem cell knowledge map data, the system can also take in external data of the system, clean, convert and integrate external data according to rules, and combine data sets after correlation. The data storage module supports data preview, export, append and share operations. The data of the shared partition can be used by other users.

(2) Data processing. Data processing module is the core of knowledge computing environment, which mainly provides algorithm management, data management, task management and other functions. The algorithm management module realizes the unified management of the system algorithm. At present, the knowledge environment integrates 30 algorithms such as machine learning, recommendation system and natural language processing, and is integrating special computing models in the stem cell.
(3) Data analysis. The data analysis module is to carry out multi-dimensional, fine-grained and multi-type visual analysis and display of knowledge computing results. Data visualization supports the display of various chart effects, such as histogram, line chart, pie chart, scatter chart, thermal chart, map, radar chart, funnel chart, word cloud, relationship chart, etc. In addition, the chart can be displayed centrally through the dashboard. The data shown in the chart on the dashboard will be dynamically updated with the changes of the data in the analyzed table. The dashboard supports links, WeChat, microblogs, QQ space and other ways to share. The platform also supports directly generating analysis reports from analysis results. The analysis report module realizes the operation functions of grouping management of analysis reports, creating, sharing and exporting analysis reports, etc.

3.2.3 Launch Stem Cell Knowledge Discovery Services

The stem cell knowledge discovery services include integrated management of stem cell scientific research data and information, "one-stop" information retrieval service and accurate knowledge retrieval and knowledge navigation based on stem cell knowledge graph, stem cell focus and frontiers detection, research profiles, which can provide basic data, information products and personalized subject knowledge discovery services for S&T Innovation and scientific research management of the institutes. The technical details are as follows.

(1) integrated management of stem cell scientific research data and information. The graph database Neo4j and the relational database MySQL are respectively used to carry out management of stem cell scientific research data and information. MySQL stores the original basic data, while Neo4j stores the triples from knowledge graph. NEO4J is a NoSQL type database, which can flexibly expand the data structure and type, and meets the requirements of knowledge graph data management.
(2) One-stop retrieval and knowledge retrieval. Solr are used to index the knowledge organization system in stem cell knowledge graph as facets. Users can carry out "one-stop" intelligent retrieval of all types of stem cell information through Solr index, and carry out accurate knowledge retrieval and navigation based on stem cell knowledge organization system as filters.
(3) Focus and frontiers topics discovery. From different levels such as international, national, and institute, a series of research topics are generated as focus or

frontiers in stem cell. These topics include related papers, patents, fundings, news, experts, and institutes with corresponding labels.

(4) Research profiles. Knowledge graph and profiling technologies are used to profile the focus and frontiers topics, scientists and research institutions from the perspectives of papers, patents, projects, news and topics, etc.

3.3 Key Technologies

The main key technologies used in SCKD include knowledge entities alignment, knowledge discovery-oriented knowledge computing technology and unified data view based on multi-dimension index.

3.3.1 Entities Alignment of Stem Cell Knowledge Graph

Entity alignment is the basis of knowledge fusion and the premise of whether a knowledge graph is standardized and scalable. Entity Alignment refers to the process of whether two or more entities in the same or different data sources point to the same object in the real world [8]. The goal of entity alignment in the knowledge graph is to be able to link data from multiple different sources with high quality and create a unified knowledge representation specification, thus helping information system to understand the data [8].

(1) Data normalization based on the unique identifiers. Improving the entity alignment quality of scientific research data is mainly through standardized data preprocessing and cleaning steps. By screening, cleaning and distinguishing entities from different sources, those representing the same object from different data sources can be merged into one entity with unified identifier and added to the stem cell knowledge graph. For example, digital object identifier (DOI), patent number, open researcher and contributor ID (ORCID), project number, concept unified identification (CUI) and other unique identifiers are used to identify journal papers, patents, researchers, projects, and knowledge concepts respectively [9].

(2) Knowledge entities alignment based on some Gold Standards. Accurate entity alignment of concepts and terms is the key to ensure the quality of stem cell knowledge graph. UMLS [7] and SemRep tool [10] are mainly used to conduct knowledge entities alignment. SemRep is one of the important achievements of the Semantic Knowledge Representation Project (SKR) of the National Library of Medicine in the United States, which is a semantic knowledge extraction and representation tool based on natural language processing technology and UMLS. It can efficiently and accurately extract knowledge entities and semantic relationships between knowledge entities from biomedical S&T literatures. Firstly, SemRep is used to automatically extract Subject-Predicate-Object triples from biomedical content, in where the Subjects and Objects are normalized concepts

in UMLS, and Predicates are the gold standard semantic type in UMLS semantic network. Then, the obtained knowledge entities are mapped into the knowledge organization system of stem cell knowledge graph. In order to ensure the data quality of stem cell knowledge graph, some data cleaning rules and manual checks are also used to verify the instances.

3.3.2 Knowledge Computing Technology for Knowledge Discovery

The primary problem faced by knowledge discovery-oriented computing is multi-source heterogeneous data. In the traditional data analysis, the data source is relatively single, the format is relatively regular, the correlation is relatively simple, and the data analysis technology is relatively mature. However, in the computing environment oriented to knowledge discovery, not only are there many kinds of data sources, but also the data formats are complex and have many correlations, which is a typical multi-source heterogeneous data. Traditional data mining algorithms cannot be directly applied to multi-source heterogeneous data.

The stem cell knowledge computing environment solves the above-mentioned knowledge computing problems of multi-source heterogeneous data by adopting an open computing interface. In addition to providing built-in data cleaning, text mining, machine learning and other algorithms, the knowledge computing environment also has good scalability, which supports users to customize and develop new algorithms, and can combine the new algorithms with the built-in algorithms to create new business models to meet specific knowledge discovery requirements.

3.3.3 Unified Data View Based on Multidimensional Indexing

The data objects in SCKD include not only scientific research data such as various S&T literatures, clinical trials, scientific research data and scientific and technological service resources. It also includes the knowledge entities in the knowledge graph and the newly generated explicit knowledge in the knowledge computing environment. Its storage forms are relational tables in relational database and RDF triples in graph database. A unified data view is needed to shield the differences in the form of data objects and provide a unified data interface to the knowledge discovery service applications. Multidimensional index is a multi-layer and multi-angle index technology for multi-form and complex data. It can be used to integrate heterogeneous information sources and provide a unified data retrieval interface. It has been widely used in knowledge fusion information system [11].

The SCKD uses multidimensional integrated index to index heterogeneous information and data to create a unified data view. Specifically, it adopts Apache Solr to index metadata, knowledge entities and knowledge organization systems. Indexed fields include the common fields for knowledge discovery, such as knowledge resource type, resource title, author, inventor, institution, applicant, institution type, publication date, source, keywords, classification, scientific instruments, experimental animals, experimental protocols, experimental reagents, methods and technologies, cells, organs, diseases, genes, scientific research activities, scientific research output, etc. [11]. According to the data update frequency of different data sources, corresponding index update mechanisms are adopted, including daily update, weekly update, monthly update, etc.

4 Achievements and Service Effect

The SCKD (https://stemcell.kmcloud.ac.cn/) adheres to the principle of "construction, service and improvement" and has achieved some achievements and service effect, bringing together a group of stable users.

4.1 Achievements

The stem cell knowledge graph covers the S&T literatures, scientific information, scientific research data and scientific service resources in stem cell, and the stem cell knowledge computing environment supports high performance data processing and rich visualization methods. SCKD implemented four core functions and provided corresponding knowledge services, which can effectively meet the different needs of librarians, researchers, S&T managers and decision makers in stem cell.

4.1.1 Stem Cell Knowledge Graph

The SCKD integrates more than 400,000 pieces of scientific information and research data in stem cell in 4 categories from 103 authoritative core data sources, including: (1) S&T literatures such as papers, patents, reports, periodicals, monographs, etc. in stem cell. (2) S&T information such as news, activities, experts, projects, policies and regulations and other S&T information in stem cell. (3) Research data such as clinical trials, drugs and scientific experiments. (4) Scientific instruments, experimental animals, experimental reagents and other S&T service resources. Furtherly, from the perspectives of scientific instruments, experimental animals, experimental protocols, experimental reagents, methods and technologies, cells, organs, diseases, genes, etc., more than 20,000 knowledge entities are extracted, which are used to semantically annotate the stem cell scientific information and research data to construct the stem cell knowledge graph. At present, there are more than **2.2 million** data in stem cell knowledge graph.

4.1.2 Stem Cell Knowledge Computing Environment

Stem cell knowledge computing environment is built based on private cloud of CLIC and Spark. Its knowledge computing framework integrates 26 data cleaning, processing and fusion rules and 30 general data mining algorithms such as natural language processing, classification regression, recommendation and result evaluation. The environment also integrates 20 stem cell-related knowledge computing models, which can provide 12 visualization methods such as histogram, line diagram, pie diagram, funnel diagram, radar diagram, word cloud diagram, etc. and an automatic report generation tool based on templates.

4.1.3 Stem Cell Knowledge Discovery Services

Based on the needs of stem cell knowledge discovery, SCKD offers four core services at present: (1) Integrated management of stem cell scientific research data and information. SCKD realized the centralize management of various types of stem cell scientific research data and information, which can quickly provide users with personalized data services. (2) "One-stop" and accurate knowledge retrieval and navigation. Through one-click retrieval, users can obtain stem cell news information, paper, patents, fundings, projects, drugs, clinic trials, policies and regulations, industrial information and other types of S&T information. Users also can search and browse from more than 20,000 stem cell knowledge entities, which can quickly and comprehensively grasp the contents of S&T literatures without reading the full text in details. (3) Research focus and frontiers detection. From different levels such as international, national and institute research and development priorities, 22 research focus and frontier topics of stem cells have been generated. (4) Scientific research profiles. Based on the stem cell knowledge graph, scientific research institutions, scientists and other scientific research innovation subjects are profiled. Some screenshots of SCKD are shown in Figs. 3, 4, 5 and 6. "Stem cell assistant", a WeChat applet corresponding to SCKD, has been developed, which extends the knowledge services to the mobile terminal (Fig. 7).

The SCKD also support other subject knowledge discovery research such as identification and evolution analysis of innovative topics in the stem cell, identification of scientific research cooperation communities, high-quality stem cell patent mining, discovery of the relationship between diseases and genes in stem cell, portrait of subject knowledge structure, etc.

Fig. 3 "One-stop" search

Fig. 4 Accurate knowledge retrieval and knowledge navigation

Fig. 5 Expert profile

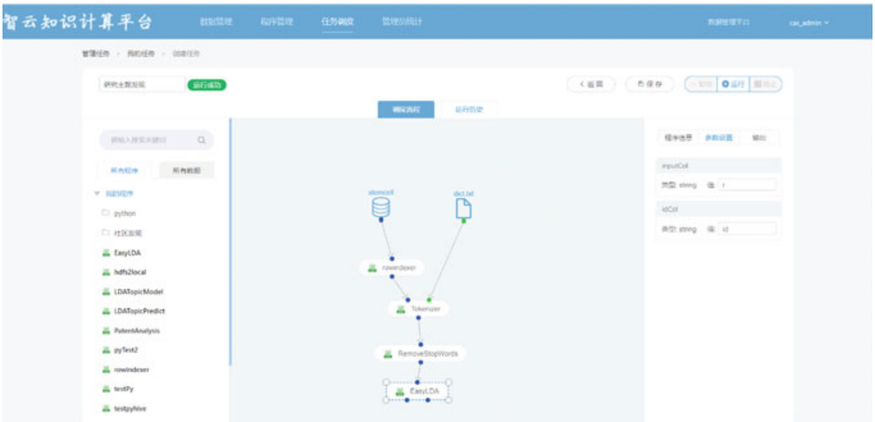


Fig. 6 Knowledge computing environment

Fig. 7 WeChat applet service

4.2 Service Effect

The SCKD was officially released on the 11th Guangzhou international conference on stem cell and regenerative medicine in November 2018. Now SCKD provides full services through contracted users. And SCKD has received extensive attention from more than 200 researchers from more than 20 institutions such as relevant research institutes of the Chinese Academy of Sciences, the Chinese Academy of Medical Sciences, Tsinghua University and Cambridge University, and the German Cancer Research Center. The number of users of the platform has steadily increased (Table 1).

Compared with similar platforms, such as Stem Cell Commons, SCKD has characteristics of "comprehensive data" and "accurate service". The platform gives full

Table 1 User access to stem cell knowledge discovery platform

Month	Total visits (number of visits)	Web page visits (hits)	Access traffic (GB)	Average online time (s)
2018.12	383	10,296	1.31	17.26
2019.01	470	11,187	1.49	18.19
2019.02	416	8908	0.94	18.26
2019.03	435	12,039	1.74	18.11
2019.04	786	16,954	2.48	17.12
2019.05	1123	35,773	5.82	17.90
2019.06	3982	87,365	8.19	39.90

play to the scientific and technological information resources and professional knowledge organization advantages of CLIC and the stem cell research advantages of GIBH, which can provide one-stop service from data, algorithms, software to applications. In addition to the public information services, SCKD also provides personalized knowledge services for scientific research, knowledge discovery and decision-making consultation of GIBH, which can effectively support the application of major scientific research projects and strategic layout in the stem cell. See Table 2 for typical cases.

5 Summary

The overall objective of SCKD is to integrate multi-source information and data, break down isolated islands of data, mine tacit knowledge, enlarge data values, integrate knowledge computing, help knowledge discovery, promote the fusion of scientific research activities and e-Science, and realize international advanced e-Science applications. With the support of the 13th Five-Year Plan e-Science Programme of Chinese Academy of Sciences, SCKD has developed a stem cell knowledge graph, an integrated professional knowledge computing tools environment, realized 4 subject knowledge discovery services, and built an e-Science application demonstration oriented to stem cell knowledge discovery. In the future, the research group will continue to follow the long-term goal of stem cell knowledge discovery, and focus on continuing to promote SCKD from three aspects: data resources, application demonstration and subject knowledge discovery.

(1) Promote the construction of a big data center for stem cell and regenerative medicine and form a subject data standard system. Integrate more professional scientific research data resources, including related pathways, proteins, and clinical trials data. More scientific information such as experimental videos, popular science, WeChat, blog data and comment information will be integrated, too. We will further research the data standards for scientific information and research data in stem cell and formulate a standardized, flexible and extensible metadata specification and data quality control standards.
(2) Further optimize the knowledge computing environment and expand application demonstration scenarios. The more knowledge computing models, methods and tools in stem cells and regenerative medicine will be integrated; a domain knowledge computing tool navigation will be added. Facing the specific needs of stem cell knowledge discovery, aiming at the scenarios of stem cell innovation topics evolution analysis, expert cooperation and scientific research community identification, the corresponding standardized process and methods will be designed, and a "one-stop" knowledge computing model is being developed.
(3) Carry out in-depth research and application of customized knowledge discovery to assist stem cell scientific research. We will further solicit the opinions of

Table 2 Typical cases of personalized knowledge services

Type	Content	Effect
Supporting the strategic layout of stem cells	Provide data support for construction plan of Guangzhou Regenerative Medicine and Health Guangdong Laboratory	The laboratory was approved to start construction in 2018
	Provide data support for construction plan of Human cell lineage Atlas Facilities	Start preliminary research
Supporting major scientific research projects	Provide data support for intellectual property analysis of the national key R&D Program of “Research on regulation and heterogeneity mechanism of pedigree fate determination in differentiation of human pluripotent stem cells”	Provide intellectual property analysis for project patent application
	Provide theory, method and technical support for CLIC to conduct the national key R&D plan “R&D and application demonstration of comprehensive science and technology service platform of Chengdu-Chongqing Urban Agglomeration”	Project is progressing smoothly
	Provide data support for GIBH to apply the national key R&D plan “application demonstration of comprehensive scientific and technological services for typical industries of Pearl River Delta Urban Agglomeration”	Project approved
Supporting information intelligence and decision consultation	Writing 4 decision-making consultation suggestions	Decision reference
	6 stem cell research reports completed	Directly Serve the Scientific Research Management of GIBH
	Editing “Stem Cell Research and Development Trends” Express Issue 16	Send it to relevant research groups and the leaders of management departments for reference.
	Support the establishment of the world’s first “human embryo gene editing” legal and regulatory database	Provide emergency data and intelligence services about ethical standards for human embryo experiments to relevant leaders

experts in stem cell, carry out customized scientific data analysis and knowledge services according to the needs of experts support scientists' knowledge discovery research in the frontiers of stem cells, and make SCKD a powerful scientific research assistant for scientific research in stem cell.

Acknowledgements We appreciate Haiyun Xu, Chunjiang Liu, Xin Zhang, Lin Peng, Wenjie Chen, Shuang Zhao, Yuan Xu from Chengdu Library and information center, Chinese Academy of Sciences and Chen Guo from Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, participated in the construction of SCKD.

References

1. Zhou Q (2013) Somatic reprogramming, challenge and hope. In: National conference of Chinese Society of Cell Biology, Wuhan, China
2. Zhang Z, Fan S (2015) On the emergence and development of subject informatics. J China Soc Sci Tech Inf 34(10):1011–1023
3. Ouyang S, He F (1999) Bioinformatics: a new field combining biological experimental data and computing technology. Chin Sci Bull 44(14):1457–1468
4. Dong J (2004) The current situation and future of medical informatics. Chin J Hosp Admin 20(4):232–235
5. Williams AJ, Harland L, Groth P et al (2012) Open PHACTS: semantic interoperability for drug discovery. Drug Discov Today 17(21–22):1188–1198
6. Brown N, Cambruzzi J, Cox PJ et al (2018) Big data in drug discovery. In: Progress in medicinal chemistry, vol 57. Elsevier, pp 277–356
7. Kashyap V, Borgida A (2003) Representing the UMLS® semantic network using OWL. In: International semantic web conference. Springer, Berlin, Heidelberg, pp 1–16
8. Zhuang Y, Li G, Feng J (2016) A survey on entity alignment of knowledge base. J Comput Res Dev 53(1):165–192
9. Wang Y, Qian L, Xie J et al (2019) Building knowledge graph with Sci-Tech Big Data. Data Anal Knowl Discov 3(1):15–26
10. Rindflesch TC, Fiszman M (2003) The interaction of domain knowledge and linguistic structure in natural language processing: interpreting hypernymic propositions in biomedical text. J Biomed Inform 36(6):462–477
11. Wen Y, Fu H, Shu F et al (2015) Amino acids industry knowledge service platform. Chin J Libr Inf Sci 8(4):78–89

Zhiqiang Zhang, Ph.D, director of Chengdu Library and Information Center, Chinese Academy of Sciences, doctoral supervisor of University of Chinese Academy of Sciences, a core professor specially employed by Chinese Academy of Sciences, member of decision-making advisory committee of Sichuan provincial Party committee and Sichuan provincial government. He has independently and cooperatively published 20 monographs, 13 translated works and over 400 papers. In addition, he has won 18 awards for scientific and technological achievements, such as provincial and ministerial awards for scientific and technological progress, and awards for outstanding achievements in social sciences. Research fields: Science and technology strategy and planning, science and technology policy and management, scientific measurement and evaluation, science, information theory, method and application, ecological economics and sustainable development, etc.

"One-Time Face Recognition System" Drives Changes in Civil Aviation Smart Security Screening Mode

Yu Shi, Xiangdong Zhou, Jun Cheng, Lijun Wang, Daijian Luo, Yong Guo, Enpeng Yang, Lijun Zhang, Lu Han, Zhenghao Li, and Changcheng Zhang

Abstract With the surge of civil aviation airport passenger traffic in recent years, the airport security and screening mode is under great pressure, which has become an obstacle for achieving the target to become a leading civil aviation power. This paper first outlines the progress and trend of civil airport security screening inside China and abroad, then explains in detail the "One-time Face Recognition System", which is an innovative achievement of the Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences in the intelligentization of civil aviation security. The system integrates various advanced technologies and innovative ideas. With the support of Civil Aviation Administration of China (CAAC) and the support of Hongguang Special Project funded by Chinese Academy of Sciences (CAS), the system has applied to Hohhot Baita International Airport, achieving great success. The system and the operating procedures have been approved by the authorities, authorizing the use of manually assisted computerized verification to replace the original manual verification, which has started a reform of the civil aviation smart security screening mode in China.

Keywords Airport security screening · Biometric recognition · Screening data chain

1 Introduction

The goal of aviation security screening is to prevent illegal interference and to maintain the safety of air transportation. For civil aviation, the security screening covers "the security screening of passengers on civil aircraft and their baggage, the personnel entering the restricted area and their belongings, air cargo and mail; security monitoring of the personnel and their belongings in the restricted area of the airport; monitoring of civil aircrafts that are performing missions" [1].

Y. Shi (✉) · X. Zhou · J. Cheng · L. Wang · D. Luo · Y. Guo · E. Yang · L. Zhang · L. Han · Z. Li · C. Zhang
Chongqing Institute of Green and Intelligent Technology, Chinese Academy of Sciences, Chongqing, China
e-mail: shiyu@cigit.ac.cn

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_23

Airport security screening is a necessary method to maintain aviation security, and its efficiency is linked to the productivity of the entire civil aviation system. Currently, airport security screenings are over-reliant on humans. A large amount of security personnel is needed, accounting for more than 50% of the airport staff, whose training and management is difficult. The work intensity is high and employees are prone to errors in long-period and high-intensity working environments. Security screening methods, such as body search and container-opening inspections, can easily lead to low passenger satisfaction with the current security screening processes. Therefore, in accordance to the requirements of the State Council, the civil aviation security screening mode innovation expert group was set up by the Civil Aviation Administration of China (CAAC) proactively to adapt to the situation. And CAAC actively promotes the research on new technology and equipment for civil aviation security screenings. Applying the new technologies in security screening processes enhances the ability to operate effectively in high-intensity work, and thus improves airport security screening safety margin, decreases human error rates, reduces staff work intensity, improves passenger satisfaction, and optimizes resource allocation.

At the end of 2016, the civil aviation security screening mode innovation expert group put forward the work requirements of "to innovate security screening modes, optimize security screening processes, actively use advanced technologies, strengthen inspections of high-risk passengers and questionable items, and improve security screening accuracy and efficiency". In the following two years, the Chongqing Institute of Green and Intelligent Technology of the Chinese Academy of Sciences carried out a series of technological integration, innovations and verification work, and jointly developed the "One-time Face Recognition System" with Civil Aviation Management Institute of China and Aviation Security Training Centre (ASTC) of International Civil Aviation Organization. This system has been applied to Hohhot Baita International Airport with the support of Hongguang Special Project (the project is supported by the key project for Transfer and Transformation of Scientific and Technological Achievements of the CAS) and CAAC, achieving remarkable results.

1.1 Development Process of Civil Aviation Security Screening

The evolution of civil aviation security screening has been event-driven for a long time. Since the world's first hijacking in Peru in 1930, there have been no fewer than a thousand hijackings and bombings worldwide. To deal with various hidden and high-tech flight safety hazards, civil aviation security screening has also been adjusted accordingly. Led by the United Kingdom and the United States, current aviation security requirements were established as a result of the pre-1970 increase in the number of hijackings. These requirements focus on the ability to find passengers carrying metallic weapons that may be used to intimidate the aircrew into changing the destination of an aircraft before boarding as safety technology inspections and quickly spread them to countries around the world. In 2001, after the "9/11" incident in the United States, countries around the world paid unprecedented attention to aviation security. Globally, airport security screenings have been comprehensively upgraded, and the civil aviation industry began to focus on the ground part in aviation security. Subsequently, a terrorist unsuccessfully attempted to denotate plastic explosives concealed within his shoes, and was subdued by passengers afterwards. And since then, airport security screenings have added new rules for removing shoes in security checks. After the liquid bomb incident in 2006, airports imposed strict restrictions on carrying liquids on board. After the plane terrorist attack in 2009 Christmas, US airports began to widely use human body scanning and imaging security gates and spared no effort to promote it to other countries (see Fig. 1) [2].

At the beginning of its establishment, China's civil aviation attached great importance to aviation safety. In 1957, then-Prime Minister Enlai Zhou issued an important directive to civil aviation: "to guarantee safety first, improve service work, and strive for smooth flights". Since then, with the changes in the aviation safety situation at home and abroad, China's civil aviation security screening has gone through a similar process as other countries.

At airports in various countries, passenger security screening is mainly carried out manually with assistance from security screening equipment. According to service content, security screening posts are divided into the following processes: explosives detection, basic order maintenance, ID verification, luggage processor, personal screening, X-ray machine screening, package hand screening, boarding gate verification, and others. For decades, this process has achieved the goals of security screening and ensured the safety of civil aviation transportation.

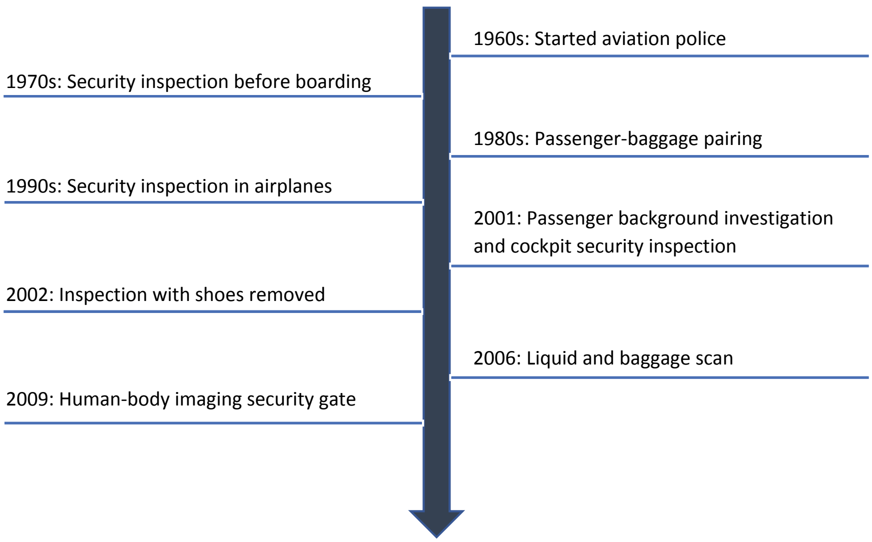


Fig. 1 History of civil aviation security screening evolution

1.2 Trade-Offs Between Safety and Efficiency

With the development of the economy, air passenger traffic is increasing day by day. The International Air Transport Association (IATA) expects a near doubling of the air travelers by 2037 [3]. The civil aviation passengers in China increased at an annual rate of more than 10% from 2014 to 2018 (see Fig. 2) [4, 5]. As a result, the pressure on airport security screenings has gradually increased. It includes on the one hand, the lack of security resources corresponding to the increase in passenger flow, and on the other hand, the length of travel time for passengers caused by security screenings. According to a 2015 survey conducted by the Travel Leaders Group, only 11.6% of consumers made negative evaluations of security screening methods, but 55.6% of consumers were dissatisfied with the security screening waiting time [6]. It suggests that the length of stay of passengers in the security screening area is an important factor affecting the travel experience of passengers. In October 2017, the International Air Transport Association (IATA) and the International Airport Association (ACI) released the New Experience Travel Technologies (NEXTT) project, which aims to use new technologies to create a "key data pillar" for passenger identity authentication and passenger flow management, thereby reduce passenger stays at airport check-in and security checkpoints. This project shows a new trend in civil aviation security screening: technology-driven, service-oriented, improving security screening efficiency and reshaping the security screening experience.

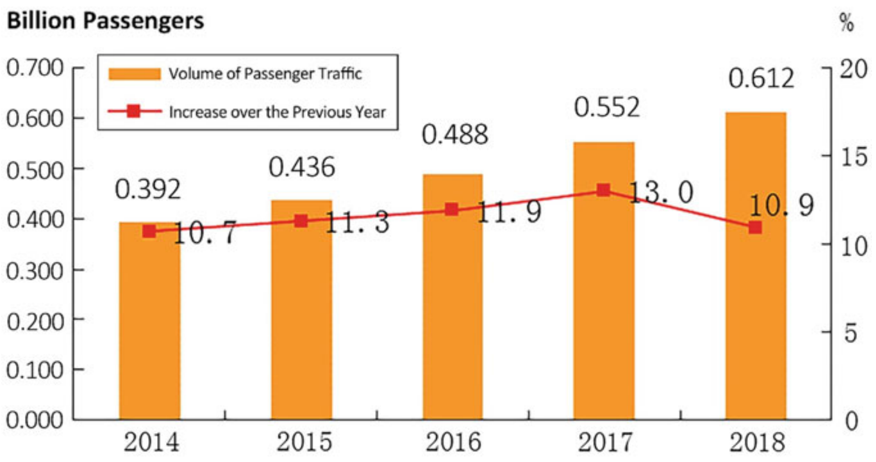


Fig. 2 Civil aviation passenger traffic (2014–2018)

1.2.1 Security Screening Information System—Differential Security Screening

Since 1998, the United States has implemented security screening information systems characterized by passenger classification and differential security screenings such as CAPPS, CAPPS II, Secure Flight, PreCheck [7]. Among them, the PreCheck system developed by the US Transportation Security Administration (TSA) provides rapid security checks for low-risk passengers identified in pre-checks to improve the security screening experience. It has been implemented in the United States since October 2011. As of January 2016, it had been applied in more than 150 airports in the United States.

In addition, in 2014, the U.S. Federal Customs and Border Protection (CBP) developed a fast-track clearance program called "Global Entry", setting up dedicated "Global Clearance" channels at airports and airlines it works with. Qualified passengers using the plan can enter the security screening quickly, without having to line up with ordinary passengers, to take off their shoes, belts, or jackets, or to take out computers and other items from their bags for scanning.

Differential civil aviation security screenings in China are mainly implemented at two levels, namely airports and passengers. At the airport level, the Civil Aviation Administration of China has divided the security screening levels into four levels for different periods, and each level corresponds to different security measures. At the passenger level, in April 2018, the Civil Aviation Administration issued "A Notice of Conducting a Pilot Study on the Development of Differential Security Screening Modes for Civil Aviation Passengers". The notice identified three key elements including differentiation of screening targets, differentiation of security screening technologies, and differentiation of security screening processes. Frequent travelers with good security credit can directly enter the "fast track", which saves the waiting time and screening time of the security screening, further improves the security screening efficiency and enhances the passenger experience.

1.2.2 Automated Border Control Based on Intelligent Identification

The automatic border control system (ABC system) is an automated control system that uses computer technology to compare fingerprints, irises, human faces, and other biological characteristics to determine passenger identity and clear the screening. Applying this system to the verification process of civil aviation passenger security screening can effectively improve verification accuracy and speed, and it has been applied to many airports around the world. Studies show that as of 2013, dozens of large airports around the world had begun to apply the system [8].

Civil aviation of China always pays close attention to the development of new technology research and application. In 2016, the Chongqing Research Institute of the Chinese Academy of Sciences and the relevant departments and units of the civil aviation industry carried out a joint study on the “Face Recognition Assisted Verification System”, which gradually realized a significant increase in the face recognition rate of passengers in the airport environment, followed by a large scale trial. As of the beginning of 2019, relevant results had been applied to 70 civil aviation airports nationwide.

1.2.3 Passenger Personal Inspection Equipment

Terahertz (THz) imaging technologies either scan subjects for natural radiation emitted by the human body (passive imaging) or expose subjects to a specific type of radiation reflected by the body (active imaging). In either case, materials such as metallic weapons or plastic explosives, which emit or reflect radiation differently from the human body, are distinguishable from the background image of the body. These screening systems generate television-like digital images that can be evaluated by image processing and analysis methods. Images are viewed by an operator trained to identify potential threat objects in these images, sometimes with the assistance of image enhancing software that highlights unusual features. The United States began to vigorously promote human body scanning imaging security gates based on THz imaging technology after the incident of the Christmas plane terrorist attack in 2009. In June 2018, the Civil Aviation Administration of China promulgated the “Internal Certification and Control Standards for Civil Aviation Millimeter Wave Human Imaging Security Screening Equipment” and the “Test Procedures for Civil Aviation Millimeter Wave Human Imaging Security Screening Equipment Prohibited Items Detection Capability”, which officially put the millimeter wave human imaging equipment into the civil aviation security screening equipment list.

1.2.4 Baggage Inspection Equipment

Prior to the advent of “apparatus for baggage inspection” that was invented by Peil in 1972, baggage inspections had been exclusively operated by manual unpacking method [9]. For nearly half a century since then, baggage security screening has been based on X-ray equipment. Currently, EU countries use multi-view X-ray equipment, while US airports use CT technology as an automatic bag explosives detection system. Both technologies can realize 3D baggage image imaging. Passengers no longer need to take out laptops and other electronic devices from their carry-on baggage separately during security screening, which can speed up the baggage inspection to a certain extent [10].

1.3 New Generation Smart Security Screening Promotion Plan

In 2013, the ACI and IATA proposed a promotion plan for Smart Security, which was supported by the International Civil Aviation Organization (ICAO) [11]. After several years of development, the plan is gradually becoming the weathervane of the global civil aviation security screening systems. At present, its improvement suggestions include: real-time monitoring and analysis of security checkpoints, and individual risk assessment of passengers on the spot; using biometric and RFID technologies to track targets, and switching subsequent security screening procedures according to risk levels unnoticeably; centralized image inspection, which enable the image inspectors to work in a non-interfering environment, and automatically separate suspicious baggage to the unpacking point inside the lane; using explosives trace detection equipment to improve security screening explosive detection capabilities, among other measures.

2 One-Time Face Recognition System

In response to the requirements of "innovating the security screening mode, optimizing the security screening process, actively using advanced technologies, strengthening the inspection of key people and items, and improving the accuracy and efficiency of security screenings" from the Civil Aviation Administration of China, the Chongqing Research Institute of the Chinese Academy of Sciences, created a data-centric, quality-targeted, and efficiency-oriented "One-time Face Recognition System", with the support of advanced concepts and technologies, making full use of new artificial intelligence technologies such as face recognition and machine learning, combined with existing facilities such as video surveillance, audible and visual alarms, and taking full advantage of big data.

2.1 Introduction of One-Time Face Recognition System

One-time Face Recognition System uses face recognition as the core technology, and effectively uses passenger identity information, facial information, and flight information to achieve efficient pass-through by presenting the identity document only once from security check to boarding, with "passing using the face". It provides comfortable, convenient and efficient security screening services during boarding procedures such as person-ID-ticket verification, security screening lane reverification, fast clearance at boarding gates, transfer connections and transits processes, which has promoted the improvement of airport safety margin, operational efficiency and passenger satisfaction, and has driven the intelligent changes of civil aviation security screening.

2.1.1 Face Recognition Technology

Face recognition is a kind of biometric recognition technology utilizing human face feature information for identity recognition. Cameras are adopted to collect images or video streams containing human faces, based on which face recognition algorithms perform face detecting, tracking and after several steps accomplish recognizing human faces.

After about 40 years of development, face recognition has been relatively mature and is currently widely used in finance, justice, public security, border inspection, education and other fields. Face recognition has the following advantages over other biometric technologies: non-contact, which means users do not need to directly contact the device; non-mandatory, that is, face images for recognition can be proactively acquired; concurrency, which means that in application, multiple faces can be sorted, judged, and identified; low detection cost, since using only a camera can accomplish information collection, which is sufficient for subsequent calculations. These advantages determine its wide application in airport security and inspection fields.

The face recognition algorithms used in the One-time Face Recognition System mainly include six modules consisting of face detection, face key point detection, face quality evaluation, face texture normalization, face feature extraction, and face feature comparison (see Fig. 3).

(1) Face detection

The purpose of face detection is to detect faces in the image and accurately calculate the location of the faces. The detection algorithm is based on convolutional neural network (CNN), where the convolutional layers of different levels and resolutions are used to obtain candidate regions of different scales in the image that may contain faces, and then the candidate regions are further judged by a fine prediction subnetwork. Finally, non-faces are discarded and faces are more accurately predicted.

Fig. 3 Flow chart of face recognition algorithm

(2) Key point detection

Face key point detection is based on the face detection results, which determines the position of each key point on the face, such as the corners of the eyes, pupils, nose, and corners of the mouth. This step uses a CNN with fewer layers to simultaneously predict the position and visibility of the key points, as well as the pose of the face, and a key point confidence evaluation module is added. It is capable of performing detection with relatively high accuracy and speed, as well as evaluating the reliability of the detected key-point locations in real-time.

(3) Quality evaluation

Face quality evaluation is mainly about evaluating the obtained face image quality for selection in the subsequent modules.

In this step, the algorithm comprehensively evaluates the quality score of a face image, by feeding the face image into a lightweight CNN, and using the quality scores of various practical factors as constraints to obtain the face image quality score of a person with comprehensive considerations.

(4) Texture normalization

Face texture normalization is to restore faces with different illumination, poses or expressions to well-lit, front-facing, and expressionless, reducing the difference between the face image to be matched and the registered face image, thereby reducing the difficulty of feature extraction and recognition. The algorithm of this step provides an end-to-end, data-driven texture regularization network and feature recognition network.

(5) Feature extraction

Face feature extraction is to extract a vector that can represent the original face from the normalized face. This step is the crucial module of face recognition and directly determines the accuracy of recognition results. Our algorithm (see Fig. 4) is based on a CNN, which fuses multi-scale information to obtain the high-level semantic feature representation. In the process of CNN training, triple face images are used as input, utilizing the complementarity of classification loss function and sorting loss function. The facial features of the same person have higher similarity, while the similarity of different facial features is lower.

(6) Feature comparison

Face feature comparison is to compare the features of the image to be identified with that of the target image, determine whether the two images are similar, and return a similarity score. In this step (see Fig. 5), the data distribution difference between different scenes is narrowed by the domain adaptation algorithm, and the face images of different scenes are projected to the same feature space with better representation capability for comparison, and then in the new feature space, the

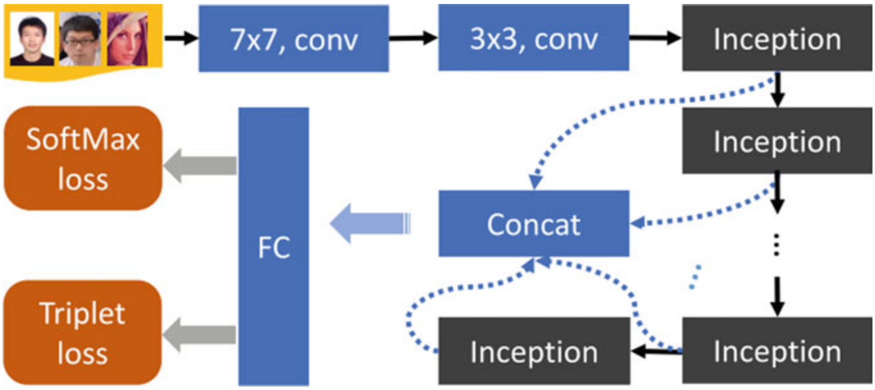


Fig. 4 Face feature extraction algorithm flow

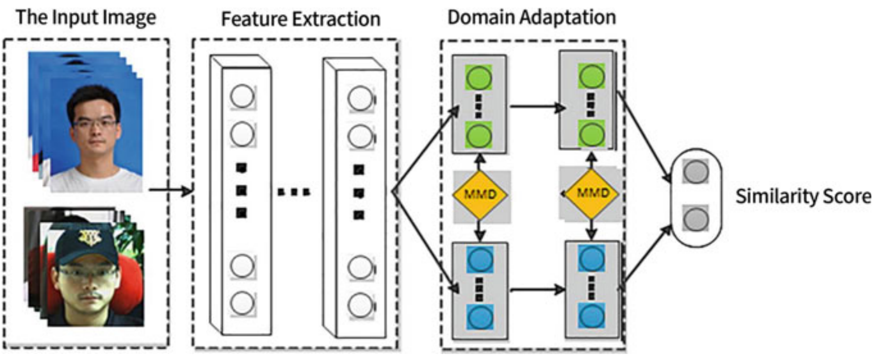


Fig. 5 Schematic for comparison of face features

discrimination between samples from different categories is maximized while the differences between samples from the same category are minimized, ensuring that the face feature representations of the same person have a higher similarity score, while similarity scores from different people are lower.

2.1.2 System Architecture

At the system architecture level, the One-time Face Recognition System is divided into three parts: server side, identification front end and query service end (see Fig. 6).

(1) Server side

The server side is composed of a server cluster consisting of a face recognition server, a data bus server, a data storage server, etc. It provides services for the front-end recognition system and query service terminals through an encrypted network

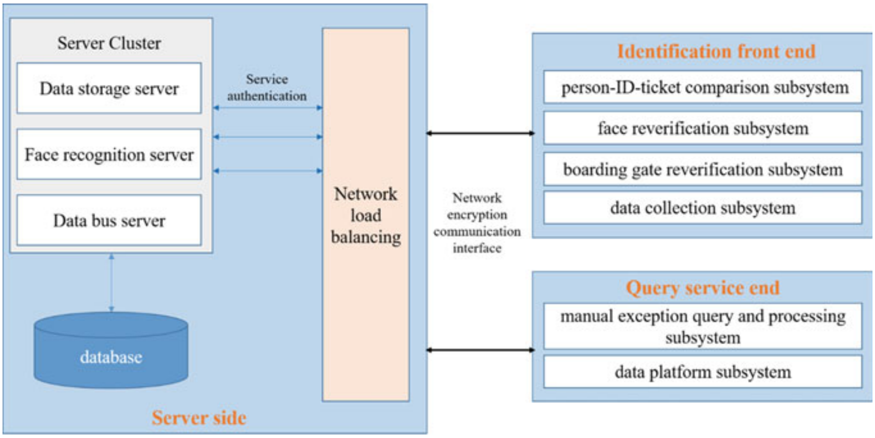


Fig. 6 System architecture

communication interface. The data storage server encrypts the stored data through encryption algorithms to prevent information leakage. The dynamic distribution of network requests in the server clusters are performed by network load balancing, ensuring fast server response and reasonable and even allocation of resources. The server receives or rejects externally initiated access requests through a unified service authentication, effectively rejecting illegal access.

(2) Identification front end

The identification front end consists of a person-ID-ticket comparison subsystem, a face reverification subsystem, a boarding gate reverification subsystem, a data collection subsystem, and other subsystems. These subsystems perform functions such as face detection, optimal face finding, feature extraction, through the camera on the front end, and upload to the server for comprehensive information discrimination. After the server returns the results, these subsystems perform on-site processing. The person-ID-ticket comparison subsystem verifies the consistency of passenger, ID document, and ticket in the verification area, and inserts into the face database after the successful verification. The face reverification subsystem performs non-contact reverification on the passengers through the face database in the security screening lane, to ensure passengers have all verification steps passed. The boarding gate reverification subsystem verifies whether the passenger passes the security verification and reverification step, as well as whether he or she is a passenger on the flight, preventing unverified passengers and passengers not for the flight from boarding. The data collection subsystem collects information of transfer passengers to facilitate face reverification at the boarding gate.

(3) Query service end

The query service end is divided into a manual exception query and processing subsystem and a screening data chain subsystem. The manual exception query and processing subsystem is mainly used to manually query and process the passenger boarding pass or identity card when an exception occurs in the identification front end; the screening data chain subsystem includes distributive storage of whole-process information, providing comprehensive, multi-angle, multi-condition query, statistics, analysis and prediction functions, providing a strong basis for airport post-event query, cause analysis, prior-event judgment and reasonable resource allocation.

2.1.3 System Function Module

According to the checkpoints involved in the security screening processes, the system is mainly used in three areas of verification, security screening and boarding (as shown in Fig. 7), and has six modules.

Part 1: Verification

(1) Person-ID-ticket self-verification gates

Person-ID-ticket self-verification gates verify whether the passenger IDs are consistent at the security check-up area of the airport terminal, that is, whether the passenger holds the ID document of himself or herself, as well as whether the passenger has checked in. If not verified, the passenger needs to use manual verification lane to (manually) verify his or her identity and air ticket information.

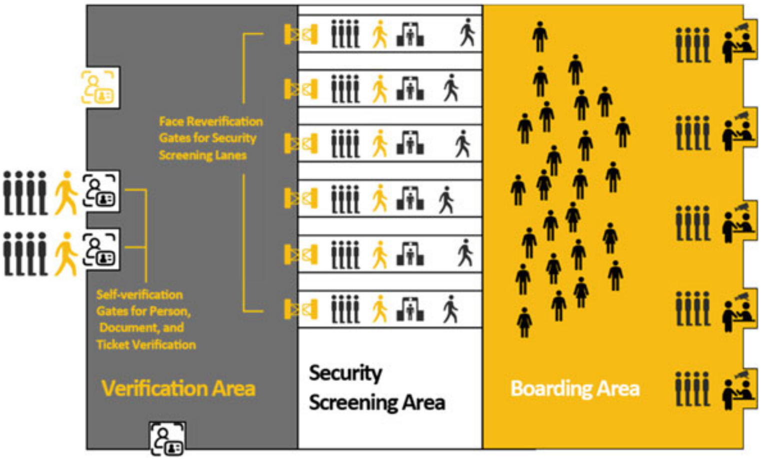


Fig. 7 Passenger security clearance process

The system collects identity information of passengers, verifies the check-in information, collects face images, and verifies the identity information. It ensures that all passengers entering the security screening waiting area have their identities and check-in status verified, and prevent unverified people from tailgating. It increases efficiency and security of the airport.

The staff adjust the number of self-verification gates to control the release rate of the security screening lanes according to the passenger queue.

(2) Face reverification gates for security screenings

Face reverification gates for security screenings are installed at the manually assisted verification counters. They read ID card information with assistance from the security staff. The system automatically collects face images and compares them against the photos from the ID cards, intercepting invalid IDs and black-listed passengers, enabling re-examination, history query, switching video, among other functions. It makes the security verification efficient, intelligent, and convenient.

(3) Terminals for collecting identities of transfer/transit/diversion passengers

The terminals for collecting identities of transfer/transit/diversion collect identity information and face images of transfer/transit/diversion passengers, and provide passenger-related data for boarding gate reverification for the next part of the journey.

Part 2: Security screenings

(4) Face reverification gates for security screening lanes

Face reverification gates for security screening lanes are used in companion with the person-ID-ticket self-verification gates or assisted verification terminals for security screenings. Passengers queue at the reverification gates at the entrance of the security screening lanes, and are reverified automatically before they enter the lanes. The system verifies whether the passengers passed security screenings in part 1 to avoid missed inspections. In case of failed reverification, manual interference is required.

Staff may adjust the passenger flow rate by switching on and off the gates according to the queueing of passengers.

Part 3: Boarding

(5) Terminals for face reverification at boarding gates

Terminals for face reverification at boarding gates identify boarding passengers through face recognition, and compare the information with the security verification information or information collected during transfer, transit, or diversion, to determine whether the passenger is a valid passenger for the flight. When the comparison fails, airport staff verify the information manually and decide whether to let go or

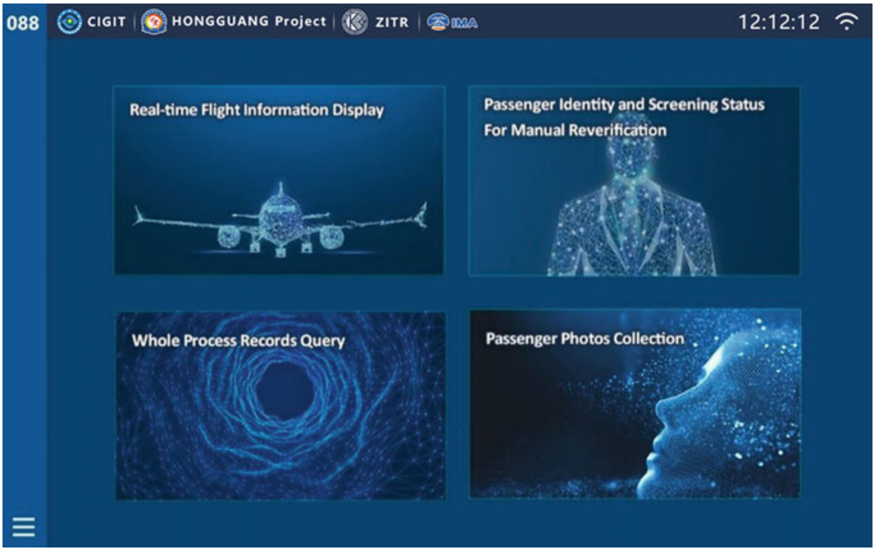


Fig. 8 Security screening whole process intelligent query system interface

stop the passenger. It effectively avoids missed inspections and passengers going to the wrong boarding gates.

(6) Terminals for whole-process information query

Information for security screenings whole-process use distributive storage and management to ensure data safety, accuracy, freshness, and validity. Terminals for whole-process information query integrate four submodules of real-time flight information display, passenger identity and screening status for manual reverification, whole-process record query, and passenger photo collection (see Fig. 8).

- Real-time flight information display

This module displays real-time information about flight including basic flight information, boarding status, passenger number and status, and real-time passenger reverification results.

- Passenger identity and screening status for manual reverification

This module is used to quickly verify the identity and screening status of passengers when the automatic reverification fails. Passenger information is checked by scanning boarding passes or ID cards, and the information is queried in the system for the staff to verify the information manually. Staff may choose to let go or stop the passenger depending on whether the passenger is a valid passenger on the flight and whether the information is valid and complete.

- Whole process records query

The system organizes and analyzes real-time information about the passengers intelligently, and pushes related verification, reverification, device-collected information to the corresponding boarding gates for the whole-process record. This module can query all records of security screening, transit, transfer, diversion at the boarding gate.

- Passenger photos collection

This module is used in companion with information collection terminals to collect the face information and identity information of transit and transfer passengers for boarding reverification in the next part of the journey.

2.2 Driving Reforms in Civil Aviation Security Screening Modes

The One-time Face Recognition System utilizes intelligent function modules such as person-ID-ticket self-verification gates, face reverification gates for security screening lanes, and boarding gates face reverification to develop a modern whole-process smart security screening mode, which changes the existing security screening mode in airports, and can effectively enhance the safety margin of airport security screenings, improve the security screening efficiency, save labor costs, and drive the reforms of civil aviation security screening mode.

2.2.1 Current Status of Civil Aviation Security Screening Mode

The civil aviation airports' existing security screening mode already has an information-based foundation, but still requires a lot of manual operations. For example, security inspectors and ground service personnel at places like the security screening lanes, boarding gates, still manually verify the passengers' identities. This mode has become a bottleneck of rapid development of intelligent technologies in civil aviation.

Current security screening mode lacks mature intelligent system support. Most of the verification at airports are still performed manually, and only the air ticket and ID card or passport information is verified, which is part of "ticket-validation". They are not capable of verifying the identity of the passengers, which is "person-validation". With the complexity of situations in passengers, at present, security technology is insufficient, and the feedback of control information is not timely and accurate, and the increase of manpower cannot effectively improve the level of security. As a result, security loopholes are prone to occur. With the rapid growth of aviation industry, the flights and passenger traffic has been increasing rapidly, which leads to insufficient staffing and high work intensity.

2.2.2 Comparison of One-Time Face Recognition System and Existing Security Screening Mode Processes

Processes in One-time Face Recognition System security screening mode is as follows: (1) After arriving at the airport and finishing check-in, the passenger only needs to take an ID card and pass the person-ID-ticket self-verification gate at the pre-security checkpoint to verify the information without printing a boarding pass. The passenger then goes to the least crowded security screening lane guided by the gate. (2) After entering the security screening lane, the dynamic face reverification system of the security screening lane automatically reverifies the passenger, and intercepts or raises alarm if the passenger has not passed the person-ID-ticket verification at pre-security screening. (3) The passenger's identity is tied to the baggage tray and the passenger baggage through face recognition, facilitating quick re-examination of the carry-on baggage information. (4) After arriving at the boarding gate, the passenger only needs to pass through the boarding gate dynamic face verification system to board without showing any ID documents or tickets.

Processes in existing security check mode is as follows: (1) After arriving at the airport to complete the check-in, the passenger prints the boarding pass and then wait in line for the security screening, and the guide manually directs the passenger to the security screening lane with fewer people in the queue. (2) Frontline security inspector manually checks the passenger's boarding pass and related information on the identity card, and verifies that the passenger, the ID document, and the ticket, are consistent. (3) The passenger's identity is not tied to the passenger's baggage. To determine the owner of the baggage, one can only manually review the surveillance video of each security screening lane one by one. (4) After arriving at the boarding gate, the passenger is required to show the boarding pass, and the boarding can only be carried out after passing the manual verification.

The shortcomings of the processes of the existing security screening mode are: (1) Printing boarding passes when checking-in is environmentally unfriendly. In addition, excessive time is consumed due to queuing of passengers, and the workload of the airport ground service staff is increased. (2) Extended work time leads to visual fatigue of the security inspectors, which means the accuracy and efficiency of the verification can't be guaranteed, and missed screenings are prone to happen. (3) The security screening guide manually directs the passengers to the queue with fewer people. However, the guide's estimation of the queue length may be inaccurate, which leads to some queues being too long. (4) Because security screening lanes are crowded and busy, missed screenings are prone to occur when passengers are checked manually. (5) The pairing of passenger and baggage is performed by manually inspecting security surveillance monitoring videos for each lane, causing

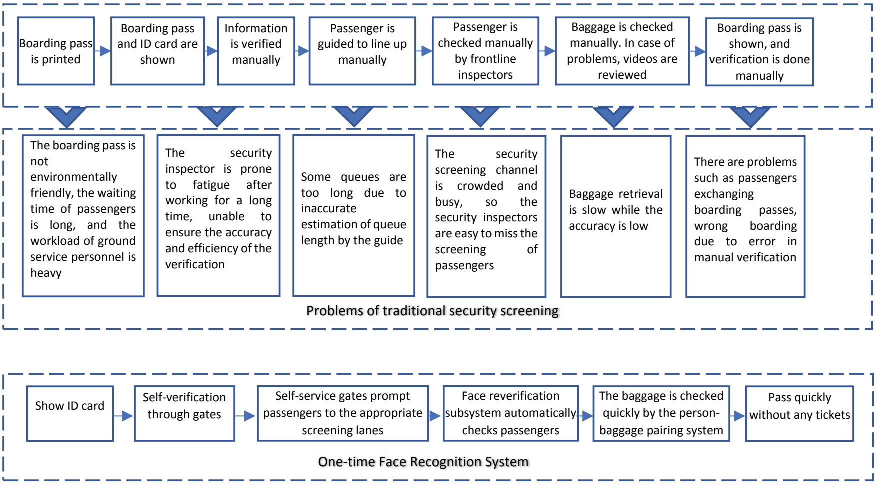


Fig. 9 The processes comparison between the two modes of One-time Face Recognition System and the existing security screening

low re-examination speed and low accuracy. (6) Manually checking passenger identities at the boarding gates does not prevent problems such as passengers exchanging boarding passes, or passengers boarding the wrong plane.

The processes comparison between the two modes of One-time Face Recognition System and the existing security screening is shown in Fig. 9.

The pros and cons comparison in practical applications of One-time Face Recognition System mode and the existing security screening mode is shown in Table 1.

2.2.3 Features of the One-Time Face Recognition System

The “One-time Face Recognition System” will completely change the manual security screenings of traditional civil aviation airports with new technological means, develop a modern whole-process smart security screening mode, actualize the whole-process smart security clearance of civil airport security screenings, improve airport security, improve screening efficiency, reduce the costs, improve service quality, and pilot innovation and reform.

The One-time Face Recognition System has the following advantages:

(1) Passenger information flight information is fully utilized, and face information is used as electronic certificate, which avoids the problems of being denied for boarding because passengers lose certificates due to loss of boarding pass or mobile phone. (2) Person-ID-ticket verification gates are used to verify the identity of the passengers, which increases security margin, improves the efficiency of security screenings, and saves the labor costs of the verification posts. (3) Self-help prompt

Table 1 Analysis of the pros and cons in One-time Face Recognition System mode and the existing security screening mode

One-time Face Recognition System	Existing security screening mode
Pro: non-contact. Passengers do not need to contact the devices directly, the recognized face image information can be proactively acquired, and the passenger only needs to show the ID card once, and then can use the face through reverification and boarding, which significantly improves the passenger experience	Con: certificates such as ID card or boarding pass need to be shown at each node, making the process tedious, and the passenger experience is not ideal
Pro: objectivity. Reduced dependence on human factors, high inspection efficiency and accuracy	Con: the inspection work mainly depends on the subjective judgment of security personnel, and personnel are easily affected by work state and business capabilities
Pro: information collection and dataization. It is beneficial to precise passenger management and personnel management in the future	None
Pro: low cost. Only one camera is required to complete the information collection, which is sufficient for subsequent calculations	Con: high labor costs, high workload during peak hours, uneven work quality, and hidden safety problems
Pro: high recognition rate in verification and reverification, which can diminish the problems such as missed screenings that may arise from the original mode	None
Con: special people cannot pass the self-service verification gate	Pro: personnel can deal with uncommon situations flexibly and respond quickly to emergencies
Con: it takes a certain amount of time and an open mind for passengers to get accustomed	Pro: passengers are already familiar with and accustomed to the original mode

function of the gates is used to intelligently guide passengers to the appropriate security screening lanes, which can improve the utilization rate of the security screening lanes. (4) Automatic reverification of passengers at security screening lanes and boarding gates can effectively avoid missed screenings caused by human factors. (5) The pairing of passengers and baggage can improve the efficiency of baggage inspections, and passengers at high risk are checked extensively.

The practical application of the One-time Face Recognition System at airports can achieve the following goals:

(1) Efficient management by an intelligent system

The application of this system will further deepen the concept of airport zoning and hierarchical security management, strengthen the identity check of boarding passengers, effectively control personnel in the restricted area, and make security work more efficient and thorough.

(2) Comprehensive security coverage

The One-time Face Recognition System forms a whole process smart security screening mode that will further improve security screening efficiency. Through self-service verification, manually assisted verification, face reverification, boarding reverification, among other measures, to achieve intelligent full coverage of passenger security screening processes, and to reduce dependence on human factors.

(3) Significant improvement in passenger experience

The One-time Face Recognition System forms a whole process smart security screening mode that collects, verifies, and reverifies information through non-contact methods, providing higher-quality and more efficient business processes, eliminating the tediousness and complexity of the original mode, greatly improving the efficiency of passenger traffic, saving the travel time of passengers, making the passenger experience more convenient, and significantly improving travel satisfaction.

(4) Optimization of human resource allocation

It introduces systematic working methods, guides optimized allocation of human resources through the system, lays a foundation for the deployment of special emergency forces for security screenings without increasing costs, and creates a scientific, effective and reliable security screening working mechanism. The use of this system can effectively reduce the work intensity of security screening staff, help them improve their work efficiency, save labor costs, and make human resources arrangements more flexible and reasonable.

3 Proof of the Application of One-Time Face Recognition System

At the beginning of 2019, with the official approval of the CAAC, the One-time Face Recognition System has been fully evaluated at all security screening lanes of Hohhot Baita International Airport (Fig. 10), and has been tested at a portion of security screening lanes at some international airports such as Beijing Capital, Shanghai Pudong, Guangzhou Baiyun, Changsha Huanghua. It is currently the only whole-process "face-only" security clearance system that is officially approved to test run by CAAC. The application of this system has effectively promoted the reform of civil aviation smart security screenings, promoted the construction process of smart airports in China, and made substantial scientific and technological contributions to the construction strategy of a strong civil aviation country. The effects of the application demonstration in Hohhot Baita International Airport are summarized as follows.

Fig. 10 Hohhot Baita international airport security screening entrance

3.1 Analysis of Application Demonstration at Hohhot Baita International Airport

The application demonstration of One-time Face Recognition System at Hohhot Baita International Airport mainly includes person-ID-ticket self-verification, face reverification at security screening lanes, and face reverification at boarding gates.

3.1.1 Person-ID-Ticket Self-Verification

(1) Installation and deployment environment

Person-ID-ticket self-verification system, namely the double-door multi-lane person-ID-ticket self-verification gates, are installed near the security screening area between every two check-in islands at the airport.

The number of lanes for the gates is determined by the number of security screening lanes it contains. According to regulations, currently 4 sets of gates are deployed on-site at Hohhot Baita International Airport, with a total of 8 self-verification lanes.

A certain area needs to be reserved between the gates and the entrance of security screening lanes, and a hard isolation facility should be set up to form a verification area.

(2) Staffing standards

Self-verification gates require order-maintenance personnel. A set of gates in operation requires one order-maintainer, who guides passengers to use the gates properly. Order-maintainers are supposed to be familiar with the protocols of the use of the devices, and check routinely to ensure the normal operations of the devices.

(3) Operational data analysis

Through the test of the actual operation of the system, in the process of person-ID-ticket self-verification gates part, the pass rate of the verification is not less than 99.56%, and the average time for a single person to pass the gates is less than 4.88 s. True tailgating alert rate is stable at 100% (Table 2).

Table 2 Person-ID-ticket self-verification gates server operation data

Date	Number of passengers to be verified by self-verification gates (number of ID cards presented)	Number of successful verifications (number of passengers passing through Gate A)	Number of passes for person-ID comparison (number of passengers passing through Gate B)	Success rate for person-ID comparison (%)	Average time for one person to pass through the gate (s)	True tailgating alert rate (%)
20190514	13,752	12,845	12,800	99.65	4.73	100
20190515	14,779	14,477	14,423	99.63	4.79	100
20190516	14,911	14,529	14,465	99.56	4.68	100
20190517	15,401	15,034	14,987	99.69	4.82	100
20190518	14,513	13,972	13,913	99.58	4.70	100
20190519	14,559	13,996	13,949	99.66	4.87	100
20190520	15,276	14,762	14,706	99.62	4.74	100
20190521	14,430	14,022	13,971	99.64	4.76	100
20190522	14,936	14,449	14,403	99.68	4.82	100
20190523	14,748	14,340	14,290	99.65	4.88	100
20190524	15,457	15,041	14,999	99.72	4.81	100

Table 3 Operational data of the server for face reverification gates at security screening lanes

Date	Number of passengers in the database	Number of passengers with successful face reverification	Number of passengers released through manual reverification	Face reverification pass rate (%)	True tailgating alert rate (%)
20190514	15,059	14,584	15	99.90	100
20190515	16,449	16,054	33	99.79	100
20190516	16,346	15,859	16	99.90	100
20190517	16,931	16,412	14	99.91	100
20190518	16,322	15,800	6	99.96	100
20190519	15,787	15,376	14	99.91	100
20190520	17,071	16,369	21	99.87	100
20190521	15,851	15,071	21	99.86	100
20190522	16,459	15,465	24	99.85	100
20190523	16,511	12,376	6	99.95	100
20190524	17,074	15,708	6	99.96	100

3.1.2 Face Reverification at Security Screening Lanes

(1) Installation and deployment environment

The system devices are placed at the end of passenger security screening lanes, including single-door security screening reverification gates and security screening information query terminals.

(2) Operational data analysis

According to the actual operation data at the airport, security screening reverification gates all have the reverification pass rate not lower than 99.79%, with no more than 0.21% requiring manual release. True tailgating alert rate is stable at 100% (Table 3).

At the same time, the average number of passengers going through this process is 157, and the average passenger queues for about 7.15 min, as shown in Table 4.

3.1.3 Face Reverification at Boarding Gates

(1) Installation and deployment environment

The equipment is placed at the boarding gates of the airport terminal. Each boarding gate is equipped with a boarding gate face reverification terminal and a whole-process information query terminal.

Table 4 Server operation data for passengers passing through the check

Date	Average passenger queueing time (min)	Number of passengers in reverification database	Average number of passengers going through the check
20190514	6.54	14,584	147
20190515	6.54	16,054	165
20190516	6.54	15,859	146
20190517	8.06	16,412	157
20190518	7.86	15,800	159
20190519	6.85	15,376	163
20190520	8.28	16,369	141
20190521	6.59	15,071	163
20190522	7.31	15,465	161
20190523	7.3	12,376	162
20190524	6.78	15,708	164

(2) Operational data analysis

Through the analysis of single-day data at the boarding gate verification stage, we found that the average single day comprehensive pass rate of the boarding gate intelligent clearance system was 84.30% (Table 5).

It can be seen from the analysis of the above data that through the seamless integration of four stepwise verification processes of gates, manual assistance, security screening lanes reverification, and boarding gates reverification, the time-consuming single person-ID comparison and verification is distributed in a multi-step, procedural verification with differentiating algorithms, gaining passenger behavior data in the first three verification steps, providing effective data support for further risk analysis, not only improving the quality and reliability of security screening verification but also providing passengers with better and more efficient travel services.

In summary, the One-time Face Recognition System can avoid the unstable quality of the verification work caused by human factors, and effectively prevent passengers from using fraudulent ID documents to enter the restricted area. At the same time, a face recognition reverification system is deployed at the security screening lanes and boarding gates to effectively prevent missed screenings and fare evasion. The system fully records the identities of transfer and transit passengers, and reverifies the identities of all passengers at Hohhot Airport, greatly improving the security level and achieving full coverage of the designated control area. When passengers are familiar with the self-service verification processes, only a small number of on-site order-maintenance personnel are required to ensure the smooth completion of security screening verification processes. At present, the average pass time of a group of person-ID-ticket self-verification gates is 6 s, which has greatly improved the efficiency of person-ID comparison. The existing passing time of 6 s is the average

Table 5 Statistics of fast pass verification of boarding gate

xxxxxx (some flight) date: 2019/x (month)/x (day)		
Total boarding number	143	Departure information system shows departure from HET: 143 (baby 0, child 0) Transit: 0
Number of people recorded in the system	134	Number of people recorded in the system = number of people checking in the security screening information system + number of people collected from (emergency counters, transfer counters, and transit counters) Security screening information system found people: 127; transfer: 7 people; transit: 0 people
Number of people in the database	121	Number of people in the database = inserted by security screening lanes + inserted through (transfer, transit) Inserted by security screening lanes: 114 Inserted by transfer: 7
Difference in number of people	13	No faces were detected for 13 passengers
Successfully verified	102	
Failed to be verified	26	1. No qualified face information is acquired in verification stage 2. Recognition score for side face is lower than the threshold
Failed to detect faces	15	No faces were detected in VIP channels, passengers with side faces, looking down at cell phones, occlusion
Comprehensive pass rate	84.30%	Comprehensive pass rate = successfully verified/number of people in the database
Transfer pass rate	71.43%	Transfer pass rate = number of recognized transfer passengers 5/number of transfer passengers in the database 7
Transit pass rate	None	Transit pass rate = number of recognized transit passengers/number of transit passengers in the database

time of the trial phase. Increased passenger familiarity and subsequent optimization of the gate and the processes can shorten the passage time to 5 s, making it maximally efficient. Both the security screening lanes and the boarding gates use non-perceptual identification, where passengers can complete the reverification at walking speed. With improved security, passengers can quickly and efficiently pass security screenings.

3.2 *Effectiveness of Application Demonstration of the Services*

The application of One-time Face Recognition System at the airport has effectively promoted the improvement of airport safety margin, operation efficiency and passenger satisfaction, and its whole-process smart security screening has changed the airport's long-time original security screening mode based on manual inspection. This has saved labor costs and laid a good foundation for smart security at airports in China, which will effectively promote the intelligent transformation of civil aviation security screening and safety measures. The application demonstration of One-time Face Recognition System at the airport has achieved "three improvements" and "two reductions".

3.2.1 Three Improvements

The first is to increase the aviation security safety margin. The One-time Face Recognition System security clearance mode further enhances the security screening process control capabilities. Through the security clearance mode enhanced with various reverifications, the coverage of security screenings and verifications have been expanded. The identity check of boarding passengers is strengthened, and boarding passengers are effectively controlled. It improves the accuracy rate of person-ID identification and reduces the dependence on human factors. Through the integration with the security screening information management system and the optimization of a series of functions such as the deployment of front-line control personnel, it has achieved a high degree of integration of manual security and technological security, effectively improving the airport's safety management and control margin.

The second is to improve the efficiency of security screening operations. It can be seen from the operational data that the values are greatly improved compared to the traditional manual mode. The average time for checking and releasing passengers at each person-ID-ticket self-verification gates is about 6 s, which is 6–14 s less than the traditional manual verification of 12–20 s. The average waiting time of passengers in front of the lanes was shortened from the previous 8.8–7.15 min, which was shortened by 1.65 min on average. The average passenger passing rate increased from 129 people per hour to 157 people per hour, with an increase of 28 people per hour.

The third is to improve the passenger experience in security screenings. Manually assisted verification security clearance mode automatically collects, verifies, and reverifies information on passengers through "contactless" security measures, eliminating the tedious and complicated processes of the original manual mode, and passengers are no longer "gazed" in the verification, which greatly saves time for passengers to pass through the checks. What's more important is that the passengers' "non-perceptual" experience is more comfortable, which greatly improves passengers' experience of passing security screenings, and will increase passengers' travel satisfaction.

3.2.2 Two Reductions

The first is to reduce the probability of a passenger complaining about security screenings. The change of the One-time Face Recognition System security clearance mode reduces the degree of intrusion of the verification staff in the passenger verification process. Passengers change from "passive verification" to "self-service verification", allowing passengers to have improved experience using security clearance services. The improvements greatly reduce the number of dissatisfactions with the security screening processes, and effectively reduce the probability of passengers complaining about security screenings.

The second is to reduce labor costs. The implementation of the One-time Face Recognition System workflow can guide the optimal allocation of human resources and create a scientific, effective and reliable security screening working mechanism without increasing costs. By partially canceling the verification counters, reducing the configuration of verification security inspectors and manually assisted system verification, the work intensity is reduced and the work efficiency is improved, which provides tangible technical support for saving airport labor costs.

4 Summary and Outlook

In China's civil aviation, tremendous growth potential and industry is working with the government to achieve it. It is continuously advancing towards the established goals of airports with "four-style" of "safe, green, intelligent, and humanistic". "One-time Face Recognition System" is just a stage on the journey. With the gradual maturity of various new technologies, more advanced technologies will be incorporated into it. Foreseeable ones include "pre-queuing system as a guide", "more convenient people and package distribution system", "more user-friendly unnoticeable security screening" and so on. China is playing an important role in this demand-led, technology-driven smart security reform. China Civil Aviation is undergoing the role change from follower to parallel walker and front runner. In the future, China Civil Aviation will continue its efforts to comprehensively use new technologies such as biometrics, artificial intelligence, big data analysis, 5G integrated communications, Internet of Things, cloud computing, GIS/BIM, to ensure the safety of aviation travel, strive to improve passenger travel comfort, convenience, and efficiency, and "shape the brand-new future of the civil aviation industry with wisdom".

References

1. China Civil Aviation Administration (1999) Rules for China civil aviation safety screening. Beijing, p 1
2. Han J (2010) Responding to the crisis: new changes in US airport security checks. Air Transp Bus 2010(13):38–40
3. The International Air Transport Association. Passenger experience & facilitation [EB/OL] [2020-03-28]. https://www.iata.org/whatwedo/passenger/pages/index.aspx
4. China Civil Aviation Administration. Statistical bulletin of civil aviation industry development 2018 [EB/OL] [2020–03–28]. https://www.caac.gov.cn/XXGK/XXGK/TZTG/201905/t20190508_196035.html
5. The International Air Transport Association. IATA and ACI launch new experience in travel and technologies (NEXTT) [EB/OL] [2020–03–28]. https://www.iata.org/pressroom/pr/Pages/2017-10-24-03.aspx
6. Sakano R, Obeng K, Fuller K (2016) Airport security and screening satisfaction a case study of U.S. J Air Transp Manag 55(8):129–138
7. Nie X, Batta R, Drury C et al (2009) Passenger grouping with risk levels in an airport security system. Eur J Oper Res 194(2):574–584
8. Jose R, Daniela M, Cristina C et al (2016) Automated border control e-gates and facial recognition systems. Comput Secur 62(1):49–72
9. Sofia K, Voula P, Filipe M (2015) Future airport terminals: new technologies promise capacity gains. J Air Transp Manag 42(1):203–212
10. Zhao Y, Chen X, Zhang Y et al (2011) Automatic explosive detecting system for passenger baggage. China Civ Aviat 126(06):64–66
11. The International Air Transport Association Smart Security getting smarter [EB/OL] [2020–03–28]. https://airlines.iata.org/analysis/smart-security-getting-smarter

Yu Shi, Professorate Senior Engineer at the Chongqing Institute of Green and Intelligence Technology, Chinese Academy of Sciences, Director of Research Center for Intelligent Security Technology. His research interest covers software engineering, artificial intelligence and computer security. Relevant achievements have brought technological transformation in the fields of civil aviation and security with significant economic and social benefits. He has won Technology Promotion and Development Award of CAS in 2019.

Urban Resources, Environment, and Ecology (UREE) Big Data Platform: Construction and Application

Weiqiang Chen, Nan Li, and Yupeng Liu

Abstract The Urban Resources, Environment, and Ecology (UREE) big data platform stores and integrates comprehensive urban data and provides real-time data management and application. A seven-layer architecture UREE was established based on open and in-depth data and state-of-the-art technologies, including data acquisition and standardization, data query and visualization, data unified interface and fusion, and urban metabolism simulation. The UREE presents cutting-edge researches at the forefront of science on sustainable development. It has integrated ten urban thematic datasets from more than 300 Chinese cities and 100 global cities, most of which can be updated timely and regularly. This platform provides powerful technical support for monitoring dynamics and understanding mechanisms of urban resources, environment, and ecology. It can further provide theoretical and practical support for alleviating the Urban Problem in urbanization, such as resources structural shortage and environmental deterioration. This study is a new case on guiding data-driven discipline development, especially in the field of urban environment and ecology. Development, improvement, and application of the UREE platform can form a data-intensive science and adaptive paradigm of resources management.

Keywords Urban environment · Urban problem · Big data platform · Data fusion · Urban metabolism simulation · Sustainable development

1 Introduction

By the end of 2018, the urbanization rate of China had reached 59.58% [1]. Rapid urbanization has brought a series of social and environmental dilemmas: urban heat islands, traffic congestion, solid waste, air pollution, lack of essential services and facilities. Urban sustainability has become the most significant urban development issue in the world [2]. In 2015, the United Nations proposed the development goal of Sustainable Cities and Communities in Transforming our world: the 2030 Agenda for

W. Chen (✉) · N. Li · Y. Liu
Key Lab of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen, Fujian, China
e-mail: wqchen@iue.ac.cn

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_24

Sustainable Development, which means building inclusive, safe and resilient sustainable cities and human communities [3]. Meanwhile, documents have been published successively in China, such as *Overall plan of ecological civilization system reform*. They formulated more than 40 reform projects related to the construction of ecological civilization, covering urban society, economy, environment, ecology and many other aspects [4, 5].

The research of sustainable urban development and the construction of ecological civilization requires careful consideration of several different factors and accurate, reliable data in multiple aspects [6], including population, land, economy, resources, environment, and ecology. Traditional urban thematic data, due to its severe time lag, coarse spatial granularity and a high degree of statistical aggregation, is difficult to directly provide sufficient support for the management decisions of rapid urban development [7]. In recent years, overwhelming urban data has been produced, with the rapid development of the Internet of Things (IoT), remote sensing, cloud computing, and other technologies [8, 9]. Compared with big data in other fields, urban big data in many situations, such as natural disaster information, environmental pollution, traffic congestion, has the more robust real-time feature, the higher spatial accuracy, and more diverse sources. Besides, the multi-source heterogeneous data of urban themes brings significant challenges to the traditional data storage and processing technologies in many aspects, such as the changes and scalability of data type and storage, and operation response speed. Only by choosing the suitable big data framework and technology to improve the efficiency and processing scale of the algorithm, can UREE achieve the storage, batch and real-time computing of massive data.

Meanwhile, with the continuous accumulation of urban big data, and the extensive engineering application of machine learning, especially Artificial Intelligence (AI), urban big data can provide a data-driven perspective. It can be directly utilized to the study of sustainable development issues (urban planning, resident health, environmental pollution, transportation planning, energy consumption, etc.) [10–13]. Therefore, the research team of Chen Weiqiang, Institute of Urban Environment, Chinese Academy of Sciences, has established a full-scale urban thematic data platform (Urban Resources, Environment, and Ecology, UREE). We systematically integrated multi-source heterogeneous data and realized real-time management and application of data.

2 UREE Big Data Platform Architecture

2.1 Platform Architecture

Based on open, multi-source heterogeneous UREE data, the UREE is mainly composed of two core services: data acquisition and standardization service, business application service. As presented in Fig. 1, the UREE collects both network open data and platform log data. Then, these data would be cleaned, transformed, fused,

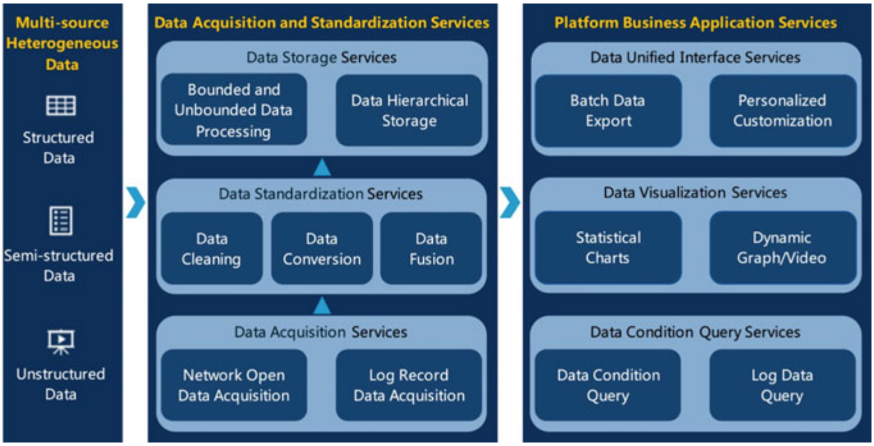


Fig. 1 Business framework of the UREE

and stored into data centers by several technologies. After that, these data would be mined and analyzed by appropriate data computing and analytics models. Finally, the UREE provides data query services, data visualization services, and data unified interface services.

2.2 Core Function

2.2.1 Data Acquisition and Standardization Services

(1) Data acquisition services

The UREE collects heterogeneous UREE data from open and various sources and the platform log data in real-time. Moreover, these data would be transmitted to the data center for storage. Data mechanisms, such as real-time updates, secure transmission, and sharing, are established to realize the rapid exchange, sharing, and storage of massive data between data centers. The datasets collected include Open Street Map (OSM), OSM Building, urban material metabolism, urban minerals, urban air quality, urban water environment, municipal solid waste, urban green space, and urban environmental events and tone. Moreover, the platform collects metadata of various urban thematic data and tracks the inheritance of fusion UREE data from the aspects of data sets and fields. Based on metadata descriptions of urban thematic data, it lays the foundation for more accurate and fast retrieval of demand data.

(2) Data standardization services

Within the UREE, we construct standardized data standards based on the core themes of the city to uniformly clean and transform the UREE data of different sources, formats and characteristics. Then, manual working and algorithm fill missing data, correct erroneous data and remove redundant data to provide complete, timely and accurate high-quality data for further data analysis and mining.

Based on the standardized UREE data, mapping rules of heterogeneous data are constructed in UREE to integrate multiple data sources into a single data source, and fuse and associate multi-source heterogeneous data. For example, based on taking the longitude and latitude information like the mapping standard, urban POI data, urban real-time population data and urban geographic information data can be fused to mine and analyze the impact of urban POI on population distribution. The primary purposes of data standardization services are to eliminate the problem of the information island between urban data and to improve the convenience of data queries and analytics.

(3) Data storage services

UREE stores both raw and fusion data. Based on data-access frequency, we design a cold-hot data layered storage mechanism to build hot data storage layers and cold data storage layers. According to the threshold of data-access frequency set in the platform, UREE stores the urban thematic data with different access frequencies to the hot storage layer or the cold storage layer. Furthermore, a hybrid index algorithm is designed for a dynamic combination of multi-source heterogeneous data, combining with the technology of bounded and unbounded data processing. It is efficient to integrate and manage heterogeneous data from various sources, such as social surveys, remote sensing, IoT, Internet, and mobile devices. Thereby, it would solve data problems in the process of multi-source heterogeneous data fusion and storage, such as dynamicity, scalability, fault tolerance, heterogeneity, consistency. It is essential to point out that we do not use platform data for commercial publishing. The purposes of data acquisition, standardization, and storage services are to provide a unified data access model for data mining and research of urban resources, environment, and ecology.

2.2.2 Platform Business Application Services

(1) Data condition query services

The UREE manages user information and permissions by building a unified authentication center. After login and authentication, users can access the urban thematic data within their permissions. Besides, they can query data online in selected themes according to the requirements. The query conditions mainly include nation, region, and time. Different thematic data have their specific query conditions. for example, the query condition of meteorological data includes the weather station, precipitation, temperature, humidity, wind direction, wind speed. Finally, the platform can return the standardized two-dimensional table data.

Meanwhile, log services are provided in the UREE to record the access frequency of datasets. Data statics and analysis with their metadata, such as granularity, time, volume, can provide a new perspective to track the research hotspots of urban resources, environment, and ecology in real-time.

(2) Data visualization services

The UREE performs the in-depth mining of data through statistical analysis and association analysis. Data fusion, analysis, and visualization are applied for various urban thematic data, including OSM Building, urban material metabolism, urban minerals, urban air quality, urban water environment, municipal solid waste, urban green space, and urban environmental events and tone, to display the potential value of data more intuitively.

The platform presents the data analysis results in the form of statistical charts, thematic charts, etc., which can be suitable for large-screen display. Meanwhile, it can show the dynamic changes of UREE data in different periods, countries and regions with dynamic graph/video. Users can view the platform data visualization through both mobile and Web. They also can send visualization requests with personalization query conditions. For example, users can view visualizations of Global Trade Data according to serval conditions, including time, country, imported goods/services, exported goods/services, goods/services amount, and goods weight. Besides, the visualization display pages can present query and download of statistical data, document data and image data.

(3) Data unified interface services

Based on the integration of massive multi-source heterogeneous urban thematic data, the UREE provides unified data access and download services and shares some urban thematic data by constructing RESTful API. The data availability and flow efficiency can be improved comprehensively by providing the RESTful API. Besides, platform users can obtain the data export permissions after application, review and other business processes. Authorized users can get the standardized fusion data in the form of batch download or offline distribution within their permissions. Moreover, authorized users can request personalization services from the UREE, including data conversion, thematic visualization, and other services. The UREE strives to provide data and technical support for researchers to analyze dynamic changes and driving mechanisms of urban resources, environment, and ecology.

2.3 Key Technology

The construction of the UREE adopted the front and back end separation. Multiple cloud resources were uniformly operated based on a hybrid cloud consisted of public cloud (Alibaba Cloud) and private cloud, as shown in Fig. 2. The UREE introduced Docker-based microservice architecture. The platform realized the DevOps by splitting various microservices and deployed both on cloud and local servers. Container services control the resources on cloud and local computers: the local computers provide the services, and the cloud processes the disaster recoveries. Load balancing can guide users to seamlessly use different virtualization services both in the public cloud and private cloud when users utilize any devices to access the UREE. When data accesses to the peak of the UREE, the cloud resources can expand horizontally and vertically by the architecture with elastic expansion automatically. Thereby, the high load functions can respond quickly.

The architecture of the UREE has seven layers based on microservice. From bottom to top, the architecture includes data sources layer, data acquisition layer, data storage layer, data analysis layer, microservice business layer, gateway load balancing layer, and business access layer, as shown in Fig. 3.

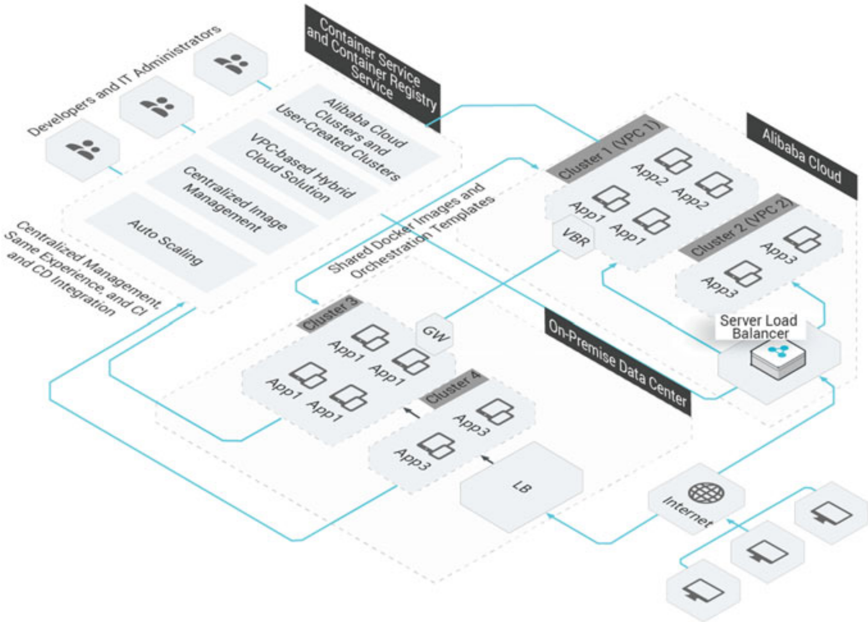


Fig. 2 Hybrid cloud architecture of public cloud (Alibaba cloud) and private cloud [14]

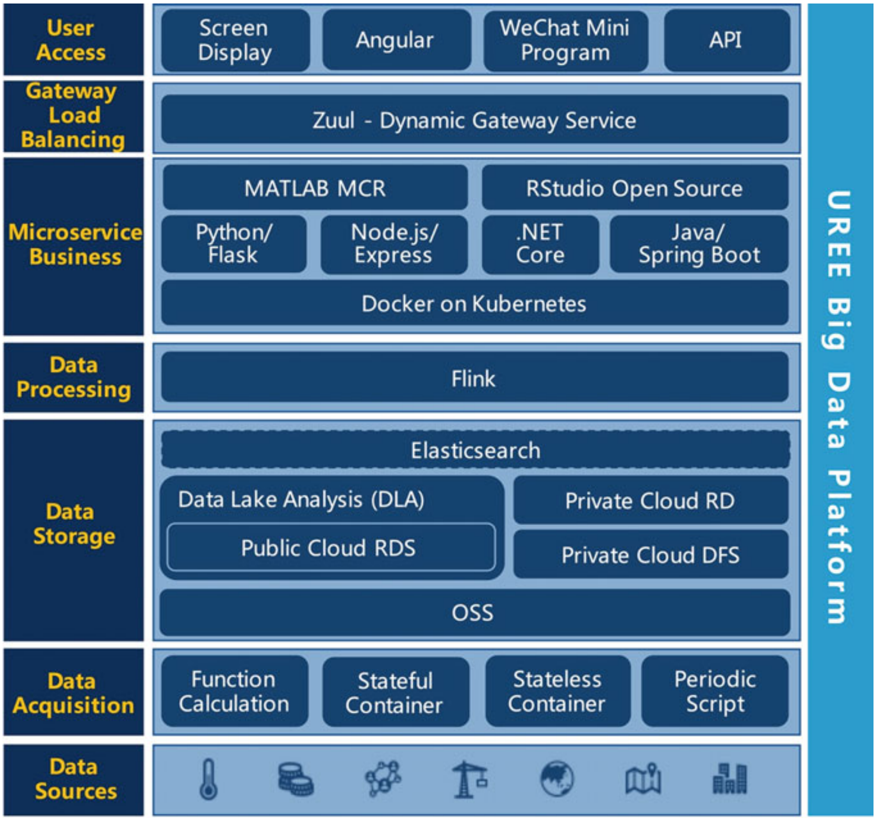


Fig. 3 Technical architecture of the UREE

2.3.1 Data Sources Layer

There are many sources of UREE data, including social surveys, remote sensing, monitoring, GIS, IoT, Internet, and mobile devices. Heterogeneous network data and statistical yearbook declared in public are the primary sources in the UREE by investigating various heterogeneous big data sources. For example, (a) the meteorological data released by China Meteorological Observatory, (b) the air monitoring data and water pollution monitoring data released by the Ministry of Ecology and Environment of the People's Republic of China, (c) the international trade data released by the United Nations, (d) the global urban road network and building data released by the OSM, (e) real-time population distribution data released by Tencent, (f) urban environmental events and tone data released by GDELT, etc.

2.3.2 Data Acquisition Layer

The acquisitions of open UREE data include real-time streaming data acquisition and periodic batch data acquisition. The UREE sets various types of data source interfaces, uploads program code to the public cloud. Furthermore, it uses function calculation of public cloud, which combined with the stateful and stateless containers, to realize event-driven automatic data acquisition without the management and operation of infrastructure (servers, etc.). These streaming data include urban real-time population distribution data, urban real-time air monitoring data, urban real-time water monitoring data, urban real-time meteorological data, global urban real-time environmental events and tone data, and the metadata of these data. Batch data automatically are collected by periodical scripts. These batch data include economic-system material flow and resource output rate data, global urban geographic information data, urban POI data, urban industrial enterprise data, global trade data, and the metadata of these data.

2.3.3 Data Storage Layer

There are many sources of urban thematic data, including social surveys, remote sensing, IoT, Internet, and mobile devices. The diversity of data sources leads to the different of data structure: (1) structured data stored in a relational database, (2) semi-structured data (CSV, XML, JSON, etc.), (3) unstructured data (document data, log data, etc.), (4) binary data (image, audio, video, etc.). Traditional storage technology has been unable to meet the needs of the explosive growth and the various formats of UREE data, which would severely restrict the potential application value of data.

The UREE stores the collected data to Object Storage Service (OSS) of Alibaba Could. Furthermore, we design a hybrid storage solution with public cloud and private cloud: (1) Data Lake Analysis (DLA) stores data in ORC format. The hierarchical storages of urban thematic cold data and hot data are constructed with the relational data services (RDS) of the public cloud. Finally, cross-database query, processing, and analysis can be conducted by standard SQL. (2) Periodic hot data (experimental data) are cached in the file system and relational database of the private cloud. At the top of the storage layer, the distributed search and analysis engine Elasticsearch (ES) services are used to get real-time full-text retrieval, ad hoc visualization and analysis of data.

2.3.4 Data Processing Layer

Within the UREE, the traditional processing mode, which needs two different architectures to process the unbounded and bounded data, is overturned by a framework and distributed processing engine (Apache Flink). Apache Flink can process the unbounded and bounded data by using just a big data processing engine or writing a systematic code. The thematic UREE data are fused and processed as bounded data streams by constructing the mapping rules of multi-source heterogeneous data, including data fusion and processing based on multiple data platforms and domains, data fusion and processing based on data characteristics, data fusion, and processing based on data semantics. The integrative processing of unbounded and bounded data in the data processing layer of the UREE can effectively reduce the delay of urban thematic data, which can reach the second or even sub-second delay. It can provide technical support to ensure the data characteristics of real-time, stability and sharing.

2.3.5 Microservice Business Layer

The UREE applied the microservice architecture to establish the business layer. It automatically deployed the docker container to build a virtual environment based on Kubernetes (k8s), which can be compatible with various programming languages, including Python, Node.js, .NET Core and Java. The UREE can select the most appropriate language and framework for different business needs. They include (1) Node.js is used to solve some highly concurrent problems, for example, visualization of massive location data uploaded by thousands of GPS devices in real-time. (2) The workflow system establishing uses the spring Boot framework of .NET Core or Java, including user information management, log information management, and other business modules. (3) Python is applied to call the TensorFlow Artificial Intelligence framework, image processing, text processing, and other services. Moreover, the UREE can deploy the R and MATLAB code in the Docker environment. Specifically, MATLAB MCR is the production environment of MATLAB, and RStudio Open Source is the production environment of R. They can implement the data model calculation and calculation result output.

2.3.6 Gateway Load Balancing Layer

The UREE constructed Zuul API Gateway of the Asynchronous Call Model. Because this framework uses an asynchronous non-blocking programming model, it can carry a large number of connections under limited resources. It is very suitable for the scenarios of IO-intensive applications in the UREE, for example, the periodic ETL of global trade data and real-time population distribution data.

2.3.7 User Access Layer

The UREE can seamlessly combine with various open-source business analysis tools and applications. Baidu Echarts and D3.js are the essential tools or applications in the platform for the visualization of various thematic data, and the visualization results can be suitable for large-screen display. The platform establishes a responsive web page with Angular. Authorized users can access the platform, which can be automatically suitable for their devices (mobile, Web, etc.). Besides, Users can utilize the Wechat Mini Program to access UREE data and data visualization.

3 UREE Big Data Platform Construction

3.1 UREE Data Resources Construction

The goal of the UREE is to collect and integrate UREE data of all cities in China and even global cities. The platform has collected 1 TB data covering ten urban themes. These data include urban meteorological data, global trade data, global urban environmental events and tone data, economic-system material flow and resource output rate data, urban industrial enterprise data, urban water monitoring data, global urban geographic information data, urban air monitoring data, urban POI data, and urban real-time population distribution data, as shown in Table 1. In future data resource construction, we will gradually collect more thematic data, such as OSM Building, urban material metabolism, urban minerals, municipal solid waste and urban green space.

Within the UREE, we collected the urban thematic data from open sources and integrated and standardized these multi-source heterogeneous data. The purpose of these data processing is to ensure data real-time availability and to improve data utilization efficiency. We do not use platform data for commercial publishing.

3.2 Platform Business Function Construction

At present, we have completed several constructions of the UREE, such as data storage mechanism design, platform prototype design and some significant functions coding. The primary functions of the platform include user registration, user login, permission management, data ETL, model calculation, simple query, condition query, data export in CSV format, and data visualization. Based on the k8s, the UREE has already automatically deployed the Docker container and established the virtual DevOps environment. It can realize the automatic environment deployment of platform infrastructures, including Alibaba Cloud function calculation, ES, OSS, DLA, RDS and other services.

Taking condition query as an example, authorized users can query the urban thematic data after logging in the platform. These data include urban water monitoring data, urban air monitoring data, urban meteorological data, urban real-time population distribution data, urban industrial enterprise data, global trade data, and urban environmental events and tone data. The platform can present the data query results in the form of tables. Users can input the conditions to query data at the top

Table 1 List of integrated databases of UREE

ID	Database name	Data description
1	Urban meteorological database	Since 1942, meteorological data of China in International Meteorological Station, including cloud height, wind direction, wind speed, air pressure, temperature, humidity, and precipitation. More than 2 million data records per year
2	Global trade database	Since 1962, the trade goods/services data of more than 170 countries/regions in the world released by the United Nations. More than 20 million data records per year
3	Global Urban environmental events and tone database	Since 1979, environmental event information and event knowledge graph data of global urban environmental themes updated daily by GDELT
4	Economic-system material flow and resource output rate database	Since 2000, total material flow (biomass, fossil fuel, metal, non-metal) and derived data of China urban/region economic system, including local mining, import and export quantity, discharge after local disposal, output rate of total material comprehensive resources, and energy output rate
5	Urban industrial enterprise database	Since 2001,data of China urban industrial enterprise (state-owned industrial enterprises, five-million-income industrial enterprises). More than 400000 data records per year
6	Urban water pollution monitoring database	Since 2006, China weekly average water quality data of 110 monitoring points in cities/regions released by the Ministry of Ecology and Environment, including PH, DO, COD, ammonia nitrogen, and water quality grades
7	Global Urban geographic information database (OSM)	Since 2012, the global urban road network data, building data and other geographic information data released by OSM
8	Urban air monitoring database	Since 2014, China hourly air monitoring data of 1511 air monitoring points in 371 cities released by the Ministry of Ecology and Environment, including PM2.5, SO_2, O_3, NO_2, and AQI
9	Urban POI database	Since 2014, POI data of more than 300 cities in China. More than 70 million data records per year

(continued)

Table 1 (continued)

ID	Database name	Data description
10	Urban real-time population distribution grid database	Since 2016, the urban real-time (every 15 min) population distribution data of China in a grid of 1 km * 1 km released by Tencent

of the table, and they can export and download the data in the CSV format, as shown in Fig. 4.

The UREE can visualize data based on the collected ten urban themes, including global trade data, urban environmental events and tone data, urban water monitoring data, and urban POI data. The visualization of urban environmental events and tone

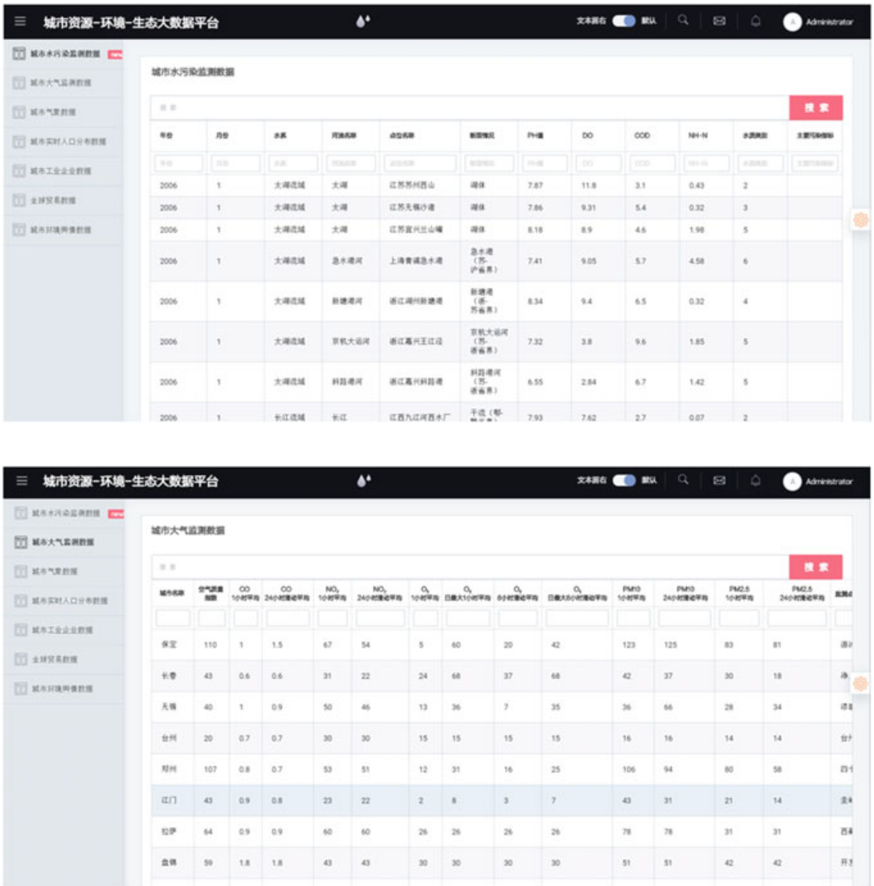


Fig. 4 Example of UREE construction: query page of urban water pollution monitoring data and urban air monitoring data

data is shown in Fig. 5. This figure presents the “tone” of environmental events in different countries and the number of times four types of events were mentioned. Also, UREE can present different visualizations according to different conditions, such as geographical area, time series, etc.

In August 2019, the internal test of the UREE will be completed. In October 2019, the UREE will be online formally. Users may register and apply for the permission of the data query and export through the portal of UREE. Furthermore, authorized

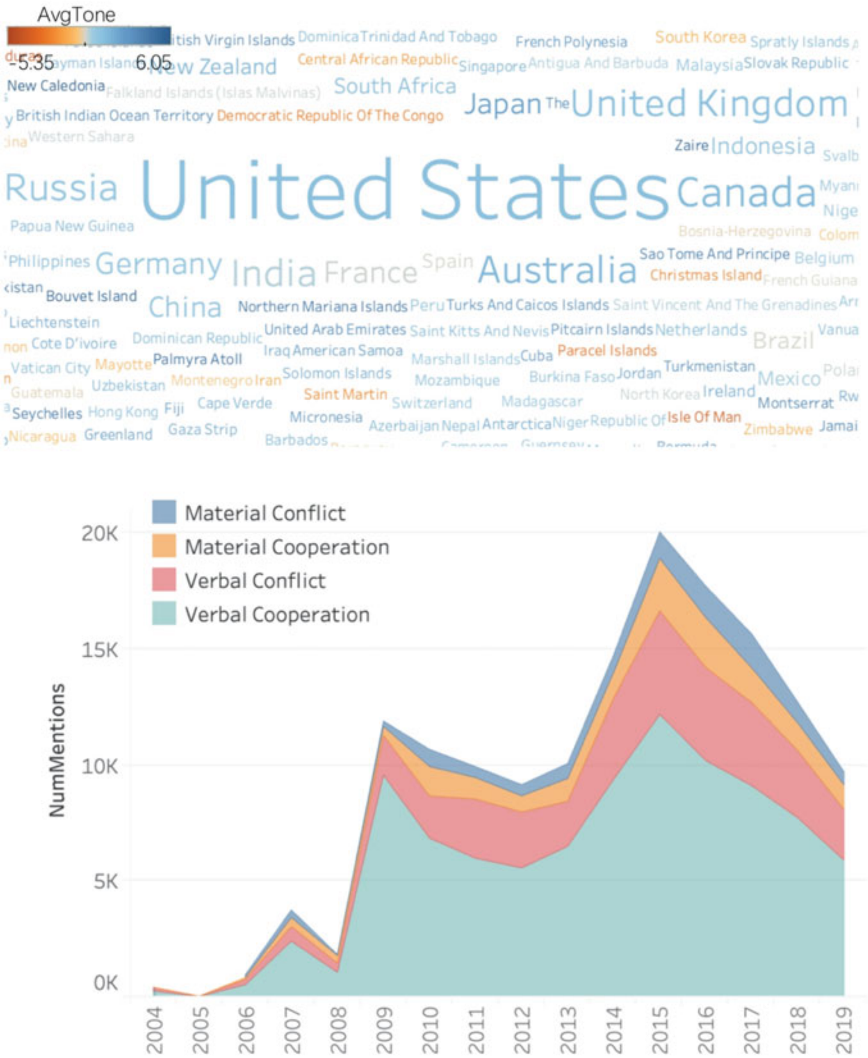


Fig. 5 Example of UREE visualization: visualization of urban environmental events and tone data

users can apply for customized services, such as data format conversion, thematic data visualization, etc.

4 Results and Prospects

The Urban Resources, Environment, and Ecology (UREE) big data platform of China is positioned as a data-driven innovation platform for urban environmental research. It hopes to provide adequate technical support for monitoring and to study the dynamic changes and driving mechanism of urban resources, environment, and ecology. The UREE can integrate massive multi-source heterogeneous data of different themes, and provide standardized data API to comprehensively improve the availability and utilization efficiency of the data for urban research. Moreover, real-time data in various urban aspects, including population, land, economy, resources, environment, and ecology, are collected, deep mined, and visually analyzed in the platform. Thereby, UREE can provide decision-making support for solving the Urban Problem (resources structural shortage, environmental deterioration, etc.) in the process of urbanization and how to build and realize sustainable cities. The UREE is an innovative platform to promote the development and research of urban environment and ecology with big data technology.

In the future, the upgrade of platform data and functions will by keeping up with the forefront of information technology and hot issues in the field of urban sustainable development research:

(1) Improve the proportion of automatic data acquisition and preprocessing, and use the heuristic algorithm to improve the efficiency of semi-automatic data acquisition and to reduce manual workload.
(2) Expand data sources of urban themes and strengthen the research and development of data fusion analysis and deep mining model algorithms across databases.
(3) Improve the accessibility of UREE data and provide multiple data and computing services, such as system access, database system access, restful API, GraphQL. Furthermore, build the UREE into the most influential sharing and computing platform for urban thematic data in the world.

References

1. National Bureau of Statistics (2019) PRC. Statstical commniuqe on the 2018 national economic and social development [R], (in Chinese)
2. Fang C, Zhou C, Gu C et al (2016) Theoretical analysis of interactive coupled effects between urbanization and eco-environment in mega-urban agglomerations [J]. Acta Geogr Sinica 71(4):531–550, (in Chinese)
3. Nation TU (2015) Transforming our world: the 2030 agenda for sustainable development [R]

4. General Office of the State Council (2015) PRC. Overall plan of ecological civilization system reform. [M], pp 5–11, (in Chinese)
5. General Office of the State Council (2015) PRC. Opinions on accelerating the construction of ecological civilization [M], pp 5–11, (in Chinese)
6. Pomázi I (2013) Green growth in cities, OECD green growth studies [M]. Paris, OECD Publishing
7. Wang Penglong, Gao Feng, Huang Chunlin et al (2018) Progresson sustainable city assessment index system for SDGs [J]. Remote Sens Technol Appl 33(5):784–792 (in Chinese)
8. Hashem IAT, Chang V, Anuar NB et al (2016) The role of big data in smart city [J]. Int J Inf Manag 36(5):748–758
9. Xiaoyu S, Jun G, Xin L (2018) Urban sustainability evaluation based on remote sensing and network data support: Progress and prospect [J]. Prog Geogr 33(10):1075–1083, (in Chinese)
10. Lim C, Kim KJ, Maglio PP (2018) Smart cities with big data: reference models, challenges, and considerations [J]. Cities S0264275117308545
11. Weiss DJ, Nelson A, Gibson HS et al (2018) A global map of travel time to cities to assess inequalities in accessibility in 2015 [J]. Nature 553:333
12. Rathore MM, Ahmad A, Paul A et al (2016) Urban planning and building smart cities based on the internet of things using big data analytics [J]. Comput Netw Int J Comput Telecommun Netw 101(C):63–80
13. Jean N, Burke M, Xie M et al (2016) Combining satellite imagery and machine learning to predict poverty [J]. Science 353(6301):790
14. Alibaba Cloud. Hybrid cloud architecture [EB/OL]. [2019-7-26]. https://help.aliyun.com/document_detail/25977.html?spm=a2c4g.11186623.6.545.dd603aa7oRkfdy

Weiqiang Chen, is currently a professor of Sustainability Science in the Institute of Urban Environment (IUE), Chinese Academy of Sciences (CAS). Before joining CAS, he was working at the Center for Industrial Ecology, Yale School of Forestry & Environmental Studies for five years. He got both his bachelor's and Ph.D. degrees in Environmental Science and Engineering from the School of Environment at Tsinghua University, Beijing. He was one of the five initiators of the "2014 Chinese Environmental Scholars Forum" held at Harvard University and the lead organizer of the "2015 Chinese Environmental Scholars Forum" held at Yale University. He is on the board of the Sustainable Urban System section of the International Society for Industrial Ecology, was the president of the Chinese Society for Industrial Ecology (July 2015-Dec. 2016), and is associate editor of Resources, Conservation, and Recycling (July. 2016-present). His research interests include industrial ecology and urban ecology, specifically on material/substance flow analysis, urban metabolism, urban mining, and global trade networks of materials. His studies have been published in PNAS, Environmental Science & Technology, and other top journals.

Present Situation and Development Prospect of the Digital Orchard Technology

Guomin Zhou and Xue Xia

Abstract Fruit industry plays a crucial role in the development of rural economy in China. The Digital Orchard is the inevitable embodiment of information technology in the fruit industry from single application to comprehensive application, and it is also the key technology and application system for comprehensive application of digital technology to study information acquisition, processing, management and utilization during orchard production, management, operation, circulation and service. This paper introduced the concept and connotation of the Digital Orchard, summarizes the research and application status of the Digital Orchard technology, and looks forward to the future trend and key development direction of the Digital Orchard.

Keywords Digital orchard · Orchard informatization · Digitization technology

1 Introduction

In China, fruit industry is the third largest planting industry that after grain and vegetable and it plays an important role in the development of rural economy. With the rapid progress of urbanization, the number of rural people engaged in agricultural production has decreased significantly, so developing industrial-scale farming and smart agriculture has become an inevitable direction to improve agricultural production efficiency and level in the future. The "China's 13th five-year plan (2016–2020)" explicitly calls for strengthening the integration of agriculture and information technology. In addition, in the "New Generation of Artificial Intelligence Development Plan" released by the state council in 2017, important arrangements were made for the major task of upgrading the intelligent agriculture industry, and pointing out that

G. Zhou (✉)
Chinese Academy of Agricultural Sciences, Beijing, China
e-mail: zhouguomin@caas.cn

X. Xia
Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing, China

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_25

a number of agricultural integration application demonstrations, including intelligent orchards, should be carried out [1].

Digital Orchard is the concrete practice and deepening of digital agriculture concept in orchard production and management, and it is also a new research direction of modern information technology and fruit tree cultivation management. Digital Orchard research has integrated digital methods to study key technologies and application systems for information acquisition, processing, management and utilization in orchard production, management, operation, circulation and service, and finally realized the digital design, visual expression and intelligent control management of orchard from production to service [2]. From the definition of the Digital Orchard, it can be seen that the Digital Orchard focuses not on the single application of information technology, but regarded the production, management and operation of the orchard as an organic and interconnected system. The digital technology was expected to be applied comprehensively and systematically to every link of orchard production and operation system to improve orchard management, and drives the orchard production and management system to develop in accordance with the goals and directions of human needs. The development of the Digital Orchard has important practical significance to the supply-side structural reform of modern fruit industry in China, and it is also an urgent need to speed up and shorten the gap with developed countries and improve the international competitiveness of the fruit industry.

From the experience of major foreign countries that producing fruit, in order to strengthen the obligation of fruit producers on fruit safety, some developed countries have implemented the Hazard Analysis & Critical Control Points (HACCP) System and put forward requirements that full monitoring from the farm to the market. In this way, the whole production process from field to table was under the safety supervision system of agricultural products. Furthermore, many countries around the world also adopt the Global Uniform Identification System (EAN.UCC) form the International Article Numbering Association to trace the whole information of fruits [3–5]. After more than 100 years of efforts, the apple industry in Japan has made the leap from "Western Apple" to "Japanese Apple", and also realized the leap from no apple country to the country where apples are plentiful. Fruit cultivation in Japan seeks to rely on machinery to save labour and costs. In terms of fruit evaluation, the successful development of the fruit selection machine with reflection integrating sphere makes the quality difference of the selected fruit within a batch very small, which is also well received by marketers and consumers [6–8].

China's fruit industry has made great progress after years of efforts and the fruit planted area and yield are ranked first in the world, which has formed certain advantages in scale. At present, the overall level of orchard production and management in China is still far behind that in advanced countries, especially the technology of digitization, information and intelligence for orchard production and management are more backward. The development of fruit industry in China is still faced with many problems, such as low yield per unit area, poor quality, single consumption pattern, hidden dangers of fruit quality and safety, weak awareness of disease and insect pest control system and comprehensive prevention, time-consuming and laborious orchard management and low level of mechanized management [9–12]. Thus construct modern fruit cultivation technology system and develop informationized and intelligent orchard by relying on the digital technology is an effective way to break through the limitation of traditional fruit industry and solve the problem of industrial development.

This paper focuses on the concept of the Digital Orchard, systematically analyzed the connotation of the Digital Orchard, then combines the research work of our team to combed and summarized the latest progress of digital orchard, and finally the development trend and application of the Digital Orchard in the future were prospected.

2 Connotation of the Digital Orchard

Digital Orchards can be thought of as the realization of comprehensive digital and network management of biological processes, environmental processes, economic processes and other processes involved in orchards. In this process, the digital models like fruit tree growth and environment model and fruit tree structure model are the core and foundation of digital orchard research. In addition, according to the link of information technology application, the research of digital orchard also focuses on the content of digital management and control of orchard production, digital management and traceability of fruit quality and safety, digital management of fruit operation and circulation, and e-commerce [13].

2.1 Digital Model of Fruit Tree

The horticultural management of orchards currently, such as pruning, water and fertilizer, flower and fruit management, are still emphasizes experience from experts that dominated production management. Through numerical simulation, however, the digital model of fruit trees can be constructed to further understand the quantitative relationship between fruit growth, environment and management measures, and reveal the growth and development mechanism of fruit trees; Morphological structure model can provide quantitative basis for solving pruning problems; Productivity features model can realize fast calculation of light distribution that meet the demand of plant design and potential production calculation; Fruit growth and the environment model can achieve the fruit growth process simulation, support the orchard production and management decisions.

2.2 Digital Management and Control of Orchard Production

From the perspective of implementation process, the digital technology of orchard production can be divided into four parts: orchard information acquisition, orchard information management and analysis, production decision analysis, and inter-orchard implementation based on decision. The current management of most orchards still treats all elements in the orchard as the same whole. However, it is objectively exist that the variation of topography, soil nutrient content, soil moisture content, orchard temperature and humidity and other microclimatic variations in the orchards make the yield differences of different fruit trees. In order to achieve precision in fertilization, irrigation, application and fruit management, the high density of information about soil, fruit trees, micro weather and the simulation operation and decision of digital management model of fruit trees are needed to pinpoint horticultural measures to each plant of orchard, then realize the digital management and control of orchard production and improve the level of precision orchard production management.

In the digital management and control of orchard production, the digital collection of orchard environment information and fruit tree growth information is the foundation, and the management of orchard archives and production history is the key. Information of orchard environment and fruit tree growth includes temperature, humidity, light intensity, effective radiation, ultraviolet intensity, rainfall, wind speed, wind direction, dew point, soil moisture, soil NPK content, image of fruit tree, etc. Only by continuously collecting the information of these factors and forming a certain amount of data sets, can the model used to guide production be built through data analysis and algorithm. Orchard archive and production record management were used to realized a series of orchard basic information and production process of digital records and documents, including the records of location and area of the orchard, fruit varieties, quantity, density, location, yield growth condition, document of all kinds of soil fertility, and fruit cultivation management record.

2.3 Product Quality and Safety Digital Management and Traceability

Fruits quality control is achieving digital management and traceability of product quality control through the determination of key quality control points, the collection of whole-process quality information, and the traceability of quality and safety information. The establishment of the fruit quality traceability system will undoubtedly contribute to the improvement of the fruit quality, and the key to the development and application of the fruit quality traceability system lies in the selection and determination of the key traceability index system of the fruit industry chain by combining the fruit GAP production technology and HACCP management system [14], and the fruit industry chain as well as to the whole process of quality traceability index data collection, storage, query and sharing. Fruit quality control of digital management is an important aspect for the technology application of the Digital Orchard. It integrated the environment information, the production process and producers information integration together, and comprehensive using technologies of database, network, coding and production to realize the prenatal management, production management, early warning, statistical analysis in production side and quality tracking, quality feedback and government regulation in consumption side.

The use of the Digital Orchard technology can facilitate the establishment of orchard production process and fruit sales records such as natural resource background, growth monitoring of fruit trees, occurrence monitoring of diseases, insect pests and weeds, inputs use and so on. Through the integration and sharing of relevant data and the analysis and mining of big data, the government can establish a regional monitoring and early warning platform for orchard production and improve the level of supervision over orchard production.

2.4 Digital Management of Fruit Circulation and E-commerce

The important characteristics of modern fruit industry are standardization, marketization, informatization and branding, which are market-oriented, economic benefit-centered and supported by scientific and technological progress, so as to realize commercial production, professional management and socialized service of fruit. For this purpose, it is necessary to study the technology of digital management and e-commerce of fruit circulation, so that fruit farmers can timely understand the dynamic information of fruit markets in domestic and foreign through various information systems, and enable fruit industry organizations to establish e-commerce channels and achieve multiple modes of fruit e-commerce such as B2B and B2C, and finally bring the management level of fruit marketing improvement and the economic benefit of orchard planting increasing.

The connotation of the Digital Orchard includes the above four aspects. From the perspective of system realization, there can be different realization modes. Figure 1 shows the implementation mode of a Cloud-Based Digital Orchard, which was using cloud technology to realized the integration of on-site information collection, digitization of orchard production management and fruit e-commerce.

3 Research and Application of the Digital Orchard Technology

In recent years, many research institutions and universities in China have been actively carrying out Digital Orchard study and have obtained preliminary research results. Citrus Research Institute of Chinese Academy of Agricultural Sciences and

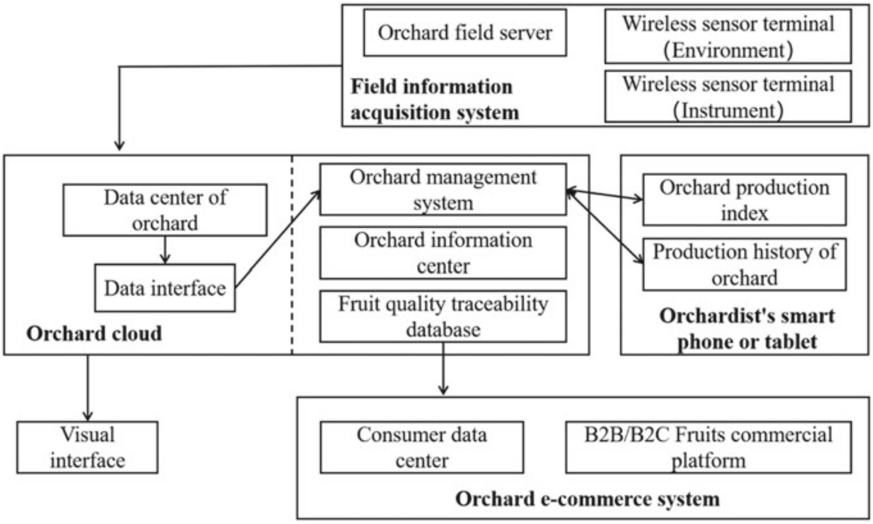


Fig. 1 Structure diagram of implementation pattern of cloud-based digital orchard

East China Jiaotong University used spectral and digital image techniques to predict fruit ripening period [15, 16]. Beijing Agricultural Information Technology Research Center has realized the construction and simulation of the morphological structure model of the main apple tree varieties [17, 18]. China Agricultural University has made a lot of achievements in the development of orchard-picking robots [19–21]. Shandong Agricultural University and Northwest Agriculture and Forestry University have made a lot of progress in the research of fruit tree growth and cultivation management mechanism [22–27].

Some research achievements of the Digital Orchard have begun to demonstrate applications. The new type of direct meteorological service for agricultural operators in Kunming city, Yunnan province is stationed in orchards to monitor air humidity, wind power, leaf temperature and soil moisture, and provide services for disease control and orchard irrigation. The organic peach base in Cixi city, Zhejiang province uses the IoT (Internet of things) technology to quickly collect planting information and environmental information, and realizes the intelligent and fine management of grapes, pears and peaches. In Jiaxing city, Zhejiang province, Zhejiang Academy of Agricultural Sciences used SMS trigger to remotely obtain the air temperature, humidity and soil moisture content in the grape greenhouses. Shandong Agricultural University used electronic tag technology to code fruit trees in Feicheng city, Shandong province, recording the species of fruit trees, varieties, responsible persons information, and logging the production information such as fertilization time, fertilizer name and application amount. Pinghe pomelo association in Fujian province utilized QR code technology to supervise the fertilization, irrigation and grafting of orchards in every link of production, realizing production records, product inquiry, quality traceability.

Focusing on the key technologies and application systems of the Digital Orchard, the Agricultural Information Institute of Chinese Academy of Agricultural Sciences has organized and carried out systematic research, which has made remarkable progress in orchard environment, fruit tree growth information acquisition, fruit tree growth model, and orchard digital management platform. The following is an introduction to the specific research.

3.1 Information on Orchard Environment and Fruit Growth

Data collection of orchard environment and fruit growth is the basis of Digital Orchard management. Climatic environmental factors such as atmosphere, temperature, light and moisture are closely related to fruit production. In terms of orchard's soil factors, soil organic matter content is an important index to evaluate orchard soil fertility; Soil moisture is the main source of water absorption by fruit trees and it's content affects the yield and quality of fruits; Heavy metal content in soil affects fruit safety. In topography factors respect, geomorphic characteristics of orchards include elevation, mountains, gradient, slope direction, height also affect the growth of fruit trees to a certain extent.

In the aspect of fruit growth information, indicators like tree vigor, branch types, the date of budding, flowering and bearing fruit, ratio of the numbers between branches and fruits, proportion of flowers and fruits are important representations of fruit tree growth state. In addition, the acquisition and prediction of pest and disease information are also important aspects of orchard management. There has been rapidly research progressing in perception of orchard environmental factors and intelligent control at home and abroad, and preliminarily realized the real-time monitoring of orchard temperature, humidity, lighting, moisture and other environmental parameters, and there are already many mature equipments to choose from. The problem, however, is that there is insufficient accumulation of microclimatic data for diverse types of orchards, as well as data on the optimal environmental parameters for the growth of different species in different ecological areas, and the lack of low-cost and highly reliable sensors for orchards [28, 29].

For orchard production management requirements, Agricultural Information Institute of Chinese Academy of Agricultural Sciences proposed a standard for the expression of collected data of fruit growth and orchard environment information based on variable acquisition indicators [30], which provides support for the integrated management of multi-site acquisition data and provides a consistent and machine-readable data interface for the upper application program. On the basis of the study on the spatial and temporal distribution characteristics of the main environmental parameters in the orchard [31] and on the orchard propagation characteristics of the 2.4 GHz wireless signal [32], a digital acquisition technology system with the whole chain of "orchard environment—growth process—operation process—orchard management" was formed [33, 34] (see Fig. 2). The equipment suite of information acquisition was developed to realizes the digital collection and real-time

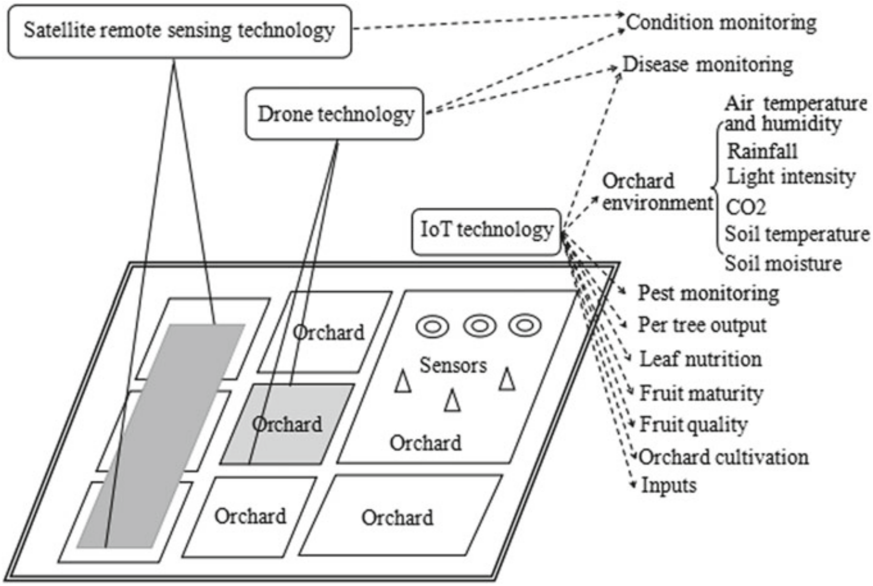


Fig. 2 Technical system diagram of orchard information collection

transmission of the orchard environment information (air temperature and humidity, rainfall, light intensity, CO_2, soil temperature, soil moisture), orchard pest monitoring information, per tree output information, orchard cultivation information, inputs information, which lays the data foundation for the orchard digital management.

Similar researches are mostly focused on the research and application of single technology or partial links, and cannot fully perceive and characterize the orchard environment and the overall situation of fruit growth. The characteristic of this study is that the digital acquisition technology covers the whole chain of orchard production and management, and realizes the standardized expression and fusion application of the collected data in different links.

3.2 Model of Fruit Production Management

The model is the core of using computer to assist accurate orchard management. The research on the production and management model of fruits involves three aspects: First, based on the entire orchard research target group of parameters, put forward the orchard the optimum parameters of the group by studying the different planting density, different tree structure, different nutrition levels, and determine the trees' light utilization rate, production efficiency, fruit quality in different growth stages, and even considering the parameters such as fruit market price, then put forward the optimal orchard group parameters through the modeling analysis. Second, the individual parameters of single fruit tree were studied, which mainly studied the related parameters such as tree structure, light utilization rate, canopy distribution, branch composition, fruit distribution and fruit quality, and proposes the management index of single fruit tree through modeling analysis. Third, taking fruits as the object, studying the relationship between the growth and development of fruit and its surrounding micro environmental factors and nutrition supply through the monitoring of fruit growth process, constructing simulation model of individual plant growth, so as to determine the tree management index based on the fruit demand [35–37].

The model-based orchard production management generally treats the fruit tree as a system, as shown in Fig. 3. The model is regarded as an objective description of the operation law of a certain aspect of the system, and the system operation state data required by the model is obtained from the system. Then the model starts operation and the target system is controlled according to the result of model operation, so that the target system is developed in a specific direction. In this model-based control process, the control effect is completely dependent on whether the model accurately describes the system operation law, and it is also implied that the system operation law can be accurately described by using the precise modeling method. However, the fruit tree system is a complex system which is in dynamic change all the time. First of all, the optimal solution based on analytic model is directly related to the hypothesis conditions and often has a strong sensitivity to the conditions. For complex system problems, if there is a difference between the hypothesis and the actual situation, the hypothesis and the actual situation will be "one false step make a great difference". Secondly, there is generally no single optimization index to solve complex system problems, while multi-level and multi-objective optimization index often results in multiple or even countless solutions. Therefore, according to the characteristics of the complex system of fruit trees and the parallel management theory [38, 39], a framework of fruit tree cultivation management model was proposed, as shown in Fig. 4. The model framework mainly includes the actual fruit tree system and the virtual fruit tree system. By comparing and analyzing the activities of the two systems, the "reference" and "prediction" of their respective future conditions are completed, then the respective management and control methods are correspondingly adjusted to achieve the purpose of implementing effective solutions, and learning and training.

Although the existing research has carried out related work on the apple tree water and fertilizer application model and the kiwi fruit climate quality evaluation model, the shortcoming of this kind of models is that the overall operation law of the

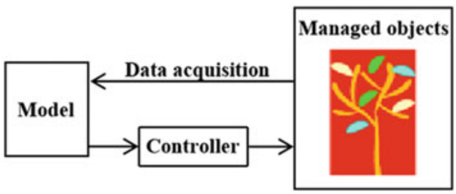


Fig. 3 Schematic diagram of model-based fruit tree management method

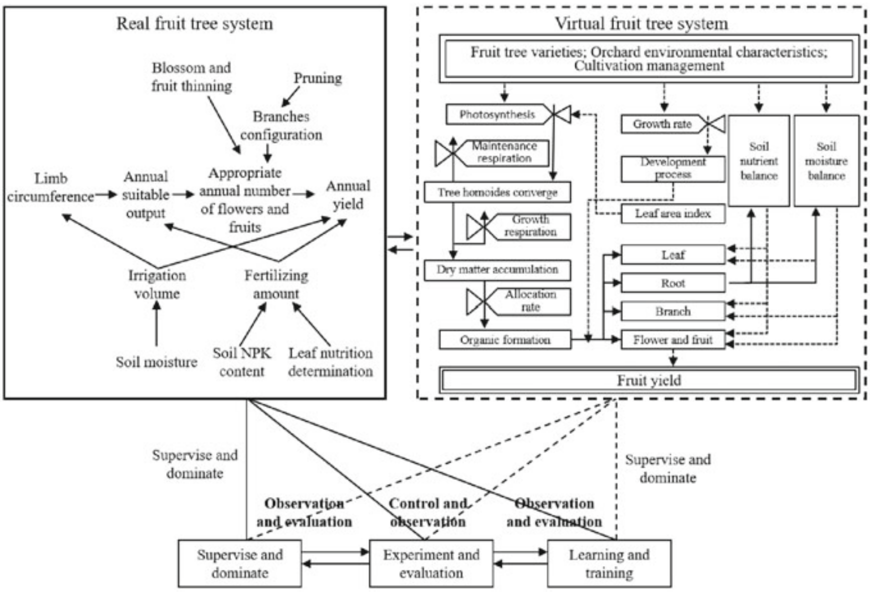


Fig. 4 Frame diagram of fruit tree cultivation management model

fruit tree growth system cannot be comprehensively described while the optimization index is relatively single. The proposed model of fruit cultivation management that based on parallel management theory, relies on the mutual reference and prediction of the actual fruit tree system and the virtual fruit tree system to evaluate the effects of different cultivation schemes, so that the model can have better adaptability and prediction performance.

3.3 Platform of Orchard Digital Management

According to the demand of orchard digital management in China, a comprehensive code system of orchard production management with five-code (orchard code, land code, operation code, input code and commodity code) interconnection was proposed with the fruit production and management as the core, and the platform of orchard digital management was developed based on it, as shown in Fig. 5, this platform has the functions of orchard monitoring, orchard production process management, expert remote diagnosis and service, fruit inventory and traceability management, which can realized the digital management of the whole fruit industry chain. The platform adopting the model of cloud computing and cloud system architecture, taking the growth management model as the core and the digitization as the foundation, linking the orchard environment, fruit tree growth, orchard production, expert guidance and government supervision together organically to serve different subjects such

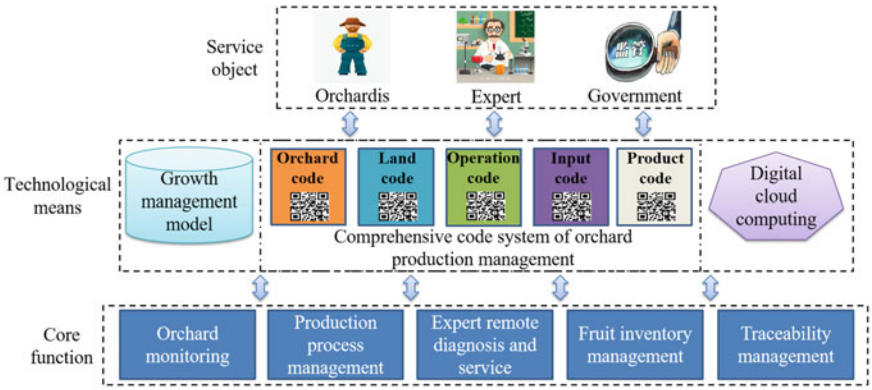


Fig. 5 System structure of the platform of digital orchard management

as orchard owners, experts and governments with data mining and model analysis, and then realizes the digitalization and network management of the orchard.

Similar platforms are mainly focus on the digital management of a single fruit variety or a single production link, and are limited to solving the problem of digital management of a link in the whole fruit industrial chain. The proposed platform of digital orchard management can comprehensively cover the whole industrial chain and multi-fruit types of orchard production, so it has advantages in platform comprehensiveness and adaptability.

3.4 Application of Research Achievements

Around the studies of information acquisition of orchard environment and fruit growth, model of fruit tree growth and platform of orchard digital management, a series of achievements of the Digital Orchard have been applied in many regions such as Liaoning, Beijing, Shandong, Shaanxi, Ningxia and Xinjiang. Various hardware equipment for remote collection of orchard information have been installed while a total of 18,900 times of systematic training and consulting services have been carried out. Meanwhile, the cumulative area of application was 1.926 million acre, and promoted 842 fruit industry cooperatives (or fruit associations) and 316,000 fruit farmers incomes increasing. The research achievements has obtained better social benefits and economic benefits. The application of the digital management platform is shown in Fig. 6. The equipment of data acquisition installed in orchards of different places are shown in Fig. 7.

Fig. 6 Application of the digital management platform

4 Development Trend of the Digital Orchard

With the rapid development of modern information technology, the research of the Digital Orchard technology will make great progress, and present the development trend of intelligence, network, mechanization, integration, and will make the orchard in the production mode and concept of revolutionary changes [40].

The intelligentization of the Digital Orchard technology will focus on the collection, storage and expression model of expert knowledge and the growth model of fruit trees, forming the core of intelligent technology and the application of basic research. The collection, storage and reasoning technology of expert knowledge of fruit tree cultivation and management will be broken through by means of a series of technological innovations such as integrated development platform, intelligent modeling tool, intelligent information acquisition tool and simple human–machine interface generation tool. Fruit tree morphological structure model and orchard intelligent management will be integrated with horticulture, ecology, physiology, computer graphics and other disciplines, and taking fruit tree organs, individuals or groups as the research object to construct the 4D morphological structure model of the fruit tree for the realization of the interaction design, geometric reconstruction and visual expression of the growth and development process of the fruit tree and its growing environment.

The networking of the Digital Orchard technology will fundamentally broken the barrier of time and space, changed the mode of fruit management and circulation, shortened the circulation link of fruit from garden to table, promoted the rapid transmission of market information such as product price, quantity and quality, eliminate the information imbalance between producers and consumers, and then entered the era of customized fruit production with consumers as the center. The information

Fig. 7 Equipment of orchards data acquisition installed in Beijing, Xinjiang, Shaanxi and Shandong

of supply and demand of fruit products will be spread through the network, which will fundamentally promote the transmission of the information about product, price and market, and promote product transaction and reduce costs. The network makes it possible for the rapid dissemination and exchange of remote consultation, remote diagnosis and technical information. The network interaction of information enables fruit farmers to easily find the required production and market information by using online media like MicroBlog and WeChat. Fruit farmers can communicate with experts online at any time, and also with fruit buyers online trading.

The mechanization of digital orchard technology will make orchard machinery precise navigation and control technology, operation decision model and operation plan real-time generation technology get a wide range of applications. The intelligent orchard equipment will realize the mechanization, intelligentization and robotization of cultivation, plant management, flowers and fruits management, fertilizer and water management, disease and insect pest control. With the continuous improvement of China's manufacturing level, small and low-cost orchard intelligent machinery will be further developed, and gradually replace human to complete orchard seedling cultivation, fertilization, pruning, bagging, disease and insect pest control, intertillage weeding, harvest and other horticultural operations.

In comprehensive development of the Digital Orchard technology, technology integration and comprehensiveness will become more and more important. The application of information technology does not have to completely replace or exclude conventional or traditional methods and technologies, but it will be more effective, practical and popular to adopt the "integration" strategy of "mutual combination and learning from each other". For example, the expert system is combined with the simulation model research, the expert system is combined with the real-time signal acquisition and processing system and even the technical and economic evaluation system, the expert system is combined with the precision agricultural machinery and tools. Therefore, the integration and combination of various agricultural information technologies will be an important development trend around the practical problems of fruit industry development.

5 Conclusion

The research and application of the Digital Orchard technology are in the ascendant, which is bringing rare opportunities to the development of fruit industry. In the future, the Digital Orchard technology will play a key role in improving the utilization rate of orchard resources and labor productivity and the progress of this technology will promote the revolutionary development of fruit production and management. Orchard production will form a virtuous cycle mode of mutual coordination of "human-resource-environment-production relationship", and the whole fruit industry will embark on a sustainable development path of high yield, stable yield, low consumption and high efficiency.

References

1. Zhao C, Yang X, Li B et al (2018) The retrospect and prospect of agricultural information technology in China. J Agri 8(1):180–186
2. Zhou G (2012) research and development of digital orchard in China. Agri Netw Inf 1:10–12
3. Han M, Ma F, Li B et al (2008) Development of Italian and French apples. Northwest Hortic 1:49–50
4. Montet D, Ramesh R (2017) Food traceability and authenticity: analytical techniques. CRC Press, Boca Raton

5. Li L, Yang J (2017) Application of computer information technology in food quality safety and detection. J Food Saf Qual 9:86–87
6. Zhang Z (2010) The enlightenment of the development trend of apple industry in aomori, Japan. Yantai Fruits 1:13–15
7. Yuan J, Mei L, Gao H et al (2008) A survey of apple in Japan. Northwest Hortic 6:45–47
8. Wu H (2016) New techniques and management of high quality cultivation of fruit trees in Japan. Fujian For 2:25–26
9. Chen X, Han M, Su G et al (2010) Discussion on today's world apple industry trends and the suggestions on sustainable and efficient development of apple industry in China. J Fruit Sci 27(4):598–604
10. Chen X, Wang N, Zhang Z et al (2019) The bottleneck problems and solutions of new and old kinetic energy conversion in China's fruit industry. China Fruits 196(2):7–10
11. Cheng C, Zhao D (2019) Development orientation and trend of the apple industry under the new situation. China Fruits 195(1):7–13
12. Deng X (2018) The supply-side reform and development trend of China's fruit industry . Mod Agric Equip 4:13–16
13. Zhou G (2015) Application situation and outlook of digital orchard. Agric Outlook 5:61–63
14. Xia Z, Guo B, Wei Y et al (2014) Key indicators screening and module establishing for fruit chain traceability. J Nucl Agricul Sci 28(5):890–896
15. Mao S, Zeng M, He S et al (2010) Application of near infrared spectra (NIRS) technology in prediction of maturity stage of fruit. Subtrop Plant Sci 39(1):82–87
16. Liu Y, Ye L, Sun X et al (2018) Maturity evaluation model of tangerine based on spectral index. Chin J Opt 11(1):83–91
17. Su H, Guo X, Lu L et al (2009) Geometry modelling and visualization of the apple inflorescence. Chin Agric Sci Bull 2:272–276
18. Wu S, Zhao C, Guo X et al (2017) Method of fruit tree canopy leaf reconstruction based on point cloud. Trans Chin Soc Agricul Eng 33(z1):212–218
19. Feng J, Liu G, Wang S et al (2013) Multi-source images registration for harvesting robot to recognize fruits. Trans Chin Soc Agricul Mach 3:197–203
20. Sun Y, Sun J, Zhao R et al (2019) Design and system performance analysis of fruit picking robot. Trans Chin Soc Agricul Mach S1:8–14
21. Zhang K, Zhao L, Sun Z et al (2013) Design and target extraction of intelligent grape bagging robot. Trans Chin Soc Agricul Mach 44(z1):240–246
22. Wang L, Zhao G, Zhu X et al (2013) Satellite remote sensing retrieval of canopy nitrogen nutritional status of apple trees at blossom stage. Chin J Appl Ecol 24(10):2863–2870
23. Cai H, Wang Q, Gao J et al (2018) Impact of branch amount on physiological characteristics of' 'Fuji' grafted on 'Y-1' dwarfing interstock . J Agricul 8:57–61
24. Chen X, Mao Z, Jiang Y et al (2019) Trinity Chinese-style apple cultivation method of wide-line, high-trunk, energy-saving and high-efficiency promotes the transformation and upgrading of China's apple industry and promotes rural revitalization. China Fruits 4:1–3
25. Li N, Li B, Wang J et al (2014) Effect of different tree forms on light interception, yield and quality of juvenile apple tree lower part. North Hortic 19:4–8
26. Shi P, Liu H, Bai H et al (2016) Study on 3D trunk model construction of peach tree canopies and light interception and space distribution of fruit quality. Acta Agricul Boreali-Occident Sinica 25(9):1371–1378
27. Zhang G, Li X, Cheng S et al (2013) Progress and prospects of 3S technology application to apple orchard informatization. Geomat Spat Inf Technol 36(8):40–44
28. Stajnko D, Berk P, Lesnik M et al (2012) Programmable ultrasonic sensing system for targeted spraying in orchards. Sensors 12(12):15500–15519
29. Tewari V, Chandel A, Nare B et al (2018) Sonar sensing predicated automatic spraying technology for orchards. Curr Sci 115(6):1115–1123
30. Zhou G, Fan J, Wu D et al (2015) Data acquisition and data format of orchard environment based on XML. Tianjin Agricul Sci 21(12):76–79
31. Guo X, Fan J, Zhou G et al (2016) Spatial and temporal distribution rule of temperature and humidity in apple tree canopy on sunny days. Chin Agric Sci Bull 35:188–192
32. Guo X, Yang X, Chen M et al (2015) A model with leaf area index and apple size parameters for 2.4 GHz radio propagation in apple orchards. Precis Agricul 16(2):180–200

33. Guo X, Zhou G, Qiu Y et al (2014) A wsn application framework for agricultural environment monitoring and controlling. J Agricul Mech Res 11:199–203
34. Xia X, Qiu Y, Hu L et al (2013) Remote automatic tele-monitoring system for orchard environment based on 3G and DDNS. Autom Instrum 28(8):18–21
35. Zude-Sasse M, Fountas S, Gemtos T et al (2016) Applications of precision agriculture in horticultural crops. Eur J Hortic Sci 81(2):78–90
36. Lentz W (1998) Model applications in horticulture: a review. Scientia Hortic 74:151–174
37. Hester S, Cacho O (2003) Modelling apple orchard systems. Agric Syst 77(2):137–154
38. Wang F (2013) Parallel control: a method for data-driven and computational control. Acta Autom Sinica 39(4):293–302
39. Wang F, Wei Q (2018) Intelligent control: from learning control to parallel control. Control Theory Appl 7:939–948
40. Zhou G, Qiu Y, Fan J et al (2018) Research progress and prospect of digital orchard techniques. China Agric Inf 30(1):10–16

Guomin Zhou, Professor, Ph.D. supervisor. He is currently a deputy director with the Department of Science and Technology Management, Chinese Academy of Agricultural Sciences. He has been engaged in agricultural information technology research for many years, and has carried out in-depth research on orchard digital management technology and system, agricultural science data sharing technology and system. He has presided over and participated in a number of research projects such as the National High Technology Research and Development Program and the Special Project of Platform for Basic Conditions of Science and Technology, and got more than 10 awards for scientific and technological achievements. He has won the “top 10 young people” of the ministry of agriculture and rural affairs and “top 10 annual figures of China’s agricultural informatization” honorary titles. Moreover. He was selected into the “Beijing science and technology nova” plan and the outstanding talent of Chinese Academy of Agricultural Sciences.

Study on Intelligent Management and Control Technology for Apiculture Production

Shengping Liu, Yeping Zhu, Yue E., Jie Zhang, Chunyang Lv, and Xiuming Guo

Abstract Facing the main battlefield of modern agricultural construction, Wisdom bee industry platform carries out key technology research and product development of the bee industry production intelligent management and control in beekeeping industry. Based on the comprehensive application of agricultural information technology such as mobile internet technology, artificial intelligence technology and Internet of Things technology, this paper focuses on the information collection and monitoring of the Internet of Things in the bee yard, the key equipment of the intelligent beehive, the quality and safety control system of the bee products and the big data analysis technology of the whole industry chain. Through intelligent sensing, interconnection and big data analysis, this paper implements bee industry information intelligent management, and provides users with standardized, automated and intelligent bee industry information service mode.

Keywords Smart apiculture · Intelligent beehive · Bee product traceability · Internet of things · Big data

China is a major beekeeping country in the world, with a long history and a long history. It is also China's most traditional and historically characteristic industry and one of the most representative agricultural industries. The number of bee colonies, the output of bee products and the export volume are among the highest in the world. China's current honey output is over 400,000 tons, accounting for more than a quarter of the world's total honey output. At the same time, bee pollination provides more than 12% of the output value of modern agricultural development and is an important part of modern agriculture [1]. However, in terms of industrial information application and development, China's beekeeping industry lags behind other agricultural industries in the following aspects: (1) Beekeeping is relatively small, and the main producers are individuals and families. The traditional bee production methods have largely restricted the development of beekeeping to specialization, standardization

S. Liu (✉) · Y. Zhu · Y. E. · J. Zhang · C. Lv · X. Guo
Agricultural Information Institute, Chinese Academy of Agriculture Sciences, Beijing, China
e-mail: liushengping@caas.cn

Key Laboratory of Agro-information Services Technology, Ministry of Agriculture, Beijing, China

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_26

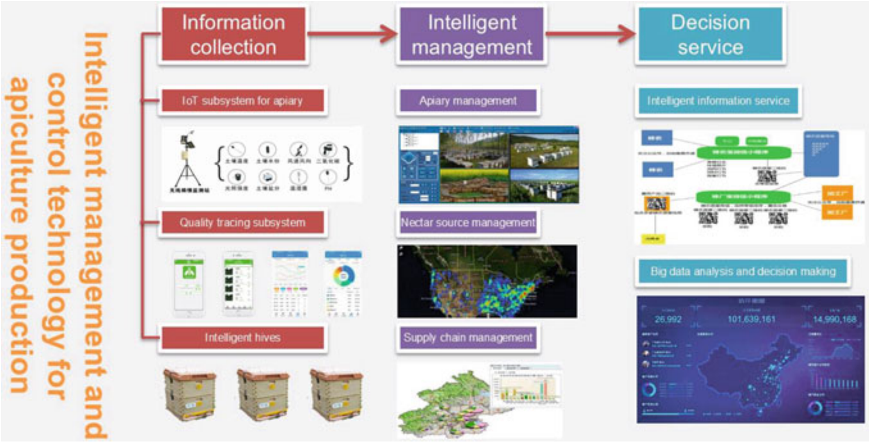


Fig. 1 Intelligent management and control technology for apiculture production

and modernization [2]. (2) The production of beekeeping is scattered, and is greatly affected by nature and the market. Due to the difficulty in achieving large-scale production requirements for scattered farmers, many standardized production technologies are difficult to implement on a large scale [3]. (3) The increasing cost of beekeeping and the lack of protective policies will bring about some quality and safety problems. In order to make up for the problems existing in the informatization and modernization of China's bee industry, this paper carries out research and development of key technologies and equipment for intelligent management and control of bee production, and is applied to the development of modern bee industry, intelligent equipment control and digital production management that have entered the bee farm [4]. The research results will help to increase the scale development level of the bee industry, realize the informatization and transparency of bee industry production, increase the income of bee farmers, and effectively increase the economic and social benefits of society.

Comprehensive application of Internet of Things technology, mobile Internet technology, artificial intelligence technology and modern Internet technology, bee industry production intelligent control technology serves bee colony breeding, bee product processing, bee product quality and safety, realizes the saving of bee industry breeding resources, and improves bee product quality and earnings. With the continuous development of smart agriculture and the specific practice in many fields of agriculture, it has provided sufficient theoretical basis and effective technical support for the proposal and development of intelligent production in apiculture. As shown in Fig. 1, starting from the entire industry chain process of information collection, intelligent management and decision-making services, the intelligent management and control technology of bee production includes the collection and monitoring of apiary environment information based on the Internet of Things, the development of key equipment for intelligent beehives, the development of bee product quality and safety control systems, and the visual analysis platform for big data of the entire bee industry chain Build etc. The study takes the key technology of intelligent management and control of the whole industry chain informatization of the bee industry as an entry point to improve the quality and safety process control level and safety monitoring capabilities of honey products as a whole, providing technical support for the modernization of the bee industry.

1 Apiary Environmental Information Collection and Monitoring Based on the Internet of Things

Although China is a large honey farming country, its research on intelligent management of apiary is relatively backward, and the technology is still immature, especially in the modernization and informatization of apiary management. In recent years, with the rapid development of information technology and changes in the social and economic structure, the entire agricultural industry has undergone tremendous changes in production methods, consumption levels, technical systems and management models [5]. The trend of the transformation of the development mode of the bee industry is very obvious. From the perspective of production management mode, the development of the bee industry shows a trend of intelligent equipment and digital management. The production of intelligent bee industry has been widely concerned all over the world, and some western developed countries have successively promoted the whole process of bee production and the supervision and management of the whole process. As an export-oriented industry, China's beekeeping industry is actively adapting to the above changes is the current development trend of bee industry.

Using agricultural Internet of Things technology, sensor technology, the apiary environmental information collection and monitoring technology is mainly used in local production of apiary [6]. The main contents of the apiary environmental information collection and monitoring technology include the development of apiary environmental information collection equipment and apiary safety monitoring equipment, and the development of automatic collection and monitoring technologies and systems for the atmosphere, water quality, light, temperature and humidity of the bee production area [7]. Through the implementation and application of equipment installation, it can realize the timing and fixed-point automatic sampling and monitoring of standard bee production environment information. At the same time, with the help of video surveillance technology, the system supports 24-h monitoring, historical playback and remote control, real-time monitoring of bee farm security and beehive status. Figure 2 shows the apiary environmental information automatic collection equipment, apiary safety monitoring equipment and supporting software system developed by the team. The equipment and system were applied in Miyun District, Beijing, and achieved good demonstration results.

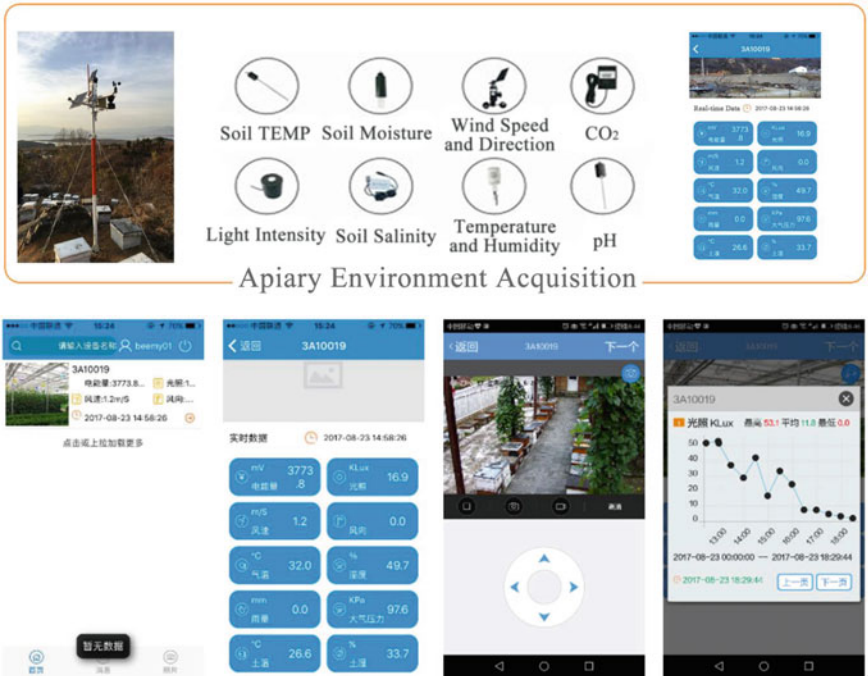


Fig. 2 Software and hardware for collecting and monitoring apiary environment information based on Internet of Things

The apiary environmental information collection and monitoring technology assists beekeepers in scientific beekeeping through modern technical methods, which can improve the efficiency of beekeeper management and reduce the labor intensity of beekeepers. The application of Internet of Things equipment can provide meteorological forecasts for beekeepers, so that beekeepers can determine the time to breed bees and plan the site in advance. The application of monitoring equipment will greatly reduce the labor intensity of bee breeding. With the transformation of the apiary production management model, beekeepers will invest less manpower to improve the operating environment of bees, and realize the transformation and improvement of apiary management from traditional extensive management to information and automated management.

2 Development of Key Equipment for Intelligent Beehive

China's beekeeping lags behind other industries in agriculture. Overall, the degree of mechanization is low and the facilities and equipment are backward. The management of bee farms by beekeepers is still relatively rugged. The traditional manual detection method has the problems of inconvenient implementation, high labor intensity, low efficiency, weak operability, and deviation from the production and living habits of the bee colony. And because of traffic restrictions, it is difficult to make full use of honey sources in some remote areas. Therefore, the implementation of intelligent equipment and digital management to enhance the scale development level of the bee industry is a technology that must be overcome in the development of bee industry in China [8].

Intelligent beehive is defined as a beehive that can use modern advanced detection technology to obtain the relevant data of the bee colony and scientifically process the data in order to complete the scientific management of the bee colony [9]. The main performance is that without opening the beehive cover, not disturbing the bees, and not visualizing the activities of the bee colony, using modern technology can understand the state of the bee colony very accurately and timely and apply intelligent disposal. The intelligent beehive will have the functions of regularly monitoring the life production environment of bees, timely reflecting the parameters in the beehive, intelligently identifying the number of bees in and out, and automatically monitoring the sound of the beehive [10]. This study developed a supporting software system that can support user viewing, statistical analysis, automatic warning, and remote control.

Figure 3 is a schematic diagram of the structure of the smart beehive developed by the team. The intelligent beehive equipment includes solar cells, sound information collection module, camera, weighing instrument, infrared probe, temperature sensor, fan, automatic feeding equipment, beehive control module, system main control module and other devices. The intelligent beehive can collect the temperature, humidity, sound information, beehive monitoring photos and video information, beehive weight information and other content in the beehive, and realize automatic feeding, bee colony behavior monitoring, honey collecting statistical analysis, beehive cooling and bee theft analysis Features. The system can notify the user in real

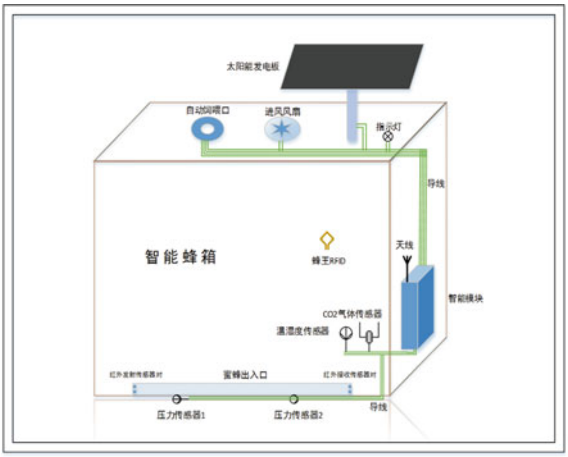


Fig. 3 Schematic diagram of smart beehive structure

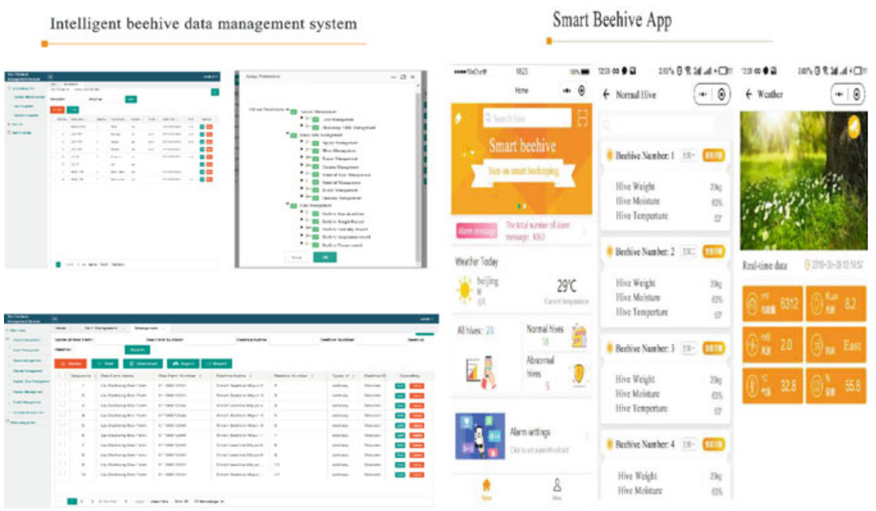


Fig. 4 Smart beehive data acquisition software system

time of all relevant information through curves, state diagrams, alarm information, etc. The system can also help beekeepers to realize real-time monitoring, remote monitoring, and smart monitoring of the environment and bee colonies in the box, thereby improving the efficiency of the farmer's work, reducing the interference to bees, and increasing the output and quality of bee products [11].

Figure 4 shows the data management and display system developed by the team for intelligent beehive hardware product information collection. Based on the obtained data, this system can realize intelligent management of intelligent beehive data, and develop a combination application of beehive and bee colony breeding models. Using the beehive management app, beekeepers can remotely check the status of beehives and real-time production warning to improve the efficiency of beekeepers and realize the modernization and informatization of bee colony breeding [12, 13]. The terminal application technology greatly reduces the repetitive and laborious inspection work of the farmer, and greatly liberates the farmer's labor force to produce more economic benefits [14].

3 Research and Development of Bee Product Quality and Safety Control System

At present, the quality and safety problems of bee products in China mainly have the following two characteristics: (1) more than 60% of bee product safety problems occur in the production and processing links; (2) 75.50% of bee product safety events are caused by human factors. On the one hand, the reason is that China's bee product market is a discrete market composed of many small producers and operators, and it is difficult to put integrity supervision in place. On the other hand, more than 90% of China's bee product processing enterprises are non-scale enterprises, and many factors such as the supply chain of bee products, the length of the chain, the difficulty of management and quality and safety monitoring, and other factors have led to a series of pesticide residues in bee products exceeding the standard and the deterioration of the environment Incidents have occurred frequently, and traditional methods have been difficult to meet the society's need to monitor the quality of bee products. Affected by factors such as the environmental pollution of the place of origin, the use of bee diseases and adulteration, and other factors, the quality and safety of bee products in China have been widely concerned over the past 15 years. The establishment of a traceability system for agricultural products has become an important means of agricultural product quality and safety supervision in China. However, in the application process of agricultural product quality traceability information system, there are problems that the information record is not accurate enough, and the traceability information collection process is easily interfered by human factors. In particular, China's beekeeping industry has the characteristics of scattered production and operation, strong liquidity, small apiary scale, distinct origin attributes, and unique consumer demand. This often leads to the inability of the traceability information system to accurately determine responsibility after quality and safety problems.

At present, the agricultural product safety research of various domestic units mainly conducts research on key technologies for traceability in the construction of traceability systems and management systems. The acquisition of traceability information is dominated by manual records, which often results in the absence, error and concealment of information records in the production process. In view of the above-mentioned characteristics existing in the bee industry, this study proposes a method for quality control of bee products that is perceptual, intelligent and collaborative. This method provides a set of control and traceability technologies and systems for bee products with structural dispersion, small individuals, and mixed processing. The system solution is shown in Fig. 5, and designs an agent-based agricultural product quality control system framework, which can add corresponding control links and conditions according to the production and processing characteristics of different bee products, adapting to different goals and types of bee product control. At the same time, the established traceability information collection system of the bee product supply chain covers the four links of bee product production, acquisition, processing and sales. For the key control points of quality safety in the quality traceability process of bee products, this system records the traceability information. The subsystems include the information collection system of the origin of the bee product, the traceability information management system of the bee product acquisition, the traceability information management system of the bee product processing enterprise and the bee product sales traceability information management system. The application of the system realizes the collection of traceability information of the production and circulation of bee products, and provides the basic support of the traceability system for the quality and safety control of bee products.

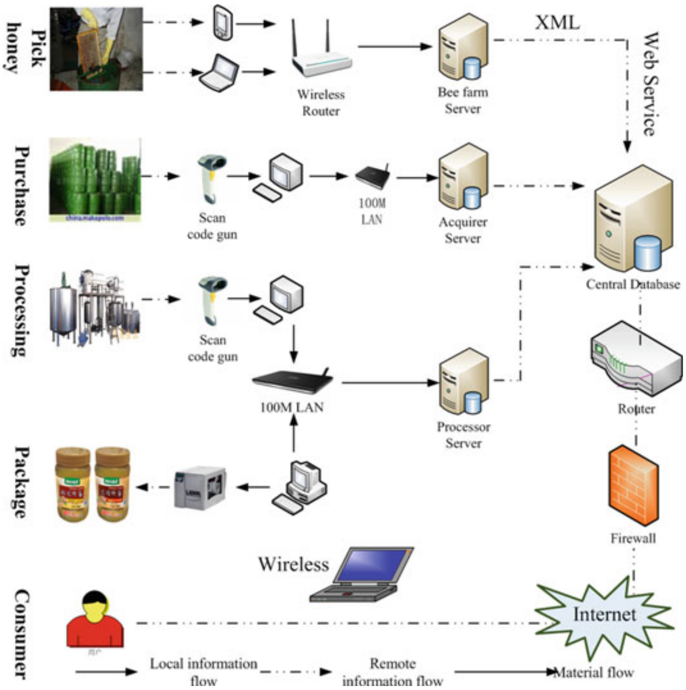


Fig. 5 Bee product quality and safety control system solution

The system analyzes the quality and safety factors of bee products and the characteristics of safe production of bee products based on the concepts of good agricultural production practices (GAP) and good manufacturing practices (GMP). It designs and establishes internal breeding files for beekeeping bases and acquisition management files for purchasers processing plant production and processing files and sellers' sales and circulation records requirements. The research team proposed various forms of bee product information tracking and tracing solutions and developed a bee product quality and safety control system. It has built a traceability information collection system for the entire supply chain of bee products, one for one item, one for anti-counterfeiting and anti-tampering, and completely opened up beekeepers, purchasers, processing plants and consumers. The team also researched and built a web service-based bee product quality traceability information network platform, which realized the real-time, automatic and intelligent supervision and management of traced collection information. Figure 6 shows the APP system interface of bee product quality tracing in the system platform, providing convenient data collection services for beekeepers, processors, purchasers and other users.

The developed platform was applied and promoted in the apiary managed by the Beijing Sericulture Beekeeping Management Station. Six demonstration bases for honey traceability monitoring technology were built, and 2,000 bee colonies

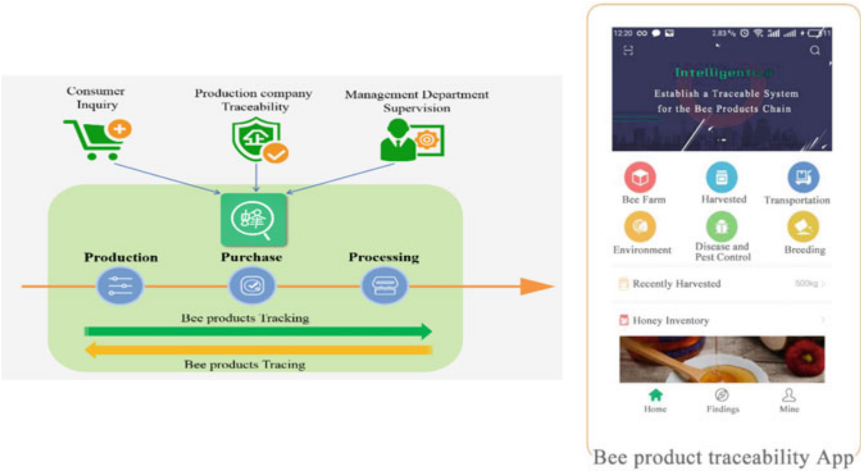


Fig. 6 Bee product quality traceability App interface

were invested, which basically realized the traceability of the entire honey production process. Figure 7 is a system screenshot of the Bee Product Quality Traceability Management Platform. The platform includes traceability query, bee farm video viewing, GIS display, news browsing, organization query and other functions. The platform provides one-stop services for Beijing bee product quality and safety

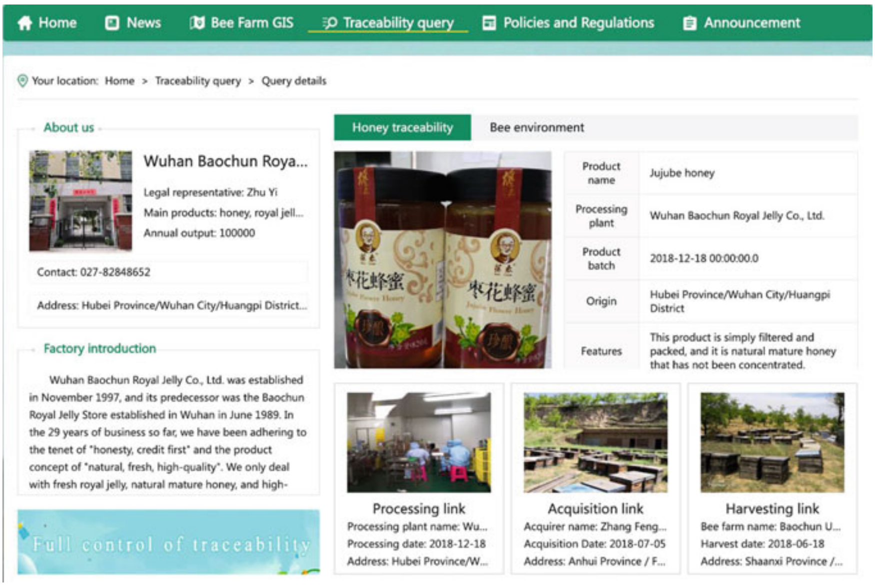


Fig. 7 Bee product quality traceability management platform

content management, organization introduction, news display, traceability management, etc., effectively improving industry transparency, enhancing high-quality bee product brand promotion, and reducing the circulation of counterfeit and shoddy products.

4 Visual Analysis of the Whole Industry Chain of Bee Products

With the development of mobile Internet technology, agriculture has officially entered the era of big data. In 2016, the Central Document No. 1 put forward the development plan of "Internet +" modern agriculture. Big data technology gradually began to be applied to agricultural production and promoted the transformation of China's industrial chain upgrade. It is necessary to start from the agricultural production process, expand the scope and scale of experiments in agricultural production such as Internet of Things technology and big data technology, and promote the construction of big data in the entire agricultural product industry chain. This research is based on the collected bee farm environmental conditions, video surveillance, smart beehive bee colony activities, and all aspects of the bee product industry chain, analyzing the quality and safety data of bee products, building a visual analysis model of bee industry data, and building a strong scalability and stability visual analysis platform for quality control data of bee products. As shown in Fig. 8, the platform is mainly for different users and industrial chain links such as beekeepers, purchasers, processing plants, and consumers in the bee production process, providing production management, process analysis, intelligent scheduling, product traceability and other services. Realize the government's strengthened supervision, quality control and protection of farmers' rights and interests.

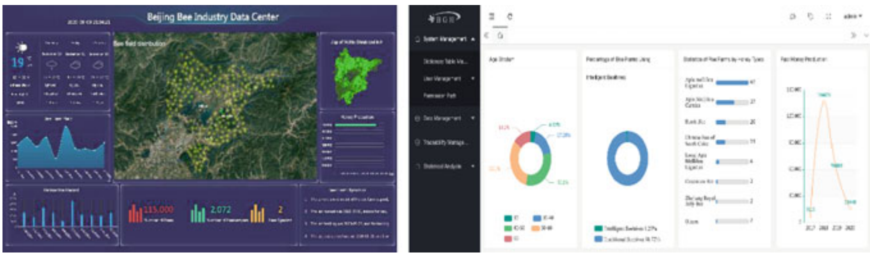


Fig. 8 Data visualization analysis platform of bee product quality control

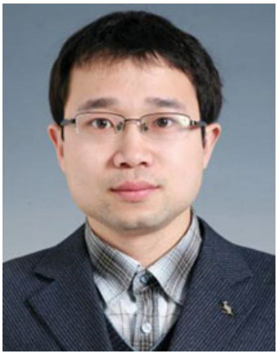
5 Conclusion

This paper carries out research on key technologies for intelligent management and control of bee production, developed a digital apiary software service system covering the entire process of honey production, and developed a smart bee system, intelligent beehive equipment, and a big data visualization analysis system for the entire industrial chain of bee products. The research results provide effective auxiliary decision-making information through effective data management and analysis, and help relevant practitioners carry out bee production activities with more efficient information and intelligent means. It will be beneficial to the overall improvement of the management and control technology of the bee industry, the economic benefits of the bee farmers, and the social benefits of the bee industry. At the same time, the application of key technologies and platforms will reduce the labor intensity of bee farming to the greatest extent and improve the raw material quality of bee products, which can not only save costs, but also increase honey output. The research content can vigorously promote the modernization development of bee industry in China, which is of great significance to promote China's transformation from a large beekeeping country to a strong beekeeping country.

References

1. You ZT, Yu YJ, Kong YG (2014) Current situation and development strategy on Chinese Smart Apiculture. Acta Agriculturae Zhejiangensis 26(4):1111–1115
2. Mahaman BD, Harizanis P, Filis I et al (2002) A diagnostic expert system for honeybee pests. Comput Electron Agric 36(1):17–31
3. Mcclure JE, Tomasko M, Collison CH (1993) BEE AWARE, an expert system for honey bee diseases, parasites, pests and predators. Comput Electron Agric 9(2):111–122
4. Zacepins A, Brusbardis V, Meitalovs J, et al. (2015) Challenges in the development of Precision Beekeeping. Biosyst Eng 130:60–71
5. Danka RG, Beaman LD (2014) Flight activity of USDA–ARS Russian honey bees (Hymenoptera: Apidae) during pollination of lowbush blueberries in maine. J Econ Entomol 100(2):267–272
6. Chen W S, Wang C H, Jiang J A, et al. (2015) Development of a monitoring system for honeybee activities. In: 2015 9th international conference on sensing technology (ICST). IEEE, pp 745–750
7. Jiang JA, Wang CH, Chen CH et al (2016) A WSN-based automatic monitoring system for the foraging behavior of honey bees and environmental factors of beehives. Comput Electron Agric 123:304–318
8. Chen C, Yang EC, Jiang JA et al (2012) An imaging system for monitoring the in-and-out activity of honey bees. Comput Electron Agric 89:100–109
9. Chazette L, Becker M, Szczerbicka H (2016) Basic algorithms for bee hive monitoring and laser-based mite control. In: 2016 IEEE symposium series on computational intelligence (SSCI). IEEE, pp 1–8
10. Zacepins A, Karasha T (2013) Application of temperature measurements for the bee colony monitoring: a review. In: Proceedings of the 12th international scientific conference "Engineering for Rural Development", pp 126–131

11. Zhu X, Wen X, Zhou S et al (2019) The temperature increase at one position in the colony can predict honey bee swarming (Apis cerana). J Apic Res 58(4):489–491
12. Watson K, Stallins JA (2016) Honey bees and colony collapse disorder: A pluralistic reframing. Geography Compass 10(5):222–236
13. Gil-Lebrero S, Quiles-Latorre FJ, Ortiz-López M et al. (2017) Honey bee colonies remote monitoring system. Sensors 17(1):55
14. Meikle WG, Holst N, Colin T, et al. (2018) Using within-day hive weight changes to measure environmental effects on honey bee colonies. PloS one 13(5)



Shengping Liu, Ph.D, is the deputy director of the Intelligent Agricultural Technology Research Office of the Institute of Agricultural Information, Chinese Academy of Agricultural Sciences, also the first outstanding young person of the Agricultural Information Research Institute. His research interests include agricultural product quality and safety control technology, crop simulation model technology and agricultural GIS application. He proposed a framework of crop production management including field data acquisition, agricultural IOT application, crop model simulation, production plan formulation, production process monitoring and product quality control.

He has presided over, participated in and completed the National Science and Technology Support Program, 863 Program, the Ministry of Agriculture key projects, the National Natural Science Foundation project and other types of scientific research topics more than 10. He took a major role in the completion of wheat, corn planting management system, Chinese agricultural economic electronic map, bee product quality traceability system and other scientific research results, gained 9 institutional, provincial and ministerial scientific Research awards. He has published over 10 papers in international and domestic journals and conferences, gained 9 provincial and ministerial research awards, attained 1 technical invention letter of patent and 18 software copyright registrations.

Machine Vision Based Phenotype Recognition of Plant and Animal

Qixin Sun, Xue Xia, and Xiujuan Chai

Abstract Phenotype recognition is very important for plant and animal breeding, variety selection, genomics and phenomics. It is likewise a key technology to support the high-quality development of modern agriculture. Machine vision, as a key branch of artificial intelligence technology, is an important mode and means in the field of phenotype recognition. This paper summarizes the concepts and connotations, the technology and application examples of phenotypic visual perception and recognition of plant and animal at present in China, and discusses the problems and challenges of plant and animal phenotype recognition. Finally, this paper looks forward to the future development direction and application of vision-based phenotype recognition technology. The aim of this paper is to provide references for promoting the leap-forward development of agricultural phenotype recognition in China.

Keywords Machine vision · Phenotypes of plant and animal · Phenotype recognition technology · Agricultural phenomics

1 Introduction

From ancient times, agriculture has been the foundation of China's national economy, the source of food and clothing for human beings, and the basis for survival. China has been exploring the development path of modern agriculture and has made great achievements in many aspects, such as agricultural science and technology, economy and society. However, there are still many problems in China's modern agriculture, like the declining cultivated land resources, the degradation of cultivated land quality, the overall backward production technology, and the increasingly serious environmental problems, which seriously restrict the development of agriculture. China's modern agriculture needs to use digital technology to improve production efficiency

Q. Sun · X. Xia · X. Chai (✉)
Agricultural Information Institute, Chinese Academy of Agricultural Sciences, Beijing, China
e-mail: chaixiujuan@caas.cn

Key Laboratory of Agricultural Big Data, Ministry of Agriculture and Rural Affairs, Beijing, China

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_27

and achieve breakthroughs urgently. With big data and artificial intelligence moving into agriculture, the power of science and technology has a profound influence on the traditional agricultural production. The industrial revolution is coming. For example, artificial intelligence is used to predict the value of agricultural products in combination with soil and climate data. Pattern classification is used to identify plant pests and diseases in the acquired images. Through the information acquisition and analysis of crops or orchard growth, precision agricultural production can be realized.

Agricultural phenotypic data contain a large amount of information and a high degree of complexity. The traditional research methods are inefficient and subjective. Only a few phenotypic characteristics can be studied at a time. Affected by traditional breeding mode in animal husbandry, it is difficult to consider the differences among individual livestock in the production process, and to conduct precise breeding management of livestock. Machine vision, because of its advantages of automation, objectivity, non-contact, high-accuracy, and strong environmental adaptability, has become the research focus in artificial intelligence. It also gives a strong impetus to phenotype recognition and production refinement in the field of agricultural plant and animal. Nowadays, using machine vision to study the phenotypes of agricultural plant and animal has become a fundamental method for plant and animal breeding, variety selection, genomics and phenotype omics. Therefore, this paper focuses on the technologies and applications based on machine vision in agricultural plant and animal phenotype recognition, summarizes and analyzes the research actuality at home and abroad, and gives the future research directions.

2 Machine Vision Based Plant Phenotype Recognition

Phenotype is all or part of the interaction between a genotype and the environment. In other words, the phenotype is all or part of the identifiable characteristics and traits produced by the interaction between a genotype and the environment. Phenomics is a field of systematic study concerned with the phenome characterization of organisms or cells, which are characteristics of phenotypes that generate arise via the interaction of the genome with the environment. The concept is coined by S. Steven A. Garan, Director of the Aging Research Centre (ARC) at a guest lecture he gave at the University of Waterloo in 1996. Plant phenome is defined as all of the physical, physiological and biochemical characteristics and traits which are decided or influenced by genome and environments, which can reflect the plant structures and compositions, or reflect the processes and results of plant growth and development [1].

With the rapid development of high-throughput sequencing technology and plant functional genomics, more and more plant and trait parameters need to be measured quickly and accurately. The traditional methods of plant phenotype identification, which are inefficient, inaccurate and objective, have greatly restricted the development of fundamental plant biology research, including genetics and gene function research. Therefore, the foreign and domestic researchers had done a lot of researches on automatic perception of vision-based plant phenotyping and made a series of progresses. In recent years, there are many reviews in this field [2–7]. Among them, Walter et al. [5] summarize the image-based phenotype recognition methods, including spectral analysis and thermal imaging analysis, by taking the phenotypic study of roots as an example.

In vision-based plant phenotype recognition, the commonly used image acquisition methods include visible light imaging, chlorophyll fluorescence imaging, thermal imaging, near-infrared imaging, hyperspectral imaging, 3D imaging, laser imaging, magnetic resonance imaging, tomography and remote sensing images. Visible light imaging is the most widely used imaging method in plant phenotype recognition. It is often used to analyze visible biomass on the surface of the earth. It is simple, direct, and low equipment cost, but strongly affected by ambient light. Chlorophyll fluorescence imaging is not sensitive to the environmental light changes and mainly used to study plant resistance to pathogenic bacteria. Thermal imaging has great potential for low-cost and high-throughput field phenotyping. There are also a variety of recognition methods used in vision-based plant phenotype recognition. Traditional methods include threshold segmentation [8], feature classification [9], edge tracking [10], and other methods, as well as new deep learning methods such as Faster R-CNN [11] and YOLO [12]. Arvidsson et al. [13] use basic image processing methods (including binarization, hole filling, growth and shrinkage) to help researchers obtain phenotypic parameters, such as salient regions, convex bodies and compactness in comparing growth behaviors of different genotypes of Arabidopsis thaliana. Lu et al. [14] use a deep convolutional neural network to model the local visual characteristics of in-field corn and then regress the number of local corn tassels to solve the counting problem. Many international institutions and organizations have also carried out plant phenotypic profiling, such as the Australian Plant Phenomics Facility.

Taking the detection of phenotypic characteristics of corn, one of the three major food crops in China, as an example, traditional corn phenotype recognition is mainly performed by software analysis after manual measuring and taking photos. However, manual measurement is time-consuming, low accuracy, tedious, and heavy workload. These disadvantages greatly limit the efficiency of large-scale genetic breeding screening based on phenotypic traits. Therefore, it is very important to use modern technology to conduct accurate and efficient data collection and analysis of corn phenotypes.

The Institute of Crop Sciences of the Chinese Academy of Agricultural Sciences regularly recorded more than 900 corn plants in the whole growth cycle from three viewpoints, and obtained about 90,000 multimodal phenotypic data such as RGB images, fluorescent images and near-infrared images. Based on this, the Agricultural Information Institute of the Chinese Academy of Agricultural Sciences has carried out research on the detection of corn phenotypes, which mainly included: calculation of the projection area, mean and variance for the color values, detection and calculation of plant height and internode, identification and statistics of the number of leaves and detection and calculation of thickness of corn stem.

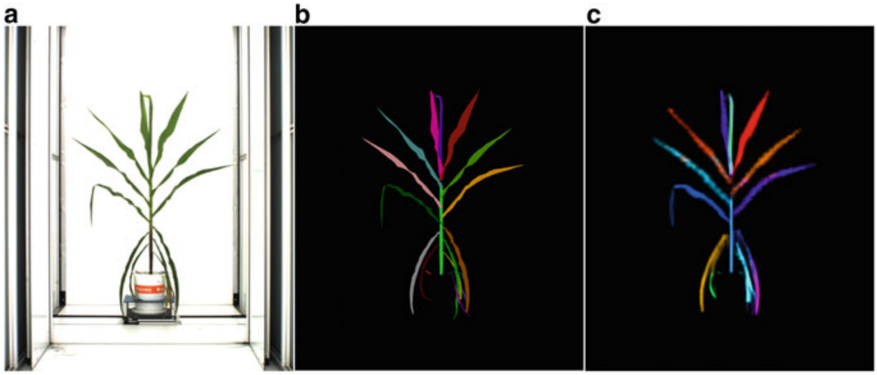


Fig. 1 An example from corn leaves recognition. **a** RGB image as input. **b** Multi-instance annotation. **c** Multi-instance object segmentation result

(1) In the problem of recognizing and counting corn leaves, not only the outline and position of the entire corn plant should be recognized from the image, but also the boundary and identity of each leaf should be determined at the pixel level. This can be formalized as multi-instance object segmentation, a typical problem in computer vision. By using the deep convolution neural network to transfer learning and fine-tuning parameters in the corn dataset, a good instance segmentation can be achieved under the condition that only more than 100 multi-instance segmentation annotation images are used in training. An example for the visualized recognition result is shown in Fig. 1.
(2) In phenomics, the internode of corn is the distance between leaf sheath points. The problems of measuring internode distance and corn height can be regarded as the problems of detecting leaf sheath points, bottom and top points of corn plants. Aiming at the corn dataset with high-resolution images and small target objects, a small object detection method guided by prior knowledge from coarse to fine is proposed in our previous work [15], as shown in Fig. 2. In the first stage, the original image is down-sampled to the appropriate size to reduce the burden on GPU memory and to roughly find the areas that may contain objects. On the down-sampled image, first, different feature maps of objects of different sizes are calculated through the backbone network. Then, the prior knowledge is used to guide the region generation network to generate a region of interest. Since the region of interest is relatively rough, in the second stage, features of small regions of interest are extracted from high-resolution images for fine classification, which improves the detection accuracy of leaf sheath points, bottom points and top points of corn plants. The corn internode distance and height detection results are shown in Fig. 3.

Using machine vision technology to obtain and detect phenotypic information of corn can efficiently and reliably extract plant appearance information, dig out

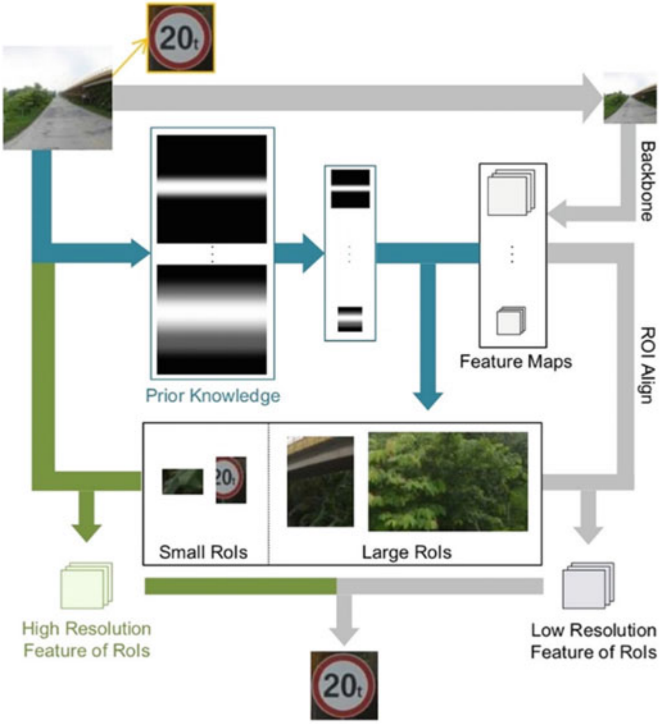


Fig. 2 The overall framework of corn detection method

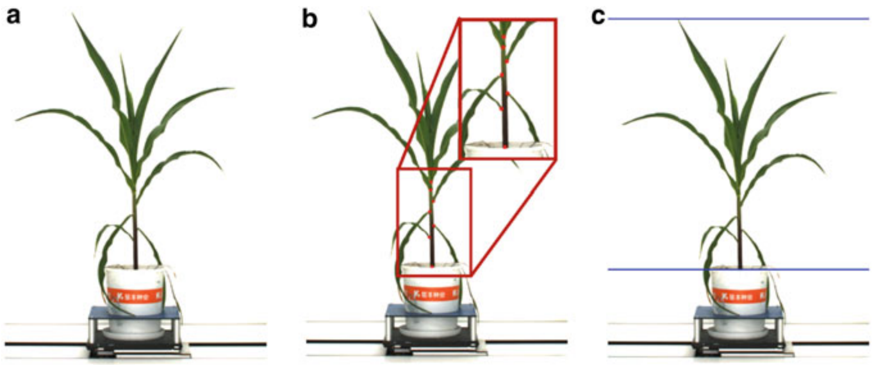


Fig. 3 A detection example of corn internode distance and height. **a** RGB image as input. **b** Result of leaf sheath points location. **c** Result of corn height detection

meaningful plant attribute characteristics, and assist in solving problems such as genetics and breeding.

Fruit tree is an important cash crop in China and plays an important role in agricultural production. Faced with the increasing planting area of fruit tree, the seasonal shortage of rural labor and the rising labor cost, the existing traditional fruit tree production methods of manual fruit production are difficult to meet the sustainable development needs of the fruit production industry. Using modern phenotype recognition technology to solve practical problems in fruit tree production becomes more important.

Focusing on fruit tree phenotypic identification, the Agricultural Information Institute of the Chinese Academy of Agricultural Sciences has carried out research on apple tree phenotype recognition. In terms of data accumulation, nearly 100,000 apple tree phenotype images were acquired based on high-definition image acquisition equipment, and a large-scale multi-modal apple phenotyping image dataset was constructed, including apple tree buds, flowers, young fruits, mature fruits and other organs. At the same time, the data were organized and annotated. Self-developed software tools were used to annotate the target apples in the image to form phenotypic labels, which provided data foundation for further research. In terms of detection models, deep convolutional neural networks can independently learn the objects to be recognized. At present, the widely used convolutional neural network methods for object recognition and detection can be divided into two categories. One is region-based object recognition, such as Faster R-CNN and Mask R-CNN. These methods are effective in small object detection, but slow in speed. The other is regression-based object recognition, such as SDD and YOLO. These methods use an end-to-end learning for object detection and recognition, which are much faster than the region-based methods and can meet the real-time requirements. Therefore, combined with the actual needs of fast and high-precision apple detection in the orchard scene, the YOLOv3 deep learning algorithm with fast detection speed in the regression-based methods is selected for this study. YOLOv3 uses a single neural network to predict the bounding box and class probability directly from the full image in one inference. It divides the input image into grids, and each grid cell predicts k bounding boxes and confidence scores of the bounding boxes, as well as the conditional probability of the C class. YOLOv3 has greatly improved on the basis of YOLO. It introduces the anchor concept from Faster R-CNN into the original framework, which improves the network performance. Some examples from the detection results of the model trained in this research are shown in Fig. 4, which demonstrates that the model can detect the apple well. Through performance testing and analysis of the model, the results show that the model in this research has a better fruit detection effect under different fruit numbers and light angles, with the accuracy rate and recall rate reaching 91% and 87% respectively.

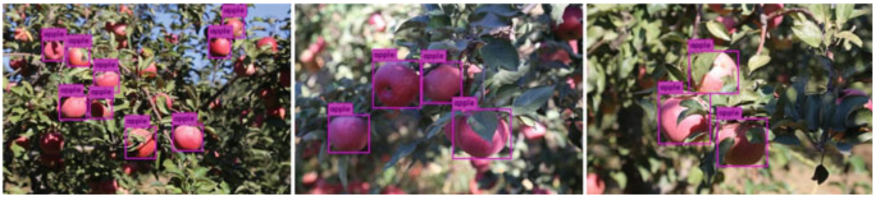


Fig. 4 The visualized results of apple detection

3 Machine Vision Based Animal Phenotype Recognition

Animal phenotypes can refer to any measurable biological traits, including individual behavior, appearance, and developmental characteristics. Large livestock in animal husbandry such as cattle, pigs and sheep are widely concerned by animal husbandry research because of their high economic value. Although the market value of the animal husbandry industry is huge, the cost of animal husbandry in China is very high. The unit feed cost and unit labor cost are significantly higher than developed countries. On the whole, China's animal husbandry industry has entered a period of reform, from a traditional free-range model to a large-scale and intensive model. Nowadays most of the breeding industry in China is extensive management, which does not consider the differences among livestock individuals in feeding preferences, food intake and health status. However, this traditional breeding method is not suitable for modern breeding industry's refined breeding standards. The construction of modern intelligent farm must provide individualized management of livestock and intelligent perception of their behaviors. Identification is the data entrance for intelligent farm. Health monitoring data, nutrition data, and breeding data of livestock individuals need to be related with themselves.

Traditional livestock identification mainly includes ear tags, implanted chips, printed patterns and other methods, which have obvious disadvantages. Whether ear tags, ear scissors, or implanted chips are invasive, difficult to prevent damage to livestock, and violate the principle of welfare feeding. Patterns on livestock cannot avoid staining, which affects the identification partly. In contrast, vision-based livestock phenotype identification has the advantages of low cost, non-contact, non-invasive, and no consumables, which has attracted a large number of scholars and industry personnel.

The researches based on machine vision in intelligent farm mainly focus on livestock identification, behavior detection, quality determination and disease diagnosis. Since livestock are mostly group-raised, the most important problem to study about livestock individuals is identification. For the problem of livestock identification, the existing researches mainly focus on two aspects: object detection and identity recognition. The main task of object detection is to detect a single animal or a group of animals in the image, so that the behavior and state of the animal can be judged depending on the characteristics of the animal image. The identity recognition is to distinguish each animal in the image and identify its identity, so as to facilitate the continuous observation and personalized management of individual livestock.

Currently, the main methods used for vision-based livestock object detection are threshold segmentation, feature classification, contour key points, statistical models, deep learning, etc. Lind et al. [16] realize the recognition and tracking of individual pigs by automatic threshold segmentation and image subtraction. Yang et al. [17] use an improved full convolution network to detect sows in the delivery houses. This method can reduce the interference in detection of sows with uneven color, unstructured lighting conditions and piglets. Vision-based identity recognition methods include color marking, template matching, feature transformation and deep learning. Ahrendt et al. [18] propose a pig tracking algorithm based on support graph, which can identify and lock individual pigs in pigsty in real-time. This study lays the foundation for subsequent real-time analysis of pig behavior, but when pigs move quickly or multiple pigs are too close, the system would fail and lose the ability to track the pigs. Yang et al. [19] study the individual identification method of pigs based on the Faster R-CNN model. The image of the pig's back with a label is taken by an overhead camera. The position of each pig is detected by using the Faster R-CNN model and identify the identity of each pig at the same time. However, the identification in this method is still achieved by marking on the pig.

In recent years, face recognition, as an important research direction in image recognition, has been successfully commercialized. Face recognition is usually used for non-invasive access control and monitoring, which is very similar to the application scenario of intelligent farm. Therefore, in theory, related technologies in face recognition can be transferred to animal identification.

The Agricultural Information Institute of the Chinese Academy of Agricultural Sciences studies the problem of automatic pig detection and identification. In terms of data collection, video data were collected from large-scale specialized pig farms, which cover a wide variety of types and sizes. The data includes pig face and pig body, which includes two kinds of scenarios of positioning bar single feeding and captive feeding. Among them, the captive feeding data is used for the pig face detection and pig body detection, and the positioning bar single feeding data is used for the pig face recognition.

First, the raw video data of pig was preprocessed. Pig images were extracted from the video with fixed frame intervals, and the fuzzy, similar and objectless images were manually filtered out. Next, the face and body positions of pigs in the captive feeding data were manually labeled for the training of detection algorithm. Then, the advanced FaceBoxes detection algorithm is selected to learn the data of pig face and pig body, so as to realize the real-time detection. After that, the real-time detection algorithm is used to detect the pig face in the positioning bar single feeding data, and the detection results are put into the neural network for the identity recognition. Combined with a large number of face data on the open source network, there is a potential connection between human face and pig face. The human face recognition model can be transferred to help the pig face recognition.

The pig detection algorithm used in this study can accurately detect the pig faces in the images. The detection result is shown in Fig. 5. The average detection speed

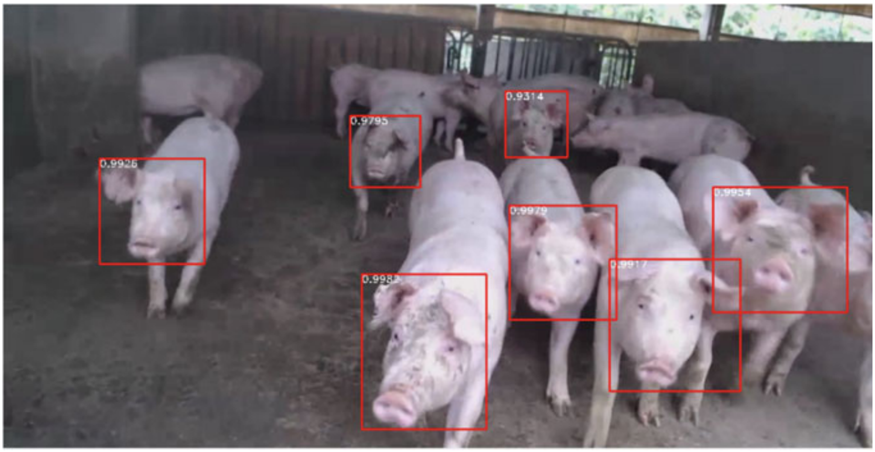


Fig. 5 The result of pig face detection

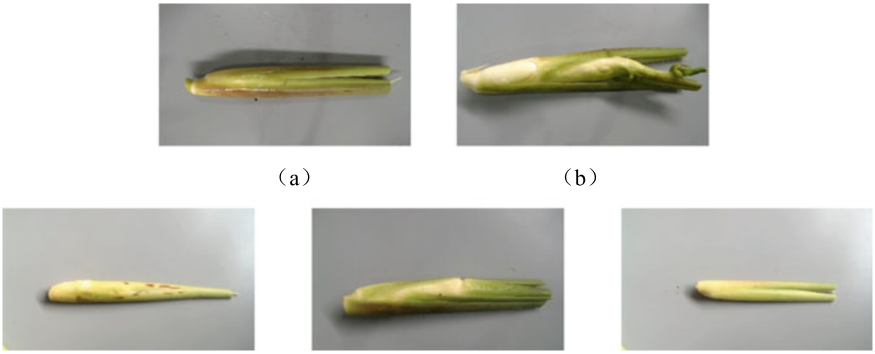
of images in different sizes is 3.1 ms per image and the average detection accuracy of high-resolution images is more than 90%. For pig identity recognition, using human face recognition with transfer learning, is tested on a complex dataset that contains different pose, illumination, occlusion and other factors. By comparing with the results of direct training and fine-tuning of the convolutional neural network, the results show that the model used in this study performs well in pig face detection and recognition in complex scenes under unrestricted conditions. The recognition accuracy of nearly 200 pig images of 80 classes reaches 99.48%, and the performance is obviously better than the existing methods.

4 Machine Vision Based Agricultural Equipment and Robots

The intellectualization of agricultural machinery and production tools reflects the transformation process of agricultural development from traditional to modern. Intelligent agricultural machinery and equipment synthetically use modern information technologies such as mechatronics, sensor control, communication technology and machine vision. It can complete a variety of tasks efficiently, safely and reliably, and apply the agricultural phenotype recognition results to actual production, such as agricultural products automatic picking, agricultural products automatic classification, and pest and disease monitoring and prevention. Japan is one of the earliest countries to explore agricultural intelligent devices and robots. Now, the research, design and manufacturing of robots in Japan are all at the world's advanced level, and its agricultural picking robots have been commercialized. The United States, as a world power in the development and manufacturing of intelligent equipment and robots, has developed and manufactured a large number of automatic intelligent devices and robots in agriculture. At present, the "agricultural Internet of Things + Robotics" technology is available on 80% of the nation's large farms. With the help of highly automated large-scale agricultural robots, a 10,000 acres of land can be managed and harvested by just three persons. The development of intelligent agricultural devices and robots in the United Kingdom is also very rapid. For different kinds of crops, there are corresponding robots for production. Harper Adams University in the UK has completed the world's first attempt to grow and harvest grain by machine. They use drones to monitor crop growth and develop agricultural machinery to grow, manage and harvest crops on their own. Using a combination of global positioning system (GPS) and smart multi-purpose tractors, the German research institute has developed a weeding machine that can move quickly between farm plots and remove weeds accurately. France has developed a fully functional grapery robot that can do almost all the work of plantation workers, including pruning vines, cutting young shoots and monitoring the health of the soil and vines. Belgium has developed a strawberry-picking robot, which uses 3D visual sensing to identify the ripeness of strawberries, and uses 3D printed "palms" to quickly harvest strawberries in the field without damaging them.

Research on intelligent agricultural devices and robots starts late in China, but has made significant progress after more than a decade of development. The upgrading of agricultural industry driven by intelligent equipment combining with phenotype recognition is gradually changing the farmers' lifestyle. At the same time, the state pays great attention to its development. The "13th Five-Year Plan for the Development of the Robot Industry" focuses on breakthroughs in service robots and the next generation of intelligent robots [20]. The National Engineering Research Center for Information Equipment in Agriculture integrates laser active measurement and visual servo technology to realize the automatic targeting and picking robot for greenhouse tomato [21]. China Agricultural University develops a vegetable grafting device combining multi-sensor fusion technology to realize grafting positioning, grasping, cutting and combining of vegetable seedlings [22, 23]. The aquaculture robot developed by the Nanjing Institute of Software is composed of a hull, a bait feeder and a control system. The aquaculture robot can implement aquaculture functions such as spraying water, sprinkle medicine and feeding food. Ying et al. [24, 25] from Zhejiang University design a fruit surface feature recognition algorithm based on machine vision and develop a fruit quality detecting and grading device that can grade fruit quality in a single channel. The cotton-picking intelligent equipment developed by Nanjing Agricultural University can identify the grade of seed cotton and pick mature cotton [26].

For example, the quality sorting of wild rice stem is mainly done by manual labor these days, but due to the gradual reduction of rural labor, it is difficult to recruit labor at low wages. At the same time, manual sorting also has the problem of low precision and slow efficiency. To reduce the labor cost and improve the classification accuracy of wild rice stem sorting, it is urgent to propose an automatic wild rice stem grading algorithm. It combines with hardware equipment can achieve the automatic grading, sorting and packaging of wild rice stem. Thus, the manual work burden and



(a) (b) (c) (d) (e)

Fig. 6 Examples from wild rice stem grading image dataset. **a** High-quality, **b** irregular shape, **c** pests and diseases, **d** aging, **e** small size

the production cost can be reduced, the "machine replacement" can be realized, and the economic benefits of wild rice stem can be improved.

The Agricultural Information Institute of the Chinese Academy of Agricultural Sciences collected 1,640 high-quality wild rice stem samples from the production base, took double-sided images for each sample to obtain 3,280 high-quality images. Besides, 1,536 inferior wild rice stems with aging, irregular shape, pests and diseases, and small size were collected. Double-sided images were taken for each sample to obtain 3,072 inferior images. These images were constructed as the wild rice stem grading image dataset. Some examples are shown in Fig. 6.

The research is mainly divided into the following three steps: ① Data collection and preprocessing. The research collects wild rice stem images and records the grade of each image, such as high quality and low quality, as the label of the image. The dataset is partitioned into training sets and test sets according to the proportion. ② Based on the deep convolutional neural network, it trains the grading model parameters of wild rice stem according to the corresponding relationship between images and labels. ③ The trained wild rice stem grading model can extract the depth features of the image to be detected, thereby determine the wild rice stem quality level.

This study is based on deep learning, extracts the phenotypic characteristics of wild rice stem, and realizes the automatic wild rice stem grading, which distinguishes wild rice stem between high quality and low quality. The grading accuracy can reach 99.6% on the collected wild rice stem dataset. Besides, the Agricultural Information Institute of the Chinese Academy of Agricultural Sciences integrates the automatic grading algorithm and hardware to produce a wild rice stem grading equipment, as shown in Fig. 7. The device uses visual grading technology to realize the grading of stick-shaped agricultural products, and its main functions include automatic loading, transmission, grading and packaging. This device can effectively save human and material resources, improve production efficiency, and promote the development of agricultural intelligent equipment to a certain extent.

Fig. 7 The wild rice stem grading equipment

5 Problems and Challenges

Although researches based on phenotypes have been conducted for more than 20 years, visual perception techniques for agricultural phenotypes are still immature. In general, it is still in laboratory conditions and small-scale simple applications. The application of machine vision technology in agricultural phenotype recognition still has the following problems:

(1) Lack of data. Currently there is still short of a unified and open large-scale general database for plant phenotype research or animal phenotype research. For example, since the harsh and complex growing environment of livestock, changes in the characteristics and activities of livestock groups/individuals will reduce the stability of data collection and increase the difficulty of the database construction. To promote the intensive research at home and abroad on the identification of plant and animal phenotypes, a unified platform for research is urgently needed.
(2) Low reliability. China's agricultural production environment is complex, and each region is interrelated and mutually restricted. However, because of various natural factors such as light, temperature, latitude, and altitude, there are obvious regional and seasonal differences. The processing results of image or video information based on machine vision are easily affected by measurement conditions, environment and other factors, which makes the machine vision technology has the features of poor environmental adaptability and reliability and great limitations in agricultural phenotype recognition.
(3) Weak generalization. In addition to the complex production environment in China's agriculture, the objects are also complex and variable. Current research is still at the stage of designing different algorithms for different agricultural research objects according to their characteristics, which lacks generalization.

Besides, the algorithms are mostly designed for a single or a small number of data modalities, with limited versatility and scalability.

(4) Poor real-time performance. Due to the limitations of mechanical control system and image analysis algorithm, most of the current researches are based on static and two-dimensional images. In general, the original plant phenotype images have high-resolution, which makes it difficult for some complex machine vision algorithms to meet the real-time detection requirements. For animal phenotype recognition, it is more necessary to analyze the video frame by frame to monitor the animal behavior. These all put forward higher requirements on the real-time performance of the machine vision algorithms.

(5) Inferior operability. Owing to the low level of integration and intelligence of machine vision based agricultural equipment and the complexity of the operation, most of the machine vision based agricultural research is still at the experimental stage. There are still many practical problems that need to be solved in large scale promotion. Besides, agriculture phenotype research belongs to the multidisciplinary field, which objectively increases the difficulty.

6 Summary and Prospect

China is active in many aspects of machine vision based agricultural phenotype research, mainly concentrating in crop growth information detection and recognition, automatic picking, quality grading, pest and disease monitoring, agricultural machinery navigation, livestock recognition and behavior detection and so on. However, due to the complex agricultural production environment and diverse agricultural research objects, machine vision based agricultural phenotypic recognition still needs a long period of development to be mature. From the perspective of research trends, the following aspects will be the key points of future research:

(1) Data is an important foundation for scientific research. The methods based on machine vision heavily depend on the size and distribution of the training samples. At present, there are no mature and open large-scale agricultural phenotypic databases for scientific research at home and abroad. Therefore, building a large-scale dataset with more agricultural objects, more phenotypic types, greater differences is very necessary for effective application and further study in this field.

(2) Deep learning is the key to the study of machine vision. How to improve the accuracy and processing efficiency of existing algorithms in the specific agricultural issues, enhance the robustness of algorithms, provide real-time support technique is still the development direction of machine vision based agricultural phenotypic recognition in the future.

(3) Research on integrated systems with tighter structures, lower cost and faster processing speed is an important development direction for machine vision based agricultural applications in the future, which is also conducive to the

popularization of agricultural machinery equipment or intelligent equipment based on machine vision system in the further.

(4) The machine vision system integrating various technologies is also a research hotspot. For example, the integration of machine vision system and BeiDou Navigation Satellite System can achieve high precision and low cost of farmland navigation. The integration of 3D imaging technology, neural network technology and intelligent control technology can make the farmland robots more intelligent.

In many crucial agricultural processes, such as breeding, spraying, weeding, picking and post-processing, the applications of machine vision based phenotype recognition can replace manual work, avoid the harm of adverse environment to human, reduce the influence of human factors, and improve production efficiency. Therefore, machine vision based phenotype recognition will be a critical technology for the intelligent and precise agriculture in the future and help agricultural production move towards modernization, automation and intelligence.

References

1. Pan Y (2015) Analysis of concepts and categories of plant phenome and phenomics. Acta Agron Sin 41(2):175–186 (in Chinese)
2. Yang W, Duan L, Chen G et al (2013) Plant phenomics and high-throughput phenotyping: accelerating rice functional genomics using multidisciplinary technologies. Curr Opin Plant Biol 16(2):180–187
3. Granier C, Vile D (2014) Phenotyping and beyond: modelling the relationships between traits. Curr Opin Plant Biol 18:96–102
4. Li L, Zhang Q, Huang D (2014) A review of imaging techniques for plant phenotyping. Sensors 14(11):20078–20111
5. Walter A, Liebisch F, Hund A (2015) Plant phenotyping: from bean weighing to image analysis. Plant Methods 11(1):14
6. Liu J, Zhao C, Yang G et al (2016) Review of field-based phenotyping by unmanned aerial vehicle remote sensing platform. Trans Chin Soc Agric Eng 32(24):98–106 (in Chinese)
7. Gao Y, Gao J, Li H et al (2017) Research progress and development countermeasures of plant phenotype monitoring technology. Jiangsu Agric Sci 45(11):5–10 (in Chinese)
8. Jidong L, De-An Z, Wei J et al (2016) Recognition of apple fruit in natural environment. Opt Int J Light Electron Opt 127(3):1354–1362
9. Lu J, Lee WS, Gan H et al (2018) Immature citrus fruit detection based on local binary pattern feature and hierarchical contour analysis. Biosyst Eng 171:78–90
10. Xu L, Lv J (2018) Recognition method for apple fruit based on SUSAN and PCNN. Multimed Tools Appl 77(6):7205–7219
11. Wang D, He D (2019) Recognition of apple targets before fruits thinning by robot based on R-FCN deep convolution neural network. Trans Chin Soc Agric Eng 35(3):156–163 (in Chinese)
12. Tian Y, Yang G, Wang Z et al (2019) Apple detection during different growth stages in orchards using the improved YOLO-V3 model. Comput Electron Agric 157:417–426
13. Arvidsson S, Pérez-Rodríguez P, Mueller-Roeber B (2011) A growth phenotyping pipeline for Arabidopsis thaliana integrating image analysis and rosette area modeling for robust quantification of genotype effects. New Phytol 191(3):895–907

14. Lu H, Cao Z, Xiao Y et al (2017) TasselNet: counting maize tassels in the wild via local counts regression network. Plant Methods 13(1):79
15. Yang Z, Chai X, Wang R et al (2019) Prior knowledge guided small object detection on high-resolution images. In: IEEE international conference on image processing, Taipei
16. Lind NM, Vinther M, Hemmingsen RP et al (2005) Validation of a digital video tracking system for recording pig locomotor behaviour. J Neurosci Methods 143(2):123–132
17. Yang A, Huang H, Zheng C et al (2018) High-accuracy image segmentation for lactating sows using a fully convolutional network. Biosys Eng 176:36–47
18. Ahrendt P, Gregersen T, Karstoft H (2011) Development of a real-time computer vision system for tracking loose-housed pigs. Comput Electron Agric 76(2):169–174
19. Yang Q, Xiao D, Lin S (2018) Feeding behavior recognition for group-housed pigs with the Faster R-CNN. Comput Electron Agric 155:453–460
20. Ministry of Industry and Information Technology, National Development and Reform Commission, Ministry of Finance et al (2016) Development plan of robot industry (2016–2020) 2016–03–21 (in Chinese)
21. Wang X, Wu P, Feng Q et al (2016) Design and test of tomatoes harvesting robot. J Agric Mech Res 38(4):94–98 (in Chinese)
22. Zhang T, Wei J (1999a) Study on vision system of vegetable grafting robot (I)—searching for growth point of cucurbit seedling with morphological method. J China Agric Univ 04:45–47 (in Chinese)
23. Zhang T, Wei J (1999b) Study on vision system of vegetable grafting robot (II)—ssearching for growth point of cucurbit seedling with geometrical method. J China Agric Univ 04:48–50 (in Chinese)
24. Rong D, Rao X, Ying Y (2017) Computer vision detection of surface defect on oranges by means of a sliding comparison window local segmentation algorithm. Comput Electron Agric 137:59–68
25. Rong D, Ying Y, Rao X (2017) Embedded vision detection of defective orange by fast adaptive lightness correction algorithm. Comput Electron Agric 138:48–59
26. Zhu R, Zhu Y, Wang L et al (2016) Cotton positioning technique based on binocular vision with implementation of scale-invariant feature transform algorithm. Trans Chin Soc Agric Eng 32(06):182–188 (in Chinese)

Xiujuan Chai, Professor, doctoral supervisor, and Director of the agricultural information technology department in the Agricultural Information Institute, Chinese Academy of Agricultural Sciences. Her research interests cover computer vision, pattern recognition and artificial intelligence. She has presided over and participated in more than 20 programs from National Natural Science Foundation of China (Youth Foundation, General Foundation, Key Program and Major Program), 2011 special plan of the Ministry of Education, Innovation Project of Academy of Agricultural Sciences, basic scientific research expenses and enterprise. She has published more than 60 academic papers and has been cited by Google scholar more than 1,500 times. Her research results have won many scientific research awards, including one second prize of the National Natural Science Award, one science and technology award of China Simulation Federation, and two first prizes in international academic competitions.

Research and Application of Key Technologies for Request Prediction and Assignment on Ridesharing Platforms

Bo Zhang, Jieping Ye, Xiaohu Qie, Guobin Wu, Liheng Tuo, and Yiping Meng

Abstract DiDi ridesharing platform provides users with accurate and high-precision travel services in real-time by using data-driven technology, machine learning method, large-scale distributed computing, operations optimization and other technique. The platform has made great breakthroughs in the key technologies of intelligent prediction and dispatch of travel platforms: Estimated Time of Arrival (ETA), Intelligent Dispatching and Supply and Demand Forecasting. We proposed a novel deep learning solution to predict the vehicle travel time based on floating-car data. We also present an order dispatch algorithm in large-scale on-demand ride-hailing platforms. While traditional order dispatch approaches usually focus on immediate customer satisfaction, the proposed algorithm is designed to provide a more efficient way to optimize resource utilization and user experience in a global and more farsighted view. We deploy the spatiotemporal multi-graph convolution network (ST-MGCN), a novel deep learning model for demand forecasting.

Keywords Estimated time of arrival · Order dispatch · Ride-hailing demand forecasting

1 Introduction

China is home to the world's largest and most diverse mobility market, resulting in immense market complexity. The country has an urban population of 800 million people, with adult workers accounting for roughly 70% of that, or 560 million. Every day there are roughly 1.1 billion trips on average. With an urban population density greater than most other countries, as well as a total number of big cities far greater than said countries, issues such as congested city streets and low vehicle utilization have severely affected the quality and efficiency of the commuter experience. China's large population, differentiated socioeconomic conditions, and diverse mobility needs have only increased the difficulties surrounding real-time

B. Zhang (✉) · J. Ye · X. Qie · G. Wu · L. Tuo · Y. Meng
Didi Chuxing, Beijing, China
e-mail: bob@didiglobal.com

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_28

mobility prediction, planning, and request assignment. People's mobility problem and how to effectively meet people's mobility needs have become a global problem, yet DiDi Chuxing is committed to making traveling better through the utilization of technology. Since its founding in 2012, the company has steadily built out a ridesharing network covering more than 1,000 cities in five countries. Over the past several years, DiDi's ridesharing platform has developed multiple key technologies, including world-class smart request prediction and assignment technology.

Smart request prediction and assignment technology makes up the core of any ridesharing platform. This technology mainly encompasses estimated time of arrival (ETA), smart request assignment, and the prediction of demand and supply volumes. The ETA function is an essential service offered by ridesharing platforms, and highly precise ETA capabilities are crucial when it comes to pre-trip pickup time estimation, price estimation, request assignment, vehicle dispatch, carpool options, or ETA during the trip. The request assignment—the process whereby the platform assigns ride requests to online drivers—is to satisfy society's greater mobility needs, which means most riders' needs are met. Demand volume prediction is one of the core issues facing the entire service industry. For a ride-hailing service provider, an accurate demand volume predictive algorithm can help the platform improve vehicle utilization, optimize request assignment, dispatch drivers more effectively, manage traffic, and circumvent congested streets. By utilizing big data and proprietary AI technologies, DiDi has made major breakthroughs in the key technology field of smart request prediction and assignment.

2 Progress of Relevant Research

Many ETA techniques have already been widely researched in geographic information systems (GIS), and standard solutions have been established in the industry. These existing solutions can be categorized into two types: route-based solutions and data-driven solutions.

Route-based solutions use intuitive physical models to represent travel time, where the total travel time for a given route is the aggregate of time needed to travel each route section plus the delay incurred at each intersection. Travel time estimate is represented as

$$\hat{y} = \sum_i \hat{t}_i + \sum_j \hat{c}_j$$

t_i is the estimated travel time for route section i, and c_j is the estimated wait time at traffic light j. Figure 1 might be a better illustration for such solutions. Route-based solutions divide the estimated travel time for the entire route into several sub-problems, each including the estimated travel time for each route section and the estimated delay incurred at each intersection. Relevant GIS data is used to estimate time needed to travel from one route section to the next. Machine learning approaches,

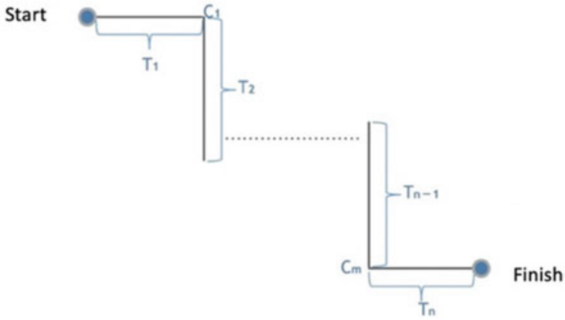


Fig. 1 Route-based solutions

such as regressions and tensor decompositions, have also been used to estimate travel time. Some researchers focus on using regularized sub-routes to approximate the travel time for a given route. Even though route-based solutions are relatively straightforward, with limited online computation needed, the following drawbacks exist:

① Even with a significant increase in available datasets and data diversity, the temporal sparsity reduces the efficacy of predicting real-time traffic conditions for the entire route.
② The capability to physically model mobility modes of a dynamic transportation system is insufficient, thus it is difficult to predict the traffic conditions for a specific section of the route at a specific future time, not to mention the state of traffic signals at a specific intersection. As such, a high degree of estimation precision for t_i and c_j cannot be guaranteed.
③ As shown in the above-mentioned equation, the division of travel time in sections will result in cumulative error.
④ The uniqueness of travel time estimates is ignored, i.e. how travel time might vary from driver to driver.

The second approach used to estimate travel time is data-driven solutions. Recently, the extension of data warehouses has enabled machine learning to serve as a powerful tool in predictive analytics. Machine learning has been applied in route-based solutions to predict traffic speed and travel times for each route section. In addition, attempts have been made to directly predict travel time for the entire route based on travel history. The problem of estimated time of arrival (ETA) can be described as a multivariable time series predictive problem. For example, it has been proposed that the weighted mean of neighboring trips can be used to estimate the time needed to travel the route in question, provided the trips are similar in their origins and destinations. Data-driven solutions have better extensibility. These data-driven solutions, however, are not without drawbacks their own, including:

① The issue of insufficient data coverage still exists. It is difficult to acquire drive times across all historical time periods for either the route in question or any similar routes. Therefore, this approach is more suited for highways, where data coverage is better due to the limited number of route sections and more stable traffic conditions.

② Travel time estimation is restricted to just a few fixed routes, with poor applicability to routes not included in the training data, thus limiting extensibility.

③ A lot of key information, such as traffic and personalized information, is neglected in data-driven approaches, which leads to low predictive precision.

With the widespread use of large-scale historical data in combination with machine learning tools, the boundaries between route-based and data-driven approaches are blurring, yet fundamental weaknesses—such as insufficient data coverage, poor generalization capability, and insufficient information utilization—continue to limit the effectiveness of these current approaches.

Smart request assignment technology based on reinforcement learning (RL): Request assignment is a process by which the platform assigns requests, sent by riders to drivers that are online at that moment in time [9]. The goal of request assignment is to satisfy society's greater mobility needs, which means most riders' needs are met. A simple assignment approach might consider using the principle of proximity, namely assigning the request to the closest driver. It is not always so simple, however, and many factors will usually go into the actual request assignment process, including fluctuations in demand and supply throughout the day, the distance between the driver and the rider, the time series of request issuance, and the direction the driver is headed in at the time of pickup. In regions where drivers and riders are sufficiently dense, optimization can be achieved through the principle of proximity, but most often supply and demand levels do not match, and so during these times the principle of proximity will fail to achieve optimal request assignment.

Demand volume predictive model based on Multi-View Spatio-Temporal convolutional neural network (CNN): Demand volume prediction is a typical spatio-temporal predictive problem. Temporally, this problem has a clear gradation and periodicity; spatially, the distributions of requests share clear regional similarities. Traditional deep learning approaches use CNN+RNN (recurrent neural network) to model space and time. For example, a deep spatio-temporal network uses stacked CNNs to extract spatial features, while a deep multi-view spatio-temporal network (DMVST) enables multi-view learning through simultaneously introducing both RNN and CNN. Under these approaches, the RNN modeling of time series is overly smooth; in terms of spatial modeling, CNN cannot handle non-Euclidean geometry, nor can it introduce multimodal city data so to conduct more precise and accurate modeling. All these issues have led to relatively low predictive accuracy and stability.

3 ETA Techniques

DiDi's ridesharing platform makes predictions about a user's future trips based on vast amounts of historical data and actual user demand, effectively assigning requests to optimize user experience and enhance overall platform efficiency. ETA systems need to consider both spatial characteristics (e.g. the number of traffic lights on a specific route, speed limits, whether or not a faster route is available) and temporal characteristics (e.g. traffic patterns in morning and evening rush hours, traffic flow congestion, congestion caused by traffic accidents, etc.) within any given transportation system. On the other hand, the operation of a transportation system cannot occur without people and vehicles, and is subject to the influence of external factors, hence the need to model personalized and external features to solve the problem's that accompany ETA forecasting. Personal and external features include variables such as driver behavior, and the impact weather—and subsequent driving conditions—have on driving speeds. As such, calculating ETA is an extremely complicated and challenging problem.

DiDi's ridesharing platform uses a systematic ETA technique, in combination with machine learning solutions, to overcome the drawbacks of the current approaches. A rich and effective feature system has been developed for location-based data, including floating car data, street network data, and user behavior information. DiDi has constructed a new deep learning model that integrates the advantages of wide linear models, deep neural networks, and recurrent neural networks to accurately gauge ETA [1].

3.1 Feature Engineering

DiDi has built a systematic feature set in the map domain—including space, time, traffic, personalized information, and extended information—and comprehensively covering a number of trip factors, including all potential routes from start to finish, relevant route sections, intersections, traffic lights, points of interest (POI), corresponding temporal properties, real-time road conditions, driver behavior, weather conditions, and temporary traffic restrictions [2]. Feature engineering architecture is shown in Fig. 2.

Feature engineering outputs three types of features:

- Dense Features: trip-level real number features (e.g. spherical distance between origin and destination, GPS coordinates for origin and destination).
- Sparse Features: trip-level discrete features (e.g. time slice numbers, days of the week, types of weather).
- Sequential Features: street-level features. Real number features can be directly inputted into the model, while sequential features need to be embedded before they can be inputted. Please note that it is no longer one feature vector per trip, but one feature vector per section of each trip (e.g. lengths of route sections, number of lanes, functional levels, and real-time traffic speeds).

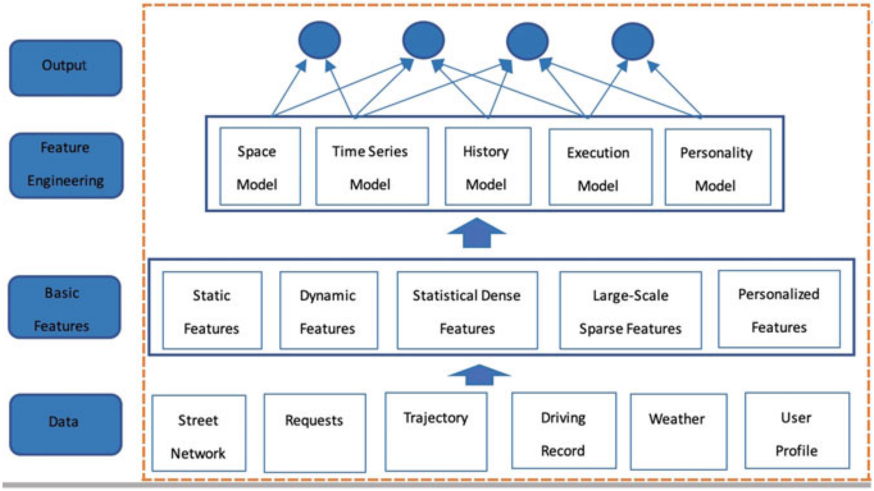


Fig. 2 Architecture of feature engineering

3.2 Target Optimization

Considering that user sensitivity to error has to do with relative values, Mean Absolute Percentage Error (MAPE) has been selected as the target function, with the corresponding optimization problem of

$$\min_f \sum_{i=1}^{N} \frac{|y_i - f(x_i)|}{y_i}$$

y_i is the actual time of arrival (ATA), $\hat{y}_i = \mathbf{f}(x_i)$ is the ETA, and $f(x_i)$ represents the regression model. Regularizer $\mathbf{\Omega}(f)$ has been added to prevent overfitting. DiDi has designed a deep learning model structure called a Wide-Deep-Recurrent Network (WDR) to perform regression on the target function $f(x_i)$.

$$\min_f \sum_{i=1}^{N} \frac{|y_i - f(x_i)|}{y_i} + \Omega(f)$$

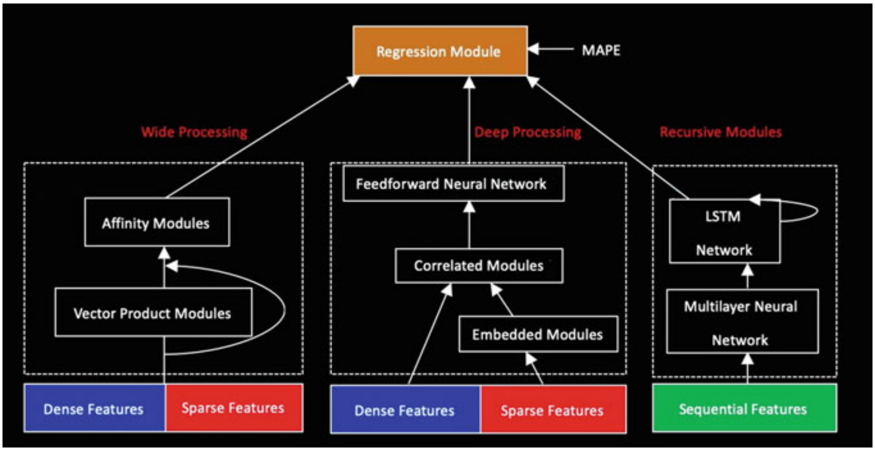


Fig. 3 Wide-deep-recurrent network architecture

3.3 Deep Learning Model Structure (Wide-Deep-Recurrent Network, WDR)

Deep learning model uses multiple basic models to fuse dense features with large-scale sparse features, surpassing the ETA technique offered by a deep neural network's general non-linear models. The new model not only boasts the advantages of its predecessor, but can also effectively use low dimensional dense features, high dimensional sparse features, and local features of the route section sequence, as well as Adaptive Moment Estimation (Adam)—a random gradient descent approach that features adaptive step sizes and momentum—to effectively improve the model's performance. The MAPE figure, a key indicator for the accuracy of ETA, is close to 11.5%, meaning the ETA error is within eight minutes for a one-hour trip, thus providing a precise real-time ETA for both the driver and the rider. Wide-Deep-Recurrent Network Architecture is shown in Fig. 3.

3.4 Experiment Result

The offline experiment used four NVIDIA Tesla P100 GUPs to perform parallel training, with the dataset shown in Table 1. Comparisons were made against a route-based method for estimated time of arrival called route-ETA and machine learning algorithms (GBDT, FM, WDR, WD-MLP, PTTE [3]). Comparisons have also been made between WDR and TEMP [4]. The result demonstrates that WDR is significantly better for both Pickup (driver picking up rider) and Trip (driver taking rider to the destination) (Tables 2 and 3).

Table 1 Offline datasets

	Date	Pickup (meter)	Trip (meter)
Training set	1.1 # 5.10 (2017)	48	51
Validation set	5.11 # 5.17 (2017)	3	4
Test set	5.18 # 5.31 (2017)	6	7
Unique links	#	0.5	0.5
Unique drivers	#	0.4	0.4

Table 2 Test results on pickup datasets

	MAPE	MAE (second)	MSE (second)
Route-ETA	31.27%	88.0	16555.8
PTTE	29.35%	83.3	13153.5
GBDT	23.64%	68.3	11674.2
FM	21.41%	63.6	9664.2
WD-MLP	21.58%	64.1	9816.3
WDR	**20.83%**	**59.9**	**9078.2**
*TEMP_{rel}	35.30%	79.7	15480.7
*WDR	**21.68**	**55.6**	**7962.0**

Table 3 Test results on trip datasets

	MAPE	MAE (second)	MSE (second)
Route-ETA	15.01%	153.2	66789.8
PTTE	14.78%	150.1	66141.2
GBDT	14.01%	133.2	52006.6
FM	12.87%	122.5	47457.3
WD-MLP	13.43%	127.5	50012.8
WDR	**11.66%**	**112.4**	**37482.7**
*TEMP_{rel}	23.53%	166.8	88767.1
*WDR	**12.27**	**87.2**	**20929.3**

4 Reinforcement Learning-Based Smart Request Assignment Technique

Efficient dispatch is the core of Didi's ridesharing platform. DiDi's technology team has originated a framework that merges RL with combinatorial optimization [5, 6].

During rush hours, if a driver receives a request in a request-dense region, and finishes the trip also in a request-dense region, then they can expect to quickly receive another request in the region where the last trip ended, for a greater-than-expected income. If the same driver receives a request in a request-dense region, but finishes the trip in a request-sparse region, then the driver will not be able to receive the next ride request quickly, thus incurring costs on both time and fuel while lowering expected income. That means when the platform makes request assignment decisions, whether or not a certain request is sent to a certain driver will impact the state of that driver in the next moment, hence impacting the driver's income for the whole day. This constitutes a semi-Markov process. DiDi's own spatio-temporal value model is based on a semi-Markov process, and enables the maximization of the driver's income. At any given point in time, there exists the problem of matching all drivers with all riders. DiDi utilizes its smart request assignment technique, which merges RL with combinatorial optimization, to maximize the income aggregate for all drivers while ensuring quality rider experiences.

4.1 Spatio-Temporal Value Model

The spatio-temporal value is the value of a driver showing up at a specific time/space, meaning a change in spatial/temporal states would result in a change in the driver's spatio-temporal value. A driver is seen as a smart entity, their state determined by the time and space they are in, while their actions define their completion of requests or downtime. To finish a request, a driver would need to go through many steps, such as driving to the pickup location, waiting for the rider, and getting the rider to their destination. The action of completing a request will cause an automatic change in the driver's spatio-temporal state, as well as bring them a reward in the form of payment for the request.

4.2 Matching Strategy

For each turn (which lasts two seconds) of request assignment, the platform will acquire the state (s) of each driver waiting to be assigned a request, and set all requests to be assigned as one of the executable actions by the drivers. The optimization goal of this problem is to maximize the income aggregate for all drivers while ensuring quality rider experiences. The driver-rider matching problem has been converted to a bipartite graph matching problem, with the KM (Kuhn-Munkres) algorithm used to solve the problem.

As the bipartite graph is being made, the edge weights of a certain driver and a certain request represent the expected income for this driver, who executes the Action (a) of completing the request in the State (s), and this is the action-state value function $Q(s, a)$ in RL. Bipartite Graph Driver-Rider Matching Approach is shown in Fig. 4.

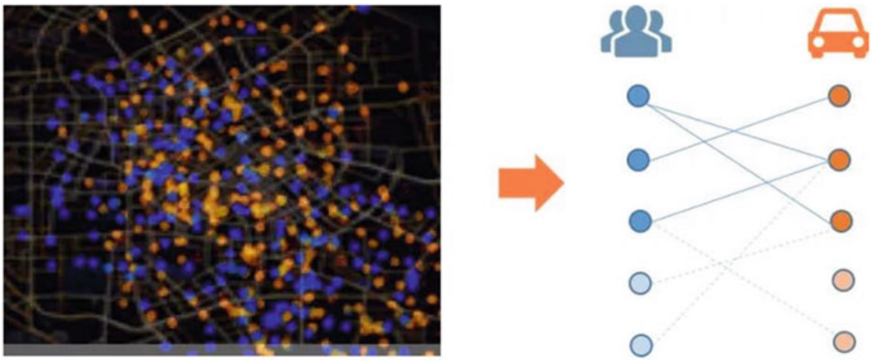


Fig. 4 Bipartite graph driver-rider matching approach

4.3 The Algorithmic Procedure

Step 1 Collect lists of drivers and requests to be assigned;

Step 2 Calculate the state-action values that correspond to each driver-rider pair, and use these values as weights to create a bipartite graph;

Step 3 Use the matched values in step 2 as weights and embed them into the KM algorithm, while fully considering relevant factors such as pickup distances and service points, to find the optimal matches before entering the final assignment stage.

4.4 Experiment Result

In order to thoroughly validate the effectiveness of the algorithm, it was evaluated across three different environments: a toy example, a simulator, and a real-world online environment. The toy example was set up to perform driver-rider matching within a square area with fixed sides and 20 temporal step sizes. Within each temporal step, the driver can only make vertical or horizontal movements. The simulator performed matching based on request data from a day in the past, and assessed the response rate and Gross Merchandise Volume (GMV) of the driver. Toy Example Results Comparison across Three Algorithms (Distance, Myopic and MDP) is shown in Fig. 5. Comparison of simulator results is shown in Table 4.

As indicated by the results above, this request assignment algorithm, based on RL and combinatorial optimization, can significantly boost a driver's income while maintaining a quality rider experience. Currently, this algorithm has been successfully deployed on the DiDi platform in more than twenty key cities, meeting the mobility needs of our vast numbers of users.

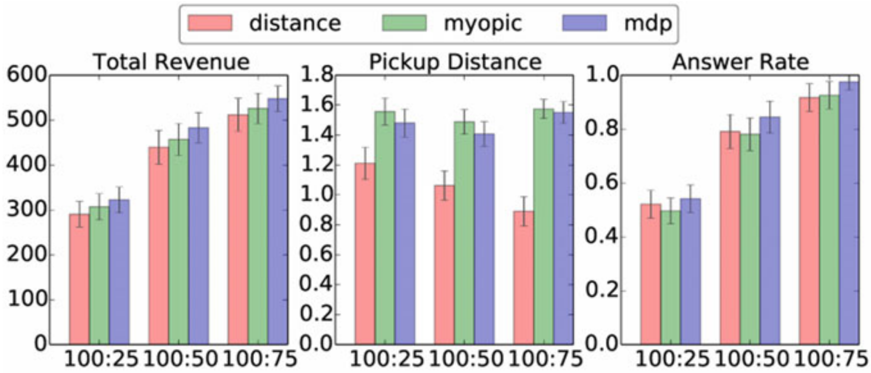


Fig. 5 Toy example results comparison across three algorithms (distance, myopic and MDP)

Table 4 Comparison of simulator results

City	MDP V_O		MDP $V_{converge}$		Days to converge
Metric	GMV (%)	Rate (pp)	GMV (%)	Rate (percentage point)	
City C	+0.6	+0.5	+0.8	+0.9	4
City D	+0.9	+1.0	+1.4	+1.5	8
City E	−0.1	+0.1	+0.5	+0.5	13
City F	+0.8	+0.7	+1.2	+1.1	7

5 Deep Learning-Based Spatio-Temporal Predictive Technique

Demand volume prediction is a typical spatio-temporal predictive problem [7]. Temporally, this problem has a clear gradation and periodicity, while spatially, the distributions of requests share clear regional similarities. Traditional deep learning approaches use CNN+RNN to model space and time. For example, a deep spatio-temporal network uses stacked CNNs to extract spatial features, while a deep multi-view spatio-temporal network (DMVST) enables multi-view learning by simultaneously introducing both RNN and CNN. Of these approaches, however, the RNN's modeling of time series is overly smooth; while terms of spatial modeling, the CNN cannot handle non-Euclidean geometry, nor can it introduce multimodal city data to conduct more precise and accurate modeling. These issues have led to relatively low predictive accuracy and stability [8].

5.1 Demand Volume Predictive Model Based on Multi-graph Spatio-Temporal CNN

DiDi's ridesharing platform uses a demand volume predictive model, based on a multi-graph spatio-temporal CNN. This model improves upon the current deep spatio-temporal model in terms of both time and space. In temporal modeling, a gating mechanism based on temporal context has been added into the RNN (CGRNN). This gating mechanism filters historical data, weighted by importance, making the time series inputted into the RNN no longer smooth, and the predictive result aggregated by the RNN more sensitive to sudden fluctuations. In terms of spatial modeling, the CNN has been replaced by a GCN (graph convolutional neural network), which, unlike a CNN, can model on non-Euclidean data structures. To capitalize on this advantage, three sets of relationships are introduced (geographic proximity, POI similarities, and route connectivity) to provide additional information for spatial modeling. The latter two sets of relationships can only be represented through non-Euclidean structures, and are unable to be handled by the original CNN. By using three GCNs to model these three sets of relationships respectively before fusing the models, we make the predictive algorithm more accurate and stable. Network structure model is illustrated in Fig. 6. The model consists of four parts: graph generation, temporal modeling based on CGRNN, spatial modeling based on MGCN, and model fusion.

Three sets of spatial relationships (geographic proximity, POI similarities, and route connectivity) are introduced to generate graphs, and regions in the predictive problem are turned into nodes on the graph. For each graph, time series modeling

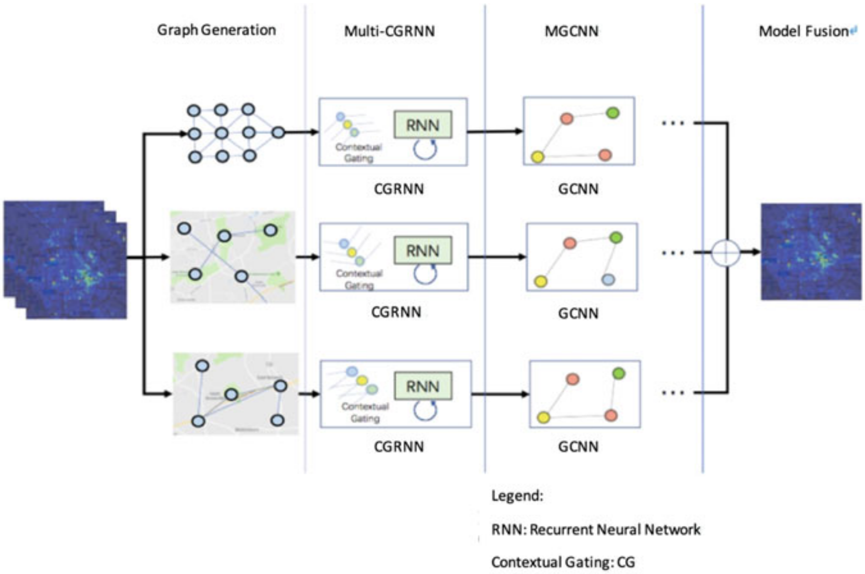


Fig. 6 Network structure model

is performed using the spatial relationships that correspond to the CGRNN-based temporal model; local spatial features are extracted from the MGCN-based spatial model through stacked GCNs. Finally, the spatio-temporal features learned by these three graphs are fused to generate the final predictive value.

5.2 Graph Generation Module

During the graph generation phase, we introduce multimodal data into the spatio-temporal prediction problem to assist with the modeling of spatio-temporal data. Specifically, three sets of relationships are introduced into the request volume prediction problem. Intuitively, we believe the distribution of requests among regions might have to do with these relationships:

(1) Geographic Proximity. The First Law of Geography states that “everything is related to everything else, but near things are more related than distant things”. A sparsity graph is made for geographic proximity among regions, according to this law.
(2) POI similarities. The similarities of POI among regions might reflect similarities in demand volumes. For instance, two residential regions might both have high demand volumes during the morning rush hour, resulting in a graph that shares POI similarities.
(3) Street Network Connectivity. Street networks are the foundation for the demand for DiDi’s services. Regions connected by street networks might share similar features in terms of their demand volumes. Therefore, integrating the connective relationships of a city’s highways and arterial roads into the spatial model will be conducive to the extraction of spatial features. Therefore, a street network connectivity graph is built based on information extracted from a city’s arterial roads, e.g. highways and overpasses.

5.3 Spatio-Temporal Modeling

The CGRNN is the model’s first step in nonlinear transformation of inputted spatio-temporal data, but the goal is to model the temporal dimension. A self-attention mechanism has been designed for the RNN, and a graph-based gating mechanism has also been introduced so that the extraction of temporal features will consider graph information and historical data information simultaneously. Additionally, multiple graph convolutional neural networks (GCNs) have been generated by the MGCN-based spatial model, thus extracting spatial features in non-Euclidean spaces from multiple inter-regional relationships.

We fused the above models, and the resulting model can help the platform predict the distribution of demand and supply for all regions across any given time period, thus enhancing platform efficiency while maximizing the utilization of traffic

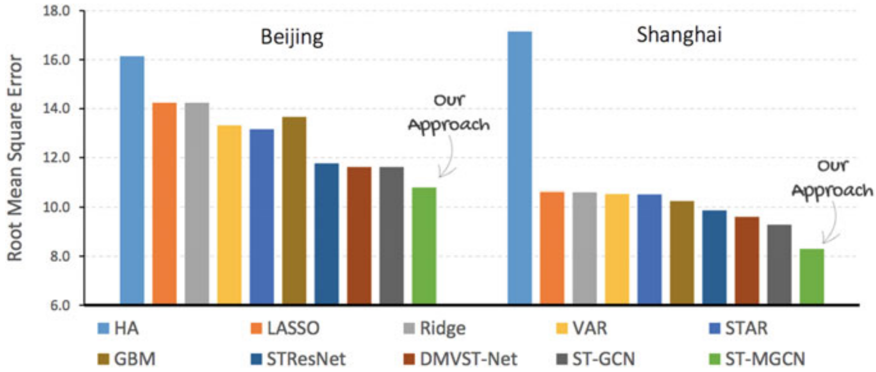


Fig. 7 Comparison with other models

resources to relieve congestion in city streets. The demand prediction accuracy for 15 min into the future has now reached 85%.

5.4 Experiment Result

An experiment on the model was performed based on DiDi Chuxing's historical request data from Beijing and Shanghai. DiDi's model is significantly superior when compared with other predictive models, including Historical Average, regression models (LASSO, Ridge), autoregressive models (VAR, STAR), and Gradient Boosting Machines (GBMs, ST-Res Net, DMVST-NET, DCRNN and ST-GCN). See Fig. 7.

6 Conclusion and Outlook

DiDi's key technologies in the field of smart request prediction and assignment have played a major role in improving the operational efficiency of its ridesharing platform. In terms of time estimation techniques, breakthroughs have been made to enable the key technology of real-time and accurate multi-strategy route planning, enabling the handling of large request volumes. The (reinforcement learning) RL and combinatorial optimization-based smart request assignment technique have been shown to significantly boost a driver's income, while sacrificing nothing in terms of rider experience. As a result, the ETA error is within five minutes for a one-hour trip. Thanks to the predictive accuracy technology, the accuracy of predicting demand and destinations of 15 min into the future has reached 85% and 90% respectively. DiDi's proprietary "Research and Application of Key Technologies for Smart

Request Prediction and Assignment on Ridesharing Platforms" has recently been authenticated by experts as a leading technology in the field.

With the help of key technologies in smart request prediction and assignment, DiDi's ridesharing platform provides a full range of ride services, including Taxi, Express and Premier, to over 550 million users worldwide. Together with other seven leading ridesharing companies Grab, Lyft, Ola, Uber, 99, Taxify and Careem, DiDi has built a global mobility services network, serving more than 80% of the global population. DiDi's ridesharing platform provides great convenience for Chinese people in their travels abroad, as well as foreign tourists visiting China, thus contributing to the booming of the mobile internet business and large-scale application of ridesharing products, as well as generating considerable economic and social benefits for the country. In April 2019, Chilean President Sebastián Piñera and DiDi's CEO Cheng Wei met to discuss the prospect of DiDi entering the Chilean market to offer translated and localized ridesharing services to the Chilean people. DiDi will continue participating in the Belt and Road Initiative proactively, and enhance the global mobility market by exporting its ridesharing platform solutions and technologies overseas.

References

1. Wang Z, Fu K, Ye J (2018) Learning to estimate the travel time. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining (KDD'18). ACM, New York, NY, USA, pp 858–866
2. Li Y, Fu K, Wang Z, Shahabi C, Ye J, Liu Y (2018) Multi-task representation learning for travel time estimation. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining (KDD'18). ACM, New York, NY, USA, pp 1695–1704
3. Wang Y, Zheng Y, Xue Y (2014). Travel time estimation of a path using sparse trajectories. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining (KDD'14)
4. Wang H, Kuo Y-H, Kifer D, Li Z (2019) A simple baseline for travel time estimation using large-scale trip data. In: Proceedings of the 24th ACM SIGSPATIAL international conference on advances in geographic information systems, vol 10, no 2, pp 1–22
5. Zhang L, Ai W, Yuan C, Zhang Y, Ye J (2018) Taxi or hitchhiking: predicting passenger's preferred service on ride sharing platforms. In: The 41st international ACM SIGIR conference on research & development in information retrieval (SIGIR'18). ACM, New York, NY, USA, pp 1041–1044
6. Wang Z, Qin Z, Tang X, Ye J, Zhu H (2018) Deep reinforcement learning with knowledge transfer for online rides order dispatching. In: 2018 IEEE international conference on data mining (ICDM), Singapore, pp 617–626
7. Xu Z, Li Z, Guan Q, Zhang D, Li Q, Nan J, Liu C, Bian W, Ye J (2018) Large-scale order dispatch in on-demand ride-hailing platforms: a learning and planning approach. In: Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining (KDD'18). ACM, New York, NY, USA, pp 905–913

8. Geng X, Li Y, Wang L, Zhang L, Yang Q, Ye J, Liu Y (2019) Spatiotemporal multi-graph convolution network for ride-hailing demand forecasting. In: Proceedings of the AAAI conference on artificial intelligence, vol 33, no 1, pp 3656–3663
9. Zhang L, Hu T, Min Y, Wu G, Zhang J, Feng P, Gong P, Ye J (2017) A taxi order dispatch model based on combinatorial optimization. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining (KDD'17). ACM, New York, NY, USA, pp 2151–2159

Bo Zhang, Co-founder, Chief Technology Officer, Didi Chuxing, CEO of DiDi's Autonomous Driving Companyco-founder and CTO of DiDi Chuxing Technology Co, is primarily responsible for research into human-machine interaction and artificial intelligence. He helped build up DiDi's product, technology, and big data systems from the ground floor, and oversaw the company's evolution from a simple taxi service business to a mobility and transportation services platform offering a wide array of services, including taxis, Express, Premier, Luxe, buses, substitute driving, Enterprise, bicycle sharing, electric bicycle sharing, car sharing and food delivery. Mr. Zhang led DiDi's technology team to a number of notable achievements, including great success at utilizing cutting-edge technologies in fields such as smart transportation, machine learning, data mining, and big data. He has also led the "DiDi Brain-Based Demonstration Project on City Smart Transportation Collaborative Management and Ridesharing", a major engineering project run by the National Development and Reform Commission (NDRC), as well as participated in the "Big Data Analytics and Application Technology with the National Engineering Laboratory" project, again run by the NDRC, and the "Machine Learning-Based Smart City Transportation Management Technology and Services with Beijing Engineering Laboratory" project, run by the Beijing Municipal Commission of Development and Reform. Sciences in 2008, with a research focusing on human-machine interaction and artificial intelligence. Since August 2019, Mr. Zhang begins to serve as the CEO of DiDi's Autonomous Driving Company, pushing forward the achievement of autonomous driving together with the whole team.

Construction of a Scientific Research Integrated Management Information Service Platform Integration in a Form of Cross-Platform and Multi-disciplinary Organization

Binjian Qiang, Rui Zhang, Yugang Chen, and Tongtong Zhang

Abstract In the era of big science, the scientific research is basically characterized by cross-platform, multi-discipline, and large-scale collaboration. Based on the abovementioned characteristics and by organically combining information with scientific research and integrating new generation information technologies (e.g., Internet of Things, Cloud Computing, Big Data and Mobile Internet), this study aimed to reconstruct the cloud center-based technical architecture and build an integrated scientific research management information service platform with scientists as center, focusing on scientific research activities and covering the integrated solutions which involved human resources, finance, condition guarantee, and daily office. The platform was expected to have the following integrated information service functions: the management collaboration of scientific research projects], the digitalization of scientific research activities, the compliance of scientific research funds, the sharing of equipment and instruments, the servitization of talent teams, the e-commercialization of supporting activities, the cyberization of scientific communication, and systematization of scientific and technological think tanks. Besides, the platform was expected to achieve the following objectives: to realize the horizontal close-loop management and vertical business interconnection of research units by human-centered application experience; to greatly improve the compliance of business and reduce the management risk of research units by refined full-cost control; to provide information support for improving the management system, optimizing the corporate governance structure and promoting the multi-disciplinary and cross-field big science cooperation and innovation.

Keywords Scientific research project · Scientific research integration · Management informatization · Digital collaboration platform

B. Qiang (✉)
Beijing Advanced Sciences and Innovation Center of CAS, Beijing, China
e-mail: qiangbinjian@basic.cas.cn

B. Qiang · R. Zhang · Y. Chen · T. Zhang
CAS Xunlian Intelligence Supply Chain Network Technology (Beijing) Co., Ltd, Beijing, China

China's e-Science Blue Book 2020,
https://doi.org/10.1007/978-981-15-8342-1_29

1 Background of Platform Construction

1.1 Construction Planning of National Laboratory for Physical Sciences

At the Fifth Plenary Session of the Eighteenth Central Committee of the Communist Party of China, it was clearly proposed that a number of national laboratories would be established in major innovation fields. In the beginning of 2017, the Party Central Committee of China made another major strategic decision of building the integrated multi-disciplinary national laboratories firstly in the fields of network information, energy resources, marine sciences, physical sciences, aerospace information, population and health and the national defense-related fields, with the following goals: to fully implement the innovation-driven development strategy; to deepen the reform of the scientific and technological system; to enhance the national innovation capability; and to build a powerful country of science and technology. The construction of national laboratories has great practical significance and a profound historical impact in realizing the "China Dream" of the great rejuvenation of the Chinese nation.

After several Chinese generations of efforts in the research of physical sciences, many gratifying achievements have been made in the construction of big science research infrastructure and research facilities, the output of scientific research results, and the construction of talent teams. However, due to the limitations from the traditional discipline-based research institution set-up, the decentralized layout of big science infrastructure, and a variety of operation management and organization patterns, there is no high-benefit and high-efficiency research support available for the frontier research of physical sciences with the increasing discipline crossing and integration and the continuous development and integration of experimental methods. Therefore, to provide solid support for the improvement of innovation capability of physical science research in China and for the rapid development of national physical science research fields and other key innovation fields, breaking down the discipline boundaries and the barriers to the organization mode of physical science research, innovating the organization mode of physical science research in China, and establishing a national laboratory for physical sciences mainly characterized by the cluster of first-class big science facilities and the "whole-chain" research team of physical sciences are a major and urgent task faced by the scientific and technological development in China as well as a glorious historical mission of scientific researchers and managers in the current generation endowed by the Communist Party of China (hereafter referred to the "Party") and State. To this end, Chinese Academy of Sciences (CAS) and the People's Government of Beijing Municipality have decided to establish Huairou Physical Science Laboratory in Huairou Science City and thus finish the full preparation for the construction of National Laboratory for Physical Sciences, with a purpose of striving to make Huairou Physical Science Laboratory approved as a national laboratory by the State at the end of 2020.

Aiming at becoming the source and support of scientific and technological innovation in the national key fields, the National Laboratory for Physical Sciences is a talent highland and think tank of physical science research in China and an iconic scientific research organization of Beijing Science and Technology Innovation Center.

1.2 Current Situation of Information Construction in Chinese Scientific Research Institutes

In terms of the current situation about the information construction, most of Chinese scientific research institutes have clearly recognized the importance of information construction to their management and development and listed the information construction as key work. The continuous perfection and optimization are always done in the establishment of information construction organization structure, the improvement of information construction working system, the research and development of information system, and the management and maintenance of information system.

Through the comparative analysis on the information construction of scientific research institutes and advanced enterprises, the qualitative and quantitative investigation was conducted from many perspectives, such as information infrastructure construction, network safety, top-down design design and information sharing collaboration. The results have shown that the information construction of scientific research institutes is still at a relatively backward level, with a gap to the advanced information enterprises; the input of information construction is not proportional to its value; and the information construction has no great contribution to strengthening the comprehensive capability and to the innovative development of these institutes. Hence, the scientific research institutes shall further intensify the research and practice of information construction, boost the promotion of informatization level and make full use of information construction during development in the new period [1].

1.3 Demands for the Operation Management Information Construction in National Laboratories

The distributed technical architecture, isolated-island business process and fragmented IT governance architecture of traditional management information systems cannot meet the operation management informatization demands of new national laboratories as a source of scientific and technological innovation in the national key fields. Especially, a new business architecture is needed to support the organizational reform and the reconstruction of cross-regional scientific research units, and the multi-disciplinary project collaboration and organization crossing corporate units.

With the major purpose of supporting the scientific, collaborative, normalized and efficient management of scientific research institutes using information and technology, and focusing on the relationship clearing, collaboration strengthening and efficiency improvement, the operation management information construction of national laboratories aims to transform the functions, break down the old mindset and work inertia, promote the discipline crossing and integration, and provide powerful support for the organization guidance system and central management. The main demands include:

Management and service of scientific research projects: It is a complete management process penetrating the whole life circle of scientific research projects, including planning, approval, budget, implementation and acceptance. There are the following tasks: to form a global scientific research project management center from an brand-new perspective of top-down management; to relate the beforehand, halfway and afterward full-process management services concerning the scientific research projects, and to link up with the intellectual property management service of technology service networks.

Management and service of human resources: It, based on the demands for the construction of national laboratories, aims to realize the career-long student source management of human resources from the business perspective, optimize the performance assessment management system and improve the information resource service capability.

Comprehensive financial management and service: It is a financial budge application with general accounting and payables as the core and integrated with such business applications as human resources, general finance, scientific research projects and conditions, and aims to establish the uniform budget management and financial accounting system.

Condition guarantee management and service: It aims to promptly, comprehensively and accurately reflect the assets demand and status in scientific research activities and promptly extract the data related to a variety of assets management for scientific management allocation.

Collaborative office management and service: It aims to realize the online and mobile convenient communication and collaboration of scientific researchers and various special applications as well as the on-line handling of routine work, and implement the process control of tasks.

Based on the demands for the operation and management information construction of national laboratories, an integrated management information service platform of scientific research was constructed in a form of cross-platform and multi-disciplinary organization. This platform aimed to explore the new management information construction mode, bring about new services by virtue of new-generation information technology, and improve the efficiency of scientific research management; to create new performance and promote the smart scientific research management relying on intelligence; to adapt to the operation mechanism of new organization mode; and eventually to realize the scientific research intelligence, the integration of science and education, the convergence and integration and intelligent analysis of digital resources, and the deep integration and orderly opening of basic resources of scientific research information, and promote the upgrade and update of management informatization in national laboratories.

2 Construction Principles and Objectives

The construction of an integrated management information service platform of scientific research completely abided by the strategies and principles (e.g., human orientation, core application guidance, ubiquitization and big digitization), and aimed to promote the upgrade of scientific research management information service in a form of cross-platform and multi-disciplinary organization.

2.1 Construction Principles

Human orientation—service strengthening: The platform focused on strengthening the information service of new scientific research units, the individualized information service of scientific research institutes, the collaboration service of big science, and the management decision-making service, and establishing the information governance system matching the new service informalization.

Core application guidance—key breakthroughs: The platform was under the guidance of core application, that was to say, it was to assure the compliance of business development based on the applications and services with a core of the intelligence management of scientific research projects, condition guarantee, financial control and collaborative office under the two-level corporate governance system.

Ubiquitization—extensive connection and collaboration: The platform adapted to the mobile and Internet era, fully supported the location-free and time-free service access of scientific researchers for collaborative office, and provided the real-time connection and data sharing of scientific instruments and big devices through Internet of Things to realize the service ubiquitization.

Big digitization—data integration: The platform was designed for integrated applications, and aimed to establish the data governance system of national laboratories, realize the integration of big data (e.g., data acquisition, sharing, analysis and service) through the planning of data architecture, build up the data resource sharing mechanism and develop the intelligent services.

2.2 Construction Objectives

The integrated management information service platform of scientific research aimed to fuse various technologies (e.g., cloud computing and artificial intelligence), combine the overall business and construction planning of national laboratories, integrate the traditional information management systems, realize the interconnectivity between various businesses and Internet services during the operation management process of national laboratories, comprehensively promote the integrated management with the scientific research projects as the main line, support the overall process management and service of scientific research projects, and meet the multi-disciplinary, cross-regional and cross-organizational collaborative innovations of national laboratories. Besides, it also aimed to fully guarantee the full-cost closed loop dynamic control and ensure the full-life circle digital collaboration of scientific research projects.

3 Overall Architecture Design of Platform

The architectural design of the integrated management information service platform of scientific research was made using the methods of enterprise architecture to unify the platform construction, realize the modular function componentization, review the informatization planning from a strategic perspective, avoid the disjunction between business and IT, achieve the interconnectivity, integrate the applications and organization of various business lines, and guarantee the centralization of information resource on the support of strategic vision on the basis of overall situation [2].

3.1 Application Architecture

The business scope of the integrated management information service platform of scientific research includes human and financial application with project management as the core. All foreground businesses of the platform were developed in the servitization form and fragmented (fully decoupled) into mobile applications. The background businesses adopted the SOA architecture, and they are accessed by PC through browser on the Internet.

The application services have two dimensions in accordance with the business value chain and the user operation and use. The business dimension includes: scientific research project management, talent resource management, financial management, scientific research condition management (asset management) and office collaboration.

From the perspective of user use, there are foreground service and background support business. The foreground service is accessed by (user role) scientists, scientific researchers and managers in various business departments. The background support business is an application module for personnel in the functional systems (e.g., finance, assets and logistics) to provide the foreground service for first-line scientific research.

The integrated management information service platform of scientific research is human-oriented and fully integrates the social collaboration services to establish the deeply integrated business collaboration service. The overall architecture is shown in Fig. 1, the architecture of functional modules is shown in Fig. 2.

Combining the social collaborative services and by virtue of such applications as role navigation, schedule navigation, backlog and message notification, the platform

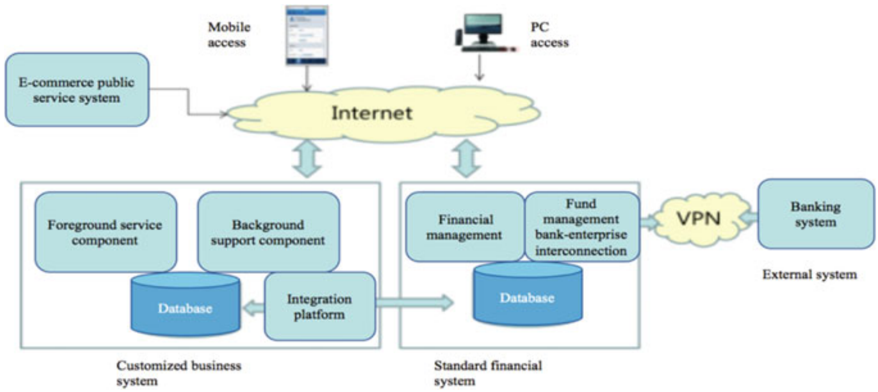


Fig. 1 Overall architecture of the integrated management information service platform of scientific research

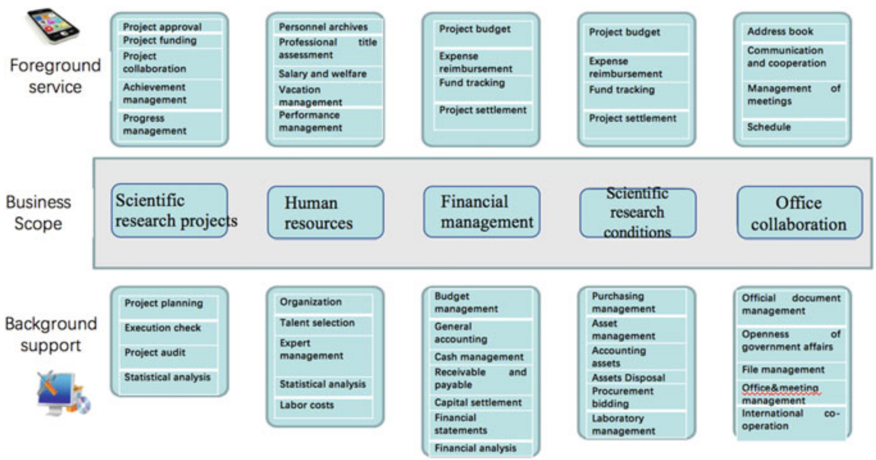


Fig. 2 Architecture of functional modules in the integrated management information service platform of scientific research

aimed to break down the traditional inflexible and mechanical business modes and realize the whole-scene online collaboration of various process links, such as project planning, declaration, approval, implementation and acceptance in the management of scientific research projects [3]. See Fig. 3.

Fig. 3 An human-oriented integrated management information service platform of scientific research

3.2 *Technical Architecture*

Different from the traditional highly coupled technical architecture, the technical architecture of the integrated management information service platform of scientific research is based on the distributed technology of cloud computing to realize the large-scale concurrent virtual intensive management mode. The technical architecture design of the platform includes the following main characteristics.

3.2.1 Platform Access in Various Ways

The platform supports multi-terminal access (smartphone, tablet computer and PC): 3G/4G network access of mobile terminal (mobile phone or PAD) through telecom operators; local area network (LAN) access of computer terminal or portable computer/mobile terminal of all institutes via WiFi and CSTNET; background system access by the operation and maintenance personnel in each institute or network center via CSTNET or Internet background for system operation and maintenance.

3.2.2 Cloud Computing Architecture

The calculation and storage infrastructure, basic platform, data and application services of the platform are uniformly deployed by the cloud computing center to realize the virtual application service of corporate units and independent accounting units through the use of virtualization technology.

3.2.3 Open Technical Routes

The open technical routes are used for the technical architecture and software development of the platform. The overall platform is developed based on open source and domestic independent and controllable infrastructure and basic software platform, independent of specific suppliers, and in accordance with open standards and open API architecture, the platform can be flexibly expanded.

The integrated management information service platform of scientific research is subject to the cloud computing deployment architecture, and the application and data services of the cloud computing center are accessed directly by users in each institute via Internet [4]. See Fig. 4.

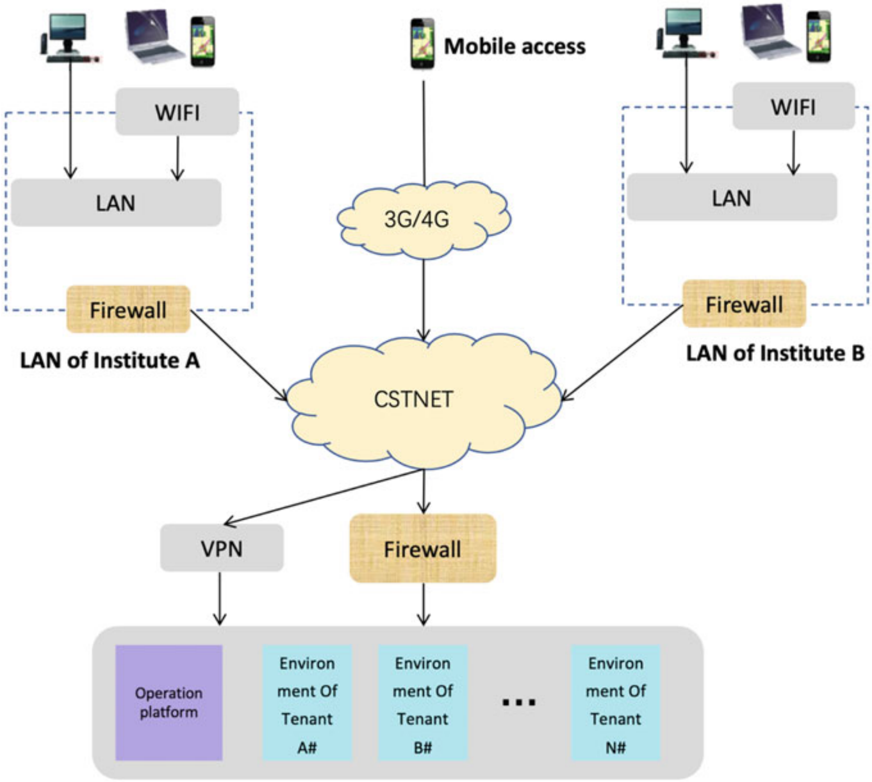


Fig. 4 Network architecture of the integrated management information service platform of scientific research

3.3 Data Architecture

The databases of the platform include business database with SQL server and social document database with NoSQL server, for which different architecture strategies are used.

The data management of the platform is the core of integration architecture design as well as a factor mostly involved in system performance and expansibility. According to the data management function model proposed by DAMA (Data Management Association), the data management focuses on data governance and relates 9 aspects: data architecture, data development, data operation, data security, reference data and master data, data warehouse, document and content management, metadata management, quantity and quality. See Fig. 5.

For the platform, the classification management is implemented in accordance with the data characteristics and by using different distributed architecture strategies. Based on data type, there are three types of data: basic data, business data and system data.

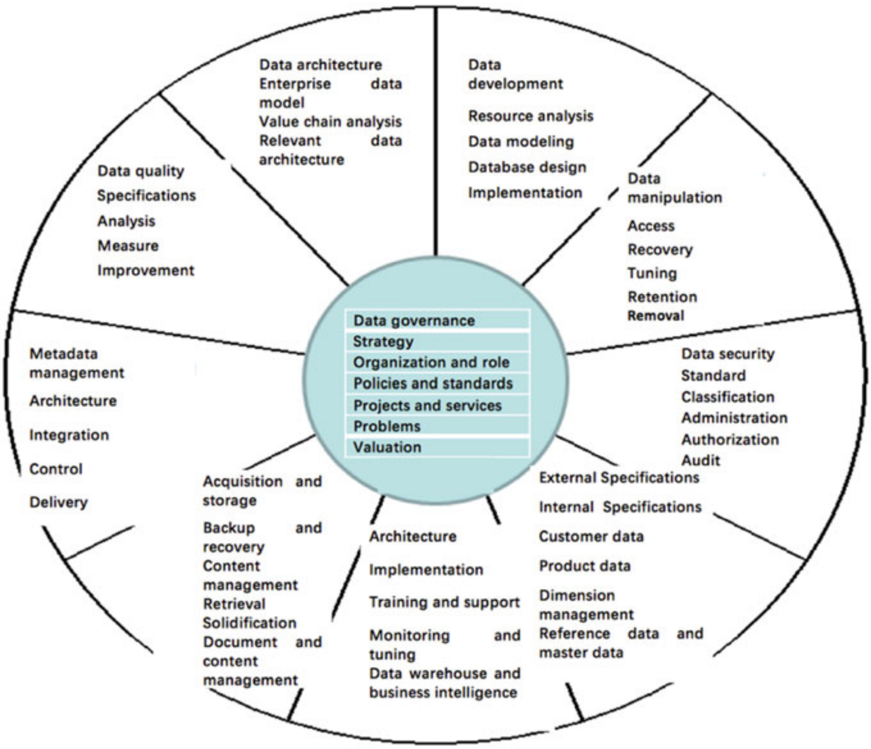


Fig. 5 Functional model for the data management of the integrated management information service platform of scientific research

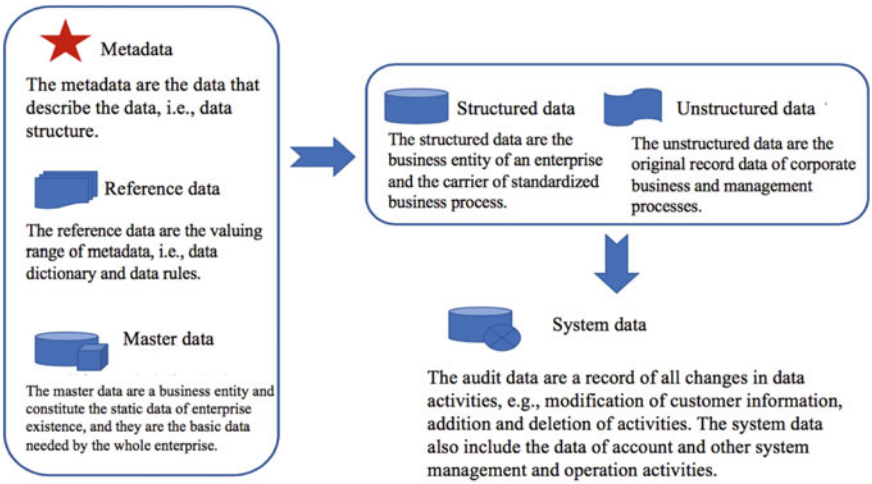


Fig. 6 Classification model for the data management of the integrated management information service platform of scientific research

The basic data include metadata, reference data and master data. These data are subject to the centralized sharing management, namely, they are shared by all systems in the national laboratories. The basic data of a global management platform can be called through API in principle, and other copies cannot be saved, which is a radical governance measure to guarantee the dada consistency and improve the data quality. The master data are the most important ones.

The business data are divided into structured data and unstructured data. In the system data, the audit data and account data are the most important ones. The classification model for the data management is shown in Fig. 6.

The platform takes cloud infrastructure as the core and builds a comprehensive solution for the future. In terms of storage and management, the basic data form a sharing data center, while the business data form a distributed data center. The data management architecture is shown in Fig. 7.

The key data under the management of sharing data center are as follows:

Account data: The account database is subject to the uniform whole-institute management mode and can meet the requirements of rapid processing under single-case performance. In order to guarantee the reliability, the dual fault-tolerant design was made in accordance with the client/server architecture. The centralized management of account data is the premise to realize the uniform identity management, uniform authentication and one-time log-in.

Master data: To integrate and maintain the unique, compete and accurate master data (organization, employees, scientific research projects and assets, as well as their corresponding relation data) within the whole platform scope, the detailed and reliable (flexible) master data management strategy needs to be intensively and

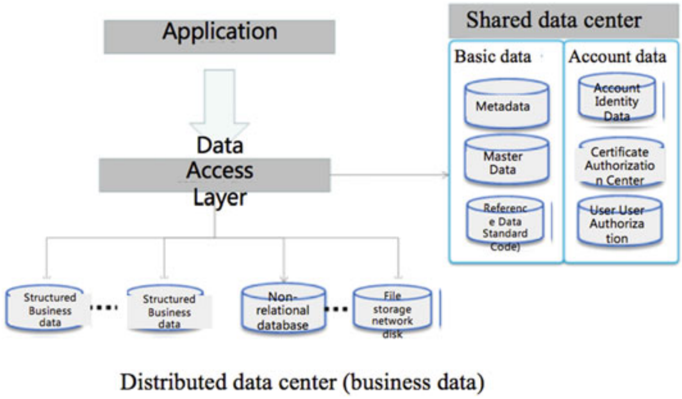


Fig. 7 Data management architecture of the integrated management information service platform of scientific research

comprehensively maintained; the master data information are shared to all application systems as necessary. The master data management model is shown in Fig. 8.

The management scheme and strategy of master data are described as follows:

(1) Mater data integration: The master data from different systems are merged to form the master data in single version for business processing and decision analysis. In principle, the application generated primarily by master data serves as the unique maintenance interface. Sharing of master data: The master data of various systems are intensively managed and their contents are being continuously enriched.

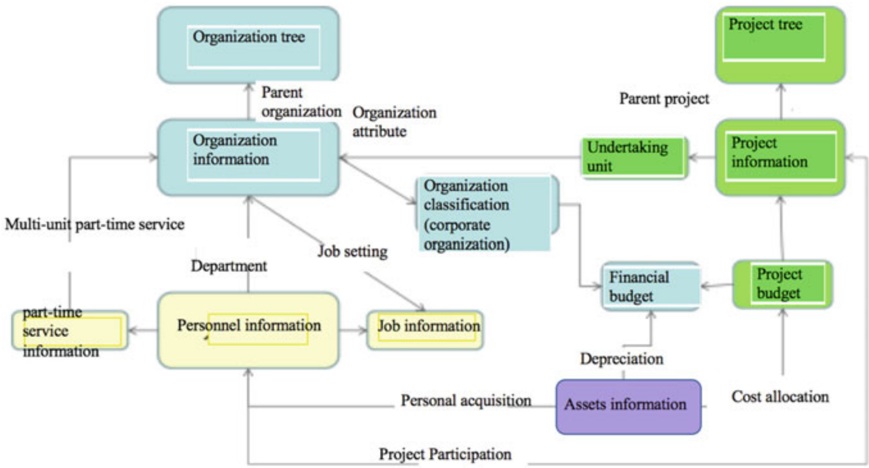


Fig. 8 Master data management of the integrated management information service platform of scientific research

(2) Intensive master data management: The master data are intensively maintained to keep a single version and shared to other application systems via API. In principle, the mode of multiple copies shall not be adopted to reduce the influence on the data consistency.

Reference data: The reference data refer to the normalized management of code data. Generally, the standard code data are managed by code-name correspondingly; the data with national standard codes and industrial standard codes are managed according to the corresponding standards; for the management of other data, the institute-level standards shall be formulated. The standard code data shall be uniform in the whole institutes, and the uniform coding system shall be used in other systems.

Business data: The business data (including finance, assets, salary and performance) are saved into SQL database. The business database is intensively managed physically and provided with authority according to the organizational architecture.

The unstructured database (including documents, communication and social collaboration) is managed using noSQL database management system, centrally saved, and provided with user and organization authority [5].

3.4 Security Architecture

As the integrated management information service platform of scientific research is a core business system serving the scientists and scientific research managers at all levels, the information security is the basis to guarantee the system functions. The complete information security architecture is the basic method to implement the information security management [6]. The information security architecture is shown in Fig. 9.

The platform has a private cloud architecture, and its confidentiality, integrity and availability are the key attributes of cloud service. In order to guarantee the user data security and business continuity, a cloud service security guarantee system of "technology + management, prevention domination, and defense in depth", by using the advanced Internet security technology and referring to ISO27001 international information security standards, the security rating and protection standards for national information systems, and the security guidelines for the key fields of CAS cloud computing, was constructed for users from such perspectives as compliance, user privacy and data, business application, infrastructure, disaster recovery, business continuity, organization and personnel, management specifications and workflows.

The platform is based on the safe technical architecture, and a complete information security governance system and a sound internal control and operation audit mechanism were established for it to ensure the user privacy and data security.

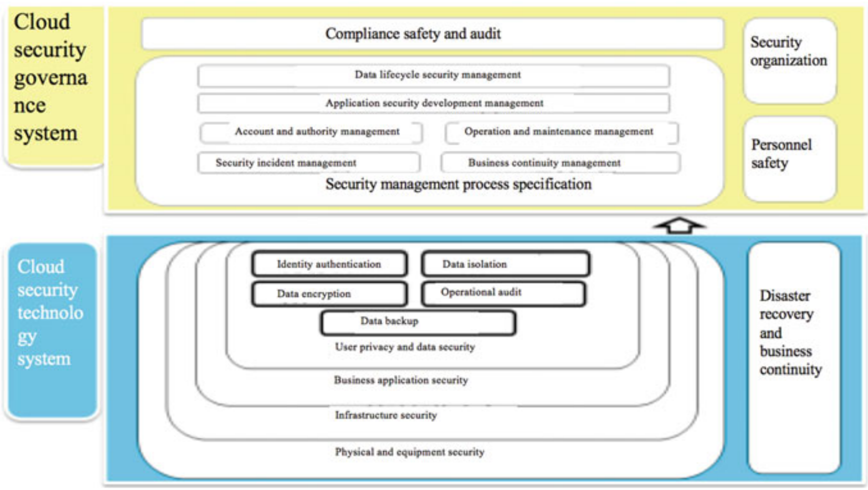


Fig. 9 Information security architecture of the integrated management information service platform of scientific research

4 Application Demonstration and Key Works in the Future

4.1 Pre-implementation Approval System of Beijing Advanced Sciences and Innovation Centre

Beijing Advanced Sciences and Innovation Centre, as the supporting service institution for CAS and the People's Government of Beijing Municipality to jointly establish Huairou National Comprehensive Science Center and Huairou Science City, is mainly to give assistance to the planning and construction of major national science and technology infrastructure and crossover research platforms, the planning and implementation of six science centers (including physical sciences and space science) in Huairou Science City, and the comprehensive services for the work extension of Beijing research institutes in Huairou Science City.

In order to support the comprehensive management of Beijing Advanced Sciences and Innovation Centre, the platform was wholly launched to support the daily work and the whole-process management (involving 6 application approval forms, including equipment, subject, contract, business trip, seal use and expense) of the Centre. It addresses the efficient cross-organization and trans-department cooperation of the National Laboratory for Physical Sciences and the demands of scientific research groups for flexible and convenient approval in the approval link, connects itself to ARP system for facilitating the online and offline application approval, and enables the budget control. The business workflow of system is shown in Fig. 10.

The platform can sufficiently satisfy the efficient cross-organization and trans-department cooperation of the National Laboratory for Physical Sciences and meet

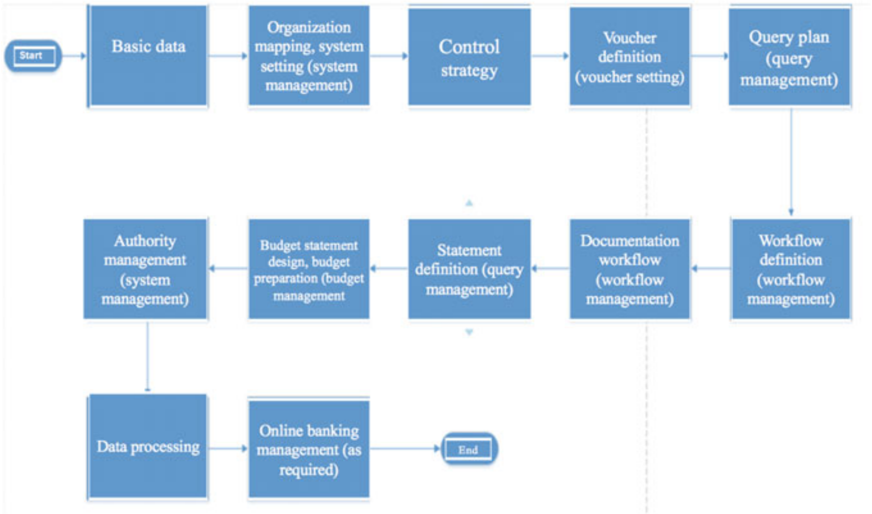


Fig. 10 Pre-implementation approval system of Beijing Advanced Sciences and Innovation Centre—business workflow

the demands of scientific research groups for flexible and convenient approval in the approval link. In addition, it enables the high configuration of application approval forms of human resource management and financial budget management, and supports the customized configuration of voucher elements, fields and approval workflows, thus covering all kinds of business scenarios. Furthermore, it has the following functions: facilitating the traceable management of trans-department, cross-functional and trans-regional work cooperation, realizing the paper-free application approval in the whole link that substitutes the previous offline approval procedures, and supporting the multi-terminal access (e.g., PC and mobile phone) and the real-time application approval, which improves the approval efficiency; achieving the budget control of application approvals, which controls the budget from the source, thus avoiding the process rejections and greatly reducing the repetitive labor; and enabling the highly controlled data access permission, the management of modules available for users, and the restriction of data accessible to users, which ensures the system and data security. The task management page is shown in Fig. 11, the approval filling page is shown in Fig. 12.

4.2 Online Platform of University of Chinese Academy of Sciences (UCAS)

By virtue of CAS advantages in science and technology, education and talents, UCAS School of Innovation and Entrepreneurship is designed to integrate the high-quality

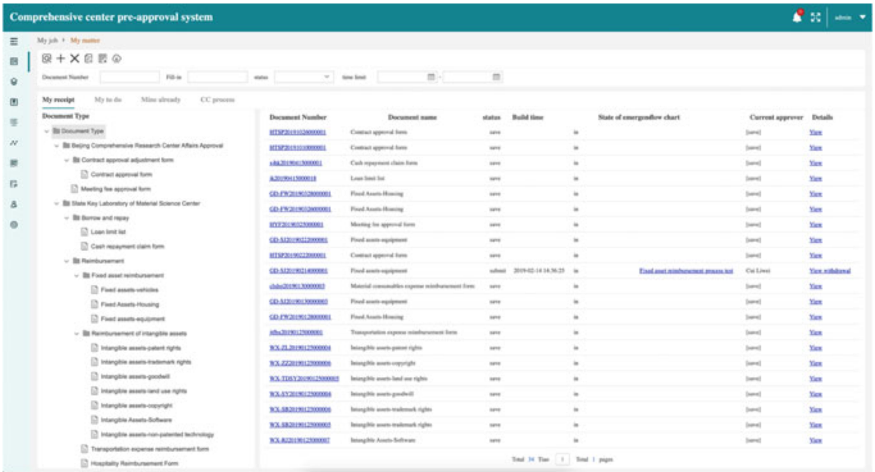


Fig. 11 Pre-implementation approval system of Beijing Advanced Sciences and Innovation Centre—task management

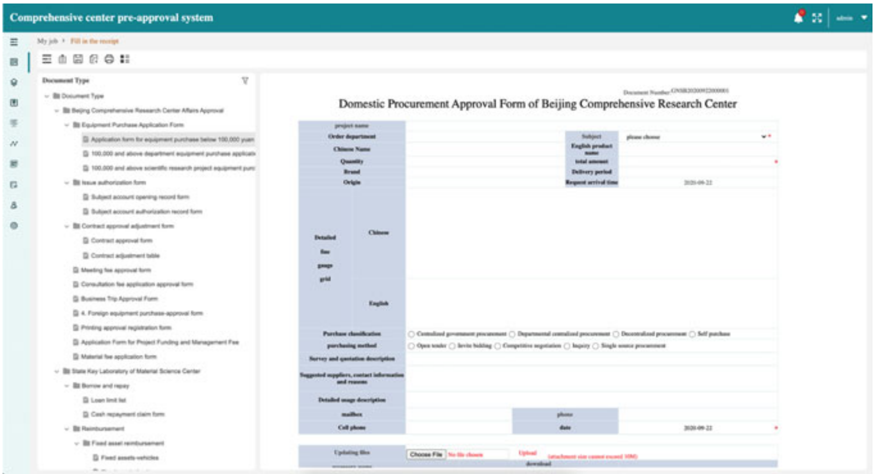


Fig. 12 Pre-implementation approval system of Beijing Advanced Sciences and Innovation Centre—approval filling and application

social resources, build a multi-disciplinary, collaborative and innovative education incubation platform, cultivate the excellent talents with innovative spirit and an entrepreneurship capability, facilitate the transfer and incubation of scientific and technological achievements, and form the open ecological system of innovation and entrepreneurship to serve the innovative national strategies.

Through using the system, UCAS School of Innovation and Entrepreneurship has built an online comprehensive service platform integrating the talent cultivation,

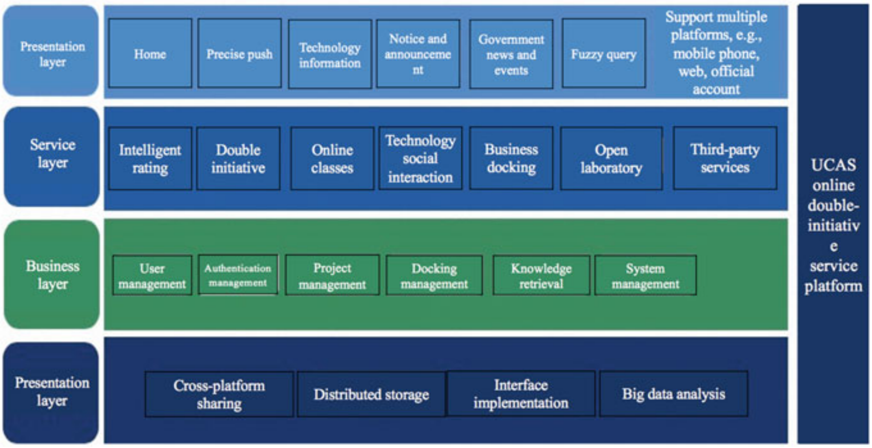


Fig. 13 UCAS online double-initiative service platform—overall architecture

science and technology services, and technological docking, which is provided for the teachers and students in the whole school. The overall architecture is shown in Fig. 13.

The platform, by constructing the one-stop comprehensive collaborative innovation services of industry, academy and scientific research, links the industry, academy, scientific research, government, finance, service and other resources to form a relation tie. With scientific research activities and achievement transformation as the core, the platform serves the construction of double-initiative education and activities and efficiently supports the transfer and transformation services of innovative achievements which integrate the demand, funds and supporting services, so as to connect all links during the transformation process of scientific research achievements and form a complete service chain for the transformation of scientific research achievements from the industry/scientific research matching to the incubation and transformation, and then to market promotion, and thus achieve the seamless transfer and transformation of innovative achievements. The platform home page is shown in Fig. 14, the scientific and technological think tank page is shown in Fig. 15.

4.3 Management Service Platform of CAS Key Laboratories

The key laboratory system is an important part of CAS scientific and technological innovation system and also the core strength of CAS fundamental research, fundamental research of applications, and frontier exploration of leading advanced technologies. The management service platform of CAS key laboratories was developed to achieve the management informatization of these laboratories, thus enabling the CAS leaders and the directors of CAS departments, subordinate units and laboratories

Fig. 14 UCAS online double-initiative service platform—home

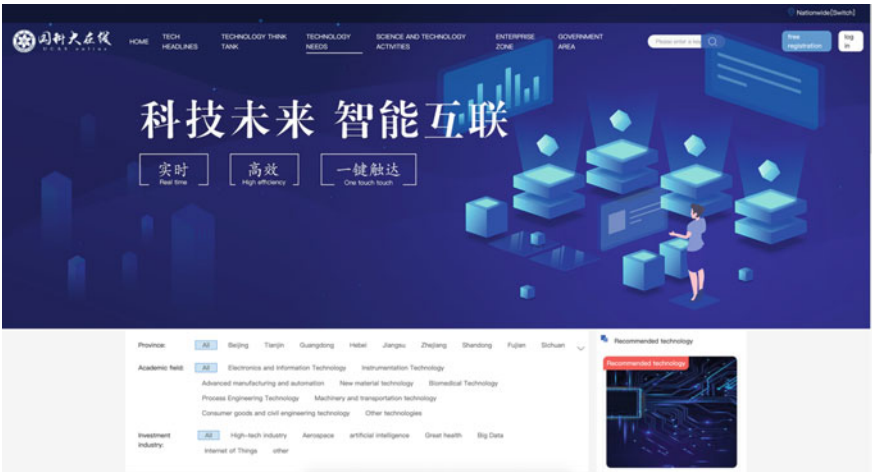


Fig. 15 UCAS online double-initiative service platform—scientific and technological think tank

to better understand and grasp the research level, operation organization and management of laboratories, and further supporting the implementation of innovation-driven development strategies in CAS. The statistical yearbook page is shown in Fig. 16.

The management service platform, starting from the actual demand of laboratory management work, provides an online information filling, reporting and approval procedure designed for the laboratory personnel and daily management workflows, and offers a big data system of scientific research information to achieve the management visualization and data analyzability of scientific research projects in the laboratories.

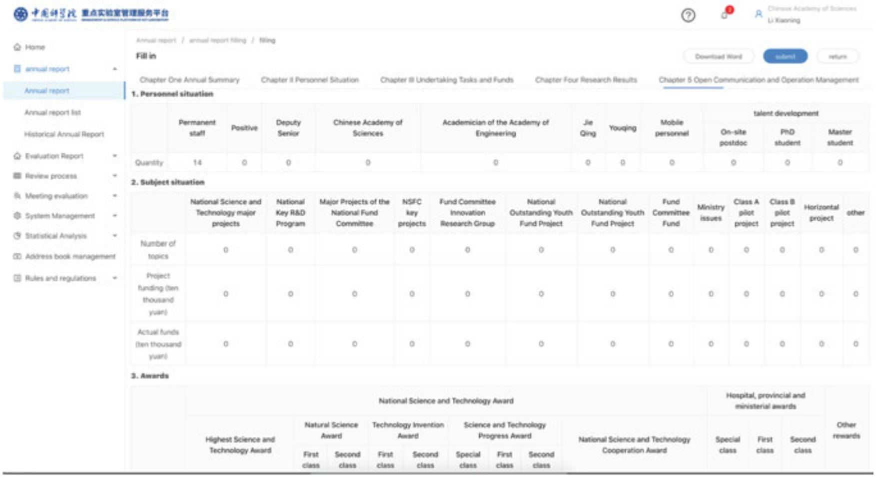


Fig. 16 Management service platform of key laboratories—statistical yearbook

4.4 Key Works in the Future

Under the strategic development background of global informatization, the scientific research institutes shall clearly comprehend the importance of information construction, carry out the overall information construction by combining the actual situations, achieve the promotion from the idea and methods to the contents and practice, and realize the improvement of informatization level.

The integrated management information service platform of scientific research efficiently supports the deepening of key applications through the business mode focusing on the management of scientific research projects and by guiding and driving the development of various businesses with the scientific research projects. Combining the administrative organization system and project management system of scientific research institutes, it provides a flexible architecture for human resources, finance and materials of the scientific research innovation system, which effectively facilitates the strategic implementation of scientific research system deepening and reform. With the establishment of human-oriented application experience, it offers mobile support to the scientific research and management personnel through the information and application services, thus greatly increasing the business operation efficiency. Furthermore, the platform with a cloud center architecture supports the transverse closed-loop management for the budget, accounting and settlement of scientific research, human resources, finance, materials, instruments and equipment, and consumables in the research units, the integrated information environment, and the interconnection of vertical businesses from the institutes to their departments/branches. And it greatly improves the business compliance and reduces the management risk of research units through the refined full-cost control, real-time electronic process management, mobile approval business control and intelligent policy guidance. The independently controllable and open platform, which is based on the cloud center architecture, efficiently lowers the system construction and operation and maintenance costs, increases the information security level, and reduces the information security risk, thus serving as an extensible platform for the informatization of scientific research management.

References

1. Duan F (2012) Brief Analysis on information and scientific research management. J Shenyang Offi (01)
2. Wu S (2015) Practice and innovation of scientific research management informatization in the new period. Human Res Manage (10)
3. Qi J (2015) Discussion on information security during the management of scientific research projects in universities and colleges. Course Educ Res (25)
4. Wen X, Chen X, Wang Q, Li T, Heslet, Wang T (2014) Scientific project management information system for research institutes based on B/S mode. Agri Netw Inf (5):57–59
5. Qian O, Chao M (2017) Exploration of information construction tendency of archive management in Scientific Research Institutes under the background of big data. Arch Electromech Warships 5:80–82
6. Chen Y (2012) Brief analysis on development and innovation of scientific research management. Legal Econ (the first ten days) (03)

Binjian Qiang, previously president assistant of University of Chinese Academy of Sciences (UCAS), secretary-general of Postgraduate School and director of Assets Management Office of UCAS, and presently deputy director of Beijing Advanced Sciences and Innovation Centre of CAS, deputy director of "CAS Huairou Science City" Special Project Office, and director of Steering Committee Office for "CAS Huairou Science City" Achievement Transformation Work Coordination.