Lecture Notes in Electrical Engineering 555

Electronic Navigation Research Institute
Editor

Air Traffic Management and Systems III

Selected Papers of the 5th ENRI International Workshop on ATM/CNS (EIWAC2017)

Springer

Lecture Notes in Electrical Engineering

Volume 555

The book series *Lecture Notes in Electrical Engineering* (LNEE) publishes the latest developments in Electrical Engineering - quickly, informally and in high quality. While original research reported in proceedings and monographs has traditionally formed the core of LNEE, we also encourage authors to submit books devoted to supporting student education and professional training in the various fields and applications areas of electrical engineering. The series cover classical and emerging topics concerning:

- Communication Engineering, Information Theory and Networks
- Electronics Engineering and Microelectronics
- Signal, Image and Speech Processing
- Wireless and Mobile Communication
- Circuits and Systems
- Energy Systems, Power Electronics and Electrical Machines
- Electro-optical Engineering
- Instrumentation Engineering
- Avionics Engineering
- Control Systems
- Internet-of-Things and Cybersecurity
- Biomedical Devices, MEMS and NEMS

For general information about this book series, comments or suggestions, please contact leontina.dicecco@springer.com.

To submit a proposal or request further information, please contact the Publishing Editor in your country:

China
Jasmine Dou, Associate Editor (jasmine.dou@springer.com)

India
Swati Meherishi, Executive Editor (swati.meherishi@springer.com)
Aninda Bose, Senior Editor (aninda.bose@springer.com)

Japan
Takeyuki Yonezawa, Editorial Director (takeyuki.yonezawa@springer.com)

South Korea
Smith (Ahram) Chae, Editor (smith.chae@springer.com)

Southeast Asia
Ramesh Nath Premnath, Editor (ramesh.premnath@springer.com)

USA, Canada:
Michael Luby, Senior Editor (michael.luby@springer.com)

All other Countries:
Leontina Di Cecco, Senior Editor (leontina.dicecco@springer.com)
Christoph Baumann, Executive Editor (christoph.baumann@springer.com)

**** Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, SCOPUS, MetaPress, Web of Science and Springerlink ****

More information about this series at http://www.springer.com/series/7818

Electronic Navigation Research Institute
Editor

Air Traffic Management and Systems III

Selected Papers of the 5th ENRI International Workshop on ATM/CNS (EIWAC2017)

Springer

Editor
Electronic Navigation Research Institute
Chofu, Tokyo, Japan

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-13-7085-4 ISBN 978-981-13-7086-1 (eBook)
https://doi.org/10.1007/978-981-13-7086-1

Library of Congress Control Number: 2019933856

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

Committee Members' List

EIWAC2017 Technical Program Committee

Editor in Chief

Tadashi Koga, Electronic Navigation Research Institute (ENRI), Japan

Editors

Jean-Marc Loscos, Direction des Services de la navigation aérienne (DSNA/DTI), France
Keisuke Matsunaga, ENRI, Japan
Low Kin Huat, Nanyang Technological University (NTU), Singapore
Masatoshi Harigae, Japan Aerospace Exploration Agency (JAXA), Japan
Mohamed Faisal Bin Mohamed Salleh, NTU/Air Traffic Management Research Institute (ATMRI), Singapore
Terumitsu Hirata, Ibaraki University, Japan
William C. Johnson, National Aeronautics and Space Administration (NASA), USA
Naruto Yonemoto, ENRI, Japan
Navinda Kithmal Wickramasinghe, ENRI, Japan
Noboru Takeichi, Tokyo Metropolitan University, Japan
Osamu Amai, ENRI, Japan
Ryota Mori, ENRI, Japan
Takayuki Yoshihara, ENRI, Japan
Yasuyuki Kakubari, ENRI, Japan
Yoichi Nakamura, ENRI, Japan

Secretaries

Atsushi Senoguchi, ENRI, Japan
Shunichi Futatsumori, ENRI, Japan

Preface

This is the third book compiled by the Electronic Navigation Research Institute (ENRI) after the completion of its workshops. ENRI is a national laboratory in Japan that specializes in air traffic management (ATM) and communication, navigation, and surveillance (CNS) for aviation. Since 2009, ENRI has organized biannual international workshops titled "ENRI International Workshop on ATM/CNS (EIWAC)." EIWAC aims to contribute to the development of civil aviation by facilitating the exchange and sharing the updated information about ATM/CNS in the world. The fifth workshop, EIWAC2017, was held in November 2017 in Tokyo. The main theme of EIWAC2017 was "Drafting the future skies." It has been a great honor for ENRI to provide opportunities to *draft the future skies* in the presence of key players from international organizations, civil aviation authorities, aviation industries, and academic institutions from all over the world.

This book is being published to share the essence of EIWAC2017 with people who are involved in CNS/ATM R&D worldwide. This book comprises four parts. Part I introduces the overview of EIWAC2017 and the summaries of the keynote and special speeches. Parts II, III, and IV comprise the selected papers from among the academia presented at EIWAC2017. Parts II and III discuss the cutting-edge research related to ATM and CNS, respectively. Part IV discusses the fundamental technologies that will be applied in future CNS/ATM systems.

Each chapter in Parts II, III, and IV comprises the selected papers that have passed through two selection stages. First, the technical program committee (TPC) conducted on-site evaluations with the help of professionals and experts who participated in EIWAC2017. They sincerely evaluated the quality of the presentations and manuscripts. Further, the papers that achieved high scores were nominated as candidate papers. A specific period was assigned to the candidates to revise and improve their papers to achieve the required publishing quality. Further, the revised papers were subjected to a review process by professionals and experts. The review is similar to the process that has been adopted by several academic journals. Multiple anonymous reviewers reviewed each paper and judged the quality of each paper.

I believe that this book presents novel ideas and great findings for professionals, experts, and researchers. Further, I hope that this book will stimulate the R&D activities in the field of CNS/ATM.

Finally, I wish to express my sincere appreciation to the EIWAC2017 TPC members, the associate editor, and the reviewers. They voluntarily supported us while selecting, reviewing, and compiling processes. Note that the TPC members have been separately listed in this book, and special appreciation has been attributed to them.

Tokyo, Japan Tadashi Koga

Contents

Part I
Introduction

Introduction to the Fifth ENRI International Workshop on ATM/CNS (EIWAC2017)

Shigeru Ozeki(✉), Tadashi Koga(✉), Takeyasu Sakai, and Xiaodong Lu

Electronic Navigation Research Institute (ENRI), 7-42-23, Jindaiji-Higashi-Machi, Chofu, Tokyo 182-0012, Japan
{ozeki,koga}@mpat.go.jp

Abstract. In this chapter, an overview of the fifth ENRI International Workshop on ATM/CNS (EIWAC2017) is provided, together with summaries of presentations at keynote sessions and special speeches. This chapter also explains the approach of the Electronic Navigation Research Institute toward the organization of EIWAC. In this workshop, various aspects of air traffic management (ATM) and its enablers in the fields of communication, navigation, and surveillance (CNS) were discussed. EIWAC2017 was held in Nakano, Tokyo, from November 14 to 16, 2017.

Keywords: Global air safety plan · Global air navigation plan · ATM · CNS · Standardization

1 Introduction

This introduction explains the approach taken by the Electronic Navigation Research Institute (ENRI) toward the organization of the ENRI International Workshop on ATM/CNS (EIWAC). EIWAC is the forum for exchanging up-to-date information from participants working in the fields of air traffic management, communication, navigation, and surveillance (ATM/CNS) concerning the strategies employed by each part of the aviation community. ENRI intends to offer all participants a forum to review their R&D strategies with reference to other parts of the aviation community. Therefore, ENRI invites participants from the wider aviation community—such as policy-makers from regulatory authorities, deployment managers, managing experts from operating fields, project managers from industry, researchers from R&D institutes, and members of academia.

The EIWAC conference is one approach taken by ENRI to fulfill its roles and functions. ENRI has been conducting research, development, and tests on electronic navigation systems for about half a century. ENRI is now the only research institute in Japan specialized in ATM and CNS. ENRI also conducts R&D projects to provide a choice of solutions for timely improvements to national, regional, and global aviation systems. Therefore, ENRI is also expected to contribute to the standardization of current and emerging ATM/CNS operations and enablers. ENRI sends its researchers to standardization meetings such as ICAO, RTCA, EUROCAE, and domestic committees, thanking them for giving ENRI a chance to make a contribution.

Electronic Navigation Research Institute, *Air Traffic Management and Systems III*, Lecture Notes in Electrical Engineering 555,
https://doi.org/10.1007/978-981-13-7086-1_1

One purpose of EIWAC is to facilitate the application of R&D results in the future. Thus, EIWAC has both keynote sessions on international standardization strategy and technical sessions on operations and R&D. In those sessions, participants share comprehensive up-to-date information from various layers of strategy between policy and R&D. The updates offer participants a chance to review their R&D strategy with a full understanding of the strategy in each layer. This review will improve the conformity of strategies among layers while adjusting the technical requirements and time to deliver the R&D results to be more realistic. Conformity will thus make R&D results easier to apply in the future. ENRI hopes that discussions in the EIWAC conference will help in putting these strategies in perspective and in improving the conformity of the activities of the participants with proper understandings.

The EIWAC conference is also intended to contribute to the modernization of aviation systems by introducing or furnishing the technical basics of harmonized global standards. The spread of standardized technologies and procedures will accelerate the modernization of aviation systems with global harmonization. The role of national research institutes such as ENRI is to harmonize standards while maintaining interoperability, even under regional conditions. EIWAC will contribute to the global harmonization of standards by providing a forum for sharing technical and operational experiences in regional conditions, forming the basis from which the standard requirements will be extracted.

The strong demand for an R&D meeting like EIWAC comes from the members of the aviation community who are concerned with the modernization of ATM/CNS systems to accommodate more traffic in the future because air traffic throughout the world is increasing steadily and is expected to maintain this trend for decades according to the ICAO forecast. Mitigation of congestion and reduction of environmental impact (while maintaining safety and efficiency) have become common interests around the world. In particular, in the Asia-Pacific region, increasing air traffic capacity, efficiency, and safety is essential because this region has the world's highest growth rate in air traffic. That is one reason why EIWAC is held in Tokyo, one of the gateways to the Asia-Pacific region.

The next section will explain some facts about EIWAC. This will be followed by reports on keynote sessions and some invited presentations.

2 Overview of EIWACs

ENRI would like to thank the members of the EIWAC Technical Program Committee (EIWAC-TPC) for their great contributions to making the workshop more attractive to potential participants. Members from other institutes offered comments to improve EIWAC by including viewpoints from outside of ENRI.

EIWAC has sessions for keynote speakers, as well as other sessions for technical discussions on operations and R&D. The keynote sessions of EIWAC are organized to share strategic updates among participants, and they are scheduled on the first day in most cases. The technical sessions offer participants chances to review operational facts and the progress of R&D with reference to updates from the keynote sessions.

Table 1 presents a brief summary with listing indices on the growth of EIWAC. It started in 2009 as a 2-day workshop and later expanded to include one more day for technical sessions to accommodate more presentations. EIWAC is growing as international participants offer more presentations and side meetings [1–5].

Table 1. Summary of the EIWAC series

Meeting#	1st	2nd	3rd	4th	5th
Name	EIWAC2009	EIWAC2010	EIWAC2013	EIWAC2015	EIWAC2017
Date, Year	March 5–6, 2009	November 10–12, 2010	February 19–21, 2013	November 17–19, 2015	November 14–16, 2017
Venue	Ohtemachi Sankei Plaza, Ohtemachi, Tokyo	Akihabara convention Hall, Akihabara, Tokyo	Odaiba Miraikan Hall, Odaiba, Tokyo	Ryogoku KFC Hall and Rooms, Ryogoku, Tokyo	CongresSquare Nakano, Nakano, Tokyo
Theme	Toward future ATM/CNS	Safety, efficiency, and environment	Drafting future sky	Global harmonization for future sky	Drafting future skies
Keynote speakers	4	7	9	13	5
Panel session	NA	"Future of automation in ATM," six panelists	"Future ATM: Centralized, de-centralized or best mixed," four panelists	NA	NA
Other sessions			Poster, Tutorial	Poster	Poster, Tutorial
Technical sessions	6	19	17	18	17
Presentations	22 w. incl. 9 from Japan	45 w. incl. 12 from Japan	46 w. incl. 13 from Japan	70 w. incl. 30 from Japan	71 w. incl. 25 from Japan
Participants	480	550	539	744	630
In first day	N/A	N/A	238	259	205
Foreigners	20	60	80	174	180
Countries	7	14	13	17	13

One of the advisors to ENRI suggested that excellent papers for EIWAC should be more visible to more researchers and students in the aviation community to encourage the next generation. ENRI reacted to this comment in 2012 by engaging an editorial team to compose the first book, "Air Traffic Management Systems," collecting selected papers from EIWAC2013 [6]. This was followed up by "Air Traffic Management

Systems—II," which collected selected papers from EIWAC2015 [7]. These books were published by Springer Japan in 2014 and 2017, respectively.

EIWAC2017 is the first international workshop to take place following a change in the corporate structure of ENRI. ENRI is now a part of the National Institute of Maritime, Port, and Aviation Technology (MPAT). MPAT was established in 2016 by uniting research institutes affiliated with the Ministry of Land Infrastructure, Transport and Tourism (MLIT) Japan. This change in corporate structure may offer ENRI chances to work with more researchers in MPAT and their colleagues for R&D on airports, safety analysis, and other areas. This change has not reduced the role of ENRI in the aviation community, and ENRI has continued to host the EIWAC conference.

3 Keynote Speeches

The keynote speeches in EIWAC2017 are summarized in this section. These speeches, on the whole, gave the authors of this section the impression that needed changes to the CNS architecture will be announced soon. The keyword "digital" is found many times in these sessions, indicating the application of network-based information management to aviation. The internet applications that have revolutionized many areas of business have not been widely introduced to aviation as yet, especially not in cockpits, because changes in aviation are slowed by the need to assess operational safety and to resolve legal issues arising from new operations and their enablers.

3.1 Stephen P. Creamer, "Global Plans and Importance of Global Interoperability"

Mr. Creamer is the director of Air Navigation Bureau (ANB) of ICAO. He presented updates from ICAO. First, he pointed to the result of the 39th ICAO triennial assembly with a revised Global Air Navigation Plan (GANP) and a new ICAO document for the Global Air Safety Plan (GASP) [8]. These are also known as ICAO Doc-9750 and Doc-10004, respectively. GANP has been revised with an extension to its Aviation System Block Upgrade (ASBU) roadmap. He emphasized the importance of spreading technology in such ICAO provisions with proper understandings and acknowledgements and of talent to catch up them. ICAO Doc-9750 and Doc-10004 will be revised every 3 years to remain up-to-date.

He explained ICAO's proposal for the enhancement of GASP. Annex 19 to the Chicago convention has already been proven effective and that the ICAO Safety Management Manual will be published as a web-based document to be updated regularly. ICAO is ready to provide various tools and consultation services to support member states in building and implementing State Safety Plans.

He raised concerns about how well the GANP has spread around the world, as well as about how ASBU is understood at planning. Because of aircraft lifetimes as long as 20 years, Block 4 in ASBU, which will start in 2037, should already be a focus of planning. Mr. Creamer asserted the need to plan based on estimates of the future density and size of airspace.

Mr. Creamer also explained ICAO proposals for the enhancement of GANP by creating a managerial structure with global, regional, and national layers. The global layer has a technical sublayer to support frameworks such as ASBU and Basic Building Blocks (BBBs) while also providing performance-based decision-making methods. This technical sublayer will control the global interoperability of products from regional R&D and the implementation after national plans and deployment. This enhancement will be proposed for the next update of the GANP in 2019.

Mr. Creamer also reported some developments at ICAO. ICAO provides a data-driven decision-making tool, iMPLEMENT, to improve safety performance in airspace. He also focused on the need for the evolution of standardization to bring about innovation on the flight deck, at the control position, and in other fields in aviation where people are using standardized tools. They are the first to know how they can be improved. Timely standardization is the key for the implementation of the elements of ASBU to upgrade the tools.

Therefore, ICAO has started to discuss SARPs-ready proposals from standards-making organizations and other aviation stakeholders. In addition, ICAO is searching for candidates for further proof-of-concept work. These actions may happen at the Standard Round Table, which is an interface meeting between standards-making organizations and ICAO. The round table also discusses the standardization roadmap. Innovation will be accelerated by timely provision of standards with increasing references from industry standards with supporting performance-based manner. This evolution in standardization is expected to improve information management, for example, by standardizing the connection between various aviation systems.

Finally, Mr. Creamer introduced some topics of interest as current and emerging issues. These include cyber safety, flights higher than 60,000 ft, and other considerations. ICAO is planning meetings to discuss what is next. The Global Air Navigation Industry Symposium (GANIS) and the Safety and Air Navigation Implementation Symposium (SANIS) have been introduced as forums for discussions leading up to the 13th Air Navigation Conference.

After his presentation, some questions were asked by participants. One question concerned the possibility for standardization as an ICAO provision with inclusion of the effects of human resources upon safety measures. Mr. Creamer answered by pointing out the example of standardization and certification of education. He mentioned that this example may be applicable to standardizing systems using artificial intelligence (AI).

3.2 Florian Guillermet, "Towards a Digital Aviation Infrastructure"

Mr. Guillermet is the executive director of SESAR Joint Undertaking. He presented the activities that SESAR is undertaking (and will undertake) "towards a digital aviation infrastructure," as well as the motivation and process for these activities.

First, he discussed why a digital aviation infrastructure is necessary for SESAR. He pointed out five major game changers reshaping the future of aviation, including (1) growing world populations, (2) technological disruption and entry players from the digital industry, (3) reshaping of customer expectations by hyper-connectivity and the

"always-on" reality, (4) automation of traditional jobs, the changing role of humans, and the race for talent, and (5) reshaping of global competition by global leaders (China and India) and the growing influence of non-state players (e.g., Google, Amazon, and the like). In addition to the above changes, there is currently a large gap between technologies on the cutting edge of IoT systems/devices and the technologies currently operational in CNS/ATM systems. This means that there is a significant room to improve current systems, and gaps must be filled.

Second, Mr. Guillermet discussed what SESAR is doing to establish a digital aviation infrastructure. He showed six building blocks of SESAR's technological vision for accomplishing this goal. These include (1) automation support, (2) flight- and flow-centric operation, (3) sharing of information through system-wide information management (SWIM), (4) integrated systems, (5) integration of all vehicles, and (6) visualization.

Third, he discussed how SESAR is proceeding with these developments. He explained the structure of SESAR, which is a unique public–private partnership consisting of 15 industry members, 100+ companies, and 3,000+ ATM experts. SESAR has many partnerships with academia, the R&D community, and other domains. The SESAR2020 program has set three domains of researches, namely, (1) exploratory research, (2) industrial research, and (3) very large-scale demonstrations. These three domains are ongoing in parallel but sometimes interact with one another. SESAR has developed 63 solutions, of which 24 are already being deployed across Europe. The details of the solutions are explained in the SESAR SOLUTIONS catalogue available on the web. The SESAR outcomes contribute not only to European states but also to a global interoperability context. Mr. Guillermet pointed out four key features to the success of SESAR, namely, (1) public–private partnership, (2) blending of corporate and academic values, (3) acceleration of the innovation lifecycle, and (4) worldwide outreach.

In addition to the above discussions, Mr. Guillermet mentioned two takeaways. The first takeaway is the game changers that are shaping the future of aviation. New service providers in other commercial domains will also change aviation, just as ICT/IoT/AI technologies have been changing our daily lives for decades. The second takeaway is the appearance of new services in airspace. There will be a significant value in "flying-something services" in the future. New mobility services, such as drones, would better assist citizens. Many investments are also being made today. In 2050, 30 billion euros of annual value are expected to be enabled by the digital aviation infrastructure in addition to the current commercial air services.

In conclusion, he discussed the "digital sky," i.e., the next necessary step to building an infrastructure suited for traffic expansion. Today, airspace is occupied mainly by traditional manned aviation. However, in the future, hundreds of thousands of connected flying devices will be found in the sky. The digital aviation infrastructure will enable all air operations. From the viewpoints of industry, innovation, and sovereignty, adaptation of the current system is not an option for continuing the aviation success story.

3.3 Pamela Whitley, "Aviation: A Global Digital Journey"

Ms. Whitley is the assistant administrator for NextGEN, FAA. She presented updates from the FAA with a focus on the digital revolution in aviation. The digital revolution comprises changes in data acquisition and its summary as useful information, as well as communication and information management. The advancement of the digital revolution for aeronautics serves to support stakeholders globally in terms of strategic and operational technical decision-making. On the other hand, improvements in mobile communications technology, such as cell phones or smartphones, are proceeding faster than those for aeronautics. Because aviation has been a series of learning from new experiences and stakeholders step forward after estimating risks before new challenges.

Ms. Whitley referred to the history of aviation, aeronautical radio systems, and air traffic control, which have affected each other as they have grown, and technologies for aviation are always at the cutting edge.

She pointed out certain aspects of aviation today. It is the responsibility of aviation to meet the strong demand for increased traffic in the future through global collaboration and proper strategic and tactical planning. One approach to addressing this challenge is through information management and sharing. Ms. Whitley pointed out that the aviation industry is rich in data but not in information because aviation works with systems before information management and focuses primarily upon safety. The approach that she suggests to stepping forward is "information." Information management will benefit aviation, and there is a great deal of experience with the technology in other business areas with stringent requirements for safety and security.

Ms. Whitley reported the FAA's decision to invest in infrastructure for aeronautical communication and information management with NextGEN. NextGEN has projects focusing on digital data and information, such as ADS-B, data communications, and SWIM. She introduced some examples of trials with these projects, such as a program at Charlotte Airport in North Carolina. A decision support tool for ground controllers works by sharing information concerning aircraft movements and intended paths from airlines; this tool helps controllers in creating smooth flows of ground traffic by synchronizing them with shared information. This improves the flow capacity of ground traffic, even with restrictions on the layout of taxiways at airports. This system is connected to the Traffic Flow Management System in the US and used to share the information of arrival time at departure fix with other controllers. Further trials are also planned for incorporating meteorological information and more.

Ms. Whitley expects improvements to the global harmonized interoperability at the next revision of GANP and GASP. She explained that the US FAA is working closely with ICAO, Europe, and Japan to bring about global harmonization in tandem with NextGEN.

She also reported some views on digital transformations. Such transformations are possible through implementation on various scales depending on the environment of the region under discussion and the needs specific to it. We can choose pieces of applications step by step when safety is assured, although support is needed from a wide range of stakeholders in the aviation community. This is why Ms. Whitley emphasized the importance of international cooperation.

She concluded her presentation by referring again to the history of aviation to point out the importance of global collaboration in safety and information management.

Some questions were raised concerning the activities related to standards and certifications for software in information management systems and the time for their delivery. Her answer was that NextGEN has projects with including the development of the standards for software installed in SWIM as well as in the Standard Terminal Automation Replacement System (STARS).

3.4 Yasuhiro Iijima, "CARATS Long-Term Vision for Air Traffic Systems—the Challenge for Realization of Future Technology—"

Mr. Iijima is the director general of air navigation services department of JCAB. He spoke about three topics of Japan's long-term vision for future air traffic systems, Collaborative Actions for Renovation of Air Traffic Systems (CARATS).

The first topic was the background and status of CARATS. CARATS aim at the renovation of Air Traffic Systems to meet the growing demands and needs for air travel. Japan has witnessed a growth in air traffic in line with the global growth of air traffic. The number of aircraft flying over the Flight Information Region (FIR) of Japan is expected to increase 1.5 times in the next two decades. Hence, air traffic control should have sufficient capacity for supporting the growing demand. Moreover, Japan is expected to accommodate a large amount of air traffic during the Olympic and Paralympics Games to be held in Tokyo in 2020. The Japan Air Navigation Service (JANS) needs to renew and improve the current ATM systems.

The second topic was CARATS' actions in collaboration with ENRI. He remarked that ENRI plays a major role in this research and development. He introduced four CARATS actions based on cooperation with ENRI: AeroMacs, Ground-Based Augmentation System (GBAS), Dynamic Airborne Reroute Procedures (DARP), and System-Wide Information Management (SWIM). In AeroMacs and GBAS projects, ENRI has developed prototype systems and it conducts evaluations and validations in operational environments. In DARP, ENRI verifies the effect of DARP based on actual flight path data provided by airlines. In SWIM, JANS developed a SWIM test bed in collaboration with ENRI and participated in international experiment programs for SWIM such as the Mini-Global Demonstration trials with Australia, Canada, Singapore, the US, and other nations.

The third topic was the importance of the Plan–Do–Check–Act (PDCA) cycle in CARATS. The CARATS measures need to be changed on a regular basis to adjust to the changing external environment and user needs. In the process of a PDCA cycle, all stakeholders should be involved through the collaboration among governments, academia, and the industry. As one practical example of PDCA, he introduced efforts on time-based traffic management. The JANS commenced trajectory-based operation trials in 2011. The ATC system calculates the time of the passing points, while the aircraft are flying enroute. Air traffic flow was managed by designating the computed time. However, there existed large differences between the time estimated by aircraft avionics and that estimated by the ATC system. The trials were suspended in August 2014. CARATS organized a new WG, "Air Traffic Time Management Study WG," for dealing with operational and technological issues in time-based traffic management.

The trial operations will be restarted in 2019. ENRI plays an essential role in verifying the improvement method examined by the WG and studying the prediction models.

In conclusion, he spoke that planning, practice, verification, and improvement is essential for the enhancement of air traffic systems in CARATS. In particular, JANS recognizes that cooperation with ENRI is indispensable.

3.5 Masahiro Kudo, "Perspective on the Role of R&D Toward the Future Sky"

Mr. Kudo is the director general of ENRI. He spoke about facts concerning the roles and strategies of the Electronic Navigation Research Institute (ENRI).

First, he spoke about ENRI's history and organization. ENRI was established in 1967 as a research institute of the Ministry of Transport, Japan, and in 2001, owing to the governmental reorganization act, it became an independent administrative agency. In 2016, the Ministry of Land, Infrastructure, and Transport (MLIT) merged three national research institutes for transportation and established one national research institute—the National Institute of Maritime, Port, and Aviation Technology (MPAT). Figure 1 shows the organizational structure of MPAT. MPAT has one governing body and three research institutes. These institutes include the National Maritime Research Institute (NMRI), the Port and Airport Research Institute (PARI), and ENRI. ENRI has three departments: one branch-navigation-systems department (Nav), one surveillance and communications department (SC), and one air traffic management department (ATM). The Iwanuma branch office is located at Sendai airport. As a center for field and flight experiments, the branch has been achieving essential missions for the last 40 years. The Great East Japan earthquake and tsunami destroyed the branch in 2011, but it has now fully recovered from this damage. Many prototype CNS/ATM systems have been installed for operational tests.

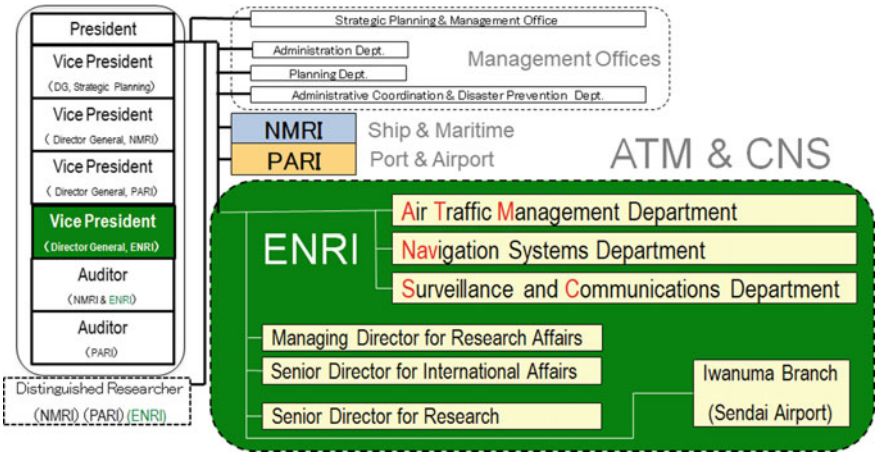


Fig. 1. MPAT organization

Second, Mr. Kudo discussed ENRI's roles. He mentioned that its fundamental role is to support government undertakings for aviation, mainly achieved through the Japan Civil Aviation Bureau (JCAB). JCAB has developed its own plan to support future air traffic in Japan in accordance with the ICAO Global Air Navigation Plan (GANP) and Aviation System Block Upgrades (ASBU). ENRI actively supports JCAB missions for GANP and ASBU and contributes to aviation societies through R&D activities.

Third, Mr. Kudo spoke about strategy. He emphasized the necessity of collaboration. Many missions and various R&D tasks are needed to achieve the objectives of GANP and ASBU. On the other hand, ENRI's research resources—personnel and budget—are limited. Nevertheless, we remain willing to provide practical and useful solutions to aviation societies. Our answer is to collaborate with external research sources in academia and industry. Such collaborations will build a firm foundation to support R&D activities at ENRI.

Mr. Kudo concluded by noting that our R&D work would firmly support safe and comfortable air travel all over the world.

4 Invited Talks

4.1 Paul Bosman, "Drafting Future Skies—From SWIM to a Fully Digitized ATM"

Mr. Bosman is the head of ATM strategies division, EUROCONTROL. He first described the background and problems with information management in current ATM systems. The present-day ATM system comprises a wide variety of applications developed over time for specific purposes. It is characterized by many custom communication protocols, each with its own self-contained information systems. Each of these interfaces is custom designed, developed, managed, and maintained individually and locally at significant cost. Moreover, the way ATM information is defined/structured and the ways it is provided and used are specific to most ATM systems. Legacy information products, such as NOTAMs, are limited in functionality and usability because of their textual nature. Paper-product AIs are also limited in usability and interoperability without a digital format.

Mr. Bosman then explained the requirements and concept of System Wide Information Management (SWIM). The expected increase in aviation capacity demands, economic pressure, and attention to environmental impact require ever more accurate and timely information. Global improvements in information management are intended to integrate the ATM network in the information sense, not just in the system sense, and are envisioned for application as a SWIM concept. This concept provides not only a system architecture for the delivery of information services to meet the expectations of the ATM community in different key performance areas but also a common understanding of different information domains.

Next, he introduced the current status and development plan for SWIM implementation in Europe. EUROCONTROL has been identified as providing the technical specifications for the foundational-information-service-implementation-standardization area in support of SWIM deployment. The goal of SWIM is to foster interoperability

between European ATM stakeholders in a networked environment. The means to achieve this goal are information services and service-orientation practice.

Consequently, draft EUROCONTROL specifications for SWIM have therefore been developed with the objective of providing foundational requirements applicable to the deployment of SWIM in support of interoperability within the European ATM network. The EUROCONTROL specifications for SWIM consistently address the description of information services, the definition of the exchanged information, and the technical infrastructure on which the services are implemented.

Finally, Mr. Bosman outlined some challenges for achieving global interoperability and harmonization by SWIM. The interoperability envisaged by SWIM requires standards to be agreed upon at the global, regional, and local levels. It is fair to assume that not all information needs to be exchanged among all in a highly interoperable way at all levels. The operational context plays an important role in setting the scope of the required interoperability, making it necessary to delineate precisely the problem/solution space. Furthermore, since levels of interoperability are correlated with levels of complexity, it is good practice to consider requiring interoperability only where needed (within an identified context) and with the right level of complexity.

4.2 Patrick Souchu, "Benefits of Standardization—EUROCAE Activities"

Patrick Souchu is the SESAR program director of DSNA, and a EUROCAE council member and the former council chair. He introduced the main characteristics of the EUROCAE association and then spoke about recent changes and current activities of EUROCAE.

First, he introduced the main characteristics of EUROCAE, which is a standardization body based in the EU; it has 220 members, 36 inactive working groups, and 1,400 experts. Then, he introduced two coordination groups recently established by EUROCAE: the European ATM Standards Coordination Group (EASCG) and the European UAS Standards Coordination Group (EUSCG). The groups aim to accommodate the increased need for new standards arising from the deployment of SESAR, the transition to performance-based regulation, and growing demands for UAS in a wide variety of fields. The EASCGS/EUSCG goals are to accomplish harmonized implementation in support of the regulatory frame, establish timely availability, streamline relevant standardization activities in EUROPE, set new standards arising from the deployment and implementation process, and implement performance-based regulation.

Second, Mr. Souchu spoke about recent changes in EUROCAE, namely, the new office, new work infrastructure, and new financial structure. These changes are expected to improve the efficiency of works and processes in EUROCAE and to bring many benefits not only for members but also for aviation stakeholders.

To conclude, Mr. Souchu discussed the status of WGs in EUROCAE. He showed the list of all active WGs. Currently, the WGs of GNSS, A-SMGCS, ADS-B, Aeronautical System security, Interoperability of ATM validation platforms, VDL Mode 2, Remote Tower WG, and SWIM services are hot topics for discussion. Other WGs are also being steadily discussed for standardization.

4.3 Satoshi Kogure, "Japanese Space-Based PNT System, QZSS—Services, System and Applications"

Satoshi Kogure is the executive director of the Quasi-Zenith Satellite System (QZSS) development, National Space Policy Secretariat (NSPS), Cabinet Office in the Government of Japan. He has reported the status of QZSS, which is a Japanese space-based position, navigation, and timing (PNT) system that is under development.

The first satellite, called "Michibiki," was launched in 2010 as a demonstration satellite based on collaboration between several Japanese research institutes; further, various technologies related to satellite navigation were validated by performing technical experiments and application demonstrations.

In 2011, the Government of Japan decided to establish their own satellite navigation system and initiated the development of QZSS. In 2017, three satellites were successfully launched. Further, the operational service provision will commence in 2018.

QZSS will provide three kinds of services, including GPS complimentary, GNSS augmentation, and messaging service, in Japan and the surrounding areas, Eastern Asia, and the West Pacific Rim. The report of Kogure includes the following topics: Services to be provided through QZSS; How QZSS provides the services: System configuration; and Expected applications to be developed.

5 Conclusions

EIWAC may have a hard time in the near future changing its name to reflect changes in aviation infrastructure. The keyword "digital" may lead the merger of CNS into an integrated infrastructure for information management; then, the C in EIWAC should be changed to stand for a keyword other than CNS.

This integration will be a result of the technical trend toward "performance-based standardization," as well as the shortage of aeronautical radio spectrum availability. The performance-based standard will allow the introduction of a new technology or architecture more easily, with reuse of many parts of safety assessments for existing systems focusing on essential performance parameters, if this is possible for the considered case.

Integration will also result from shortages of the aeronautical radio spectrum. This spectrum should be managed and guarded to prepare for the strong demands for communication bandwidth in the future by improving spectral efficiency. This will impose changes to the aeronautical CNS architecture to compete with other mobile communications at spectrum assignments. One example of such architecture is found in the military radio system, which has a 4D reference for position and time through GNSS, a multipurpose datalink with CNS functions for friendly mobile devices, sensors and voice communication for suspicious or incompatible mobile devices, and network infrastructure for information management. This architecture may improve the radio spectrum efficiency for CNS functions by integrating these functions into an efficient radio communication system. This topic will be discussed at the next EIWAC conference, regardless of its name.

References

1. EIWAC2009 homepage. https://www.enri.go.jp/eiwac/2009/en/index.html
2. EIWAC2010 homepage. https://www.enri.go.jp/eiwac/en_EIWAC2010siryou.html
3. EIWAC2013 homepage. https://www.enri.go.jp/eiwac/en_EIWAC2013siryou.html
4. EIWAC2015 homepage. https://www.enri.go.jp/eiwac/eiwac_2015_eng.html
5. EIWAC2017 homepage. https://www.enri.go.jp/eiwac/eiwac_2017_eng.html
6. Electronic Navigation Research Institute. Ed. (2014) Air traffic management and systems. Springer, pp 3–14
7. Electronic Navigation Research Institute. Ed. (2017) Air traffic management and systems II. Springer, pp 3–24
8. International Civil Aviation Organization (2014) 2013–2028 Global air navigation capacity & efficiency plan. 2014–2016 Triennium Edition DOC 9750: 10–11

Part II
Air Traffic Managements

Future Airspace Design by Dynamic Sectorization

M. Schultz(✉), I. Gerdes, T. Standfuß, and A. Temme

Institute of Flight Guidance, Department of Air Transportation, German Aerospace Center (DLR), Braunschweig, Germany
{michael.schultz, ingrid.gerdes, thomas.standfuss, annette.temme}@dlr.de

Abstract. The future airspace has to provide a reliable infrastructure and operational concept to ensure efficient and safe operations considering both flight-centric operations and the integration of new entrants. We propose an approach for a dynamic sectorization to manage the air traffic demand and flow appropriately. Our dynamic sectorization results in enhancements of the current operational structure (less deviation in controller task load) and leads to a significantly lower controller task load for the newly created airspace. Since future 4D trajectory management demands an efficient consideration of operational (e.g., temporally restricted areas), ecological (e.g., contrail prevention), and economic (e.g., functional airspace blocks) constraints, our dynamic sectorization method contributes to the highly flexible use of current and future airspace. In this paper, we provide an overview of several use cases and describe the working principle of our approach: fuzzy clustering of air traffic, Voronoi diagram for initial structures, and evolutionary algorithms for optimization.

Keywords: Airspace efficiency · Dynamic sectorization · Clustering · Evolutionary algorithms · Optimization

1 Introduction

Due to political restrictions, more specifically the federal air sovereignty, airspace structure in Europe is determined by a high level of fragmentation. Furthermore, air space areas are generally not determined by the dynamic traffic flow over the day, but by static constraints (e.g., national borders). The resulting inefficiencies (negative operational and economic effects) are caused by coordination efforts as well as asynchronies between operational concepts and capacity management of the different decision-making units. Concepts such as dynamic capacity balancing target this problem, but do not treat the root cause of having an insufficient airspace design. Based on the liberalization of the aviation markets and as a part of the Single European Sky concept to create a legislative framework for European aviation, the Functional Airspace Block (FAB) approach was introduced by the European Commission in order to restructure the European air space. The management of European Air Traffic within a Functional Airspace Block system is one of the main pillars in the Single European Sky concept for meeting future air traffic requirements.

Electronic Navigation Research Institute, *Air Traffic Management and Systems III*, Lecture Notes in Electrical Engineering 555,
https://doi.org/10.1007/978-981-13-7086-1_2

The sectorization of the airspace considers requirements of air traffic control (safety, capacity, and efficiency), users (unhindered access), and environment (restricted areas over cities, residential areas, etc.) as far as possible. Particularly, Air Traffic Control (ATC) demands a sufficient airspace by designing appropriate procedures, routes, or holding areas, and considering operational demands such as handling of mixed traffic, balanced workload, or efficient arrival/departure sequences. The current allocation of air space areas systematically leads to operational inefficiencies and demonstrates that the task load of the air traffic controllers significantly deviates depending on the air space sector and the traffic demand/flow over the day.

To ensure a more efficient allocation and a harmonized task load distribution, we consequently changed the current paradigm of traffic flow, which is determined by airspace structure, to a dynamic approach of a structure which is adjusted to the traffic flow sequentially. We contribute to the flight-centered Air Traffic Management with a specific approach to dynamically optimize the airspace focusing on the sector structure and resource allocation, considering both operational and economic efficiency targets (e.g., task load, fragmentation of flights by sectors). Therefore, we develop a four-step approach containing (1) economic performance evaluation, (2) fuzzy clustering of traffic flows on the day of operations, (3) generation of a new sector structure based on Voronoi diagrams, and (4) evolutionary algorithms in order to adjust and optimize the new sector structure depending on dynamic demands over the day of operations (see [1, 2]).

Our dynamic sectorization concept enables sector adaptation in a time-dependent way by dynamically adapting the position and shape of the sectors with respect to actual necessities and restrictions (e.g., capacity, task load, controller availability, or stability/resilience). Due to operational determinants and restrictions, the scope will cover upper airspaces only. The developed tool AutoSec is an essential and superior approach to combine the idea of completely unstructured airspace and today's rigid structures to achieve both a balanced and more efficient use of the airspace. Additionally, the dynamic sectorization may enable efficient strategies to cope with severe weather conditions, temporal closures (e.g., space vehicles, intruders, military), and integration of new entrants (e.g., personal or unmanned air vehicles).

In the context of SESAR's flight-centric operations, we developed a Dynamic Airspace Sectorization (DAS) approach, which differs from the Dynamic Airspace Configuration (DAC) where predefined airspace blocks are combined to superior structures [1]. We are assuming a continuous airspace that will be separated without a specific demand for underlying structures but considering both the current/future air traffic flows and the controller's ability to manage all assigned aircraft (e.g., measured by task/workload). We fundamentally changed the common method of "traffic flow is following the provided structure" to "a structure follows the traffic flow" paradigm. Furthermore, our approach bridges the gap between structured and unstructured airspace designs and will be a fundamental key element for efficient air traffic operations taking into account both regular and disruptive events.

1.1 Objectives and Structure of the Document

This paper provides an overview of our research focusing on the efficient sectorization of the airspace with a special emphasis on a dynamic adaptation of operational air

traffic demands and flows. In the first section, we introduce the common challenge of an efficient airspace design using examples from the European airspace, the consideration of volcanic ash clouds, the integration of space vehicles in highly segregated airspace, as well as the air traffic control in urban environments. In the second chapter, we introduce our methodology in detail, addressing the approaches of fuzzy clustering, Voronoi diagrams, and evolutionary algorithms to analyze, structure, and optimize the airspace, respectively. In the following, we introduce our tool environment and results of the exemplary application of dynamic sectorization to a part of the European upper airspace. This paper finishes with a conclusion and an outlook for our future research, implementation, and validation tasks.

2 Airspace Structure

Air traffic demands efficient airspace management, which ensures safe operations considering economic (e.g., cost-efficient air traffic service), environmental (e.g., contrail prevention [3]), and operational (e.g., convective weather cells) requirements. In this section, we focus exemplarily on several topics connected to efficient airspace structures with a clear indication that a dynamic sectorization could provide significant benefits to the specific challenge.

2.1 Analysis of European Airspace

Due to the legal and political framework, airspace structure in Europe is characterized by a high level of fragmentation. Basically, boundaries of airspaces, and partly sectors, are determined according to national territories. The current structure of European ATM is expected to lead to inefficiencies caused by coordination efforts, inconsistencies between the ANSPs' strategies and capacity restrictions [4]. Furthermore, the growing number of flight movements, as well as the rising market share of Low-Cost Carriers in Europe, pose new operational challenges to Air Navigation Service Providers (ANSPs) and lead to an increasing cost pressure. Consequently, productivity and cost efficiency of ANSPs have gained increasing attention recently [5].

Following the liberalization of the air transport markets and as a part of the Single European Sky (SES) concept, the Functional Airspace Block (FAB, see Fig. 1) approach was introduced by the European Commission (EC) in order to restructure the European airspace [6]. Through the consolidation of airspaces and a subsequent adjustment of procedures and technical equipment, nine FABs were constructed to meet future air traffic requirements.

The management of European Air Traffic within a Functional Airspace Block system is one of the main pillars in the Single European Sky concept. However, over the past years, significant problems in implementing this concept could be observed for some FABs. First adopted in 2004, all FABs were scheduled to be operating by 2012. However, just two of the nine FABs have been implemented today. A key underlying rationale of calls for intensified cooperation among European ANSPs and for a consolidation of this sector is the assumption that the provision of air navigation services is characterized by significant economies of scale. However, previous studies show that

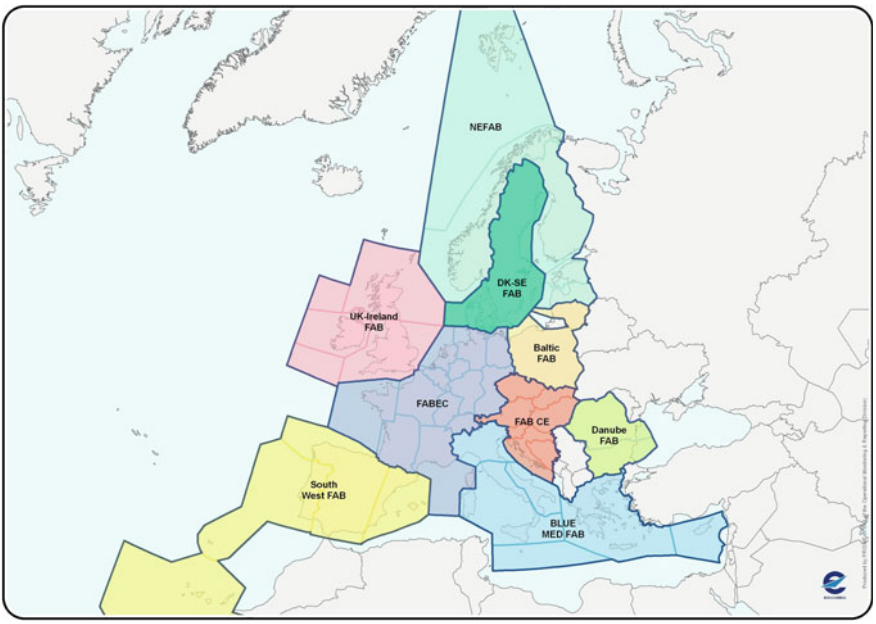


Fig. 1. Functional airspace blocks (EUROPEAN Commission MEMO/13/525)

there is a turning point when diseconomies of scale occur [7, 8]. The reasons derived concern in particular operational procedures, cultural and legal differences, an inappropriate allocation of ANSPs to one FAB, and missing incentives in implementation.

Since the fragmented airspace shows inefficiencies as well as the consolidation into FABs, an alternative restructuring of the European ATM is mandatory. Based on the lessons learned through the previous studies, a dynamic sectorization was identified as an appropriate candidate for further improving the ATM efficiency [2].

2.2 Volcanic Ash Cloud

The eruption of the Eyjafjallajokull volcano in 2010 seriously affected the European air traffic system. Airspaces under the control of different Air Navigation Service Providers were affected, such as all Italian flight information regions, Croatia, Albania, Macedonia (former Yugoslav Republic of Macedonia), Serbia, Slovenia, Hungary, Romania, Bulgaria, Bosnia, Herzegovina, Malta, East of Greece, Cyprus, Turkey, Bulgaria, Ukraine, and Moldavia.

This specific non-regular situation emphasizes the need for a harmonized and cooperative approach without insistence on national sovereignty of airspaces or predefined airspace structures. In order to predict and prevent similar effects on the European airspace, a test scenario was created. This scenario includes a potential eruption of the Stromboli volcano and contains the ash distribution created for test purposes by the organization committee of the ICAO's regular volcanic ash exercise (VOLCEX 14/01). The simulation was performed using a fast-time simulation tool and

the traffic data as well as ash distribution characteristics for 1st and 2nd of April were simulated (see Fig. 2).

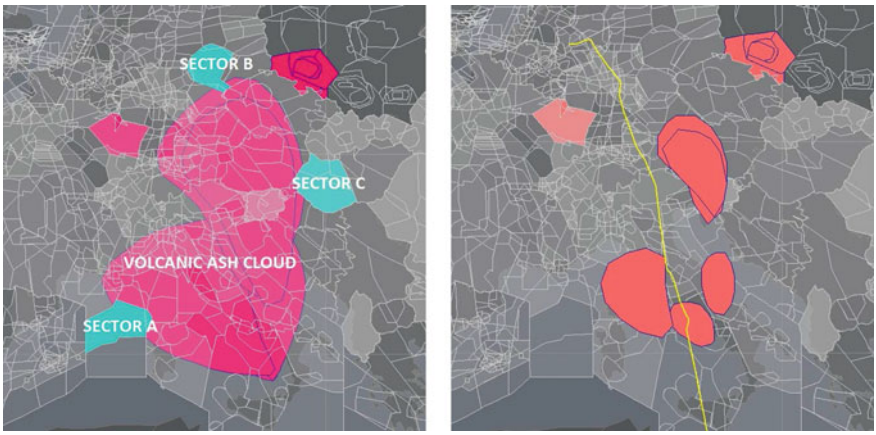


Fig. 2. Volcanic ash cloud in European airspace (left) and an exemplarily rerouted flight considering a safety risk assessment (right) [9]

Beside the cross-border effect of the volcanic ash cloud, both the prediction (development of area in time and space) and the specific particle concentration of the affected areas are essential in order to efficiently manage this non-regular situation. To generate reliable flight trajectories around the volcanic ash cloud, an active airspace management is needed to efficiently use the available, scarce airspace capacity. Furthermore, this will lead to the necessity of restructuring the airspace to adapt the sectorization to the new trajectories with the target of distributing the controller task load equally between all sectors. The high complexity of this disruptive scenario is primarily driven by the uncertainties in the predictions of the ash cloud and the different handlings of regular and non-regular situations. Our approach of a dynamic airspace sectorization is immanently designed to handle these kinds of disruptive scenarios in the same way as regular days of operations.

2.3 Interoperable Air and Space Traffic Management

The expected increasing demand for commercial space transportation will require an interoperable air and space management, where the incoming and outgoing space flights challenge the current segregated airspace in Europe [10]. A seamless and safe integration into the civil airspace is needed to handle a regular pass-through of space vehicles (see Fig. 3). There are several approaches of future air/space operations, which have to be covered by a comprehensive air/space traffic management, ranging from rocket launch operations up to suborbital flights with capabilities of operating in normal airport environments (e.g., arrival/departure areas, standard runway operations). To enable a reliable mixed air and space traffic management in Europe, the operational

airspace concept has to be timely and spatially flexible in order to adapt upcoming changes in the flight path of the space vehicle. In this context, the dynamic airspace sectorization is a key technology in enabling future European landing sides.

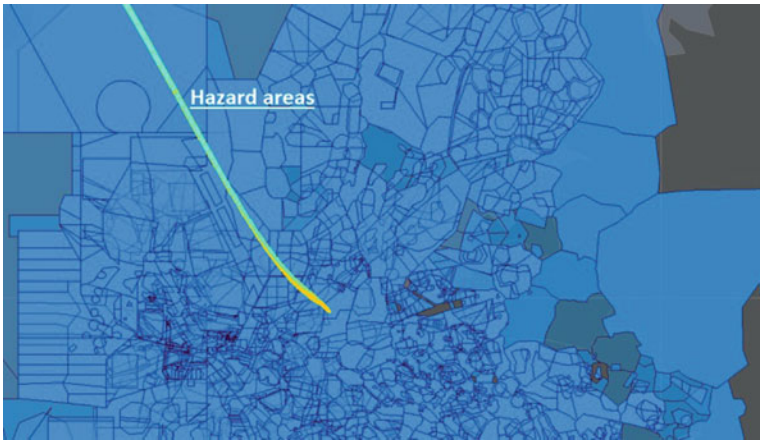


Fig. 3. Integration of space flights into European airspace [11]

2.4 Urban Environment

From a ground-based traffic point of view, the traffic theory indicates three phases of traffic: free flow, bound flow (synchronized traffic), and congestion (wide moving jam) with zones of phase transitions [12]. In the context of traffic management, these zones call for different kinds of management solutions. In comparison to the air traffic, an almost free flow phase could be understood as flights in upper airspace and the synchronized/congested phase is found in the terminal area and on the runway. The different kinds of traffic result in different controller strategies. As an example, controllers like to grant directs to the runway if the traffic density is low, but with increasing traffic, the controllers switch to specific arrival patterns (e.g., trombone structures or holding areas) to efficiently handle the increasing task load.

Due to increasing traffic demand by personal air vehicles, it is expected that a future urban traffic management will primarily have to cope with the synchronized and congested phases, which demands specific adaptation of the accompanied airspace. In the project Metropolis, different kinds of potential airspace designs were tested regarding airspace efficiency [13]. In Fig. 4, four different approaches are shown with an increasing airspace structure: free flight (individual conflict solution), layered structure (altitude bands corresponding to heading ranges), zones (today's airspace design), and tubes (road-based concept, separation by speed, direction, and vehicle types). In the context of the project Metropolis, the layer concept offers the best balance with regard to safety, airspace capacity, and efficiency [13].

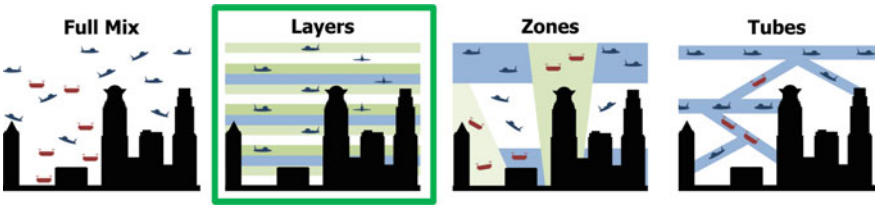


Fig. 4. Different airspace designs with increasing structure to integrate personal and unmanned aerial vehicles into an urban environment [14]

If, in the future, the urban area will consist of a significant amount of movements of personal air vehicles, the frequency of traffic will follow the daily time-dependent demand for transportation. As shown for the case of Berlin in Fig. 5, there is a clear indication of a highly used infrastructure, which changes over the day (e.g., morning and evening peaks).

In Fig. 5, a morning peak is shown in the Berlin road transport system, where red indicates congestion and green is used for free flow conditions. An appropriate airspace design and operational concept have to essentially consider the expected high amount of synchronized air traffic and should be able to handle the changing air traffic demand through a dynamic and adaptive approach.

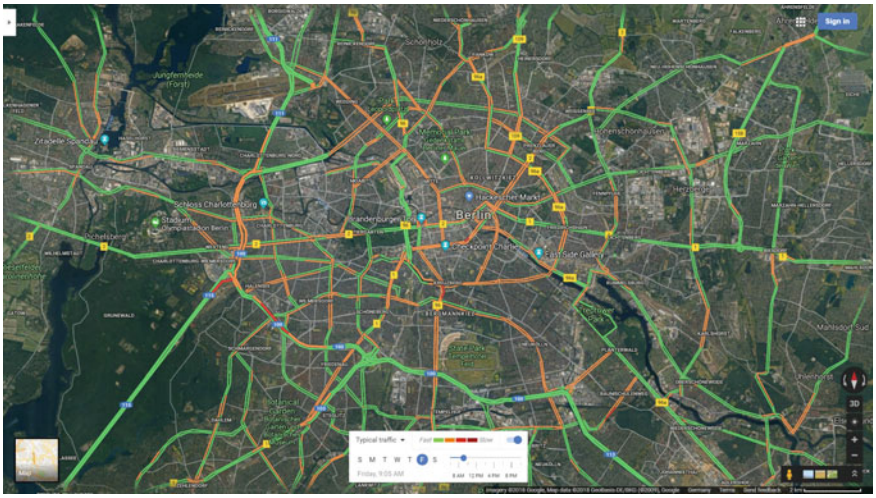


Fig. 5. Traffic densities on Friday morning in Berlin using Google Maps' traffic situation display

2.5 Unmanned Freight Operations

In the context of the DLR research project Unmanned Freight Operations (UFO) [15], we analyzed the implementation of unmanned cargo aircraft flying in a formation for the transport of relief goods starting in Europe and flying to Africa. We assumed that a specific, temporary airspace structure could provide an efficient handling of this formation. In a first step, we used a segregated airspace area for the unmanned cargo relief flight as is done today for military and relief flights as well. This allowed us to evaluate the impact on airspace capacity of such flights [16]. Dynamic adaptation of the sector structure is a prerequisite for allowing more flexibility in airspace segregation. Instead of reserving an airspace area for the whole flight duration, the adaptation facilitates a restructuring of airspace during the segregated flight so only a specified area around this flight has to be reserved. This reserved area moves according to the actual flight's trajectory, and the sectorization of the surrounding unrestricted airspace is dynamically adapted with respect to the residual traffic demand.

3 Dynamic Airspace Sectorization

In Sect. 2, we emphasize the need for a comprehensive airspace management, using examples of special traffic patterns, air traffic influenced by disruptive events, or new operational concepts necessary for future urban environments. All these changes require a restructuring of the airspace in a more flexible and time-dependent way. The necessity to adapt the current and future ATM system results in new concepts, such as the Dynamic Airspace Configuration (DAC) [17] or Dynamic Airspace Sectorization (DAS) [18]. For a DAC system, the airspace is structured into smaller airspace blocks, which are subsequently combined to form sectors in dependence of the actual traffic situation. The DAS system is not built on a substructure. A completely new airspace structure is designed for each specific traffic situation (continuous in space and time).

Both approaches focus on a harmonized, uniformly distributed task load for and between the airspace controller(s). Thereby, the task load depends on the density of air traffic (movements per area) and on the complexity of the air traffic movements (e.g., level changes, relative speeds). The proposed more flexible and especially time-dependent concepts for air traffic control structures (sectors) will be able to cope with the identified future requirements, even when there is no expert knowledge for future operations. Both the DAC and DAS approaches will provide an efficient support to human controllers or controlling systems. It is assumed that flights in urban environments will fulfill similar requirements to today's flight planning procedures (e.g., communication of flight plan) and the future negotiation of full 4D business trajectories.

3.1 Theoretical Background

In this section, we introduce the basic elements of our DAS approach, which consists of three major steps: clustering of air traffic movements (fuzzy clustering), design of an initial airspace structure using Voronoi diagrams, and optimization of the structure with evolutionary algorithms [1, 2].

3.1.1 Fuzzy Clustering

The aim of clustering techniques in general is to find a partition of a given dataset. In a fuzzy partition, a datum is not necessarily assigned to a unique class or cluster. Instead, membership degrees are associated with each datum and each cluster. These membership degrees provide information about the ambiguity of the classification. Fuzzy clustering techniques can adapt to both noisy data and not well-separated classes. Although noise will not be a major issue with radar data of flight trajectories which are initially used for the DAS approach, the clear separation of classes might be difficult.

This and our earlier promising results with fuzzy clustering techniques (see [19]) allow us to choose fuzzy clustering techniques for our DAS approach, which is based on optimizing an objective function. The clustering solution consists of cluster centers, i.e., the center of gravity of a cluster with regard to a characteristic distance function, and membership degrees. For an overview of fuzzy clustering and its applications, see e.g., [19] or [20]. For our approach, we use the fuzzy-c-means clustering approach [21] that is based on the Euclidean distance measure. The resulting centers are used as initialization for the following structuring with Voronoi diagrams.

3.1.2 Structuring Approach: Voronoi Diagram

A Voronoi diagram [22] is a possibility for structuring an area, depending on a certain number of so-called center points, into sections where every point belongs to the section with the nearest center point. Edges are created of all points which belong to two center points, i.e., the points of an edge have the same distance to the two center points, and vertices are those points which are associated with three different center points. An area containing all points with the same nearest point is called “face”.

At first, faces at the boundary of the area of interest are not limited at their outside. A convex hull is added once the faces have been determined. Often this convex hull simply has the form of a rectangle. A more detailed description of our structuring approach is given in [1]. In Fig. 6, the area of Norway is used to derive an exemplary airspace design. Starting with a given traffic sample, seven cluster centers were identified and surrounded by Voronoi structures. The Voronoi diagram provides an inner structure defined by the blue dots in Fig. 6 (right).

As mentioned before, the application of Voronoi diagrams is limited to areas with a convex hull. In the case of airspace areas, this methodological limitation has to be overcome to provide an operational concept for air traffic control. Therefore, the determination of the initial airspace structure has been extended in such a way that a Voronoi diagram can be adapted to arbitrary boundaries. Hereby, the line-segment-intersecting method was adapted to the problem of intersecting the Voronoi diagram and the boundary of the selected airspace area [1, 22]. Thus, the intersections of edges with the national border provide an additional set of structural elements (green dots) which, together with the blue dots, are subject to the following optimization. Whereas the blue dots could be rearranged inside the boundaries, the green dots can only be moved on the boundary.

3.1.3 Evolutionary Algorithms

The common principles of evolutionary algorithms follow the idea of biological evolution (cf. [23, 24]). In nature, a group of individuals mix their genetic material,

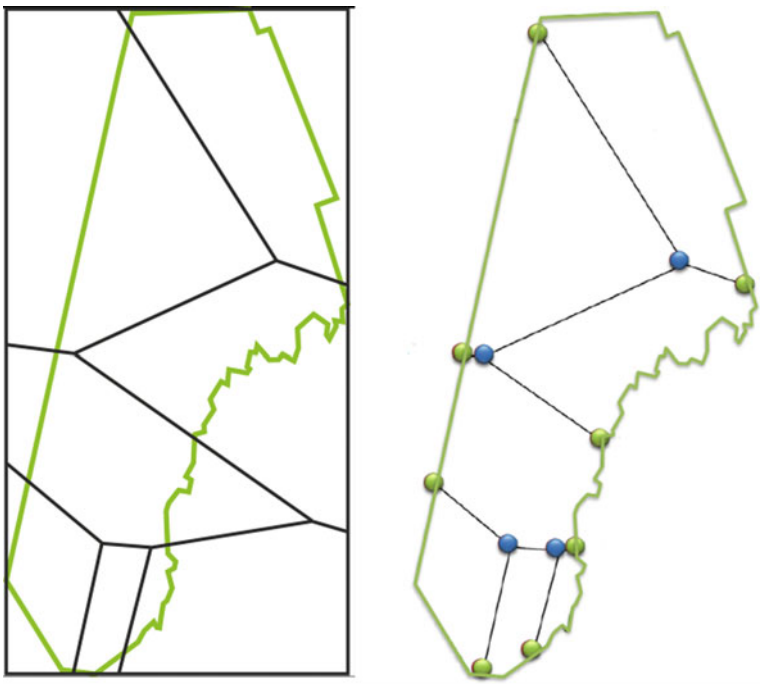


Fig. 6. Example of a Voronoi diagram (black) for Norway with seven center points, a square as convex hull (black), country/sector boundary in green. Additional to the blue dots (inner structure), the green dots arise from the intersection of boundaries and edges of Voronoi diagram

especially the information coded in chromosomes, to obtain a better chance of survival in a hostile environment through a higher degree of adaptation. For an evolutionary algorithm, a population of solutions for an artificial problem is coded as sequence (chromosome) of parameters (genes) describing a problem solution. These fundamental principles are transferred to the optimization of a technical/operational environment (cf. [1]). Using a set of available solutions, specific components and parameters are mixed and stochastically mutated to generate a new set of solutions.

The assessment of these new solutions concerning their fitness with regard to an underlying evaluation function results in a hierarchical order, where the most appropriate solutions are used as the parental generation for the next generation of solutions. In the context of the dynamic airspace sectorization, the idea of evolutionary algorithms is used to optimize the initial airspace structure considering different evaluation functions [1]. These functions are defined depending on the specific research/operational scenario. For the approach of functional airspace blocks, the evaluation function could consider the ANSP productivity (composite flight hours per air traffic controller-hour on duty) or controlled IFR airport movements. Aiming at a harmonized distribution of controller effort in the European network, the evaluation function could include the aggregated task load and the corresponding standard deviation for each sector over the day of operation to efficiently allocate the available expertise. For each controller task related to managing aircraft in airspace, a specific effort in seconds is allocated. The controller task load

caused by a flight in a selected sector is calculated as sum of all actions to be taken by the responsible controllers. Amount and distribution of these task load sums can be used as factor for the evaluation function (for details see [1]).

3.2 Simulation Environment, Tool Chain

The Institute for Flight Guidance of DLR has developed a basic framework of tools which implement the developed DAS concept and enable the creation of a complete new sectorization, considering the current and future air traffic flows and demands.

The new sector structure is dynamically created in three steps (see Fig. 7). First, the software tools RouGe (route generator) and PrePro (preprocessor) developed at DLR are used to generate the necessary airspace data (traffic and sector data) based on Eurocontrol's DDR-2 data, see [25]. In this step, the DDR-2 data are mainly used as a reliable source for air traffic movements. In future operational environments, the air traffic movements will be provided by online ANSP data services, which should consist of historic data (DDR-2), current data, and flight planning data (to predict the air traffic progress). Furthermore, PrePro is used to define the boundary of the airspace area of interest and to smooth the traffic to avoid high numbers of waypoints. This region could be any connected area of airspace with one continuous boundary (connectedness constraint [18]) independent of national borders.

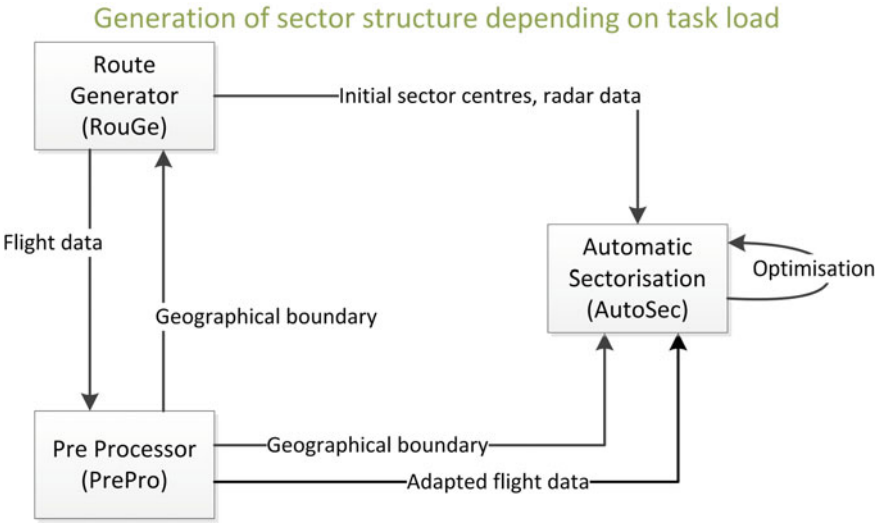


Fig. 7. Fundamental structure of framework of tools to create a sector structure (cf. [2])

RouGe uses this boundary generated with PrePro to filter the radar data of the aircraft movements. We cluster the air traffic data, use the resulting cluster centers as centers for the calculation of an initial Voronoi diagram implemented in AutoSec, and optimize the resulting diagram (i.e., an initial airspace structure) with the evolutionary algorithm with respect to the air traffic flow [1] selected with RouGe.

3.3 Evaluation Function

The most important and crucial part of our DAS approach is the definition of an appropriate evaluation function for the evolutionary algorithm to optimize the air space structure. Therefore, several evaluation functions with different factors have been comprehensively tested [1], and the following factors were selected for a future application of the developed DAS approach:

- Sum of task load in the investigation scenario,
- Standard deviation of task load (as a measure of uniformity of its distribution),
- Well-formed controlling areas, measured by the standard deviation of interior angles,
- Number of flight intervals caused by the airspace sectors (intersection of flights).

Including the number of flight intervals per sector indirectly results in avoiding short dwell times of flights in sectors, which is an important constraint seen by several authors for the creation of dynamic airspace sectorizations [26, 27]. Furthermore, the number of flight intervals per sector represents the segmentation of flights over different sectors. The ideal angle is calculated as the sum of standard deviations of interior angles for all sectors related to the optimal mean value of interior angles for the corresponding sector (polygon). Both factors influence the optimization to prefer a more convex structure of the sector. This approach covers the proposed strong convexity constraint, which should be applied to the created sectors to ensure that an aircraft visits every sector only once. However, today's operational sectors are not necessarily convex (see Fig. 6, green boundary), and our DAS approach is able to handle non-convex structures efficiently.

A weighted combination of sum and standard deviation of task load, ideal angle, and number of flights per sector is then used as evaluation function. The calculation is based on a fixed flight plan which remains unchanged during the optimization process. In our investigation scenario, no traffic simulation takes place; operational measurements, such as delay or punctuality, are not taken into account in the evaluation.

4 Applications

Within recent work, it was shown that it will be possible to replace ATM experts, who are normally responsible for arranging a system of sectors, with our DAS system, thereby improving the balance of the task load between the observed sectors. For our evaluation experiment, the airspace area EDYYDUTA with four sectors (see Fig. 8) was used from flight level 300.

In the first two steps, a Voronoi diagram was created which is shown in Fig. 9. The results for the evaluation function for the original sectorization were better (lower) than for our start solution with the Voronoi diagram (see Table 1). Nevertheless, when executing the third step and optimizing the start solution, the resulting sectorization (see Fig. 10) has much better results for all parts of the evaluation function than for initial/original sectorizations.

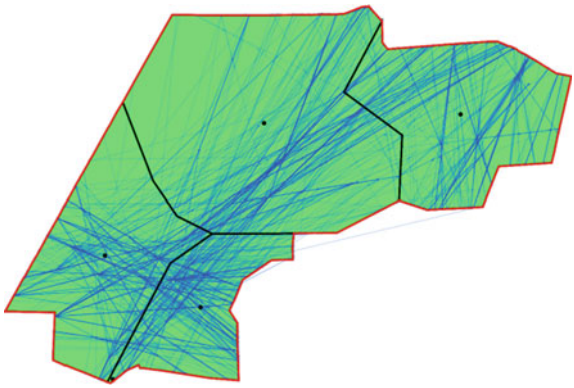


Fig. 8. Airspace EDYYDUTA with air traffic movements (aircraft trajectories) as blue lines with the current existing sectors. A darker blue color represents a higher traffic density on the trajectories

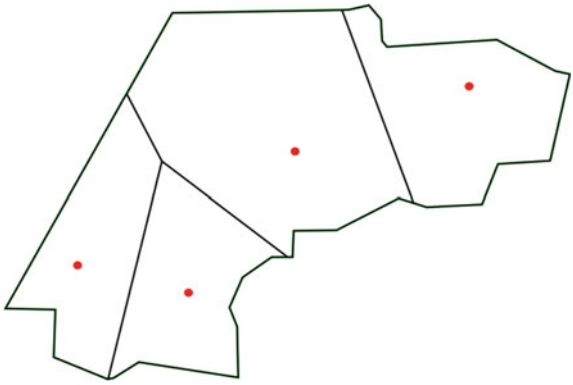


Fig. 9. Airspace EDYYDUTA with cluster centers (red dots) as Voronoi diagram adapted to the outer sector boundary

Table 1. Comparison of results for evaluation function factors Task Load (TL) and Task Load Standard Deviation (TLSD) for original, Voronoi and optimized Voronoi sectorization

Type	Original	Voronoi	Optimized
TL	231848.9	245068.4	227703.4
TLSD	14634.6	17093.3	1756.5

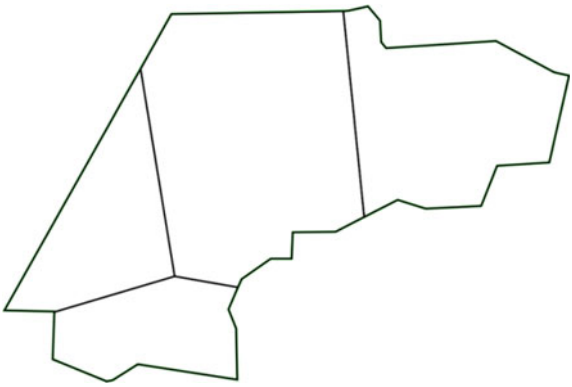


Fig. 10. Example for EDYYDUTA airspace structure after optimization with the evolutionary algorithm

5 Summary and Outlook

In our paper, we provide an overview of current and future fields of application demanding a dynamic sectorization of controlled airspaces. To enable efficient air traffic operation, we propose a DAS approach considering the air traffic flow (aircraft trajectories) and the complexity of the airspace (measured in controller task load). We implemented three steps to derive the dynamic airspace structure: fuzzy clustering to derive centers of dense traffic for air traffic movements, Voronoi diagrams for initial structures, and evolutionary algorithms to finally optimize the airspace structure with regard to specific evaluation functions. Using the exemplary airspace EDYYDUTA, we could show that our approach is able to reproduce today's structure, without bringing in any expert knowledge to the optimization process.

It is expected that the aircraft movements used already depend on the current structure and the interdependencies restrict a stepwise improvement of today's airspaces. Our DAS approach, however, provides a reliable technology for overcoming this limitation and also for enabling scalable concepts ranging from space traffic management to personal air vehicles in urban environments.

To further improve our DAS approach and to proof the concept in the operational controller environment, we have brought our expertise into a new project which is investigating the ecological impact of contrails in comparison to the operational effort needed. In this context, we will provide adaptive airspace structures indicating contrail areas and set up a real-time controller environment to enable an operational interaction between human controllers and our DAS system.

References

1. Gerdes I, Temme A, Schultz M (2016) Dynamic airspace sectorization using controller task load. SESAR innovation days, Delft

2. Standfuß T, Gerdes I, Temme A, Schultz M (2018) Dynamic airspace optimisation. CEAS Aeronaut J
3. Rosenow J, Fricke H, Luchkova T, Schultz M (2018) Minimizing contrail formation by rerouting around dynamic ice-supersaturated regions. AAOAJ 2(3):105–111
4. Eurocontrol FAA (2016) Comparison of air traffic management-related operational performance: U.S./Europe 2015
5. Buxbaum J, Standfuß T (2014) Data envelopment analysis. DFS Innovat Fokus 1:11–16
6. Regulation (EC) No549/2004, Framework for the creation of the single European sky
7. Standfuß T, Fichert F, Schultz M (2017) Efficiency gains through functional airspace blocks? An analysis of economies of scale in European air traffic management, ITEA
8. Button K, Nieva R (2013) Single European sky and the functional airspace blocks: will they improve economic efficiency, vol 33, pp 73–80
9. Luchkova T, Vujasinovic R, Lau A, Schultz M (2015) Analysis of impacts an Eruption of volcano stromboli could have on European air traffic. In: 11th USA/Europe air traffic management research and development seminar
10. Kaltenhäuser S, Morlang F, Luchkova T, Hampe J, Sippel M (2017) Facilitating sustainable commercial space transportation through an efficient integration into air traffic management. New Space 5(4):244–256
11. Luchkova T, Kaltenhäuser S, Morlang F (2016) Air traffic impact analysis design for a suborbital point-to-point passenger transport concept. In: 3rd annual space traffic management conference "Emerging Dynamics"
12. Kerner BS (1998) Experimental features of self-organization in traffic flow. Phys Rev Lett 81:3797
13. Sunil E, Hoekstra J, Ellerbroek J, Bussink F, Nieuwenhuisen D, Vidosavljevic A, Kern S (2015) Metropolis: relating airspace structure and capacity for extreme traffic densities. In: 11th USA/Europe air traffic management research and development seminar
14. Schneider O, Kern S, Knabe F, Gerdes I, Delahaye D, Vidosavljevic A, Leeuwen P van, Nieuwenhuisen D, Sunil E, Hoekstra J, Ellerbroek J (2014) METROPOLIS—urban airspace design, D 2.2—concept design (D 2.2)
15. Temme A, Helm S (2016) Unmanned freight operations. DLRK 2016, Brunswick
16. Brodersen Y, Luchkova T, Temme A, Lindner M, Rosenow J, Schultz M (2017) Entwicklung und Bewertung von Formationsflugszenarien unbemannter Frachtflugzeuge. DLRK 2017, Munich
17. Kopardekar P, Bilimoria K, Sridhar B (2007) Initial concepts for dynamic airspace configuration. In: 7th AIAA aviation technology, integration and operations conference (ATIO), Belfast
18. Flener P, Pearson J (2013) Automatic airspace sectorisation: a survey. Comput Res Reposit 1311:0653
19. Keller A (2002) Objective function based fuzzy clustering in air traffic management. Otto-von-Guericke University, Magdeburg
20. Oliviera JV de, Pedrycz W (2007) Advances in fuzzy clustering and its applications. Wiley
21. Krishnapuram R, Keller J (1993) A possibilistic approach to clustering. IEEE Trans Fuzzy Syst 2:98–110
22. Berg M, Cheong O, van Kevald M, Pvermars M (2008) Computational geometry, algorithms and applications. Springer, Berlin, Heidelberg
23. Michalewicz Z (1996) Genetic algorithms + data structures = evolution programs. Springer, Berlin, Heidelberg

24. Gerdes I, Klawonn F, Kruse R (2004) Evolutionäre Algorithmen. Vieweg, Wiesbaden
25. Eurocontrol (2017) Eurocontrol demand data repository. Eurocontrol. https://www.eurocontrol.int/ddr. Accessed 26 Sept 2017
26. Delahaye D, Schoenauer M, Alliot JM (1998) Airspace sectoring by evolutionary algorithms. In: IEEE international congress on evolutionary computation
27. Kulkarni S, Ganesan R, Sherry L (2011) Static sectorization approach to dynamic airspace configuration using approximate dynamic programming. In: ICNS conference

An Approach for Attribute- and Performance-Based Evaluation of Interdependent Critical Infrastructures

P. Förster[1(✉)], P. M. Schachtebeck[1], T. Feuerle[1], P. Hecker[1], M. Branlat[2], I. Herera[2], and R. Woltjer[3]

[1] Technische Universität Braunschweig, Institute of Flight Guidance, Braunschweig, Germany
{peter.foerster, p.schachtebeck, t.feuerle, p.hecker}@tu-braunschweig.de
[2] SINTEF Digital, Trondheim, Norway
{matthieu.branlat, ivonne.herrera}@sintef.no
[3] Swedish Defense Research Agency – FOI, Linköping, Sweden
rogier.woltjer@foi.se

Abstract. Notably, in engineering and behavioral sciences, the topic of resilience is being investigated broadly quantitatively in technical systems like infrastructure or public transportation and qualitatively with respect to social and organizational aspects. Especially with regard to disturbances and crises in complex socio-technical systems, human operators play a pivotal role in ensuring the continuation of operations by adapting to the situation. An integrated framework for quantitative assessment as well as behavioral aspects in a socio-technical system is therefore essential to measure resilience and to compare different design approaches. The combination of quantitative and qualitative approaches is presented in this paper. In the wake of a crisis often not only the system itself is affected but also interdependent systems. The resilience of those combined systems is the subject of this conceptual paper. Two objectives are pursued. First, the creation of generic resilience management guidelines which are subsequently translated into operating procedures, strategies, and practices in order to support individuals, systems, and organizations in the face of crisis and to validate their cross-domain applicability. The second objective of the paper is to contribute to closing the gap between formal descriptions of resilience in technical systems and the representation of the influence of the human operators. This is done by following an approach that combines serious gaming exercises of different scenarios, expert judgment and a simplified simulation of the involved systems which provides a quantitative assessment of resilience. The ongoing work described in this paper is being carried out within the scope of the DARWIN project which has received funding by Horizon 2020. Preliminary results of the project that address the creation of resilience management guidelines will be presented.

Keywords: Resilience in ATM · Crisis management

Electronic Navigation Research Institute, *Air Traffic Management and Systems III*, Lecture Notes in Electrical Engineering 555,
https://doi.org/10.1007/978-981-13-7086-1_3

1 Introduction

Exceptional events, such as disturbances of great magnitude or of unusual character, can pose a significant challenge to critical infrastructure systems with respect to safety and operational continuity. In the aftermath of such events, the performance of these systems is reduced or even disabled over a certain period of time. The return of the system to nominal conditions is delayed or at worst, further implications of the disturbance will be amplified.

Involved actors are often unable to cope appropriately with an unfamiliar situation. The great number of stakeholders involved in the ATM system, interacting with each other under the constraints of different goals and competition for resources, illustrates the significance of social and organizational aspects of a socio-technical system and the problems arising in the face of a crisis. This is especially valid with regard to interdependent systems, which often display a confusion of responsibilities or inadequate synchronization between them [1]. In order to mitigate the consequences of exceptional events, affecting critical infrastructures, service providers aim to increase the resilience of the system. Two main directions of assessment of resilience in complex systems are currently subject of investigation [2]. On the one hand, performance-based metrics are applied to directly measure the outcome of the system during a rebound. This is done by evaluating the system's functionality, i.e., the state of the system, which is adapted to the particular research domain and scenario. The other aspect is attribute-related. Here, system attributes are explored which increase the resilience of the system. They often represent indicators that are gathered by subjective assessment. Resilience, as an inherit property of a socio-technical system, enables to anticipate and have readiness to respond to the implications of disturbances and to graceful extend essential functionalities when surprises arrive [3]. Both directions have become part of a broad range of research in recent years.

The project DARWIN develops resilience-enhancing measures intended to improve the response to expected and unexpected crises affecting critical infrastructures. In particular, DARWIN's measures consist of innovative training modules for crisis management practitioners and a set of resilience management guidelines. Intended as cross-domain reference guidelines, the latter can help organizations to operationalize various resilience concepts depending on the specific domain in which organizations operate. Thus, developed guidelines aim to serve as a set of generic recommendations in order to enable institutions to derive particular and domain-specific instructions by applying a set of resilience concepts, found to increase resilience. The developed guidelines will be assessed by means of an integrated framework that considers attribute-based metrics as well as performance-based approaches to evaluate resilience and considers two critical infrastructure systems, air traffic management, and health care.

DARWIN will investigate its guidelines' effectiveness in two coupled domains, ATM and health care, and will do so by means of serious gaming exercises. The exercises will involve domain experts and will make use of cross-domain crisis scenarios, thus capturing relevant interdependencies between actors and resources. In order to detect bottlenecks to monitor and to perform what-if analyses, a computer-based simulation will be added. This simulation represents a simplified conceptual

model of the relevant parts of the involved systems of a particular scenario. Workflows as well as available resources are depicted here. Expert knowledge supports the creation of the conceptual model and the configuration of the implemented model into a discrete event simulation environment.

To quantitatively assess the impact of the guidelines on system resilience, the simplified model incorporates individually defined performance indicators for each particular scenario. The evolution of the chosen performance indicators over time (or the state of the particular system functionality) will serve to calculate the systems resilience, based on approaches made in [4–7]. Reference [6] proposes a resilience factor by means of a sigmoid function.

$$f(t) = \frac{1}{1 + \exp\left(\pm\frac{t-\mu}{\tau}\right)}, \tag{1}$$

Reference [7] formalizes this approach by describing the course of the system functionality by a succession of three sigmoid functions, investigating the time series of occurring delay at an airport during an exceptional storm.

Figure 1 illustrates the underlying concept of resilience quantification which is widely spread in the engineering domain and was introduced by [8]. The course of the system performance represents different phases in the aftermath of a disturbance. The stages can be described by the time constants of the particular sigmoid function. Due to a disturbance d, the functionality is being deviated from the former (static), reference state. Resilient systems would anticipate the event sooner and due to adaptive capacities, effective countermeasures c would be introduced to shorten the recovery phase as well as to limit the extent of performance loss. Depending on the resilience abilities of the system, the difference between the asymptotically reached new state and the reference state would be minimized.

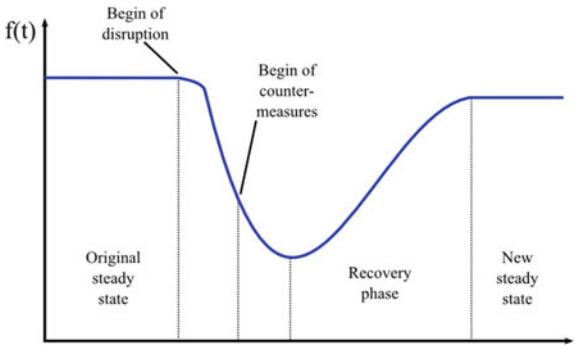


Fig. 1. Systems functionality as a measure for resilience

The paper is divided into four chapters. Chapter 2 gives a short overview of relevant resilience concepts, as well as qualitative and quantitative assessments and

modeling approaches found in resilience literature. The necessity to consider both in a framework, attribute as well as performance-based approaches will be discussed.

This chapter presents (i) the process adopted by DARWIN to generate generic resilience guidelines and (ii) discusses the framework which is used for their evaluation. Chapter 4 discusses a scenario to illustrate the approach.

2 Resilience Concepts and Motivation for the Framework

Currently, resilience is being investigated in a broad spectrum of research disciplines. This stretches from the beginnings in material sciences and psychology to ecology, environmental sciences and the engineering of technical systems.

A good overview over the distribution in different research domains is provided in [9]. Due to the variety of the domains, the terminology has not been evolved consistently over the years.

For example, the terms robustness and resilience are often used synonymously in different fields of research. An explanation of these terms, discussed in the light of the various definitions of resilience, is given in [5]. The described methodology relates to a performance-based approach in ATM and proposes a differentiation between both terms. Depending on the system functionality and a predefined reference state, the abilities of a system to absorb disturbances as well as to bounce back in case of deviation from the reference state, are discussed against the background of a performance-based ATM system. This stage of an envisaged ATM system is intended to operate on the basis of key performance indicators (KPIs). Performance-based operations envisaged by SESAR [10] and quantified by means of ICAO performance indicators [11] can therefore be used to express resilience. Thus, resilient design principles might contribute to the shape of the future ATM system. Examples of performance-based approaches of resilience in the ATM system, that incorporate KPIs, can be found in [12, 13].

As stated before, with respect to the broader research community of resilience, different concepts as well as different approaches for assessment exists today. A widely used classification of resilience is the division into ecological resilience [14], engineering resilience [14] and, being a younger discipline, resilience engineering [15]. Ecological resilience can be described as the ability of an ecosystem to maintain equilibrium in the face of a disturbance, whereas different reference states are possible. In contrast to that, resilience in technical systems, such as in transportation or power infrastructure, refers to one reference state only. The aim of the resilient system is to return to this reference state as fast and with less cost as possible. Resilience engineering investigates sociological aspects in complex socio-technical systems by ensuring the safe operation of the system. Here, human and organizational aspects play a pivotal role.

Other concepts also address the various dimensions of resilience in socio-technical systems, such as the technical, organizational, economic or social dimensions, as proposed in [8] which also states that a single measure cannot be applied on all four dimensions and that each dimension has to define appropriate means to assess resilience. Here, properties of resilience are denoted as robustness (ability to withstand a

disturbance), redundancy (available substitutes of elements), resourcefulness (ability to manage resources during a disturbance), and rapidity (ability to reach a defined system state in a given time horizon). Reference [6] provides a good survey of definitions of resilience in different research domains and suggests three capacities as fundamental properties of a resilient system. Absorptive (ability to withstand disturbances by means of proactive design of operations and resources for a minimal deviation from the system state), adaptive (ability to adapt to occurring disturbances in case absorptive capacity is exhausted by means of prediction of disturbances or restructuring), and restorative capacity (ability to return fast to the reference state) are to be optimized in order to increase the resilience of a socio-technical system. Though both latter resilience concepts [8, 6] refer to a performance-based perspective. That is, the system performance is being used to define measures of resilience with regard to the rebound from a disturbance.

With respect to resilience engineering (RE), which is a more attribute-related method of assessment of resilience, Woods [3] proposes four concepts of resilience. Resilience as rebound ("how a system rebounds from disrupting or traumatic events and returns to previous or normal activities"), resilience as robustness ("expanding the set of disturbances the system can respond to effectively"), resilience as graceful extensibility ("how a system extends performance, or brings extra adaptive capacity to bear, when surprise events challenge its boundaries"), and resilience as sustained adaptability ("the ability to adapt to future surprises as conditions continue to evolve"), advocating the latter two as foci for the development of the resilience concept in RE.

Considering the different approaches of resilience concepts, considerable overlapping of various aspects can be observed. This paragraph provides an example by means of the three capacities, introduced in the engineering domain by [6]. For instance, general resilience design principles, proposed by [16] can be associated to a particular capacity. For example, the proposed functional and physical redundancy, see [16], can be associated with absorptive capacity. Reorganization and human backup could be assigned to adaptive capacity. Reference [17] provides further examples with respect to the three capacities. Here, absorptive capacity is associated to the design principle: "Design and prepare redundancy, backup and substitution to lower the interdependency impacts" and adaptive capacity is enhanced by "Adjusting and improving the organizational and administrative structure to increase early-warning awareness". As stated in [17], recovery capacity is related to "Design advanced decision support platform to quickly find the restoration sequences and priorities, optimum resource allocation strategies". This design principle is addressed by the simulation, which is part of the DARWIN framework. During the course of the crisis, the system state is presented to the participants, supporting for example the anticipation of bottlenecks, such as resource shortages, earlier or the comparison of different strategies, by the use of predictions.

Although there is a large variety of conceptions of resilience in socio-technical systems, stemming from organizational as well as engineering domains (the worldwide systematic literature review in [18] reveals roughly 300 definitions), to the best knowledge of the authors, no generally accepted definition of the concept of resilience for socio-technical system is available to this day.

Similar to the different concepts of resilience, evolving from behavioral and engineering disciplines, the evaluation of resilient systems also displays a variety of approaches. With regard to technical systems, such as public transportation or road infrastructure (domains which are tending to engineering resilience and are therefore performance based), a great number of approaches for resilience quantification can be found, see [1, 6, 19–21]. A good overview of resilience quantification approaches is provided by [20]. A great number of approaches relate to the system functionality (see Chap. 1) and its course over time. In this case, resilience is being described by the amount of loss of performance and time constants defining the phases of disruption and recovery. A first illustration is provided in [5]. Resilience engineering approaches on the other hand assesses resilience of socio-technical systems in an attribute-related manner. A prominent, widely known approach to assess resilience consists of four so called cornerstones (see [22]), to be considered by involved stakeholders. To anticipate, monitor, respond and adapt, learn and evolve is essential in order to constitute a resilient system. An assignment to adaptive and restorative capacity can easily be deducted, though when it comes to quantification of resilience different approaches between engineering resilience and resilience engineering are taken. For example, in contrast to the measurement of resilience by means of system functionality by engineering resilience, [3] constitutes, "that is not possible to measure resilience per se, but the potential for resilience", see also [1]. That is, in the field of resilience engineering current research investigates so-called indicators which reveal potential to "remain resilient when challenging events occur" [1]. Several RE approaches have been proposed to identify and to apply those indicators of resilience potential, such as the Resilience Analysis Grid (RAG), the Functional Resonance Analysis Method (FRAM) or the Q4 Balance framework. Reference [1] provides a good summary of available tools for assessment in the field of resilience engineering (RE). A SESAR project on RE [23] applied the application of resilience engineering concepts to the ATM system using a workshop- and human-in-the-loop-simulation-based concept evaluation approach for ATM safety assessment and design. Brittleness is another aspect to measure resilience, since it describes "how rapidly a system's performance declines when it nears and reaches its boundary conditions" [3]. Though it should be noted, that a general formalization is hard to deduct and not yet presented.

In summary, it can be stated that quantifications of socio-technical systems focus rather on the rebound after a disturbance, thus neglecting influences of human operators, whereas socio-technical indicators that reveal resilience potential cannot easily be quantified. As an example of a misalignment of concepts of resilience engineering and engineering resilience might serve the recent approach to assign adaptive capacities only to the recovery phase and absorptive capacity only to the disruption phase, see [21]. Here, the adaptive capabilities of human operators during the decrease of the performance, and even further back, before a crisis, are not considered appropriately.

Since the proposed framework of DARWIN includes a simulation to support human operators, aspects of its implementation will be addressed shortly. Reference [17] discusses different modeling approaches for critical infrastructure systems (CIS). Agent-based modeling as well as flow-based techniques seem to be the preferred methods, keeping in mind CIS represent complex interdependencies of a variety of technical system, resources, and human actors. Reference [24] states that, as opposed to

topological based modeling methods, flow-based modeling methods cover all three resilient capacities. Reference [24] also states the problem which arises regarding the available quantifications of resilience of technical systems with respect to the adaptive capacity, as part of a resilient system. Adaptive capacity cannot be measured explicitly. That is, when analyzing quantitative measures of the performance of the system, the human contribution is already inherent to system. Any quantitative assessment of the system functionality will incorporate adaptive capacity.

The aim of this project is first to provide resilience management guidelines to support organizations as well as individuals in critical situation and second, to evaluate their effectiveness by attribute- and performance-based approaches. It aims to contribute shorten the gap between resilience engineering and engineering resilience, by integrating measuring and modeling approaches of both perspectives in one framework.

The following chapter presents the approach to obtain the generic resilience management guidelines in DARWIN. It also introduces the developed evaluation framework.

3 Resilience Management Guidelines

3.1 The DARWIN Project

DARWIN is a project funded by the European Union's Horizon 2020 research program. It is part of the DRS-7-2014 [25] projects in which multiple consortia investigate crises and disaster resilience in order to operationalize resilience concepts. The rationale behind the call was to enhance the preparedness, response, and recovery in the wake of man-made and natural disasters like the sinking of the Deepwater Horizon in 2010 and the eruption of the Icelandic volcano Eyjafjallajökull in 2010.

The consortium is led by SINTEF Digital from Norway. Other partners are the Technische Universität Braunschweig (Germany), Carr Communications (Ireland), Deep Blue Srl (Italy), ENAV S.p.A. (Italy), Instituto Superiore de Sanità—ISS (Italy), the Swedish Defense Research Agency—FOI (Sweden), Centre for Teaching & Research in Disaster Medicine and Traumatology—KCM (Sweden) and Ben-Gurion University of the Negev (Israel).

As introduced, the aim of the project is to improve the responses to expected and unexpected crises affecting critical infrastructures by the development of resilience management guidelines that cover the essential resilience abilities of the affected stakeholders. Adapted from Hollnagel [22, 26, 27], these abilities are to be able to:

- **Anticipate**: The ability to understand how the situation at hand develops and how certain changes might influence the system positively or negatively.
- **Monitor**: This describes the active search for signs of what might happen in the near future—be it positive or negative.
- **Respond and Adapt**: This is the preparedness to react and to possess the resources to encounter a disastrous event in an efficient and flexible way.
- **Learn and Evolve**: The ability to learn from previous experiences to be more effective the next time a similar (or other) situation arises.

3.2 Approach of the Project

Based on a systematic literature review covering multiple areas of research, a number of concepts for crisis management have been extracted [18]. This vast list has been condensed using interviews and a two-stage Delphi process with experts, both internal and external to the consortium [28]. As a result of this process, a total of 51 resilience concepts emerged which was assigned to eleven categories: (1) collaboration; (2) planning; (3) procedures; (4) training; (5) infrastructure; (6) communication; (7) governance; (8) learning lessons; (9) situation understanding; (10) resources; and (11) evaluation [15].

During the development process, these concepts were transformed into a set of requirements [29] which builds the foundation to derive the generic resilience management guidelines. These aim at guiding organizations during the development and improvement of the capabilities and processes that support resilience, as described by the requirements. Resilience management guidelines, specific to their domain of operation are used to assess, improve, and create such processes and tools, typically by roles at higher echelons in the organization. The guidelines provide the generic material based on which actors, in charge of policies and regulations in their domain develop domain-specific guidelines.

Figure 2 depicts the project's hierarchical view of guidelines. The separation of users shown in Fig. 2 is a simplification because actors often have roles at different levels (e.g., organization leaders can also be policy makers, as experts in their domain and primary stakeholders). It should also be mentioned that community leaders are part of the scope of DARWIN since community resilience is addressed (it is, however, not part of the examples discussed in this paper and therefore neglected in the description).

3.3 Concept Cards

The DARWIN guidelines are developed in the form of Concept Cards (CCs). These documents constitute mechanism by which DARWIN translates general resilience concepts into actionable guidelines. CCs propose interventions that can be implemented in order to reach the capabilities identified in crisis management practices and scientific literature. Many CCs are interrelated. The guidelines build on the CCs by organizing and relating them. This aspect of the guidelines is a consequence of the fact that resilience management capabilities are not independent. For instance, the management of adaptive capacity requires that coordination is properly supported between operational units; these two types of resilience management capabilities are different, but interdependent [30].

Each CC consists of a table that contains information on its purpose, relevant actors in the field of crisis management the CC is addressed to, expected benefits, propositions for implementation, and associated challenges.

Ways for implementation are described for utilization before, during and after a crisis and, when possible, are complemented by illustrative examples of best practices available in the literature, which provide additional guidance. Estimation of these practices' maturity in terms of technology readiness levels (TRLs) is also provided [30].

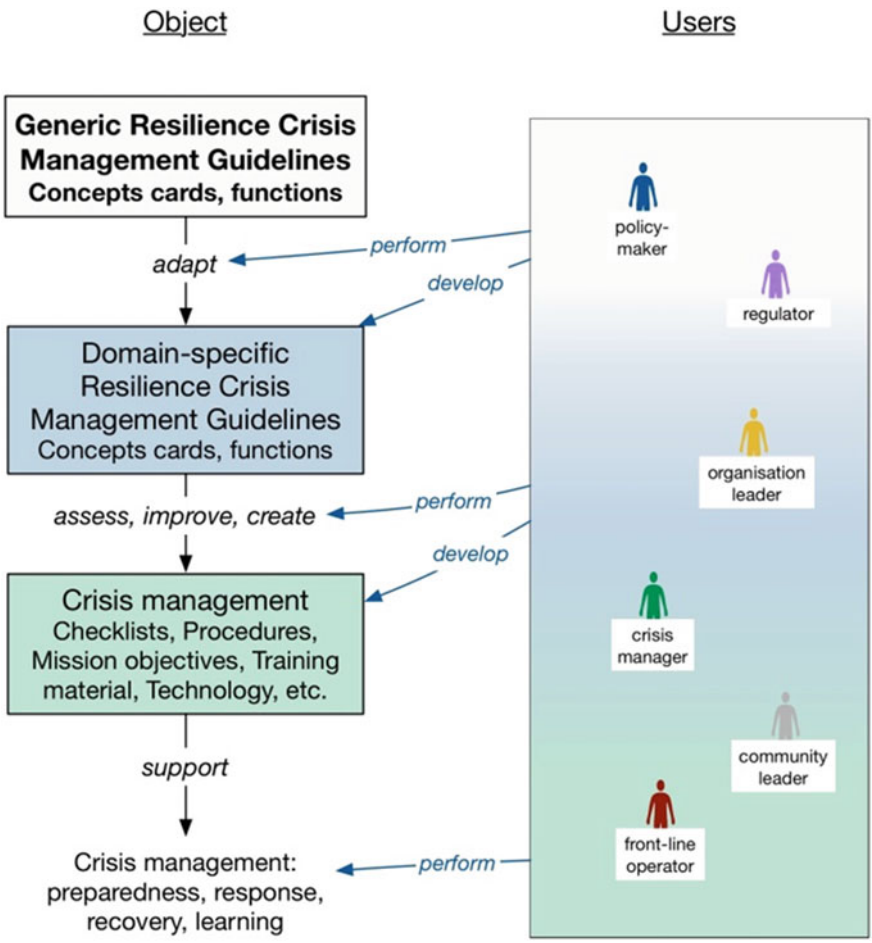


Fig. 2. Applicability of the DARWIN resilience management guidelines for crisis management [30] Copyright 2017 DARWIN

The first three Concept Cards developed include resilience management capabilities such as:

- Ensure that the actors involved in resilience management have a clear understanding of their responsibilities and the responsibilities of other involved actors.
- Noticing brittleness.
- Systematic management of policies involving policy makers and operational personnel for dealing with emergencies and disruptions.

3.4 First Evaluation of Concept Cards

DARWIN adopts a two-phase evaluation framework consisting of an initial preliminary evaluation followed by a set of crisis simulation exercises. This incremental

approach was chosen to maximize the success of the latter exercises, as these could benefit from lessons learnt and insights from the preliminary evaluation.

3.5 Evaluation Framework and Simulation

Serious gaming is being used to evaluate the resilience management guidelines, derived from the generic guidelines. The interacting of the participants from both domains paves the way to reveal emergent behavior, which serves redefining parts of the guidelines. Emergent behavior is likely to be expected due to the complex interdependencies of involved actors and organizations, roles, technical system or resources, between both investigated systems. For example, [31] indicates that in the case of a crisis emergent behavior of type 3 (multiple emergence) is likely to occur. This is characterized by "multiple negative and positive feedback loops" and unpredictability and might lead to even chaotic actions of the participants. In gaming sessions, communication and problem solving capabilities of experts, representing both domains and the first two levels of hierarchy (see Fig. 2), will be examined. The gaming sessions will make use of table-top exercises as well as of available domain-specific simulation environments. During these pilot trials, coordinated solutions for problems, arising from the different scenarios, have to be devised among the participants. The achieved results will be evaluated by means of attribute focused metrics on the one and performance-based metrics on the other hand. Here, the aim is to reflect both perspectives of resilience, resilience engineering, and engineering resilience. Due to its broad scope, represented for example by being multidisciplinary, incorporating operational experience as well as technical, human and organizational performance, or by investigating the various interdependencies in a CIS, resilience engineering though still lacks a standard terminology. As mentioned before, this is also represented concerning the measures to assess resilience. They do not refer to a distinct set of indicators but instead to attributes. With regard to the attribute-based measures, [9] proposes a differentiation between conceptual frameworks, which offer "best practices" and semi-quantitative indices, which offer "expert assessments of different qualitative aspects of resilience". The latter is characterized by evaluating different characteristics that enhance resilience by means of questions, "designed" for experts and represented on a Likert or percentage scale. The answers are later aggregated and weighted to construct an "index of resilience". For example, during the pilot trials of the crisis simulation exercises, described in Chap. 4, questionnaires will be distributed among the participants. For example, with respect to pilot number 4, the questions tackle resilience empowering aspects such as the understanding of interdependencies, capacities, and constraints, or to manage goal conflicts between actors. The questions again are derived from the concept cards. This for example includes cards such as "Understanding roles and responsibilities", "Adapting to expected and unexpected events", "Adapting plans and procedures during crises" or "Noticing brittleness". The results of the questionnaires will be evaluated by experts at a later stage, targeting to reveal problems during the collaboration and to come up with an evaluation with regard to the resilience during the exercises.

With respect to performance-based quantification, an additional simulation of the particular scenario is being developed. A simplified model which is tailored to the

particular scenario is being implemented into a discrete event simulation environment. As stated in [32], the simulation of examples in the network domain addresses schedules, capacities, resources and can reasonably be translated into a discrete event simulation (DES). Reference [32] refers to a "middle level of abstraction" which can be applied to transportation and emergency departments by elements of a DES, such as resources, entities, and flowcharts. Therefore, DES is chosen as an adequate approach to provide insights into the performance of the particular participants.

Generally, the scope of a scenario is defined by boundaries to other systems, the number of involved actors and their roles, technical system, and resources. The level of detail of those elements is also being outlined by the scenario. This conceptual model of the particular scenario aims to represent the relevant operational procedures and resources. The necessary abstraction of the particular elements of the system, to roles and resources, is being carried by the help of domain experts and can be represented by different means such as workflow depictions, i.e., flowcharts. This is done in preparation of implementing the model. The flow-based related modeling approach of the simulation denotes so-called entities as commodities which flow through the system (or network) from node to node [17], such as aircraft (for example moving at the airport) or patients (for example moving around different departments at the hospital). The availability of shared resources defines the particular process durations in which entities are being "kept up" till a further advancing in the network. Until now, TU Braunschweig has implemented the conceptual model of one scenario by means of the discrete event simulation environment SIMEVENTS. Other examples of the application of SIMEVENTS can be found in [33]. The underlying method used in this environment is transaction based modeling. It depends on events which evoke the advancing of an entity in the network, such as the beginning or the end of a process or a rule-based clearance of an involved actor. Currently, additional functionalities for monitoring or configuring (of a particular scenario) by the participants are developed. This, for example, includes functionalities to easily model interdependencies between two systems. Important dependencies represent for example physical, geographical and cyber interdependencies; see [32]. Furthermore, functionalities to enable what-if simulations and the calculation of resilience are developed. The implemented simplified representation of the particular system supports the participants of the pilot trials to monitor the current state of the system and to perform projections of different operational options during the course of the crisis. It calculates the resilience of the system (as proposed in Chap. 1), based on performance indicators selected for each scenario. The indicators are used to represent the system performance. Various time constants of the functionality will be calculated and will represent the performance-based part of the resilience evaluation. Possible performance indicators could be the delay of flights and the capacity of the affected airport or other airports in the network, which are affected as well. Also, the number of casualties among the passengers could be a reasonable performance indicator. In the case of including a number of performance indicators, weighting factors will be applied in order to define the system functionality. The simulation is intended to support the gaming sessions during the debriefing phase. Events arising during the pilot trials, such as the assigning of resources to specific processes, the giving of clearances, or the start or end of processes, will serve as input for the simulation. During the debriefing, the performance of system, as well as

development of resources over time, thus indicating possible bottlenecks in the system, will be discussed. The latter point relates to the identification of brittleness, which plays a crucial role with respect to the resilience engineering concepts, see [3].

The following chapter shortly discusses a scenario and a set of performance indicators as possible representations of the system functionality.

4 Scenario Example

Four scenarios are used in the project [30]. They consist of:

- Aircraft crash in an urban area close to Rome Fiumicino Airport shortly after take-off,
- Total loss of radar information at Rome Area Control Centre caused by a cyberattack,
- Disease outbreak during an incoming flight,
- Collision between an oil tanker and a passenger ferry on the Baltic Sea near Sweden.

This chapter will discuss the latter of the above scenarios.

In this scenario, a small oil tanker and a ferry carrying about 1900 passengers collide in the Baltic Sea near Sweden and a number of injured persons need to be evacuated and transported to hospitals on land while the ferry remains afloat and moves slowly to the nearest suitable harbor.

For the evacuation, multiple sea rescue vessels and helicopters will be used and ambulances will be dispatched to transport the injured persons from the shore to the hospitals.

4.1 Modeling of the Scenario and Simulation

The simulation described in Sect. 3 is adjusted to the scenario so that multiple distribution schemes for the differently injured persons as well as resource allocations of transport vehicles can be tested.

This results in estimates of durations of the evacuation and the usage of hospital capacities or vehicles. Thus, helping a participant to detect the bottlenecks in the system and assess the outcome of a particular course of action.

The following aspects could serve as examples of possible performance indicators to provide a quantitative assessment of resilience in this particular scenario.

- Capacity of the hospitals.
- Duration for transport of all injured persons to hospitals.
- Usage of particular vehicles

4.2 Using the Simulation as a Tool to Evaluate Concept Cards

The simulation supports the evaluation of specific concept cards with its ability to conduct what-if simulations. During after-crisis reviews, they can be used to determine

how well the *adaption relative to events* could be applied by playing out different courses of action. The ability to detect bottlenecks and lack of resources during this example of crisis management helps to *notice brittleness*.

Figure 3 shows the Graphical User Interface (GUI) of the implementation of this example. It depicts the map of the accident and related information on the available transportation resources and the patient allocation parameters.

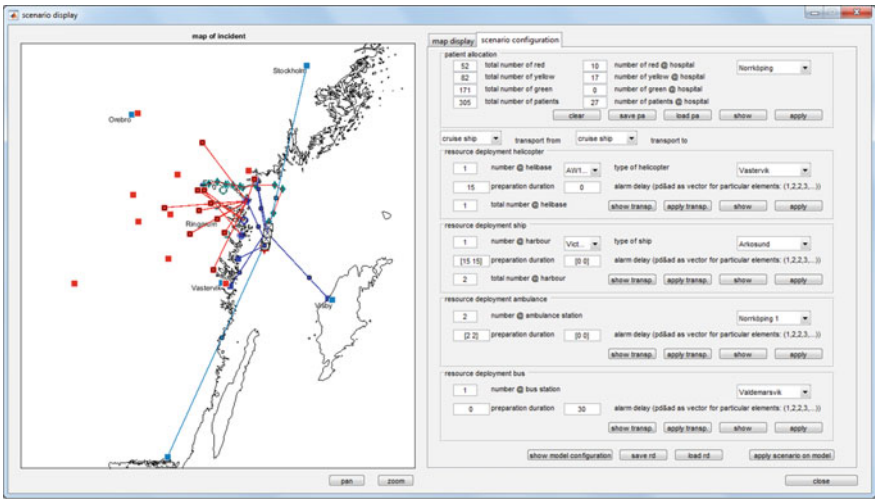


Fig. 3. Graphical user interface

5 Outlook

The next step in the project will be the application of the generic guidelines to a particular domain by operational experts to derive domain-specific guidelines. Afterwards, simulation runs of the different scenarios during the debriefing phases of particular scenarios will be performed, beginning in October.

The results will be presented to a group of external experts called "DARWIN Community of Practitioners (DCoP)". A group of experts which accompanies the consortium during the complete duration of the project and assures that the contents are closely related to practice.

Their feedback will be used to generate a revised version of the resilience management guidelines that will be published by the end of the project.

6 Summary

This conceptual paper presents first results of generic resilience management guidelines, developed in the DARWIN project. The guidelines aim to support domain experts to adapt current operational procedures in order to increase resilience. Due to the

general character of the guidelines, they will be applied to the ATM and the healthcare domain. Simulation trials, performed by operational experts in the form of gaming sessions, assess the effectiveness of the adapted operational procedures. Scenarios are developed which emphasize the organizational interdependencies between the two critical infrastructure systems. An evaluation framework is presented that aims to incorporate attribute-based as well as performance-based approaches, contributing to bridge the gap between the fields of engineering resilience and resilience engineering. With regard to engineering resilience, the performance of the system is described by the system's functionality, using designated performance indicators. A simplified conceptual model of each scenario is implemented into a discrete event simulation environment in order to investigate the functionality of the system over the course of a simulation trial. The participants of the debriefing phase of an exercise are supported by the environment by means of monitoring and what if probing to assess their different courses of action. The aim is to improve the resilience of the system by participants reflecting on their actions during crisis management.

Acknowledgements. The authors would like to thank the colleagues of the consortium working on the DARWIN project enabling this paper, especially the members of Ben-Gurion University, KMC and ISS for their contributions during the refinement of the resilience concepts and the definition of test scenarios.

The research leading to the results received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 653289. Opinions expressed in this publication reflect only the authors' view and that the Agency is not responsible for any use that may be made of the information it contains.

References

1. Herrera I, Grotan TO, Woltjer R, Nevhage B, Nilsson S, Trnka J, Adini B, Cohen O, Forsberg R, Jonson CO (2017) Applying resilience concepts in crisis management and critical infrastructures—the DARWIN project. In: Walls L, Revie M, Bedford T (eds) Risk, reliability and safety: innovating theory and practice. Taylor & Francis Group, London
2. Vugrin ED, Turnquist MA, Brown NJK (2014) Optimal recovery sequencing for enhanced resilience and service restoration in transportation networks. Int J Crit Infrastruct 10 (3/4):218–246. https://doi.org/10.1504/ijcis.2014.066356
3. Woods DD (2015) Four concepts for resilience and the implications for the future of resilience engineering. Reliab Eng Syst Saf 141:5–9. http://dx.doi.org/10.1016/j.ress.2015.03.018
4. Henry D, Emmanuel Ramirez-Marquez J (2012) Generic metrics and quantitative approaches for system resilience as a function of time. Reliab Eng Syst Saf 99:114–122. http://doi.org/10.1016/j.ress.2011.09.002
5. Gluchshenko O, Foerster P (2013) Performance based approach to investigate resilience and robustness of an ATM System. In: Tenth USA/Europe air traffic management research and development seminar (ATM2013)
6. Francis R, Bekera B (2014) A metric and frameworks for resilience analysis of engineered and infrastructure systems. Reliab Eng Syst Saf 121:90–103. https://doi.org/10.1016/j.ress.2013.07.004

7. Fürstenau N, Mittendorf M, Shumpei K (2016) Nonlinear dynamics approach for modeling of air traffic disruption and recovery. In: 7th international conference on research in air transportation (ICRAT). Philadelphia, PA, USA
8. Bruneau M, Chang SE, Eguchi RT, Lee GC, O'Rourke TD, Reinhorn AM, Shinozuka M, Tierney K, Wallace WA, von Winterfeldt D (2003) A framework to quantitatively assess and enhance the seismic resilience of communities. Earthq Spectra 19(4):733–752. https://doi.org/10.1193/1.1623497
9. Hosseini S, Barker K, Ramirez-Marquez JE (2016) A review of definitions and measures of system resilience. Reliab Eng Syst Saf 145:47–61. https://doi.org/10.1016/j.ress.2015.08.006
10. SESAR JU (2009) Concept story board. SESAR JU, Brussels, Belgium, 2 June 2009
11. ICAO (2009) Manual on global performance of the air navigation system. Doc 9883, ICAO, Montréal, Canada
12. Pascarella D, Gargiulo F, Errico A, and Filippone E (2016) An analytical approach for optimal resilience management in future ATM systems. In: Novais P, Camacho D, Analide C, El Fallah Seghrouchni A, Badica C (eds) Intelligent distributed computing IX. Studies in computational intelligence, vol 616. Springer, Cham, pp 415–425. http://doi.org/10.1007/978-3-319-25017-5
13. Jung J, Verma SA, Zelinski SJ, Kozon TE, and Sturre L (2015) Assessing resilience of scheduled performance-based navigation arrival operations. In: Eleventh USA/Europe air traffic management research and development seminar (ATM2015)
14. Holling CS (1996) Engineering resilience versus ecological resilience. In: Schulze P (ed) Engineering within ecological constraints. National Academy Press, Washington, DC, pp 31–44
15. Hollnagel E, Woods DD, Leveson N (eds) (2006) Resilience engineering: concepts and precepts. Ashgate, Aldershot
16. Madni AM, Jackson S (2009) Towards a conceptual framework for resilience engineering. IEEE Syst J 3(2)
17. Ouyang M (2014) Review on modeling and simulation of interdependent critical infrastructure systems. Reliab Eng Syst Saf 121:43–60. https://doi.org/10.1016/j.ress.2013.06.040
18. DARWIN (2015) Deliverable D1.1—Consolidation of resilience concepts and practices for crisis management. Technical report. http://h2020darwin.eu/project-deliverables. Accessed 18 Jan 2017
19. Baroud H, Barker K, Ramirez-Marquez JE, Rocco CM (2015) Inherent costs and interdependent impacts of infrastructure network resilience. Risk Anal 35(4):642–662. https://doi.org/10.1111/risa.12223
20. Ouyang M, Dueñas-Osorio L (2014) Multi-dimensional hurricane resilience assessment of electric power systems. Struct Saf 48:15–24. https://doi.org/10.1016/j.strusafe.2014.01.001
21. Nan C, Sansavini G (2017) A quantitative method for assessing resilience of interdependent infrastructures. Reliab Eng Syst Saf 157:35–53. https://doi.org/10.1016/j.ress.2016.08.013
22. Hollnagel E (2011) Prologue: the scope of resilience engineering and epilogue: RAG—the resilience analysis grid. In: Hollnagel E, Pariès J, Woods DD, Wreathall J (eds) Resilience engineering in practice: a guidebook. Ashgate, Aldershot
23. Woltjer R, Pinska-chauvin E, Laursen T, Josefsson B (2015) Towards understanding work-as-done in air traffic management safety assessment and design. Reliab Eng Syst Saf 141:115–130. https://doi.org/10.1016/j.ress.2015.03.010
24. Blom HAP, Bouarfa S (2015) Resilience. In: Cook A, Rivas D (eds) Complexity science in air traffic management. Ashgate. ISBN 978-1-4724-6037-0

25. Horizon 2020 (2015) Horizon 2020 work program 2014–2015, 14. Secure societies—protecting freedom and security of Europe and its citizens—revised. Consolidated version following European Commission Decisions C (2015)2453 of 17 April 2015, European Commission, Brussels, Belgium
26. Hollnagel E (2009) The four cornerstones of resilience engineering. In: Nemeth C, Hollnagel E, Dekker S (eds) Resilience engineering perspectives: preparation and restoration, vol 2. Ashgate, Burlington (VT), USA, pp 117–133
27. Herrera I (2016) Resilience engineering and indicators of resilience. In: IRGC, resource guide on resilience. EPFL international risk governance center, Lausanne, Switzerland
28. DARWIN (2016) Deliverable D1.2—Evaluation and selection of resilience concepts and approaches. Technical report. http://h2020darwin.eu/project-deliverables. Accessed 18 Jan 2017
29. DARWIN (2016) Deliverable D1.3—Practitioner and academic requirements for resilience management guidelines. Technical report. http://h2020darwin.eu/project-deliverables. Accessed 18 Jan 2017
30. DARWIN (2017) Deliverable D2.1—Generic resilience management guidelines
31. Bouarfa S, Blom HAP, Curran R, Everdij MHC (2013) Agent-based modeling and simulation of emergent behavior in air transportation. J Complex Adapt Syst Model 1:15. https://doi.org/10.1186/2194-3206-1-15
32. Rinaldi SM (2004) Modeling and simulating critical infrastructures and their interdependencies. In: Proceedings of the 37th Hawaii international conference on system sciences. Big Island, HI, USA, p 8. https://doi.org/10.1109/hicss.2004.1265180
33. Alsebae AA, Leeson MS and Green RJ (2013) SimEvents—based modeling and simulation study of stop—and—wait protocol. In: Proceedings of international conference on modelling, identification and control (ICMIC). IEEE, New York, NY, USA

Trajectory Mathematical Distance Applied to Airspace Major Flows Extraction

D. Delahaye[1(✉)], S. Puechmorel[1], S. Alam[2], and E. Feron[3]

[1] OPTIM, ENAC-LAB, French Civil Aviation University, 7 Avenue Edouard Belin, 31055 Toulouse, France
{daniel.delahaye, stephane.puechmorel}@enac.fr
[2] Nanyang Technological University, Singapore, Singapore
sameeralam@ntu.edu.sg
[3] Georgia Technology Institute, Atlanta, USA
eric.feron@aerospace.gatech.edu

Abstract. In this paper, the problem of aircraft trajectories representation and analysis is addressed. In many operational situations, there is a need to have a value expressing how trajectories are close to each other. Some measures have been previously defined, mainly for trajectory prediction applications, all of them being based on distance computations at given positions in space and time. The approach presented here is to consider the trajectory as a whole object belonging to a functional space and to perform all computations in this space. An efficient algorithm for computing mathematical distance between trajectories is then presented and applied to the major flows extraction in the French airspace.

Keywords: Distance · Trajectory · Homotopy · Hierarchical clustering

1 Introduction

Future Air Traffic Management relies, in part, on the use of decision support tools (DST) to provide improved service to the user community under increasing traffic demand. Furthermore, this improvement has to be validated by the mean of system performance metrics such as complexity, robustness, and capacity. As aircraft fly 4D trajectories, there is a strong need to quantify the associated trajectory accuracy in order to validate aircraft models and trackers. Such validation is usually based on a comparison between the actual trajectory and a reference by the mean of a trajectory distance. This last point is the key element of the whole process. Such trajectory distance is still needed for ATM applications and the goal of this paper is to present a new trajectory distance based on rigorous mathematical concepts. Although trajectories are well understood and studied, relatively little investigation on the precise comparison of trajectories is presented in the literature. A key issue in performance evaluation of ATM decision support tools (DST) is the distance metric that determines the

Electronic Navigation Research Institute, *Air Traffic Management and Systems III*, Lecture Notes in Electrical Engineering 555,
https://doi.org/10.1007/978-981-13-7086-1_4

similarity of trajectories. Most existing measures [1, 2] compute a mean distance of the corresponding positions of two equal duration trajectories. Supplementary statistics such as variance, median, minimum, and maximum distances are also suggested to extend the description of similarity. In [3], Needman proposed an alignment based distance metric that reveals the spatial transition and temporal shift between the given trajectories, and introduced an area based metric that calculates the total enclosed area between trajectories using trajectory intersection.

One main disadvantage of the existing approaches is that they are all limited to the equal duration (lifetime) trajectories. By duration, we refer to the number of coordinate points that constitute the trajectory. These coordinates are sampled at different instances. Since the existing measures depend on the mutual coordinate correspondences, they cannot be applied to trajectories that have different durations. Conventional distance measures assume that the temporal sampling rates of trajectories are equal. They do not cope with the uneven sampling instances, i.e., varying temporal distance between the coordinates. Therefore, there is a need to develop other alternatives that can effectively measure the difference between unrestricted trajectories. There are a lot of ATM applications where such distance between trajectories is needed.

Aircraft Model Inference

All aircraft models are based on ODEs (Ordinary Differential Equation), including tabular ones (see Fig. 1).

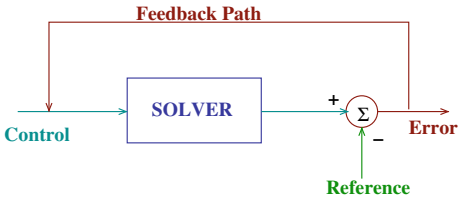


Fig. 1. General simulation model. The solver generate trajectories as close as possible to the reference trajectory thanks to the feedback path

The aircraft model inference consists in answering the following question: **Given a parametrized model and a goal trajectory, can we infer the best parameter values?** A model can be viewed as a mapping from the control space into the trajectory space. The way to answer the previous question is then given by the closest model to the goal trajectory (see Fig. 2).

In order to find the closest model in this trajectory space, a reliable trajectory distance is needed. The model inference problem has to solve the accuracy-smoothness dilemma: **Over-fitted models are generally poor predictors**. The previous construction gives the shortest path (and thus the distance) between the goal trajectory and the trajectory set which can be synthesized by the model.

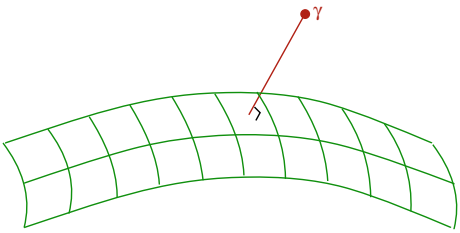


Fig. 2. Finding the best model from a given class. The green "grid" represents such class produced by the model and γ is the goal aircraft trajectory

Trajectory Prediction

Air traffic management research and development has provided a substantial collection of decision support tools that provide automated conflict detection and resolution [4–6], trial planning [7], controller advisories for metering and sequencing [8, 9], traffic load forecasting [10, 11], and weather impact assessment [12–14]. The ability to properly forecast future aircraft trajectories is central in many of those decision support tools. As a result, trajectory prediction (TP) and the treatment of trajectory prediction uncertainty are an active of research and development (e.g., [15–19]).

Accuracy of TP is generally defined as point spatial accuracy (goal attainment) or as trajectory following accuracy. The last one can be rigorously defined by the mean of trajectory space. The first one is a limit case of the second by adding a weight function in the energy functional.

When we refer to trajectory prediction errors for a specific DST, we are typically comparing the predicted trajectory for a specific DST to the actual trajectory to be experienced by an aircraft. Discrepancies between these two types of trajectories typically affect the performance of the DST.

Radar Tracker Evaluation

The goal of a radar tracker is to eliminate the residual noise coming from the radars. It is a key element of the ATM system and its accuracy is one of the factors which determines the separation norm. In order to validate such trackers, an exact reference trajectory is generated and perturbed Gaussian noise. This perturbed trajectory is then used as input of the tested tracker. The tracker generates an estimated trajectory which is compared to the reference trajectory. In order to do such comparison, a reliable trajectory distance is also needed.

Alternative Route Synthesis

Airspace congestion is related to aircraft located in the same area during the same period of time. Then, when congestion has to be minimized, algorithms have to separate aircraft in time (slot allocation), in space (route allocation) or both (bi-allocation). When route allocation is investigated, associated algorithms need alternative routes set in order to spread the traffic on them. A route is said to be alternative to another if it is different enough based on a trajectory distance.

Major Flows Definition

When radar tracks are observed on a radar screen over a long period of time in a dense area, it is very easy to see major flows connecting major airports. The expression "major flows" is often used but never rigorously defined. Based on an exact trajectory distance and a learning classifier, it is possible to answer the following questions: **Given a set of observed trajectories, can we it into "similar" trajectory classes? If yes, classes with highest number of elements will rigorously define the major flows. Given those classes and a new trajectory, can we tell if it belongs to a major flow and which one?** The principle of the major flows definition is to use shape space to represent trajectory shapes as points and to use a shape distance (the shape of a trajectory is the path followed by an aircraft, that is the projection in the 3D space of its 4D trajectory. The speed on the path has no impact).

Major flows have not to be confused with highest density in the airspace. As a matter of fact, some approaches for major flows extraction consider the accumulated traffic in the airspace and build a kind of density map for which the highest areas are considered as major flows. This approach may be completely false as shown on Fig. 3. On this figure, four artificial trajectories share a common highest density area but, as it can be seen on the figure, no aircraft is flying this "high density trajectory".

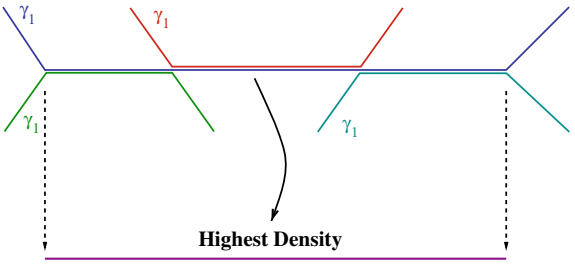


Fig. 3. Four trajectories γ_1, γ_2, γ_3, and γ_4 are sharing a common central straight line. One can identify an average of two aircraft at each point of this line and only one in the other segments. If we compute the highest density associated with those trajectories, we will extract the central segment which is flown by no aircraft

Another approach consists in extracting major flows on a set of trajectories thanks to an efficient HMI and a bundling algorithm [20]. The results produced by this kind of algorithm are quite similar to the ones presented in this paper but it is done manually and we propose to do it automatically.

As it has been shown in this section, mathematical distance between trajectories is a real need for many ATM applications. The next section of this paper presents some current trajectory distance metrics and shows their limitations. The third part gives a

detailed mathematical description of our new trajectory distance. The fourth part introduces the associated algorithms implementation. Finally, the fifth part presents the application of such algorithms to the major flows extraction of the French airspace with $\simeq$ 8000 trajectories.

2 Mathematical Distance Between Trajectories

2.1 Introduction

In a vector space, distances are very well defined. If we consider two points $\vec{P}_1 = (x_1, y_1)^T$ and $\vec{P}_2 = (x_2, y_2)^T$ in a plane (see Fig. 4), the distance between them can be computed with the classical formula of the euclidean distance (see Fig. 4):

$$d(\vec{P}_1, \vec{P}_2) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{1}$$

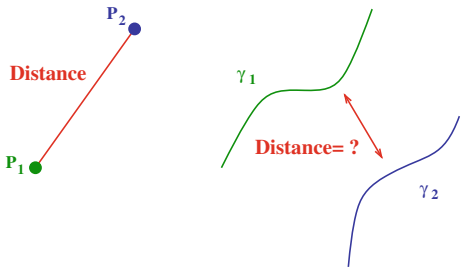


Fig. 4. On the left, two points $\vec{P}_1$ and $\vec{P}_2$ has been drawn for which the classical Euclidean distance is shown in red. On the right, two trajectories are drawn (γ_1, γ_2) for which one want to determine a mathematical distance

What is the distance, if now the points $\vec{P}_1$ and $\vec{P}_2$ are replaced by two trajectories γ_1 and γ_2? Trajectories are infinite dimension mathematical objects which are not easy to manipulate. We are looking for a mathematical distance between trajectories (γ_1 and γ_2) with the following properties:

- $d(\gamma_1(t), \gamma_2(t)) = 0 \;\Rightarrow\; \gamma_1(t) = \gamma_2(t)$
- $d(\gamma_1(t), \gamma_2(t)) = d(\gamma_2(t), \gamma_1(t))$
- $d(\gamma_1(t), \gamma_2(t)) + d(\gamma_2(t), \gamma_3(t)) \geq d(\gamma_1(t), \gamma_3(t))$

One of the main results of this paper is the establishment of such mathematical distance between aircraft trajectories.

2.2 Current Trajectory Distances

An aircraft trajectory is a time sequence of coordinates representing the aircraft path over a period of time and may be represented by a : $T = \{(x_1, y_1, z_1, t_1), (x_2, y_2, z_2, t_2), \ldots, (x_N, y_N, z_N, t_N)\}$ where N is the duration.The simplest metric used for computing the distance between a pair of trajectories is the mean of coordinate distance, which is given as

$$m_1(T^a, T^b) = \frac{1}{N}\sum_{n=1}^{N} d_n \tag{2}$$

where the displacement between the positions is calculated using the Cartesian distance

$$d_n = [(x_n^a - x_n^b)^2 + (y_n^a - y_n^b)^2 + (z_n^a - z_n^b)^2]^{\frac{1}{2}} \tag{3}$$

Note that, the mean of distance metric makes three critical assumptions:

1. the durations of both trajectories are the same: $N^a = N^b = N$
2. the coordinates are synchronized $t_n^a = t_n^b$
3. the time sampling rate is constant $t_{n+1}^a - t_n^a = t_{m+1}^a - t_m^a$

It is evident that the mean of distance is very sensitive to the partial mismatches and cannot deal with the distortions in time. To provide more descriptive information, the second-order statistics such as median, variance, minimum and maximum distance may be incorporated. For instance variance trajectory distance is defined as

$$m_2(T^a, T^b) = \frac{1}{N}\sum_{n=1}^{N}(d_n - m_1(T^a, T^b))^2 \tag{4}$$

Although these statistics supply extra information, they inherit (even amplify) the shortcomings of the ordinary mean of distance metric m_1. Besides, none of the above metrics is sufficient enough by itself to make an accurate assessment of the similarity. Another possible candidate for the distance between two trajectories γ_1 and γ_2 will simply be to take the supremum norm (see Fig. 5), that is:

$$d_\infty(\gamma_1, \gamma_2) = \sup_{s\in\mathbb{R}} \|\gamma_1(s) - \gamma_2(s)\| \tag{5}$$

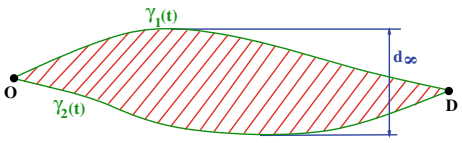


Fig. 5. Supremum norm distance

Since γ_1 and γ_2 are constant outside bounded intervals of $\mathbb{R}$, the supremum is well defined. However, this metric is not sensitive to global properties of curves. In Fig. 6, the curves γ_1 and γ_2 are at the same distance from γ_3 but have very different shapes.

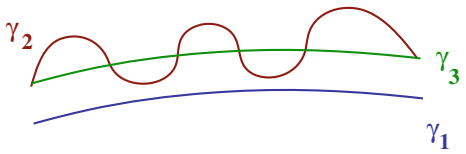


Fig. 6. Different trajectories with same sup distance

From an operational point of view, γ_1 is just a shifted copy of γ_3 while γ_2 will probably not be realistic. For trajectories γ_1, γ_2 with the same origin–destination pairs, $\gamma_1 - \gamma_2$ can be defined as a compactly supported mapping and an area distance between trajectories can be defined (see Fig. 7):

$$d_2(\gamma_1, \gamma_2) = \left(\int_{\mathbb{R}} \|\gamma_1(t) - \gamma_2(t)\|^2 dt \right)^{\frac{1}{2}} \tag{6}$$

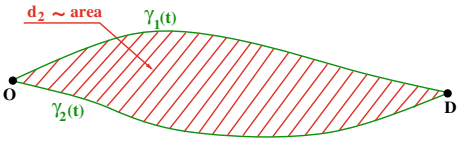


Fig. 7. Area distance between trajectories with the same origin–destination pairs

An extension of such area based distance metric is proposed in [3]. The crossing points of two paths (where $T^a(p_i) = T^b(p_j)$) are used to define regions $Q_j, j = 1, \ldots, J$ between trajectories (see Fig. 8).

For each region, a polygon model is generated and the enclosed area is found by the parameterized shape. The resulting distance is given by:

$$m_3(T^a, T^b) = \sum_{n=1}^{N} area(Q_j) \tag{7}$$

This metric can handle more complex trajectories, however it is sensitive to entanglements of the trajectory, it discards the time continuity, and fails to distinguish two trajectories in opposite directions. Furthermore, it is not adapted to 3D trajectories. In order to introduce our new mathematical distance between trajectories, one must first give some representation definitions.

2.3 Representation

Since objects of interest are aircraft trajectories, we need to find an adapted framework in which computations may be made on trajectories as a whole. There are basically two ways of understanding what a trajectory is:

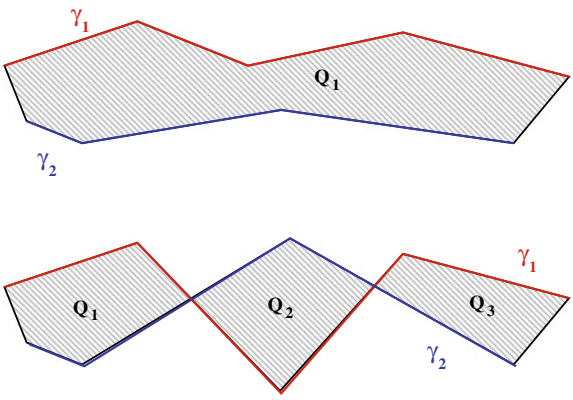


Fig. 8. Area distance between trajectories with or without crossings

- The time/position approach. In this case, a trajectory can be represented as a mapping from a bounded interval of $\mathbb{R}$ (the life time of the trajectory) to $\mathbb{R}^3$ or $\mathbb{R}^6$ depending on whether speed is part of the data or not. Since there is an explicit dependence on time, there is a need to calibrate trajectories with time shifts for all applications involving trajectory comparison. We will see in the following that there is nevertheless a mean of reducing the problem so that origin of time is automatically calibrated.
- The shape approach. Here, trajectories are understood as paths and time is not directly relevant (from a more formal point of view, we take the quotient of the trajectories understood as mappings by the group of diffeomorphisms acting on time), so that we may assume that the underlying life time of trajectories is always the interval [0, 1]. This is the right framework for dealing with major flows estimation.

2.4 Trajectories as Mappings

We will assume in the following that trajectories are given as mappings from a compact interval of $\mathbb{R}$ to $\mathbb{R}^3$. The case of mappings from $\mathbb{R}$ to $\mathbb{R}^6$ (that is with explicit speed, for example as given by radar tracking filter) can be derived with minor changes and thus will not be addressed here. Since physical trajectories are smooth unless there is a perturbing noise, we made the choice to take all trajectories as smooth mappings from a compact interval of $\mathbb{R}$ to $\mathbb{R}^3$. The first point to deal with is the necessary calibration of the origin of time for trajectories comparison. Remembering that there is an explicit dependence on time, one cannot just time shift one trajectory in time in order to make it coincident with another in order to compare them: this will result in forgetting distortions in time, that is trajectories with the same range (as mappings) but different positions at different times may become equal. The previous family of semi-distances has nice features because of the scaling ability, but since it is not a single metric, it is

difficult to use standard algorithms based on distances (for example, classification algorithms). There is thus a need for another definition of proximity between trajectories that will yield a single value while capturing interesting global characteristics. Before introducing our homotopic distance between trajectories one must introduce how do we cope with time difference between trajectories.

2.5 Parametrization Invariance

The parametrization invariance is a very feature: the shape of an object is independent on the way its contour is followed. In its seminal paper, Kendall introduced the notion of shape manifold [21]: the originality of its work was the use of a differential geometry setting to implicitly enforce the invariance with respect to shape-preserving transformations. Curves were represented as finite sequences of distinguished points, called landmarks. Some related algorithms were eventually designed for air traffic analysis applications. In a study conducted by the Mitre corporation on behalf of the Federal Aviation Authority (FAA) [22], a spectral clustering algorithm was applied to sampled trajectories. Only the distance between landmarks was used, no invariance under Euclidean transformations was imposed. Due to the high computational complexity, a random projection was first applied to the data in order to reduce the dimension of the samples. The most important limitation of this approach is that the shape of the trajectories is not taken into account when applying the clustering procedure unless a re-sampling procedure based on arc-length is applied: changing the time parametrization of the flight paths will induce a change in the classification. Methods based on times series as surveyed in [23, 24] are appealing, but are inadequate for the present application. Finally, functional data statistics [25, 26] provides a powerful framework, still lacking the re-parametrization invariance. In this section, flight paths will be modeled as points in an infinite dimensional riemanian manifold. In such space, each point is considered as a full trajectory. An intrinsic notion of distance exists in this setting and is defined as the infimum of the length of the paths connecting two points. Having this at hand allows the use of standard, distance based algorithms like k-means, k-medoids or hierarchical clustering.

2.6 Trajectories Registration

A flight path may be modeled as a smooth curve $\gamma : [a, b] \to \mathbb{R}^3$ that maps a time to a position (at each time one can determine the associated position). Two distinct trajectories γ_1, γ_2 are most of the time defined on different time intervals, say $[a_1, b_1]$ (resp. $[a_2, b_2]$) for γ_1 (resp. γ_2), making the comparison between them quite awkward. This issue is well known in the field of functional data statistics as the registration problem. In a formal sense, it amounts to find a pair (ϕ_1, ϕ_2) of strictly increasing diffeomorphisms $\phi_1 : [0, 1] \to [a_1, b_1]$, $\phi_2 : [0, 1] \to [a_2, b_2]$ such that the transformed curves $\gamma_1 \circ \phi_1$, $\gamma_2 \circ \phi_2$, defined on the common interval [0, 1], are as similar as possible. The special problem instance:

$$\min_{\phi_1,\phi_2} \int_0^1 \|\gamma_1 \circ \phi_1(t) - \gamma_2 \circ \phi_2(t)\|^2 dt \tag{8}$$

gives the Fréchet distance between γ_1, γ_2. Computing the optimal ϕ_1, ϕ_2 is a difficult task, unless the curves are assumed to be polygonal. Furthermore, as mentioned in [26], the registration procedure may remove some important features from the data: the extra degree of freedom provided by the so-called warping functions ϕ_1, ϕ_2 may have the detrimental effect of registering curves that does not need it [27]. A discrete relative to the Fréchet distance is known as dynamic time warping and may be used to compare sampled sequences. Nevertheless, it suffers from the same drawback (two trajectories having different time sampling will be considered as different which is not relevant for our application).

3 Algorithm

3.1 Distance Based on Homotopy Between Trajectories

In order to compute the distance between two trajectories (γ_1, γ_2), a time regularization is first applied to both trajectories. Then, an homotopy Φ between γ_1, γ_2 is built for which a discrete grid is built in order to minimize its associate energy. Let a be the origin of the trajectory γ. We have: $\gamma(t) = a + \int_0^t \gamma'(s)ds$, so a couple $(a, \gamma') \ (\in \mathcal{W})$ with γ' compactly supported defines a trajectory. An homotopy between (a, γ_1') and (b, γ_2') is a continuous mapping $\Phi : [0, 1] \to \mathcal{W}$ such that $\Phi(0) = (a, \gamma_1')$, $\Phi(1) = (b, \gamma_2')$. Intuitively, an homotopy is a continuous deformation between two trajectories (see Fig. 9).

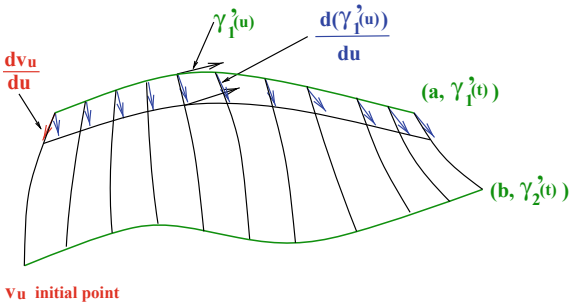


Fig. 9. Smooth path between two curves

The deformation energy between γ_1 and γ_2 is linked to the distance between those trajectories and can be computed with the energy of the homotopy between γ_1 and γ_2:

$$E(\Phi) = \int_0^1 \left(\left\| \frac{\partial v_u}{\partial u} \right\|^2 + \int_{\mathbb{R}} \left\| \frac{\partial \gamma_u'(s)}{\partial u} \right\|^2 ds \right) du \tag{9}$$

where $\frac{\partial v_u}{\partial u}$ is the normal vector (see Fig. 9). In the case of a linear homotopy (which is the simplest one), the associated energy is given by:

$$\Phi_0(u,s) = \left([(1-u).a + u.b], \left[(1-u).\gamma_1'(s) + u.\gamma_2'(s)\right]\right) \tag{10}$$

$$\boxed{E(\Phi_0) = \|b-a\|^2 + \int_{\mathbb{R}} \left\|\gamma_1'(s) - \gamma_2'(s)\right\|^2 ds} \tag{11}$$

There is an infinite number of homotopies shifting from γ_1 to γ_2 and our problem is to find the one with the minimum energy. The deformation energy of a shape homotopy is obtained with a slight change in the expression for trajectories.

$$\boxed{E(\Phi) = \int_0^1 \left(\left\| \frac{\partial v_u}{\partial u} \right\|^2 + \int_{\mathbb{R}} \left\| \frac{\partial \gamma_u'(s)}{\partial u} \right\|^2 . \left\|\gamma_u'(s)\right\| ds \right) du} \tag{12}$$

In order to compute such energy, a grid on the homotopy connecting γ_1 to γ_2 is built, as shown in Fig. 10.

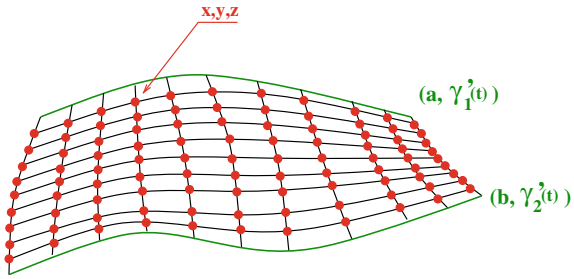


Fig. 10. Structure of the grid used for homotopy energy minimization. Each red point has 2D coordinates (x, y) for which an optimization algorithm is used for searching the z coordinates which minimize the energy of the homotopy connecting γ_1 to γ_2

This grid helps us to compute an approximation of summation used in $E(\Phi)$. The optimization algorithm is searching for the z coordinate of each grid point in order to minimize $E(\Phi)$. One can show that such problem is convex (from the optimization theory point of view) and gradient like method can be used to find the associated minimum (quadratic programming has been used to solved this problem efficiently).

3.2 Clustering Algorithm

We consider a set of trajectories extracted from the radar track database of a given airspace. Having defined a distance between trajectories, one can gather together such trajectories in order to create clusters by using an adaptive clustering algorithm (hierarchical clustering). Such a clustering algorithm aims to partition the trajectory set

into K clusters. To reach this goal, trajectories are considered as points in the associated metric space (blue points on Fig. 11). This algorithm uses two parameters, d_{min} and d_{max}, to respectively fuse clusters and create new clusters. Initially, each trajectory is considered as the centroid of a cluster. We then apply the three following principles one after the other:

- if two centroids are at a distance lower than d_{min}, we fuse them into a single cluster, of which the resulting centroid is the barycenter of the two initial centroids. The barycenter is computed the following way:

$$\mu_i = \frac{1}{N} \sum_{i=1}^{i=N} \gamma_i \tag{13}$$

- a new individual is aggregated to a cluster if its distance from the closest centroid is lower than d_{max} and in this case we compute the new global centroid.
- Otherwise, we create a new cluster containing the single trajectory.

The number of clusters is also a result of the algorithm. An example of clustering result is given in Fig. 11. For each cluster c, one can compute also the following features:

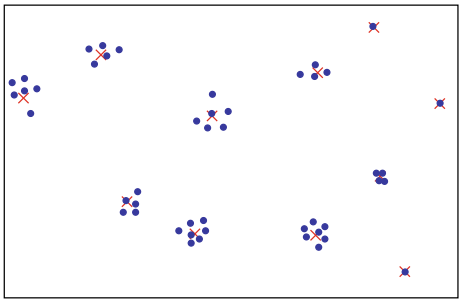


Fig. 11. In this example, the algorithm finds eleven clusters with different features

- Number of trajectories in the cluster N_c;
- Mean trajectory which is the cluster centroid (γ_c);
- Dispersion of the cluster:

$$\sum_{i=1}^{N_c} \left\| \gamma_j - \gamma_c \right\|^2 \tag{14}$$

where $\|.\|$ is the norm in the trajectory metric space.

4 Results

The algorithm has been applied to the French airspace with a heavy traffic of 8764 flights corresponding to June 27, 2015. This traffic has been extracted from the radar track database. Each trajectory being sampled every ten seconds, one has to manipulate about five million points, each them having four coordinates (x, y, z, t). The traffic is represented in Fig. 12.

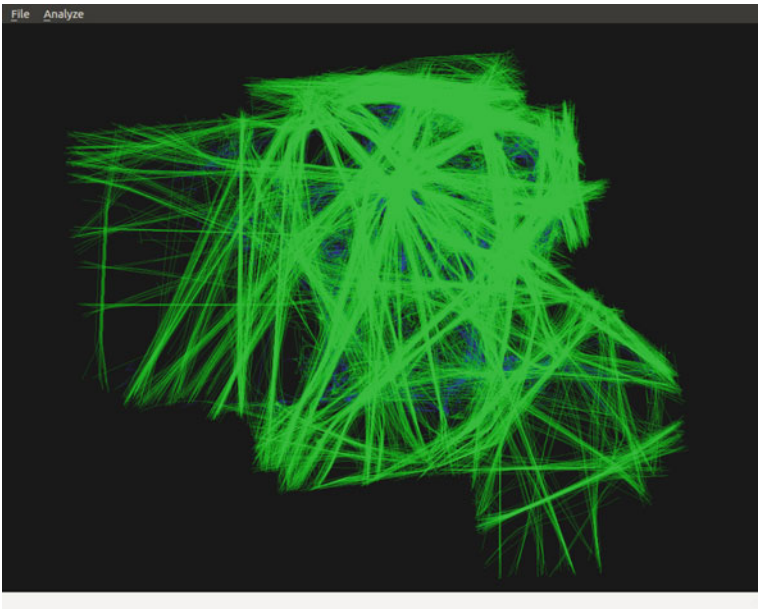


Fig. 12. Radar tracks of the France traffic of June, 27, 2015. This traffic corresponds to the upper airspace

The initial step consists in computing the trajectory registration in order to remove the absolute time dependency. Then, the d_{min} and d_{max} distance have been fixed in order to apply the hierarchical clustering algorithm. Those distances have been established during experimentations ($d_{min} = 2.30$ $d_{max} = 4.5$). Based on those distances, the hierarchical clustering algorithm has extracted 47 major flows for this day as shown in Fig. 13. The algorithm has been implemented into C++ and executed on a Intel-Xeon3.2Ghz PC computer with an executing time of 30 s for extraction the major flow associated to the 8764 flights of June 27, 2015.

5 Conclusion

This paper has shown that distance between trajectories is a real need for ATM applications. Several ways of computing distances on the space of trajectories have been presented with their limitations. This family of metrics, scale based, is mainly

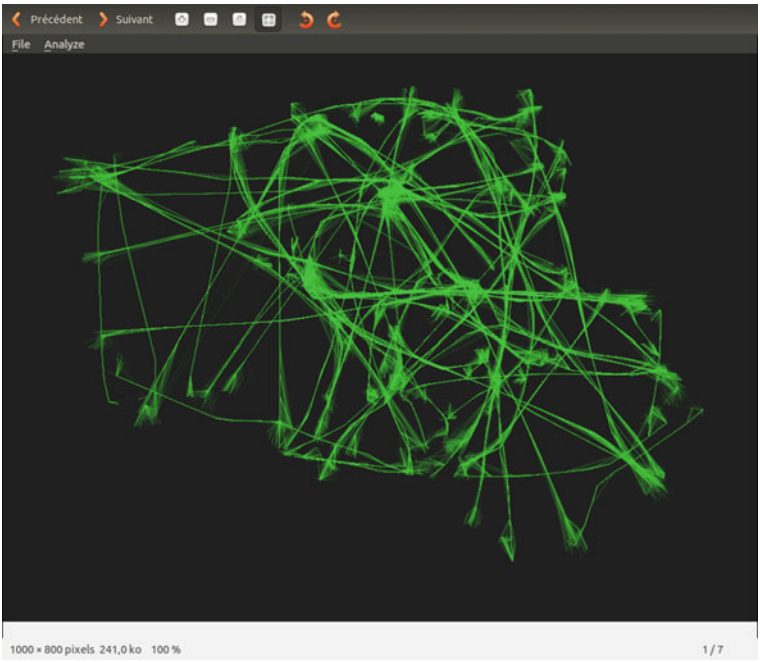


Fig. 13. Major flows extracted from the hierarchical algorithm

useful for descriptive purpose and to quickly analyze a set of trajectories (for example, as a tool complementary to standard descriptive statistics).

Then, a concept originating from functional analysis has been introduced in order to work directly on trajectories as a whole. For more in depth analysis of trajectories, a new kind of distance has been introduced that is based on the energies of homotopies joining pairs of trajectories. This yield to a variational problem that cannot be solved directly, but may be reduced to a quadratic optimization problem. This kind of distance allows computations to be done on trajectories understood as shapes (or embeddings).

Based on this new distance between trajectories, an efficient major flows extraction has been developed with nice results on the French airspace for several thousands of trajectories.

References

1. Jaynes C, Webb S, Steele R, Xiong Q (2002) An open development environment for evaluation of video surveillance system. In: Proceedings of PETS, Copenhagen
2. Senior A, Hampapur A, Tian Y, Brown L, Pankanti S, Bolle R (2001) Appearance models for occlusion handling. In: Proceedings of PETS, Hawai, Kauai
3. Needham C, Boyle R (2003) Performance evaluation metrics and statistics for positional tracker evaluation. In: Proceedings of ICVS, Graz, Austria, pp 278–289

4. Brudnicki DJ, McFarland AL (1997) User request evaluation tool (uret) conflict probe performance and benefits assessment. In: FAA/Eurocontrol ATM Seminar Saclay France
5. Erzberger H, Paielli RA, Isaacson DR, Eshowl MM (1997) Conflict detection in the presence of prediction error. In: FAA/Eurocontrol ATM seminar saclay France
6. Vink A (1997) Eatchip medium term conflict detection: Part 1 eatchip context. In: FAA/Eurocontrol ATM Seminar Saclay France
7. McNally BD, Bach RE, Chan W (1998) Field test evaluation of the ctas conflict prediction and trial planning capability. In: AIAA-1998-4480 AIAA GNC Conference Boston MA
8. Coppenbarger R, Lanier R, Sweet D, Dorsky S (2004) Design and development of the enroute descent advisor (eda) for conflict-free arrival metering. In: AIAA-2004-4875 AIAA GNC conference. Providence, RI
9. Swensen HN et al (1997) Design and operational evaluation of the traffic management advisor at the forth worth air route traffic control center. In: FAA/Eurocontrol ATM seminar Saclay France
10. Masalonis A et al (2004) Using probabilistic demand prediction for traffic flow management decision support. In: AIAA-2004-5231 AIAA GNC Conference Providence RI
11. Meckiff C, Chone R, Nicolaon JP (1998) The tactical load smoother for multi-sector planning. In: FAA/Eurocontrol ATM Seminar Orlando FL
12. Evans J et al (2003) Reducing severe weather delays in congested airspace with weather support for tactical air traffic management. In: FAA/Eurocontrol ATM seminar BUdapest Hungary
13. Kirk DB et al (2004) Problem analysis resolution and ranking (parr) development and assessment. In: FAA/Eurocontrol ATM Seminar Santa Fe NM
14. Sud V et al (2001) Air traffic flow management collaborative routing coordination tools. In: AIAA-2001-4112 AIAA GNC conference Montreal PQ
15. Mondoloni S, Paglone SM, Green S (2002) Trajectory modeling accuracy for air traffic management decision support tools. In: ICAS Congress Toronto ON
16. Ryan HF, Paglione M, Green S (2004) Review of trajectory accuracy methodology and comparison of error measure metrics. In: AIAA-2004-4787 AIAA GNC conference Providence RI
17. Slattery R, Zhao Y (1997) Trajectory synthesis for air traffic automation. AIAA J Guid Control Dyn 20:232–238
18. Mondoloni S, Bayraktutar I (2005) Impact of factors, conditions and metrics on trajectory prediction accuracy. In: Proceeding of the sixth Eurocontrol-FAA R&D Seminar, Baltimore. Eurocontrol-FAA
19. Swierstra S, Green S (2003) Common trajectory prediction capability for decision support tools. In: FAA/Eurocontrol ATM Seminar BUdapest Hungary
20. Marzuoli A, Hurter C, Féron E (2012) Data visualization techniques for airspace flow modeling. In: CIDU 2012, conference on intelligent data understanding. Boulder, United States, pp 79–86
21. Kendall David G (1984) Shape manifolds, procrustean metrics, and complex projective spaces. Bull Lond Math Soc 16(2):81–121
22. Enriquez M (2013) Identifying temporally persistent flows in the terminal airspace via spectral clustering. In: FAA-Eurocontrol, editor, ATM seminar, vol. 10
23. Warren Liao T (2005) Clustering of time series data-a survey. Pattern Recognit 38:1857–1874
24. Rani S, Sikka G (2012) Recent techniques of clustering of time series data: A survey. Int J Comput Appl 52 (15):1–9. Full text available

25. Ferraty F, Vieu P (2006) Nonparametric functional data analysis: theory and practice. Springer series in statistics. Springer, Berlin
26. Ramsay J, Silverman BW (2006) Functional data analysis. Springer series in statistics. Springer, New York
27. Ramsay JO, Li X (1998) Curve registration. J R Stat Soc: Ser B (Statistical Methodology) 60 (2):351–363

Aircraft Safety Analysis Using Generalized Polynomial Chaos

J. Diepolder(✉), P. Piprek, B. Grüter, T. Akman, and F. Holzapfel

Institute of Flight System Dynamics, Technical University of Munich, Garching, Germany
{Johannes.Diepolder, Patrick.Piprek, Benedikt.Grueter, Tugba.Akman, Florian.Holzapfel}@tum.de

Abstract. In this paper we investigate the application of generalized polynomial chaos (gPC) for optimal control based aircraft safety assessment with parameter uncertainties. The approach is based on the formulation of an appropriate optimal control problem to obtain worst case inputs. The criterion to be assessed is introduced in the cost function and the numerical solution is obtained using direct optimal control methods. In this context, we consider the case where the parameter distribution is unknown and assume a truncated uniform distribution with truncation values to be determined. The approach can be summarized as follows: First an optimization assisted bisection search algorithm is performed. This algorithm yields regions of a defined size in which a violation of the criterion occurs. In order to obtain a local explicit representation of the worst-case solution, we approximate this solution in the parameter space using a spectral representation based on gPC. This representation is then used to determine the worst case truncation limits of the uniform distribution and to estimate the exceedance probability for the criterion under investigation. The application is illustrated using a F-16 short period model with model reference adaptive controller. In this example, we estimate the exceedance probability for the maximum tracking error in the angle of attack for worst case reference command inputs and plant uncertainties in pitch damping, pitch stiffness, and control effectiveness.

Keywords: Safety analysis · Generalized polynomial chaos · Optimal control

1 Introduction

Model based testing of flight control laws is a key element in the development process to verify the safety level of aerial systems. The standard procedures in industry practice for this purpose are mainly based on Monte Carlo and gridding methods. These methods are computationally demanding and in case of gridding may miss solutions in between grid points. Therefore, it is useful to complement them by using optimization

Electronic Navigation Research Institute, *Air Traffic Management and Systems III*, Lecture Notes in Electrical Engineering 555,
https://doi.org/10.1007/978-981-13-7086-1_5

[1, 2] or optimal control methods [3, 4]. By solving a minimization problem, worst-case scenarios with respect to the criterion under investigation can be found. Specifically, optimal control based methods have shown to be highly effective in finding worst case continuous inputs and disturbances. One of the main differences between optimization-based methods, e.g., based on global optimization schemes such as genetic algorithms or evolutionary strategies, and optimal control methods is the treatment of the control input and the discretization of the dynamic system. In case of direct optimal control methods a variety of discretization methods, such as collocation methods or multiple shooting, can be used to find an approximate solution to the infinite-dimensional control problem [5, 6]. The resulting problem usually exhibits a sparse structure which can be exploited by gradient based schemes tailored to these kind of problems. Moreover, optimal control theory can be used to check the results e.g. by investigating the co-states or the switching function of the optimal solution. Contrary, global optimization schemes have shown to be effective in finding worst case control inputs which rely on a coarse discretization or parametrization of the control signal. Even though the application of the latter is very generic it is difficult to check the optimality of the solution or to define a meaningful termination criterion.

Besides the worst case control input, uncertain parameters play a major role for testing flight control systems: An important question to be investigated in this case concerns the exceedance probability of the safety criterion, when considering distributed parameters. The problem of uncertainty quantification has been addressed by various methods [7, 8] in the past. Popular approaches to obtain stochastic information about the solution are sampling-based methods. However, these methods usually exhibit slow convergence and require a large number of realizations. Especially for the optimal control based approach, the computational requirements would be prohibitive, as each sampling point requires the solution of a high-dimensional nonlinear parameter optimization problem. Therefore, the generalized polynomial chaos (gPC) approach [9] has become a popular choice for uncertainty quantification. This method is based on a spectral representation of the uncertain output using orthogonal polynomial basis functions, which exhibits fast convergence and only needs few realizations to determine the expansion coefficients for the polynomial basis. The paper is organized as follows: First we introduce the basic notions concerning optimal control based flight control law testing and provide a brief overview for the gPC method. Afterwards, the main approach for the approximation of the worst case solution in the random space is presented. The following section describes a bisection search algorithm used to identify regions where a violation of the criterion is expected. Finally, the approach presented in this paper is illustrated by investigating the exceedance probability for the angle-of-attack tracking error using a model reference adaptive controlled short period F-16 fighter model for a three-dimensional random space.

2 Theoretical Background

2.1 Optimal Control Based Flight Control Law Testing

Optimal control based flight control law testing can be used to find continuous worst case inputs $\mathbf{u}(t) \in \mathbb{R}^{n_u}$ for the dynamic closed loop system $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta})$ with the system states $\mathbf{x}(t) \in \mathbb{R}^{n_x}$ and the parameters $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$. The criterion to be tested is introduced in the cost function using a MAYER-cost $J = y_c(t_f)$, which evaluates the criterion y_c at the free final time t_f. Note that by letting the final time free, we do not need a-priori information about when the worst case occurs. Furthermore, path constraints $\mathbf{c}(\mathbf{x}(t), \mathbf{u}(t)) \leq \mathbf{0}$ can be used to include physical limits such as actuator position and rate limits. Moreover, the parameters are constrained via the box bounds $\boldsymbol{\theta}_{lb} \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}_{ub}$. The boundary conditions at the initial time $t_0 = 0$ and final time t_f are introduced by the equality constraint $\boldsymbol{\phi}(\mathbf{x}(t_0), \mathbf{x}(t_f)) = 0$. Summarizing, the optimal control problem formulation reads

$$\begin{aligned} \text{minimize} \quad & J = y_c(t_f) \\ \text{subject to} \quad & \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \boldsymbol{\theta}) = \dot{\mathbf{x}}(t), \\ & \mathbf{c}(\mathbf{x}(t), \mathbf{u}(t)) \leq \mathbf{0}, \\ & \boldsymbol{\phi}(\mathbf{x}(t_0), \mathbf{x}(t_f)) = \mathbf{0}, \\ & \boldsymbol{\theta}_{lb} \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}_{ub}. \end{aligned} \tag{1}$$

As a generic tool to solve problem (1), direct methods [5, 6] have shown to be very effective. In contrast to indirect methods, which rely on the derivation of necessary conditions and the solution of the resulting two-point boundary value problem, direct methods first discretize the problem in time. For the dynamic equation the discretization is achieved by means of transcription methods such as shooting or collocation. The discretized optimal control problem yields a nonlinear programming problem (NLP) with optimization variable vector $\mathbf{z} \in \mathbb{R}^{n_z}$, which can be efficiently solved via solvers such as IPOPT [10] or SNOPT [11]. Here, we consider trapezoidal collocation where the states and controls are fully discrectized on a time grid $\mathbf{t} = t_f \cdot \boldsymbol{\tau} = t_f \cdot [\tau_0, \tau_1, \ldots, \tau_f]$ with the normalized time grid $\boldsymbol{\tau} \in [0, 1]$, which yields the following optimization variable vector

$$\mathbf{z} = [t_f, \boldsymbol{\theta}^T, \mathbf{x}_0^T, \mathbf{u}_0^T, \ldots, \mathbf{x}_f^T, \mathbf{u}_f^T]^T, \tag{2}$$

with $\mathbf{x}_i = \mathbf{x}(t_f \cdot \tau_i)$, $\mathbf{u}_i$ respectively. The boundary conditions and the discretized dynamic constraints

$$\mathbf{x}_{i+1} - \mathbf{x}_i - t_f(\tau_{i+1} - \tau_i)\frac{\mathbf{f}_{i+1} + \mathbf{f}_i}{2} = \mathbf{0}, \tag{3}$$

with $\mathbf{f}_i = \mathbf{f}(\mathbf{x}_i, \mathbf{u}_i, \boldsymbol{\theta})$ for time steps $i = 1, \ldots, f-1$ are included in the equality constraints $\mathbf{h}(\mathbf{z}) = \mathbf{0}$ of the NLP. Moreover, the path constraints are imposed on the time grid points by introducing the inequality constraints $\mathbf{g}(\mathbf{z}) \leq \mathbf{0}$.

Note that in the problem formulation (1) the parameters can be subject to optimization or fixed to a specific value. The first type, with free parameters, is used in the bisection search algorithm outlined in Sect. 4 to test for violations in a specific region of the parameter space. In the second type, the optimal control problem becomes a parameter dependent problem and the parameters $\boldsymbol{\theta}$ are not part of the optimization variable vector $\mathbf{z}$. This type of problem is used for the computation of the expansion coefficients for the gPC approximation described in the following section.

2.2 Generalized Polynomial Chaos

The gPC approach [7] uses an orthogonal polynomial basis $\psi_m, m = 0, \ldots, d_i$, with

$$\int_{\Omega_i} \psi_m(\theta_i)\psi_n(\theta_i)\rho_i(\theta_i)d\theta_i = h_m^2 \delta_{mn}, \tag{4}$$

for the uncertain parameters $\theta_i, i = 1, \ldots, n_\theta$, with probability density functions $\rho_i(\theta_i)$ defined over the random space Ω_i to approximate the response metric $y(\mathbf{z}; \boldsymbol{\theta})$. In Eq. (4) the Kronecker delta δ_{mn} is 1 if $m = n$ and 0 otherwise and h_m is defined by the relation

$$\int_{\Omega_i} \psi_m^2(\theta_i)\rho_i(\theta_i)d\theta_i = h_m^2. \tag{5}$$

In the following we assume, without loss of generality, a normalized basis. Depending on the type of distribution of the random parameters θ_i, the basis functions ψ_m should be selected according to Table 1 in order to ensure exponential convergence. The response surface representation can then be written as a sum of the products of the expansion coefficients $\hat{y}_m(\mathbf{z})$ and the joint polynomial basis $\Psi_m(\boldsymbol{\theta})$

$$\Psi_m(\boldsymbol{\theta}) = \prod_{i=1}^{n_\theta} \psi_m(\theta_i), \tag{6}$$

by

$$\mathbf{y}(\mathbf{z}; \boldsymbol{\theta}) \approx \sum_{m=0}^{M-1} \hat{\mathbf{y}}_m(\mathbf{z})\Psi_m(\boldsymbol{\theta}). \tag{7}$$

Table 1. Polynomial basis for different continuous distribution types [7]

Distribution	Polynomial basis	Support
Gaussian	Hermite	$(-\infty, \infty)$
Gamma	Laguerre	$[0, \infty)$
Beta	Jacobi	$[-1, 1]$
Uniform	Legendre	$[-1, 1]$

In the following, the calculation of the expansion coefficients is described only for the case of a scalar parameter θ. For a description of the multivariate case please refer to [7]. The coefficients $\hat{y}_m(\mathbf{z})$ for a scalar parameter θ are defined by the integral over the random space Ω:

$$\hat{y}_m(\mathbf{z}) = \int_{\Omega} y(\mathbf{z}, \theta)\psi_m(\theta)\rho(\theta)d\theta \tag{8}$$

This integral can be efficiently approximated using a Gaussian quadrature with nodes $\theta^{(j)}$ and weights $w^{(j)}$:

$$\hat{y}_m(\mathbf{z}) = \sum_{j=1}^{M} y(\mathbf{z}, \theta^{(j)})\psi_m(\theta^{(j)})w^{(j)} \tag{9}$$

Note that for tensor grids the number of quadrature points increases exponentially, which renders the evaluation computationally expensive in higher dimensions (CURSE OF DIMENSIONALITY). For dimensions $n_\theta > 5$ it is thus useful to employ sparse grid representations.

3 General Approach

The goal of our approach for performing the safety analysis is to approximate the worst case solution for the optimal control formulation (1) in the space of the distributed parameters $\boldsymbol{\theta}$. This approach relies on a gPC approximation in the random space to cheaply evaluate stochastic information regarding the criterion. Please take into account that the worst case solution of problem (1) might change considerably depending on the values of the distributed parameters $\boldsymbol{\theta}$. A global approximation of the response surface over the whole parameter space may thus not exhibit sufficient accuracy. Furthermore, large parts of the random space may not contain a violation of the criterion under investigation. Thus, for the approximation we only want to focus our attention on regions where we expect a violation of the criterion. The main steps of the approach are:

I Search for violated regions in the random space using an optimization-based bisection algorithm.
II Perform the gPC approximation locally in the violated regions using stochastic collocation.
III Determine stochastic information with respect to the clearance criterion based on the gPC approximation.

In the first step, we search the parameter space for regions of a defined size in which the criterion under investigation is violated. The search algorithm is based on the idea of bisection in combination with optimization and is described in Sect. 4. In essence, the algorithm exploits the fact that we can clear large portions of the parameter space by searching for a worst case solution using the optimal control problem (1). If no violation is detected, the region is regarded as cleared. Otherwise, we further refine the search space until a defined cell size is reached.

Within these identified regions we then compute the approximation using gPC with the respective polynomial basis. For the parameter distributions we consider the case where the distributions are unknown. It can be shown that under reasonable assumptions [12] the worst case distribution is of type truncated uniform with unknown bounds. In practice these bounds are often simply taken to be the extremal values of the parameter space, as it is difficult to determine the true worst case bounds. This assumption is justified for many applications, as the worst case solutions are often found at the maximum or minimum values of the parameters. Nevertheless, to perform a rigorous analysis it is important to test which values for the bounds of the truncated uniform distribution actually lead to an extremal value of the criterion.

For the last step, in the context of flight control law clearance, we are especially interested in computing the probability p_c for the criterion y_c to exceed the performance threshold γ:

$$p_c = \mathcal{P}(y_c < \gamma) \tag{10}$$

4 Bisection Search Algorithm

As outlined in Sect. 3, our goal is to perform the gPC approximation only locally within the n_R regions $\mathcal{R}_i(\boldsymbol{\theta}^i_{lb}, \boldsymbol{\theta}^i_{ub})$

$$\mathcal{R}_i(\boldsymbol{\theta}^i_{lb}, \boldsymbol{\theta}^i_{ub}) = \left\{\boldsymbol{\theta} \in \mathbb{R}^{n_\theta} : \boldsymbol{\theta}^i_{lb} \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}^i_{ub}\right\}, \tag{11}$$

$i = 1, \ldots, n_R$ with maximum edge lengths $\mathbf{h}$, i.e.,

$$\left(\theta^i_{ub,j} - \theta^i_{lb,j}\right) \leq h_j, \forall j = 1, \ldots, n_\theta. \tag{12}$$

This is mainly due to the fact that several solutions of the optimal control problem at the collocation points within each region are necessary to compute the gPC expansion coefficients in Eq. (8). A division of the whole parameter space and the analysis of the response surface in all regions would be computationally very demanding, if not prohibitive. Therefore, we seek to perform the approximation only for those regions where we expect a violation of the criterion. In order to find those regions we exploit the fact that we can test, if a violation of the criterion within $\mathcal{R}_i(\boldsymbol{\theta}^i_{lb}, \boldsymbol{\theta}^i_{ub})$ can be expected, via solving the optimal control problem (1). Please note that for this optimal control problem the parameters are optimizable within the bounds $\boldsymbol{\theta}^i_{lb}$ and $\boldsymbol{\theta}^i_{ub}$. For the formal description of the algorithm the corresponding function is denoted by $\left[\hat{J}_i, \hat{\boldsymbol{\theta}}\right] = Test\left(\mathcal{R}_i(\boldsymbol{\theta}^i_{lb}, \boldsymbol{\theta}^i_{ub})\right)$.

The solution of this optimal control problem tells us if a violation $\hat{J}_i < \gamma$ has been detected or if the region can be regarded as cleared. In order to efficiently find those grid regions we employ a search algorithm based on the idea of bisection.

To formally describe the algorithm let us define three different types of regions $\mathcal{R}$. The first type, $\mathcal{R}^S$, are regions that need to be searched. The second type, $\mathcal{R}^V$, are the regions where we found a violation of the criterion but which have edge lengths greater than defined by $\mathbf{h}$. The third type, $\mathcal{R}^{V_{min}}$, are the grid cells in which we found a violation of the criterion and that are below the required size. The refinement of the search space is achieved by repeatedly bisecting regions where a violation of the criterion was found into two new regions $\mathcal{R}^V$ and $\mathcal{R}^S$: $\left[\mathcal{R}^V, \mathcal{R}^S\right] = Bisect\left(\mathcal{R}_i, \hat{\boldsymbol{\theta}}_i\right)$. Here, $\mathcal{R}^V$ is the region which contains the worst case solution violating the criterion and $\mathcal{R}^S$ is the complementary region that still needs to be tested. In order to make the bisection unique for more than one parameter, i.e., $n_\theta > 1$, we perform the bisection in those directions for which the bisection plane has the greatest normal distance from the worst case. The reasoning behind this strategy is to obtain search regions $\mathcal{R}^S$ which have a higher chance of not containing a violation of the criterion. Moreover, the region is not refined in directions in which the edge length is smaller than or equal to the maximum allowable size. The described algorithm is formalized in Algorithm 1.

Algorithm 1 $TestAndBisect\left(\mathcal{R}_i, \gamma, \mathbf{h}\right)$

$\left[\hat{J}_i, \hat{\boldsymbol{\theta}}_i\right] = Test\left(\mathcal{R}_i\right)$
if $\hat{J}_i < \gamma$ **then**
  if $isMinimumSize\left(\mathcal{R}_i, \mathbf{h}\right)$ **then**
    $\mathcal{R}^{V_{min}} \leftarrow \mathcal{R}_i$
  else
    $\left[\mathcal{R}^V, \mathcal{R}^S\right] = Bisect\left(\mathcal{R}_i, \hat{\boldsymbol{\theta}}_i, \mathbf{h}\right)$
    $TestAndBisect\left(\mathcal{R}^S, \gamma, \mathbf{h}\right)$
    if $isMinimumSize\left(\mathcal{R}^V, \mathbf{h}\right)$ **then**
      $\mathcal{R}^{V_{min}} \leftarrow \mathcal{R}^V$
    else
      $TestAndBisect\left(\mathcal{R}^V, \gamma, \mathbf{h}\right)$
    end if
  end if
end if

The search set for Algorithm 1 is initialized with the whole search space, $\mathcal{R}^S \leftarrow \mathcal{R}_0(\boldsymbol{\theta}_{lb}, \boldsymbol{\theta}_{ub})$, and the violated regions are initially an empty set, $\mathcal{R}^{V_{min}} \leftarrow \emptyset$, which starts the recursive algorithm $TestAndBisect\left(\mathcal{R}^S, \gamma, \mathbf{h}\right)$. It should be noted that the computational efficiency of Algorithm 1 largely depends on the number and location of the violated regions. In case one single region of the required minimum size contains the violated portion, the number N_1 of necessary bisections is

$$N_1 = \sum_{j=1}^{n_\theta} \lceil \log_2\left(\left(\theta_{ub,j} - \theta_{lb,j}\right)/h_j\right) \rceil. \tag{13}$$

4.1 Illustration of the Algorithm

To illustrate the algorithm we will use HIMMELBLAU's function

$$J(\theta_1, \theta_2) = (\theta_1^2 + \theta_2 - 11)^2 + (\theta_1 + \theta_2^2 - 7)^2 \tag{14}$$

depending on the parameters θ_1 and θ_2. In this example the parameter region is restricted within $-5 \leq \theta_1 \leq 5$ and $-5 \leq \theta_2 \leq 5$ in which HIMMELBLAU's function has four distinct minima, as can be seen from Fig. 1. We illustrate Algorithm 1 by estimating the portion of the search space for which we have a value $J < 1$. Note that for this example the *Test*()-function in Algorithm 1 only minimizes the function locally within the search space. For solving the associated optimization problem within each region we use the MATLAB NLP solver *fmincon*() with two randomly initialized starting points. Figure 2 illustrates the regions for the bisection search Algorithm 1.

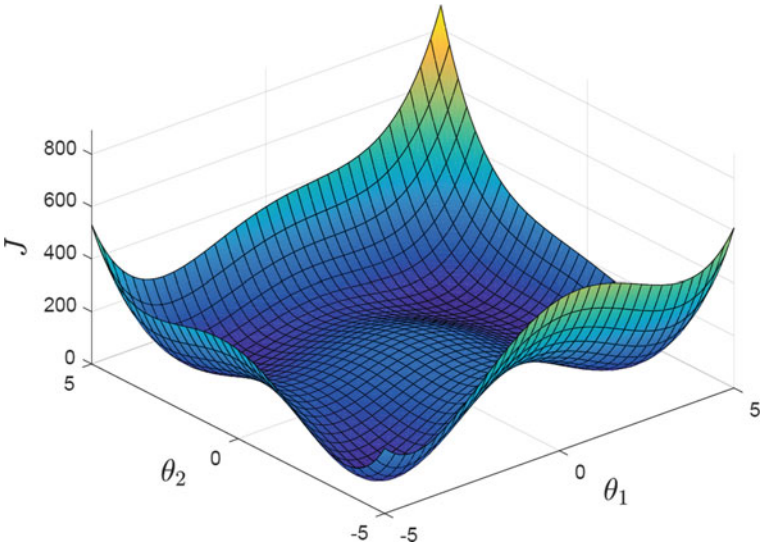


Fig. 1. HIMMELBLAU's function

Let us further investigate the question concerning the symmetric truncation limit values $\boldsymbol{\theta}_{lim}$ for which the value for the portion $J < 1$ is extremal. From Fig. 2 it can be seen that the symmetric, outer limiting boundary of the highlighted regions (dashed line) can be chosen as a first approximation for the upper bound for the truncation limits. Furthermore, the symmetric, inner limiting boundary (dash-dotted line) represents a first approximation to the lower bound where no regions $J < 1$ are contained. The solid line shows the worst case truncation values $\boldsymbol{\theta}_{lim}$ at which the portion for $J < 1$ (see brown markers in Fig. 3) is maximal. The corresponding values are summarized in Table 2.

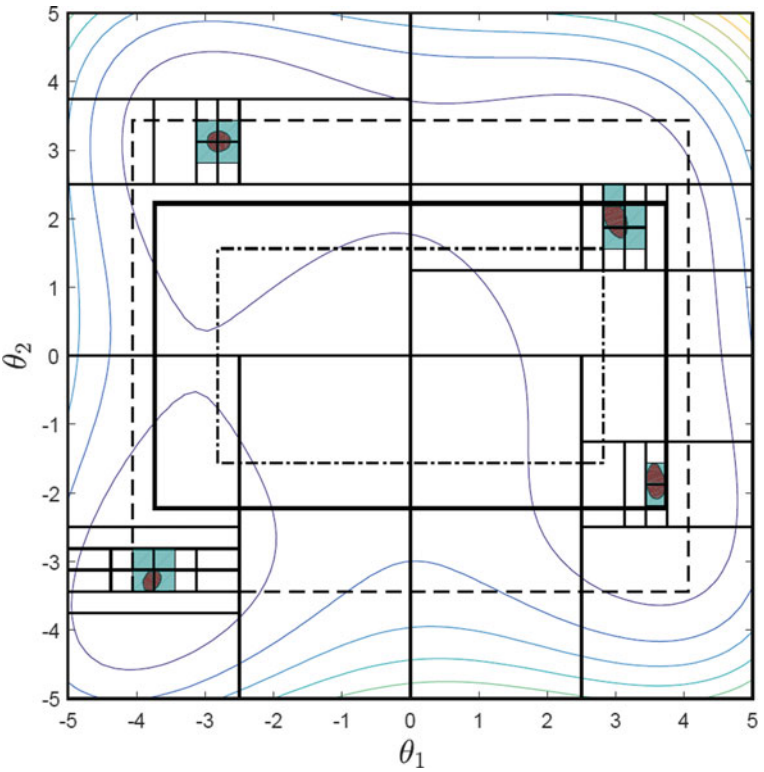


Fig. 2. Visualization of the bisection search algorithm for HIMMELBLAU's function (violated areas marked in brown, identified regions in blue, approximation of the upper truncation limits dashed line, approximation of the lower truncation limits dash-dotted line, worst case limits thick solid line)

Table 2. Limits and portions with a value $J < 1$

Case	Limits $\boldsymbol{\theta_{lim}}$	Portion $J < 1$[%]
Full	$\theta_1 \in [-5, 5]$ $\theta_2 \in [-5, 5]$	0.412
Approximation upper bound	$\theta_1 \in [-4.1, -3.4]$ $\theta_2 \in [4.1, 3.4]$	0.733
Maximum	$\theta_1 \in [-3.7, -2.2]$ $\theta_2 \in [3.7, 2.2]$	0.761

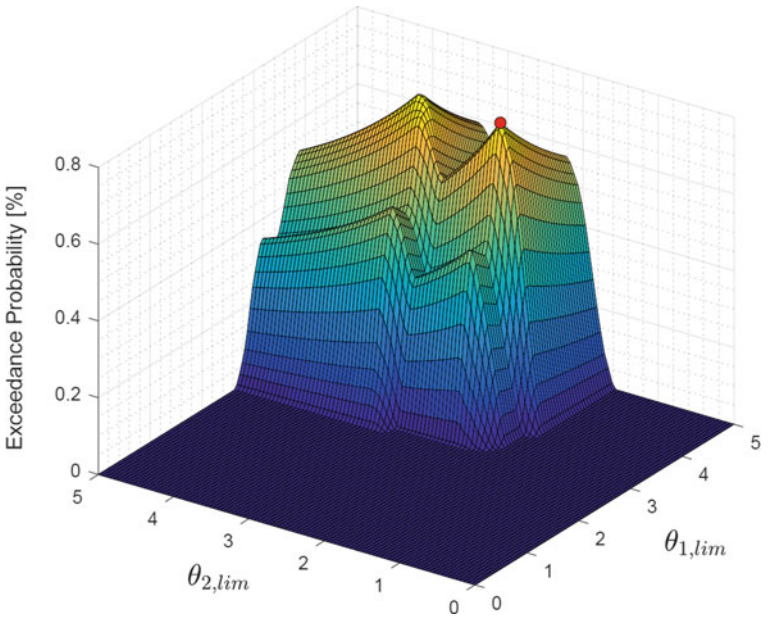


Fig. 3. Portion with $J < 1$ for different truncation values (maximum value marked red)

5 Illustrative Example

For illustration purposes we use the example from [13] for testing a model reference adaptive controller and extend it by assuming a truncated uniform distribution for the uncertain parameters. In the following, we will briefly summarize the main model equations and state the problem formulation to be investigated.

5.1 Model Equations

The model represents a short period approximation for the F-16 aircraft with model reference adaptive controller. This controller translates the reference command r (angle-of-attack) into the control u such that the plant model states $\mathbf{x}_P$ track the reference model states $\mathbf{x}_M$ asymptotically. The state space representation comprises states

$$\mathbf{x} = [e_\alpha, e_q, \alpha_M, q_M, \Theta_\alpha, \Theta_q, \Theta_r]^T, \tag{15}$$

namely the states of the tracking error dynamic $\mathbf{e}_C^T = \mathbf{x}_P^T - \mathbf{x}_M^T = [e_\alpha, e_q]$ for the angle-of-attack error e_α and pitch rate error e_q, the reference model states $\mathbf{x}_M^T = [\alpha_M, q_M]$, i.e., the angle of attack α_M and pitch rate q_M, and the states for the adaptation law $\mathbf{\Theta}^T = [\mathbf{\Theta}_\mathbf{x}^T, \Theta_r] = [\Theta_\alpha, \Theta_q, \Theta_r]$

$$\begin{aligned}\dot{\mathbf{e}}_C &= \mathbf{A}_M \mathbf{e}_C + \mathbf{B}_P \lambda_{M_\eta}(\boldsymbol{\Theta} - \boldsymbol{\Theta}^*) \cdot \boldsymbol{\omega},\\ \dot{\mathbf{x}}_M &= \mathbf{A}_M \mathbf{x}_M + \mathbf{B}_P K_r r,\\ \dot{\boldsymbol{\Theta}}^T &= -\boldsymbol{\Gamma}\boldsymbol{\omega}\mathbf{e}_C^T \mathbf{P}\mathbf{B}_P.\end{aligned} \tag{16}$$

The system matrix $\mathbf{A}_M$ of the reference model and the feed-forward gain K_r are defined as

$$\mathbf{A}_M = \begin{bmatrix} Z_\alpha & 1+Z_q \\ M_{\alpha_{des}} & M_{q_{des}} \end{bmatrix}, K_r = -1.55, \tag{17}$$

with $M_{\alpha_{des}} = -11.38$ and $M_{q_{des}} = -4.16$. The plant model system matrix $\mathbf{A}_P$ and input matrix $\mathbf{B}_P$ read:

$$\mathbf{A}_P = \begin{bmatrix} Z_\alpha & 1+Z_q \\ \lambda_{M_\alpha} M_\alpha & \lambda_{M_q} M_q \end{bmatrix}, \mathbf{B}_P = \begin{bmatrix} 0 \\ M_\eta \end{bmatrix}, \tag{18}$$

with $M_\alpha = -30.79$, $M_q = -3.75$, $Z_\alpha = -1.84$, $Z_q = -0.09$, and $M_\eta = -12.75$. The nominal plant coefficients are subject to the multiplicative uncertainties λ_{M_α}, λ_{M_q}, and λ_{M_η} which are assumed to be a truncated uniform distribution with symmetric limits to the nominal value 1, ranging between 0.5 and 1.5. Here $\boldsymbol{\Theta}^* = [\boldsymbol{\Theta}_{\mathbf{x}}^{*T}, \Theta_r^*]$ are the so called ideal parameters related to the matching conditions

$$\mathbf{A}_M = \mathbf{A}_P + \mathbf{B}_P \lambda_{M_\eta} \boldsymbol{\Theta}_{\mathbf{x}}^*, \tag{19}$$

$$\mathbf{B}_P \cdot K_r = \mathbf{B}_P \lambda_{M_\eta} \Theta_r^*, \tag{20}$$

and $\boldsymbol{\omega}^T = [\mathbf{x}_P^T, r]$ is the regressor vector. Furthermore, $\boldsymbol{\Gamma} = 200 \cdot \mathbf{I}^{3\times 3}$ and $\mathbf{P}$ are parameters of the adaptive controller, where $\mathbf{P}$ is the solution of the Lyapunov equation

$$\mathbf{A}_M^T \mathbf{P} + \mathbf{P}\mathbf{A}_M = -\mathbf{Q}, \tag{21}$$

with $\mathbf{Q} = \mathbf{I}^{2\times 2}$. Note that the matrix $\mathbf{Q}$ and the learning rate $\boldsymbol{\Gamma}$ are design parameters which effect the tracking performance of the controller. At higher values of the learning rate usually a smaller error between the reference model states and the uncertain plant can be expected (see [13] for a comparative study of different learning rates).

5.2 Problem Formulation

For the example problem formulation we investigate the probability p_c that the tracking error for the angle-of-attack e_α exceeds the limit of 5 deg at the free final time within the bounds $t_f \in [0s, 4s]$

$$p_c = \mathcal{P}\big(e_\alpha(t_f) > 5\,\mathrm{deg}\big). \tag{22}$$

The optimal control problem starts from the trim condition $\mathbf{x}(t_0) = \mathbf{0}$ and reads for this particular example

$$\begin{aligned} \text{minimize} \quad & J = -e_\alpha(t_f) \\ \text{subject to} \quad & \mathbf{f}\left(\mathbf{x}(t), r(t), \boldsymbol{\theta}\right) = \dot{\mathbf{x}}(t), \\ & \mathbf{x}(t_0) = \mathbf{0}, \\ & \boldsymbol{\theta}^i_{lb} \leq \boldsymbol{\theta} \leq \boldsymbol{\theta}^i_{ub}. \end{aligned} \tag{23}$$

In this formulation, the uncertain parameter vector $\boldsymbol{\theta}$ contains the plant uncertainties as defined in Sect. 5.1, i.e. $\boldsymbol{\theta} = \left[\lambda_{M_\alpha}, \lambda_{M_\eta}, \lambda_{M_q}\right]^T$.

The maximum allowable edge length for the bisection Algorithm 1 is set to $h = 0.25$ for all directions. For the numerical solution of problem (23) we employ a trapezoidal collocation scheme using the optimal control tool *FALCON.m* [14] with the NLP solver IPOPT.

The results of the gPC analysis are compared with the solution obtained from MONTE CARLO sampling over the optimal control solution in the random space.

5.3 Results and Discussion

First, we visualize the results for a two-dimensional grid in $0.5 \leq \lambda_{M_\alpha} \leq 1.5$, $0.5 \leq \lambda_{M_\eta} \leq 1.5$ for $\lambda_{M_q} = 1$ in Fig. 4. As can be seen three regions $\mathcal{R}^{V_{min}}$ violating the criterion have been identified for this case. The green dots in Fig. 4 represent the collocation points for evaluating the expansion coefficients for a third-order gPC approximation. The red region in Fig. 4 indicates where a violation can be expected when using the local gPC approximations. Figure 5 illustrates the critical regions for the full three-dimensional case $0.5 \leq \lambda_{M_\alpha} \leq 1.5$, $0.5 \leq \lambda_{M_\eta} \leq 1.5$, and $0.5 \leq \lambda_{M_q} \leq 1.5$ in which the identified regions of the maximum allowable size containing a violation are marked in blue. In this example the violations occur at the maximum, respectively minimum, values of the admissible parameter space which means that the worst case truncation values for the truncated uniform distributions are taken as the maximum limits.

For a comparative analysis we performed a MONTE CARLO [15] study using 5000 pseudo-random sampling points. It should be noted that for the sampling points a large number of optimal control problems must be solved which is computationally very demanding. The mean value of the exceedance probability obtained from MONTE CARLO analysis yields 6.20% and the 95% confidence interval for the chosen sample size is between 5.53 and 6.87%. In comparison the approximation using the gPC approach presented in this paper yields a probability of 6.56% which is in good agreement with the value obtained from the MONTE CARLO analysis.

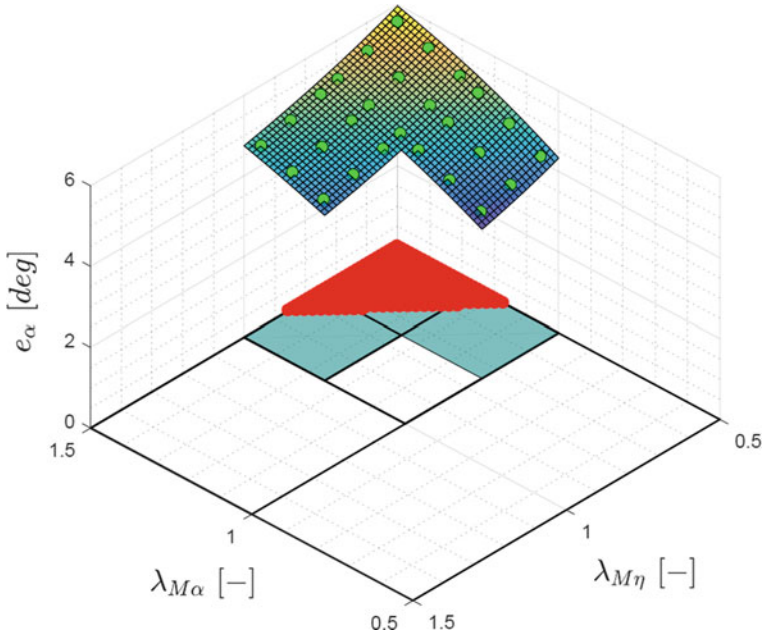


Fig. 4. Third-order gPC approximation for the violated regions in the two-dimensional parameter space in λ_{M_z} and λ_{M_η} for $\lambda_{M_q} = 1$ (violated area marked in red, collocation points for the gPC approximation are marked in green)

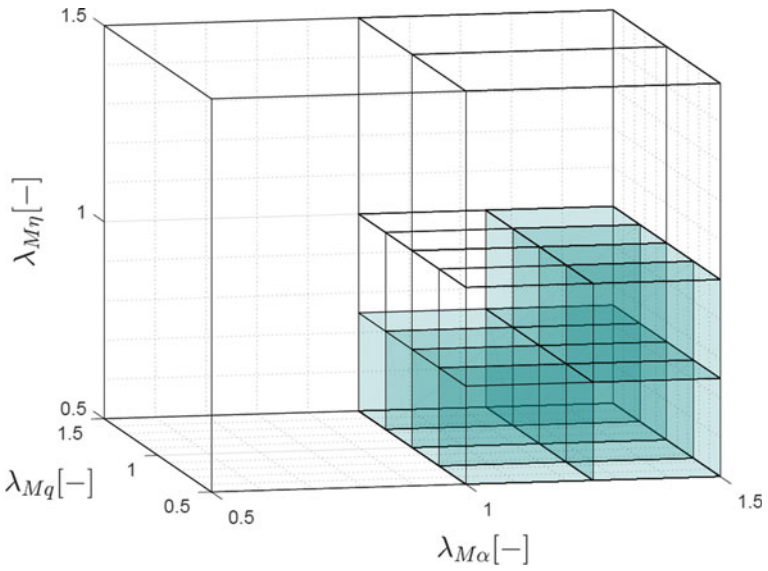


Fig. 5. Result of the bisection search algorithm for the three-dimensional parameter space λ_{M_z}, λ_{M_η}, and λ_{M_q}. Regions in which a violation was detected are highlighted in blue

6 Conclusions

We presented an approach to estimate the exceedance probability for safety criteria using optimal control methods in combination with generalized polynomial chaos. This analysis assumes a truncated uniform distribution, which is the worst case distribution when the true distribution is unknown. By means of an optimization assisted search algorithm, we first identify critical regions where we expect a violation of the criterion under investigation. The gPC representation can then be used to cheaply determine stochastic information such as the exceedance probability using sampling. It is important to mention that direct methods usually employ gradient based optimization schemes which only guarantee local optimality. Thus, good initial guesses for the optimal control problem are of high importance. For the analysis in this paper we used the approach presented in [16] and gradually built up solution candidates using first sampling to generate initial populations for a global optimization scheme (differential evolution). Solutions of the global optimization scheme are then used as initial guesses for the optimal control problem. In our experience, it is important for the robustness of the algorithm to numerically solve the optimal control problem starting from several initial guesses and to choose the worst case among the solutions. A major challenge of the approach presented here is that for high-dimensional random spaces the gPC expansion becomes increasingly costly to evaluate, as one optimal control solution is necessary per collocation point. Future work will thus be devoted on obtaining approximations of the worst case based on post-optimal sensitivity analysis. As the criterion is introduced in the cost function this analysis allows to locally approximate the solution of the optimal control problem to second order. Using this method the local evaluation of the worst case solution around a nominal point becomes much cheaper and should greatly enhance the efficiency of the proposed method.

Acknowledgements. This work was supported by the DFG grant HO4190/8-1.

References

1. Morari M, Thoma M, Varga A, Hansson A, Puyou G (2012) Optimization based clearance of flight control laws, vol 416. Springer, Heidelberg
2. Menon Prathyush P, Kim Jongrae, Bates Declan G, Postlethwaite Ian (2006) Clearance of nonlinear flight control laws using hybrid evolutionary optimization. IEEE Trans Evol Comput 10(6):689–699
3. Herrmann Avriel A, Ben-Asher Joseph Z (2016) Flight control law clearance using optimal control theory. J Aircr 53(2):515–529
4. Diepolder J, Ben-Asher JZ, Gabrys AC, Schatz SP, Bittner M, Rieck M, Grüter B, Holzapfel F (2016) Flight control law clearance using worst-case inputs. In: ICAS 30th international congress of the international council of the aeronautical sciences
5. Gerdts M (2012) Optimal Control of ODEs and DAEs. De Gruyter Textbook, De Gruyter, Berlin and Boston
6. Betts JT (2009) Practical methods for optimal control and estimation using nonlinear programming, 2nd edn. Advances in design and control. siam, society for industrial and applied mathematics. Philadelphia

7. Xiu D (2009) Fast numerical methods for stochastic computations - a review. Commun Comput Phys 5(2–4):242–272
8. Eldred MS (2009) Recent advances in non-intrusive polynomial chaos and stochastic collocation methods for uncertainty analysis and design. In: 50th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics, and materials conference. https://arc.aiaa.org/doi/10.2514/6.2009-2274
9. Xiu D, Em Karniadakis G (2002) The Wiener-Askey polynomial chaos for stochastic differential equations. SIAM J Sci Comput 24(2):619–644
10. Wächter A, Biegler LT (2006) On the implementation of a primal-dual interior point filter line search algorithm for large-scale nonlinear programming. Math Program 106(1):25–57
11. Gill PE, Murray W, Saunders MA (2002) SNOPT: An SQP algorithm for large-scale constrained optimization. SIAM J Optim 12(4):979–1006
12. Barmish BR, Lagoa CM (1996) The uniformity principle: A new tool for probabilistic robustness analysis. In: 30th conference on information science and systems, Princeton
13. Diepolder J, Heise CD, Bittner M, Rieck M, Grüter B, Holzapfel F, Ben-Asher JZ. Optimal control based tracking error estimation for model reference adaptive control. In: IACAS 57th Israel Annual Conference on Aerospace Sciences
14. Rieck M, Bittner M, Grüter B, Diepolder J (2016) FALCON.m: User Guide
15. Thomopoulos N (2013) Essentials of Monte Carlo simulation. Statistical methods for building simulation models. Springer, New York
16. Diepolder J, Saurabh S, Akkinapalli VS, Raab S, Zhang J, Bhardwaj P, Krenmayr M, Grüter B, Holzapfel F (2017) Flight control law testing using optimal control and postoptimal sensitivity analysis. In: CEAS EuroGNC

Robust Optimal Trajectory Planning Under Uncertain Winds and Convective Risk

Daniel González-Arribas[1(✉)], Manuel Soler[1], Manuel Sanjurjo-Rivo[1], Javier García-Heras[1], Daniel Sacher[2], Ulrike Gelhardt[2], Juergen Lang[2], Thomas Hauf[3], and Juan Simarro[4]

[1] Escuela Politécnica, Area of Aerospace Engineering, Universidad Carlos III de Madrid, 28911 Leganés, Madrid, Spain
{dangonza, masolera, msanjurj}@ing.uc3m.es
[2] MeteoSolutions GmbH, Wilhelminenstraße 2, 64283 Darmstadt, Germany
{daniel.sacher, ulrike.gelhardt, juergen.lang}@meteosolutions.de
[3] Leibniz University Hannover, Scientific Advisor of TBO-Met Project, Hanover, Germany
hauf@muk.uni-hannover.de
[4] Agencia Estatal de Meteorología (AEMET), Madrid, Spain
jsimarrog@aemet.es

Abstract. The existence of significant uncertainties in the models and systems required for trajectory prediction represents a major challenge for the Trajectory-Based Operations concept. Weather can be considered as one of the most relevant sources of uncertainty. Understanding and managing the impact of these uncertainties are necessary in order to increase the predictability of the ATM system. We present preliminary results on robust trajectory planning in which weather is assumed to be the unique source of uncertainty. State-of-the-art probabilistic forecasts from Ensemble Prediction Systems are employed to characterize uncertainty in the wind and potential convective areas. A robust optimal control methodology to produce efficient and predictable trajectories in the presence of these uncertainties is presented. A set of Pareto-optimal trajectories is obtained for different preferences between predictability, convective risk, and average efficiency.

Keywords: Trajectory planning · Uncertainty · Convection · Optimal control · Ensemble prediction systems

1 Introduction

A major challenge for Trajectory-Based Operations (TBO) is the existence of significant uncertainties in the models and systems required for trajectory prediction. Understanding and managing the impact of these uncertainties is necessary in order to increase the predictability of the ATM system. In turn, predictability and robustness

Electronic Navigation Research Institute, *Air Traffic Management and Systems III*, Lecture Notes in Electrical Engineering 555,
https://doi.org/10.1007/978-981-13-7086-1_6

improvements in trajectories will produce gains in the high level goals (capacity, efficiency, safety, and environmental impact) pursued within a modernized ATM system. Some examples of relevant uncertainty sources are: (1) meteorological uncertainty; (2) uncertainty in the aircraft performance model [1]; (3) uncertainty in initial mass [2] and other parameters and (4) uncertainty in the aircraft intent [3]. In this paper, the focus is on the first, i.e., meteorological uncertainty, one of the most important sources of uncertainty that affects the ATM system. Indeed, the SESAR Exploratory Research TBO-Met Project[1] focuses on the analysis of meteorological uncertainty coming from the following two sources: (1) wind, and (2) convective regions. While we will not consider additional uncertainty sources, we will note that our methodology could be extended to include them.

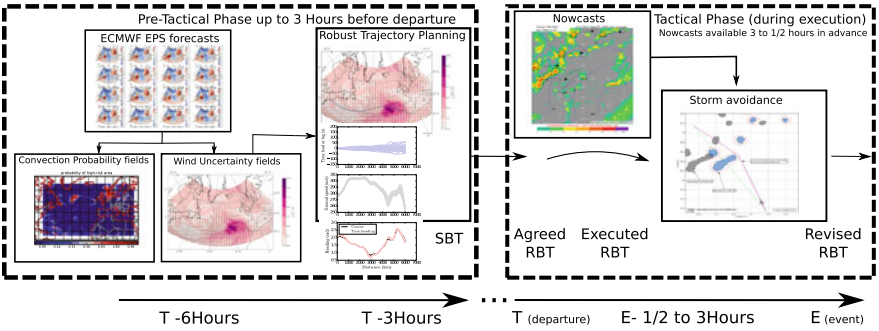


Fig. 1. TBO-Met Trajectory Planning Methodology for both pre-tactical and tactical levels. Recall that the present paper focuses on the pre-tactical level

The main contribution of this paper is to extend the methodology for robust route optimization in [4, 5] to the consideration of convection risk. The focus is on the pre-tactical level (in this context, around 3 h before departure). We make use of Ensemble Prediction Systems and optimal control techniques. Figure 1 sketches the intended methodology for the Trajectory planning problem in TBO-Met Project. The pre-tactical level is represented in the left-hand side of the figure, considering both wind uncertainty and convective phenomena.

Convective regions are defined as areas within which individual convective events, which we will refer to as "convective storms", may develop. Convective storms need a trigger mechanism and the onset and the location of those individual storms is impossible to forecast at the flight planning level. Nevertheless, one can obtain

[1] TBO-MET project (https://tbomet-h2020.com/) has received funding from the SESAR JU under grant agreement No 699294 under EU's Horizon 2020 research and innovation programme. Consortium members are UNIVERSITY OF SEVILLE (Coordinator), AEMET (Agencia Española de Meteorología), METEOSOLUTIONS GmbH, PARIS-LODRON-UNIVERSITAT SALZBURG, and UNIVERSIDAD CARLOS III DE MADRID.

forecasts with some characteristics that act as necessary (but not sufficient) conditions for the formation of storms, and that information can be employed to create an index that estimates probability of convection, i.e., an indicator of convection risk then can be used for trajectory planning.

Numerical Weather Prediction (NWP) centers developed Ensemble Prediction Systems (EPS) in order to provide probabilistic meteorological forecasts in addition to deterministic predictions. They seek to provide an estimation of the uncertainty that is inherent to the NWP process [6], a task that cannot be achieved with deterministic forecasting. In an EPS, several runs of the NWP model are launched with different characteristics in order to produce a set of (typically) 10 to 50 different forecasts or "members" of the ensemble. We refer to [7] for a review of the status of NWP as well as the relevance of EPS in a wider meteorological context.

The ATM research community has recently started to use EPS in order to study the predictability of flight plans and the sensitivity to weather prediction uncertainty. The main research effort in this direction has been undertaken within IMET, a SESAR WP-E project. It sought to develop a "probabilistic trajectory prediction" (PTP) system, where a deterministic TP is run once for each member in order to produce a trajectory ensemble. Preliminary results of this project were presented in [8] and a follow-up publication [9] showed how the information obtained with this approach could be used to improve decision-making at the pre-tactical level. Outside IMET, e.g., in [10] the authors present an analysis of the impact of uncertainty in average wind on final fuel consumption. In [4, 5], we developed (within the framework of SESAR's TBO-Met project) a robust approach to aircraft trajectory planning under wind uncertainty using EPS. In parallel to the latter, an approach based on Dynamic Programming for aircraft trajectory planning under wind uncertainties (and also using EPS) has been recently published in [11]. All in all, the focus of all these works is on wind uncertainty. To the best of authors' knowledge, the inclusion of convective indicators in robust flight planning is an unexplored field. Thus, we go beyond the state of the art by extending the approach in [4, 5] to the consideration of convection.

The paper is structured as follows: we introduce convection and its associated indicators in Sect. 2. The robust optimal control methodology is presented in Sect. 3. In Sect. 4, we present a case study, including the simulation results and a discussion. Finally, some conclusions are drawn in Sect. 5. See Fig. 2 for a graphical overview of the relationship between the different parts of the methodology. In Fig. 2, the EPS is the input to the problem, which directly provides an uncertainty quantification of both wind and temperature, and indirectly provides (via convection indicators) the convective risk (See Sect. 2). These are the elements we consider to formulate the robust trajectory optimization problem (Sect. 3) and obtain the flight plan.

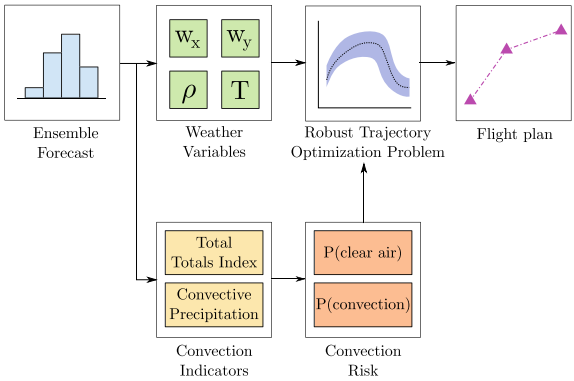


Fig. 2. Diagrammatic description of the methodology

2 Convection Modelling

2.1 Convection Indicators

Within this paper, we attempt to delimit high-risk areas due to deep convection and their respective uncertainty. The term "convective area" is defined here as an area of potentially developing storms, which comprise events such as individual storm cells, multi-cells, mesoscale convective complexes, and squall lines. The onset and the location of those individual storms are difficult to evaluate for the time being and impossible to determine in many cases. Favourable environmental characteristics and conditions for certain types, however, are known, e.g.:

- A squall line (at least in Central Europe) very often develops several hundred kilometres ahead of and parallel to an approaching cold front. It is initiated and recognized by a boundary convergence line. Many such lines often occur approx. 10 km apart, but not all of them necessarily develop into a squall line, though some of them do.
- Air mass storms preferably develop in the afternoon. The onset time of first shallow clouds and the development of deep convective clouds can be forecasted by standard meteorological procedures.
- Moderate mid-level shear enhances the storm strength, while too strong shear and no-shear environments are more likely related to weak storms.
- Long-lived storms are linked to the renewal and generation of new cells immediately ahead of a mature cell.
- Storms embedded in a cold front are out of scope of this study as they can be forecasted very well by synoptic forecasts of low pressure systems.
- The structure of the environmental temperature profile (temp) allows deriving certain features of the storm. Maritime dominated storms reveal a temp close to the moist-adiabatic implying weak updrafts, while continental storms exhibit more

potential energy to be released. The latter is defined by the area between moist adiabatic and the temperature profile.
- Environmental characteristics are used to derive empirically a range of convective indices.

Important to note is that the above characteristics are necessary conditions, but they do not allow the forecast of the precise location and onset. Convective storms need a trigger mechanism. In order to precisely forecast a storm, we therefore need to forecast the trigger mechanisms like, e.g., boundary convergence lines, tropospheric gravity waves, mountains or surface temperature inhomogeneity.

From the above, we conclude that we need an indicator to describe the necessary precondition for the potential development of convection and an indicator which comprises the essential activator in order to develop a storm which has to be avoided by aircraft. As described below, this will be done by using a combination of two convection indicators: *Total Totals Index* and *Convective Precipitation*, which are available from EPSs. When both indicators exceed certain thresholds for a high number of EPS members, the grid point is assumed to lie within the zone of high probability (low uncertainty) of convection which can be interpreted as a no-fly-zone. If only one criterion is fulfilled for a high number of EPS members, the grid point is located in a region of convective uncertainty. The boundary of uncertainty areas will delimit convective regions.

Those convective areas may have a persistence or life time of up to 60 h. Carbone et al. [12] and previous studies investigated precipitation episodes and found much longer life times of those episodes, respectively travelling convective regions, than those of the individual storms developing within. Here we pursue similar thoughts. Convective regions are perceived as areas with a high weather risk, which are characterized by always occurring and unpredictable individual storms. Convective regions, therefore, must not necessarily be avoided but require a higher weather situation awareness by pilots and controllers. Also, trajectories passing through a convective area are subject to diversions resulting in increased flight duration and delays. Thus, the intersection of a trajectory with a convective region does not imply, as already said above, that the whole area has to be circumnavigated, but rather that delays have to be expected. The dimension of the latter depends, among other factors, on the type of storms embedded in the convective area, density of cells, their orientation, the size of gaps separating the storms, and the time of onset.

We decided to combine two indicators for convection:

(1)*Total Totals Index (TT)*[2]: The sum of the vertical totals (*VT*) $VT = T_{850} - T_{500}$ (temperature gradient between 850 and 500 hPa) and the cross totals (*CT*) $CT = T_{d_{850}} - T_{500}$ (moisture content between 850 and 500 hPa by subtracting the temperature

[2] Attributable to National Weather Service Louisville, KY: http://www.weather.gov/lmk/indices http://www.weather.gov/lmk/indices, accessed July 25, 2016.

in 500 hPa from dew point temperature in 850 hPa). As a result, *TT* accounts for both static stability and 850 hPa moisture, but would be unrepresentative in situations where the low-level moisture resides below the 850 hPa level. In addition, convection may be inhibited despite a high *TT* value if a significant capping inversion is present. $VT = 40$ is close to dry adiabatic for the 850–500 hPa layer. However, *VT* generally will be much less, with values around 26 or more, representing sufficient static instability (without regard to moisture) for thunderstorm occurrence. $CT > 18$ often is necessary for convection, but it is the combined Total Totals Index that is most important. The risk of severe weather activity is operationally defined as in Table 1 (see also [13]).

Table 1. Operational taxonomy of risk of severe weather activity

	Thunderstorm activity			
TT_i	Moderate	Heavy	Severe	Tornadoes
< 44	–	–	–	–
44–45	Isolated	–	–	–
46–47	Scattered	Few	–	–
48–49	Scattered	Few	Isolated	–
50–51	*	Scattered	Few	Isolated
52–55	*	Numerous	Few/scattered	Few
> 55	*	Numerous	Scattered	Scattered

(2)*Convective Precipitation (CP)*[3]: an estimation of the precipitation coming from convective clouds. The total precipitation is the sum of the so-called large-scale precipitation and the convective precipitation.

The moist convection scheme is based on the mass-flux approach and represents deep (including *cumulus congestus*), shallow and mid-level (elevated moist layers) convection. The distinction between deep and shallow convection is made on the basis of the cloud depth (< 200 hPa for shallow). For deep convection, the mass-flux is determined by assuming that convection removes Convective Available Potential Energy (CAPE) over a given time scale. The intensity of shallow convection is based on the budget of the moist static energy, i.e., the convective flux at cloud base equals the contribution of all other physical processes when integrated over the sub-cloud layer. Finally, mid-level convection can occur for elevated moist layers, and its mass flux is set according to the large-scale vertical velocity. The scheme, originally described in Tiedtke [14], has evolved over time and among many changes includes a modified entrainment formulation leading to an improved representation of tropical variability of convection [15], and a modified CAPE closure leading to a significantly improved diurnal cycle of convection [16].

[3] ECMWF, Reading, UK, accessed July 25, 2016:
http://www.ecmwf.int/en/research/modelling-and-prediction/atmospheric-physics
http://www.ecmwf.int/en/research/modelling-and-prediction/atmospheric-physics.

2.2 Calculation of Probability of Convention/Clear Air

In order to fulfill the desired requirements, the following data processing for convection will be provided which can be applied individually or in a processing chain:

- Grid-based output of the Total Totals Index and the Convective Precipitation: using the ECMWF-ENS data, both convective indicators, TT and CP are given. The results of this workflow are the TT and CP for each member at the horizontal nodes of the desired sub-grid.
- Ensemble-based probability of convection/clear air for each grid point: with regard to flight trajectories, it is important to delimit regions of uncertain weather conditions from regions where the forecast is more reliable. Convective regions of high uncertainty can then be defined as those areas where neither convection nor clear air can be safely predicted. So, the calculation of the two quantities is suggested:
  - Probability of convection
  - Probability of clear air

Probability of convection the ensemble-based probability of convection is the fraction of ensemble members with values above the given thresholds TT_H and CP_H for all TT and CP of the ensemble members. For TT_H we suggest one of the threshold values given in Table 1. For CP_H we suggest 0; which means that any given amount of convective precipitation originates from convective events:

$$p_c = \frac{N_c}{N}, \tag{1}$$

where N is the numbers of ensemble members, $N_c = \sum_{i=1}^{N} i$, and so that $TT_i > TT_H \wedge CP_i > CP_H$.

Probability of clear air value that can show regions of clear air with low uncertainty:

$$p_{nc} = \frac{N_{nc}}{N}, \tag{2}$$

where N is the numbers of ensemble members, $N_{nc} = \sum_{i=1}^{N} i$, and so that $TT_i \leq TT_H \wedge CP_i \leq CP_H$.

Considering both values p_c and p_{nc} at each grid node, we are able to divide the focused area into 3 zones (see Fig. 3 for a schematic):

(1) Convective zones, i.e., high-risk areas with low uncertainty,
(2) Clear air zones with low uncertainty,
(3) Zones with high uncertainty.

With these two parameters (p_c & p_{nc}), further post-processing (e.g., classifications as described above) can be done.

High-risk areas for each ensemble member in order to get high-risk areas where each zone is based on the individual prediction of a single ensemble member, we look at the forecasted values of TT_i and CP_i at each horizontal grid node. In analogy to the

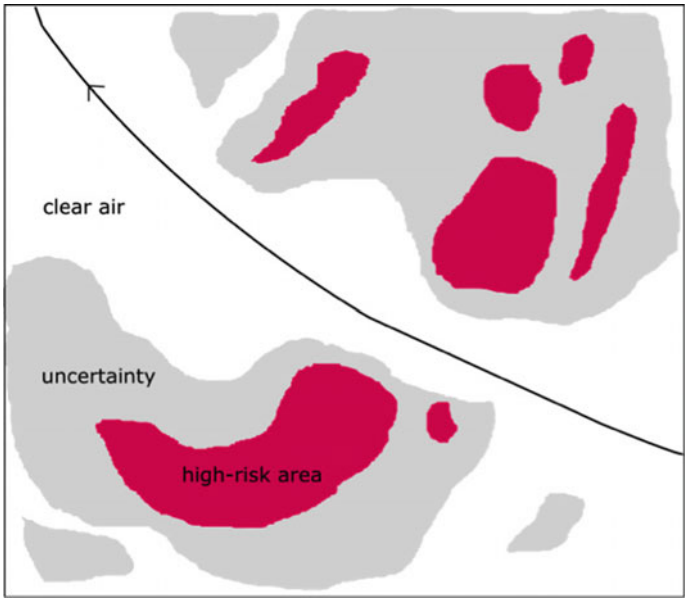


Fig. 3. Schematic illustration of the suggested classification of the focused area into 3 different zones: clear air (white), high-risk areas (pink), and uncertainty (grey)

ensemble-base probability of convection, we define a high-risk area for an ensemble member as an area where the following condition is fulfilled at each grid point: $TT_i > TT_H \vee CP_i > CP_H$.

That means that a high-risk area is delimited by the regions of low uncertainty which include the regions of high probability of convection. As the Total Totals Index is a smooth field, we suppose that we get clear structures of convective zones as well. Otherwise morphological operations can be applied to the generated field in order to eliminate unreliable singularities in the convective zones.

3 Trajectory Planning Methodology

Our approach will be based on a robust optimal control methodology for aircraft trajectory optimization problems [5]. We will here summarize the method and incorporate the developed convection metric.

3.1 Robust Optimal Control

We consider a dynamical system given by a randomly parametrized differential-algebraic equation with constraints. Uncertainty is described with the aid of a standard Kolmogorov probability space $(\Omega, \mathcal{F}, P)$; it is composed by a sample space of possible outcomes Ω, a σ-algebra of events $\mathcal{F}$ containing sets of outcomes, and the probability function P that assigns a probability to each of these events. The uncertain parameters

of the system will be modeled as a *constant* random variable $\xi : \Omega \to \mathbb{R}^{n_\xi}$. For each possible outcome $\omega \in \Omega$, the random variables take a different value $\xi(\omega)$.

We denote the state vector by $\mathbf{x} \in \mathbb{R}^{n_x}$, the control vector by $\mathbf{u} \in \mathbb{R}^{n_u}$, the algebraic variables by $\mathbf{z} \in \mathbb{R}^{n_z}$ and $t \in \mathbb{R}$ is the independent variable (usually time). For each outcome $\omega_0 \in \Omega$, there exist a unique trajectory path $t \to (\mathbf{x}(\omega_0, t), \mathbf{z}(\omega_0, t), \mathbf{u}(\omega_0, t))$ that corresponds to the realization of the random variables $\xi(\omega_0)$. The dynamics of the system are given by the functions $f : \mathbb{R}^{n_x} \times \mathbb{R}^{n_z} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_\xi} \times \mathbb{R} \to \mathbb{R}^{n_x}$, $h : \mathbb{R}^{n_x} \times \mathbb{R}^{n_z} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_\xi} \times \mathbb{R} \to \mathbb{R}^{n_h}$, and $g : \mathbb{R}^{n_x} \times \mathbb{R}^{n_z} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_\xi} \times \mathbb{R} \to \mathbb{R}^{n_g}$, such that valid trajectories fulfill the conditions almost surely (i.e., with probability 1)[4]:

$$\frac{d}{dt}\mathbf{x}(\omega, t) = f(\mathbf{x}(\omega, t), \mathbf{z}(\omega, t), \mathbf{u}(\omega, t), \xi(\omega), t), \tag{3}$$

$$h(\mathbf{x}(\omega, t), \mathbf{z}(\omega, t), \mathbf{u}(\omega, t), \xi(\omega), t) = 0, \tag{4}$$

$$g_L \leq g(\mathbf{x}(\omega, t), \mathbf{z}(\omega, t), \mathbf{u}(\omega, t), \xi(\omega), t) \leq g_U, \tag{5}$$

where $\omega \in \Omega$ is the sample point on the underlying abstract probability space. Therefore, for each possible scenario or realization of the random parameters $\xi(\omega)$, the trajectory will follow the deterministic differential equation (3) for the corresponding fixed value of ξ. We employ the notation $\mathbf{x}(\omega, t)$ and $\mathbf{u}(\omega, t)$ in order to emphasize the fact that the trajectory depends on the realization of the random parameters.

In addition to the random parameters, the trajectory is completely determined for a given control or guidance law. In some literature on robust optimal control ([17–20]), the employed control law depends only on time $\mathbf{u}(\omega, t) = \mathbf{u}_L(t)$, which corresponds to an "open-loop" control system. This formulation is not practical for general optimal control problems (including the problem that we consider), for several reasons. In first place, unstable dynamical systems would produce diverging trajectories which could enter undesirable regions of the state space. Additionally, final conditions cannot be applied and unique paths for some of the states cannot be achieved in general. These issues need to be addressed in the context of aircraft trajectory planning.

In [5], an optimal *guidance* scheme is introduced where some of the states are "tracked" (i.e. identical in all scenarios) and the controls are specific to each scenario in order to guarantee that the tracked states follow the unique computed trajectory. The associated practical concept demands that the controls can be computed online in this fashion in order to follow a guidance path; this condition is verified in aircraft trajectory planning, where the pilot or the autopilot can manage the aircraft in order to follow the route and the vertical profile.

We define the number of "control degrees of freedom" of the dynamical system as $d = n_z + n_u - n_h$. Let $q_x \leq \min\{n_x, d\}$ be the number of tracked states; without loss of generality, we can assume that the tracked states are the first q_x states (rearrange the state vector otherwise), i.e.,

[4] The $\leq$ sign applies in an element-wise fashion in Eq. (5) and analogous equations.

$$\mathbf{x} = \left[x_1 \ldots x_{q_x}\, x_{q_x+1} \ldots x_{n_x}\right]^T = \begin{bmatrix} \mathbf{x}_q \\ \mathbf{x}_r \end{bmatrix},$$

where $\mathbf{x}_q$ is the tracked part of the state vector and $\mathbf{x}_r$ is the untracked part. Let $\mathbb{I}_n$ be the identity matrix of shape $n \times n$ and $0_{n_1,n_2}$ be the zero matrix (i.e., a matrix with zeroes in all its entries) of shape $n_1 \times n_2$. We define the matrix $E_x \in \mathbb{R}^{q_x \times n_x}$ as

$$E_x = \left[\mathbb{I}_{q_x}\ 0_{q_x, n_x - q_x}\right].$$

This matrix transforms the state vector into the "tracked states" vector $\mathbf{x}_q = E_x\mathbf{x}$ that contains only the states whose evolution is equal in all scenarios. In this work, $q_x = d$ will hold so that there are enough tracked states to consume all available control degrees of freedom. This is not necessary in general.

With the aid of the tracking matrices, we can now define the tracking conditions (which, again, apply almost surely):

$$\begin{aligned} E_x(\mathbf{x}(\omega_1, t) - \mathbf{x}(\omega_2, t)) &= 0,\ \forall t,\ \forall \omega_1, \omega_2 \in \Omega, \\ E_z(\mathbf{z}(\omega_1, t) - \mathbf{z}(\omega_2, t)) &= 0,\ \forall t,\ \forall \omega_1, \omega_2 \in \Omega, \\ E_u(\mathbf{u}(\omega_1, t) - \mathbf{u}(\omega_2, t)) &= 0,\ \forall t,\ \forall \omega_1, \omega_2 \in \Omega. \end{aligned} \tag{6}$$

The tracking conditions enforce equality in the tracked variables between realizations: note that $E_x(\mathbf{x}(\omega_1, t) - \mathbf{x}(\omega_2, t))$ is the vector of differences between the tracked states in outcome ω_1 and the tracked states in outcome ω_2. The other two conditions are analogous tracking conditions for the dependent variables and the controls.

We are now ready to formulate the robust optimal control problem. Let $\mathbb{E}[\cdot]$ be the expectation operator associated to the probability space $(\Omega, \mathcal{F}, P)$. We define the terminal cost or "Mayer term" $\Phi : \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}$, the running cost or "Lagrange term" $\mathcal{L} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_z} \times \mathbb{R}^{n_u} \times \mathbb{R}^{n_\xi} \times \mathbb{R} \to \mathbb{R}$. We define the cost functional to minimize as:

$$J = \mathbb{E}\left[\Phi(t_0, t_f, \mathbf{x}(t_0), \mathbf{x}(t_f)) + \int_{t_0}^{t_f} \mathcal{L}(\mathbf{x}, \mathbf{z}, \mathbf{u}, \xi, t)dt\right]. \tag{7}$$

The initial conditions are:

$$\mathbf{x}(t_0) = \mathbf{x}_0 \tag{8}$$

We define the function $\Psi : \mathbb{R} \times \mathbb{R} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}$ that contains the final conditions:

$$\mathbb{E}\left[\Psi(t_f, \mathbf{x}(t_f))\right] = 0. \tag{9}$$

While these conditions are imposed in average, the ones that depend only on tracked states collapse to boundary conditions that are imposed exactly (as the value of the tracked states at the endpoints is unique); otherwise, they remain probabilistic constraints.

The objective J and the boundary conditions Ψ are written in terms of mean value, but they can be easily generalized to other statistics under the expected value formulation. For example, the variance of a function $G(\xi)$ can be written as $\mathbb{E}[(G - \mathbb{E}[G])^2] = \mathbb{E}[G^2] - \mathbb{E}[G]^2$ using expected values.

We also group the differential-algebraic equations and constraints (3)–(5):

$$\begin{aligned} \frac{d}{dt}\mathbf{x} &= f(\mathbf{x}, \mathbf{z}, \mathbf{u}, \xi, t), \\ h(\mathbf{x}, \mathbf{z}, \mathbf{u}, \xi, t) &= 0, \\ g_L \leq g(\mathbf{x}, \mathbf{z}, \mathbf{u}, \xi, t) &\leq g_U. \end{aligned} \tag{10}$$

The robust optimal control problem with tracking can now be defined as:

$$\left.\begin{array}{ll} \text{minimize} & J(7) \\ \text{subject to} & \text{differential-algebraic equations (10)} \\ & \text{boundary conditions (9)} \\ & \text{tracking conditions (6)} \end{array}\right\} \quad \text{(ROCT)}$$

3.2 Probabilistic Discretization

A key component of the methodology is what we call a "stochastic quadrature rule" (SQR), which discretizes the potentially continuous random variables into a (possibly weighted) discrete set of individual scenarios. We define it by a finite set of quadrature points $\{\xi_k\}$, $k \in \{1, \ldots, N\}$ and weights $\{w_k\}$, $k \in \{1, \ldots, N\}$, such that the stochastic integral $I = \int_\Omega g(\xi(\omega))d\omega$ is approximated with the sum:

$$Qg = \sum_{k=1}^{N} w_k g(\xi_k)$$

where $g(\xi)$ is an arbitrary function. By choosing the function $g(\cdot)$, we can obtain basic statistical quantities, such as averages and variances, as well as more advanced moments. A summary of the most relevant classes of stochastic quadrature rules can be found in [5]; in the present work, we employ a trivial quadrature rule (each ensemble member is a scenario with $w_k = 1/N$), as uncertainty is already characterized by discrete scenarios from EPS forecasts. However, the integration of additional sources of uncertainty may require the usage of a non-trivial SQR.

Let $\mathbf{x}_q(t) : \mathbb{R} \to \mathbb{R}^{q_x}$ define a trajectory for the tracked states, and $\mathbf{z}_q(t)$ and $\mathbf{u}_q(t)$ analogously. Suppose a SQR has been chosen, with a number of points N. For each one of these points ξ_k, the tracking trajectory $(\mathbf{x}_q, \mathbf{z}_q, \mathbf{u}_q)(t)$ defines a unique trajectory given a full set of initial conditions; we will now collect each one of these N trajectories in a *trajectory ensemble*. We define the trajectory ensemble associated to a tracking trajectory $(\mathbf{x}_q, \mathbf{z}_q, \mathbf{u}_q)(t)$ as the set of trajectories $\{(\mathbf{x}_k, \mathbf{z}_k, \mathbf{u}_k)(t)\}$ with $k \in \{1, \ldots, N\}$

such that the trajectory k is generated by the initial conditions $\mathbf{x}_k(t_0) = \mathbf{x}_0$ and the tracking trajectory with $\xi = \xi_k$, i.e.,

$$\begin{aligned} \frac{d}{dt}\mathbf{x}_k &= f(\mathbf{x}_k, \mathbf{z}_k, \mathbf{u}_k, \xi_k, t), \\ h(\mathbf{x}_k, \mathbf{z}_k, \mathbf{u}_k, \xi_k, t) &= 0, \\ g_L \leq g(\mathbf{x}_k, \mathbf{z}_k, \mathbf{u}_k, \xi_k, t) &\leq g_U, \\ E_x\mathbf{x}_k(t) &= \mathbf{x}_q(t), \\ E_z\mathbf{z}_k(t) &= \mathbf{z}_q(t), \\ E_u\mathbf{u}_k(t) &= \mathbf{u}_q(t). \end{aligned} \tag{11}$$

We will now build a virtual dynamical system where the state vectors of all the trajectories in the trajectory ensemble are merged into a single large state vector, which represents the "collective trajectory". Its state vector $\mathbf{x}_E \in \mathbb{R}^{n_x N}$ contains the state vector of all the trajectories in the ensemble (the control $\mathbf{u}_E \in \mathbb{R}^{n_u N}$ and algebraic $\mathbf{z}_E \in \mathbb{R}^{n_z N}$ vectors are analogous), and each individual trajectory follows the dynamics associated to each scenario (with $\xi = \xi_k$).

$$\mathbf{x}_E = \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_N \end{bmatrix}; \mathbf{z}_E = \begin{bmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_N \end{bmatrix}; \mathbf{u}_E = \begin{bmatrix} \mathbf{u}_1 \\ \vdots \\ \mathbf{u}_N \end{bmatrix}. \tag{12}$$

We define the differential equation, algebraic equations and inequality constraints of this augmented dynamical system as:

$$f_E(\mathbf{x}_E, \mathbf{z}_E, \mathbf{u}_E, t) = \begin{bmatrix} f(\mathbf{x}_1, \mathbf{z}_1, \mathbf{u}_1, \xi_1, t) \\ \vdots \\ f(\mathbf{x}_N, \mathbf{z}_N, \mathbf{u}_N, \xi_N, t) \end{bmatrix}, \tag{13}$$

$$h_E(\mathbf{x}_E, \mathbf{z}_E, \mathbf{u}_E, t) = \begin{bmatrix} h(\mathbf{x}_1, \mathbf{z}_1, \mathbf{u}_1, \xi_1, t) \\ \vdots \\ h(\mathbf{x}_N, \mathbf{z}_N, \mathbf{u}_N, \xi_N, t) \end{bmatrix}, \tag{14}$$

$$g_E(\mathbf{x}_E, \mathbf{z}_E, \mathbf{u}_E, t) = \begin{bmatrix} g(\mathbf{x}_1, \mathbf{z}_1, \mathbf{u}_1, \xi_1, t) \\ \vdots \\ g(\mathbf{x}_N, \mathbf{z}_N, \mathbf{u}_N, \xi_N, t) \end{bmatrix}. \tag{15}$$

This dynamical system approximates an uncertain system, but it is deterministic. This fact implies that we can use it to formulate an approximation of the robust optimal

control problem. Let us define the approximate cost functional, divided into Mayer and Lagrange terms, using the trajectory ensemble:

$$J_E = \Phi_E(\mathbf{x}_E(t_0), \mathbf{x}_E(t_f)) + \int_{t_0}^{t_f} \mathcal{L}_E(\mathbf{x}_E, \mathbf{z}_E, \mathbf{u}_E, t)dt, \tag{16}$$

$$\Phi_E(\mathbf{x}_E(t_0), \mathbf{x}_E(t_f)) = \sum_{k=1}^{N} w_k \phi(\mathbf{x}_k(t_0), \mathbf{x}_k(t_f)), \tag{17}$$

$$\mathcal{L}_E(\mathbf{x}_E, \mathbf{z}_E, \mathbf{u}_E, t) = \sum_{k=1}^{N} w_k \mathcal{L}(\mathbf{x}_k, \mathbf{z}_k, \mathbf{u}_k, t); \tag{18}$$

and discretize the boundary conditions as

$$\Psi_E(t_f, \mathbf{x}_E(\omega, t_f)) = \sum_{k=1}^{N} w_k \Psi(t_f, \mathbf{x}_k(t_f)). \tag{19}$$

For concise writing of the discretization of the tracking conditions (6), we will define the matrix $E_x^N \in \mathbb{R}^{q_x(N-1)\times n_x N}$ as:

$$E_x^N = \begin{bmatrix} E_x & & \\ & \ddots & \\ & & E_x \end{bmatrix} \begin{bmatrix} \mathbb{I}_{n_x} & -\mathbb{I}_{n_x} & & & \\ & \mathbb{I}_{n_x} & -\mathbb{I}_{n_x} & & \\ & & \ddots & \ddots & \\ & & & \mathbb{I}_{n_x} & -\mathbb{I}_{n_x} \end{bmatrix}. \tag{20}$$

$E_z^N \in \mathbb{R}^{q_z(N-1)\times n_z N}$ and $E_u^N \in \mathbb{R}^{q_u(N-1)\times n_u N}$ can be defined in analogous fashion. These matrices map the ensemble state vector to the differences in the tracked states between trajectories.

Making use of Eqs. (11)–(15) as well as Eqs. (16)–(20), we can complete now the formulation of the deterministic approximant:

$$\left.\begin{array}{ll} \text{minimize} & J_E \\ \text{subject to} & \dot{\mathbf{x}}_E = f_E(\mathbf{x}_E, \mathbf{z}_E, \mathbf{u}_E, t) \\ & h_E(\mathbf{x}_E, \mathbf{z}_E, \mathbf{u}_E, t) = 0 \\ & I_G g_L \le g_E(\mathbf{x}_E, \mathbf{z}_E, \mathbf{u}_E, t) \le I_G g_U \\ & E_x^N \mathbf{x}_E = 0 \\ & E_z^N \mathbf{z}_E = 0 \\ & E_u^N \mathbf{u}_E = 0 \\ & \psi_E(t_0, t_f, \mathbf{x}_E(\omega, t_0), \mathbf{x}_E(\omega, t_f)) = 0 \end{array}\right\} \quad \text{(DROCT)}$$

where $I_G = [\mathbb{I}_{n_g} \dots \mathbb{I}_{n_g}]^T \in \mathbb{R}^{n_g N \times n_g}$.

Note that, in practical implementation, defining and modeling the tracked variables in every scenario are not necessary as they are equal; as a consequence, implementation of the tracking constraints $E_x^N \mathbf{x}_E = 0$, $E_z^N \mathbf{z}_E = 0$ and $E_u^N \mathbf{u}_E = 0$ can also be omitted.

3.3 Application

We will now apply this methodology to find routes that minimize a weighted sum of average flight time, flight time dispersion (weighted with the "dispersion penalty" parameter *DP*) and convection risk (weighted with the "convection penalty" parameter *CP*). Let $\kappa(\phi, \lambda)$ be the smoothed and interpolated probability of convection and let the exposure to convection be:

$$EC = \int_{s_0}^{s_f} \kappa(\phi(s), \lambda(s)) ds \tag{21}$$

The cost functional to minimize is then

$$J = \mathbb{E}[t_f] + \text{DP} \underbrace{(t_{f,max} - t_{f,min})}_{\text{uncertainty in } t_f} + \text{CP} \underbrace{\int_{s_0}^{s_f} \kappa(\phi(s), \lambda(s)) ds}_{\text{exposure to convection}} \tag{22}$$

Note that $(t_{f,max} - t_{f,min})$ is the difference between the earliest and the latest arrival time. We also highlight the fact that, at constant fixed airspeed and altitude, fuel burn is mostly determined by flight time, so a "fast in average" solution (i.e., a solution with low $\mathbb{E}[t_f]$) will also be "efficient in average" (i.e., the average fuel burn will be low) and vice versa. For the same reason, a solution with high arrival time uncertainty (a high value of $t_{f,max} - t_{f,min}$) will also have high uncertainty in fuel burn.

With this cost functional, the parameters *DP* and *CP* regulate the solution flight plan depending on the preferences of the flight planner. High values of *DP* will produce more predictable trajectories absent a convection-related reroute by avoiding regions where the wind is more unpredictable. High values of *CP* will produce trajectories that are less likely to be rerouted by avoiding regions where there is high likelihood of convection. Low values of the parameters will produce flight plans that are more efficient in average, but less predictable.

We consider a free-routing airspace (the lateral profile is not restricted to a fixed set of airways) and a 3-DoF point-mass model of aircraft used widely in ATM studies, BADA 3 [21]. We will restrict ourselves to the analysis of the cruise phase for the sake of simplicity (note that the impact of wind forecast uncertainty is cumulative, and thus more important for longer flights, and the cruise phase comprises most of a medium-haul or long-haul flight). In addition, we assume constant flight level and airspeed but we will note that our methodology can be extended to full 4D problems. We consider an ellipsoidal Earth as in the WGS84 model, with radii of curvature of ellipsoid meridian and prime vertical denoted by R_M and R_N, respectively. We take the wind and temperature fields from an EPS forecast and compute density with the ideal gas law (as the pressure is determined by the flight level).

We make a further simplification by taking the heading as a control variable instead of the bank angle, thus allowing it to change instantaneously. Therefore, the dynamics of the system are described by the following system of differential equations (f in Eq. (10)):

$$\dot{\mathbf{x}} = \frac{d}{dt}\begin{bmatrix}\phi\\ \lambda\end{bmatrix} = \begin{bmatrix}(R_N+h)^{-1}(v\cos(\chi)+w_x(\phi,\lambda,t))\\ (R_M+h)^{-1}\cos^{-1}(\phi)(v\sin(\chi)+w_y(\phi,\lambda,t))\end{bmatrix}, \tag{23}$$

where ϕ is the latitude, λ is the longitude, v is the true airspeed, h is the geodetic altitude, χ is the heading, and w_x and w_y are the zonal and meridional components of the wind.[5] The control vector is composed by the heading χ.

We reformulate this dynamical system as a differential-algebraic system (DAE) (see discussion in [5]) with the addition of the ground speed v_G as an algebraic variable and the course ψ as a control variable, linked to the remaining variables by two new equality constraints (check Fig. 4 for a graphical explanation of Eq. 25). The reformulated system is given by the dynamical function:

$$\frac{d}{dt}\begin{bmatrix}\phi\\ \lambda\end{bmatrix} = \begin{bmatrix}(R_N+h)^{-1}v_G\cos\psi\\ (R_M+h)^{-1}\cos^{-1}(\phi)v_G\sin\psi\end{bmatrix}, \tag{24}$$

and the equality constraints (h in Eq. (10)):

$$\begin{aligned} v_G\cos\psi &= v\cos(\chi)+w_x(\phi,\lambda,t),\\ v_G\sin\psi &= v\sin(\chi)+w_y(\phi,\lambda,t). \end{aligned} \tag{25}$$

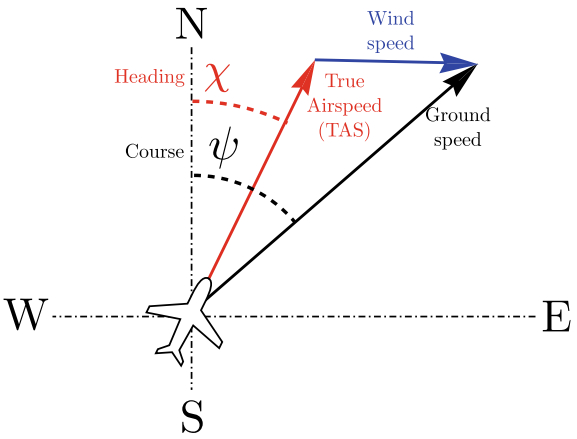


Fig. 4. Relationship between airspeed, groundspeed, wind, heading and course

[5] Contrary to the usual definition, we take w_y to be in a South to North direction.

We add the inequality constraint:

$$v_G \geq 0 \tag{26}$$

to ensure uniqueness of v_G and ψ (otherwise, $(-v_G^*, \psi^* + \pi/2)$ would produce the same left-hand side of Eq. (25) as (v_G^*, ψ^*)).

We will now perform a coordinate transformation in order to employ distance flown along the route (s) as the independent variable. This allows us to apply our methodology in a manner that is consistent with existing planning and flight procedures (again, see the discussion in [5]). As a consequence, the time t becomes a state variable and the new dynamical function can be obtained by dividing the time derivatives by $ds/dt = v_G$:

$$\frac{d}{ds}\begin{bmatrix}\phi\\ \lambda\\ t\end{bmatrix} = \begin{bmatrix}(R_N + h)^{-1}\cos\psi\\ (R_M + h)^{-1}\cos^{-1}(\phi)\sin\psi\\ v_G^{-1}\end{bmatrix}. \tag{27}$$

All constraints remain the same as in the untransformed system of differential-algebraic equations.

We can now discretize the uncertainty and create the trajectory ensemble. An ensemble forecast contains a set of ensemble members, each one defining a different wind forecast (and, therefore, different functions w_x and w_y). If the ensemble contains N members, we define N scenarios, each one having weight $w_k = 1/N$ and the wind function that corresponds to the respective member; our stochastic quadrature rule is, therefore, a simple empirical average. We will write the compact form of the trajectory ensemble directly.

We choose to track the course ψ, i.e., the function $\psi(s)$ is the same in every scenario (thus, we do not need to implement scenario-specific versions). As a consequence of (27), this implies that the evolution of the latitude ϕ is unique (as it only depends on the evolution of the unique variable ψ) and λ is also unique (as it only depends on ϕ and ψ). Therefore, the position variables act like tracked variables too, which is both relevant from the implementation point of view (because we do not need to create copies of them for each scenario) and a desired goal (since we want to obtain a unique route). Taking advantage of these manipulations, we can define the dynamical system associated to the trajectory ensemble with the dynamical function:

$$\frac{d}{dr}\begin{bmatrix}\phi\\ \lambda\\ t_1\\ \vdots\\ t_N\end{bmatrix} = \begin{bmatrix}\dfrac{\cos(\chi_G)}{R_N + h}\\ \dfrac{\sin(\chi_G)}{(R_M + h)\cos\phi}\\ 1/v_{GS,1}\\ \vdots\\ 1/v_{GS,N}\end{bmatrix} \tag{28}$$

$$\begin{bmatrix} v_{\mathrm{GS},1}\cos(\chi_{\mathrm{G}}) \\ \vdots \\ v_{\mathrm{GS},N}\cos(\chi_{\mathrm{G}}) \\ v_{\mathrm{GS},1}\sin(\chi_{\mathrm{G}}) \\ \vdots \\ v_{\mathrm{GS},N}\sin(\chi_{\mathrm{G}}) \end{bmatrix} = \begin{bmatrix} v_{\mathrm{TAS}}\cos(\chi_i) + w_{y,1}(\phi,\lambda) \\ \vdots \\ v_{\mathrm{TAS}}\cos(\chi_i) + w_{y,N}(\phi,\lambda) \\ v_{\mathrm{TAS}}\sin(\chi_i) + w_{x,1}(\phi,\lambda) \\ \vdots \\ v_{\mathrm{TAS}}\sin(\chi_i) + w_{x,N}(\phi,\lambda) \end{bmatrix} \tag{29}$$

Finally, we complete the definition of the discretized robust optimal control problem by adding the boundary conditions:

$$(\phi(0),\lambda(0)) = (\phi_0,\lambda_0) \tag{30}$$

$$(\phi(s_f),\lambda(s_f)) = (\phi_f,\lambda_f) \tag{31}$$

$$t_i(0) = 0 \forall i \in \{1,\ldots,N\} \tag{32}$$

$$t_{f,\min} \leq t_i(r_f) \leq t_{f,\max} \forall i \in \{1,\ldots,N\} \tag{33}$$

where $t_{f,\min}$ and $t_{f,\max}$ are scalar decision variables.

Note again that true airspeed and barometric altitude are assumed to be constant and thus, minimizing time is closely related to minimizing fuel consumption. Therefore, fuel consumption can be readily obtained after optimization by integrating the following differential equation:

$$\dot{m} = \eta(V)\cdot T, \tag{34}$$

where η is the thrust-specific fuel consumption and taking into account the fact that $T = D$. Both the drag and the thrust can be modeled with an aircraft performance model (in our case, BADA).

We solve this problem with standard direct collocation methods [22], discretizing the trajectory with a numerical scheme and solving the resulting nonlinear optimization problem with NLP software.

4 Case Study

4.1 Description and Statement

We consider an BADA3 A330 Aircraft model flying from the vertical of New York ($-73.8°$, $40.6°$) to the vertical of Argel ($3.2°$, $36.7°$) at constant barometric altitude 200 hPa and Mach 0.82. Initial mass has been set to 200 tons. We use a forecast for a pressure of 200 hPa 9 h in advance for the 19th of December, 2016 from the ECMWF

ensemble, elaborated by the European Center for Medium-Range Weather Forecasts (ECMWF)[6] with 51 members. We rely on the CasADi library [23] as NLP interface [24] and IPOPT [25] as NLP solver.

4.2 Results and Discussion

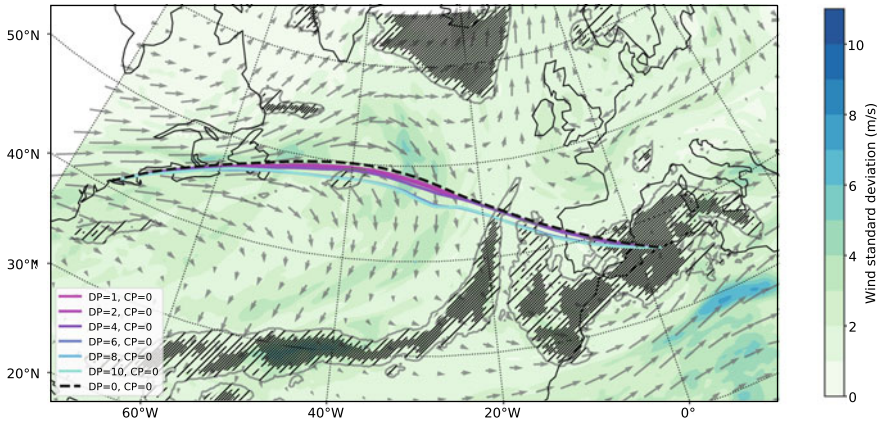
Figure 5 displays the geographical routes for different values of *DP* and *CP*. It can be seen that routes computed with higher *DP* (setting $CP = 0$) tend to avoid the high uncertainty zone in the North Atlantic in order to increase predictability, at the cost of taking a more indirect route that is longer on average. It can be also observed that routes computed with higher *CP* (setting $DP = 0$) tend to reduce the exposure to convective risk zones, again at the cost of taking a more indirect route. Four selected flight plans produced with different *CP*-*DP* pairs are shown in Fig. 6. It is important to remark that the exposure to convective risk areas can increase expected delay because of increased (tactical) ATFM regulations or ATC advisories to avoid developed storms. The characterization of this relationship is the target of future work.

The evolution of time spreads and convective exposure over the flown distance is presented in Fig. 7. Comparing the maximum average efficiency trajectory (corresponding to $DP = 0$ and $CP = 0$) with a more predictable trajectory, e.g., that of $DP = 6$ and $CP = 0$, it can be seen how the spread in the ensemble times increases slightly when the aircraft crosses the area of high uncertainty in the middle of the North Atlantic. Similarly, comparing the maximum average efficiency trajectory with another with less exposure to convection, e.g., that of $DP = 0$ and $CP = 0.01$, it can be readily seen in Fig. 7 that the exposure (the integral, or area below the curve) is reduced. For the minimum average fuel case ($DP = 0$), the time dispersion at the final fix is around 200s, whereas for the $DP = 6$ case, the time dispersion at the final fix is around 100s. In other words, around one and a half minutes reduction in *time uncertainty* could be achieved by flying a more predictable trajectory ($DP = 6$). This improvement comes at the cost of extra flight time and associated fuel burnt on average.

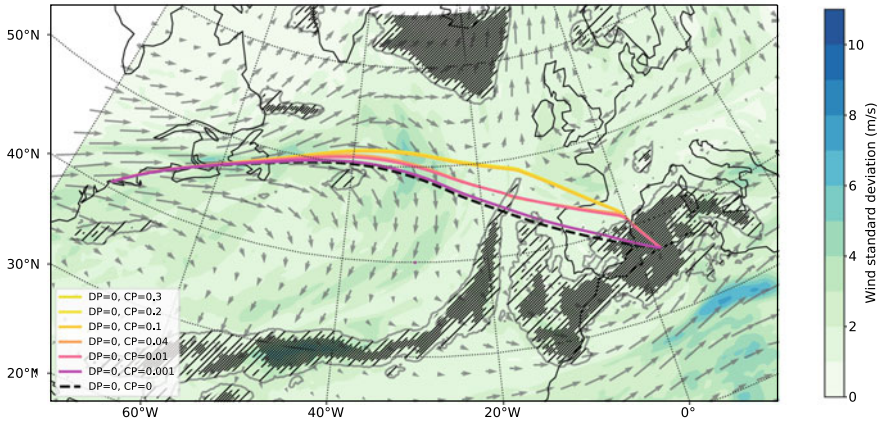
Figure 8 presents different Pareto frontiers (for different pre-set average flight times) and shows different possible solutions with trade-offs between time spread, accumulated convection, and arrival time (directly related to consumption). In other words, and in order to make visible the trade-off effects of dispersion and convection, we have solved the problem with an additional constraint that enforces aircraft's final average time to be a pre-set one. For the sake of illustration, we present result for average final times set equal to 395; 400; and 405 min.

Quantitative values of flight dispersion and exposure to convection can be readily extracted out of Fig. 8. Consider the flight constraint to reach the final fix at an average flight time of 400 min (orange line in Fig. 8a): for a dispersion of roughly 1.5 min the exposure to convective areas would be of roughly 1000 e-km (equivalent kilometers, i.e., kilometers flown at probability of convection equal to one). If ones want to reduce this exposure to convection to roughly 500 e-km, it comes at a cost of an extra 1.5 min (3

[6] http://apps.ecmwf.int/datasets/.



(a) Optimal paths with different DP values (CP=0).



(b) Optimal paths with different CP values (DP=0).

Fig. 5. Optimal trajectories for *DP* and *CP* values. Higher brightness in the trajectory color indicates higher values of the penalty. Color contour scale indicates wind uncertainty characterized as $\sqrt{\sigma_u^2 + \sigma_v^2}$, with σ_u being the standard deviation of the u component of wind across different members and σ_v analogous for the v-component. Dashed regions indicate regions of convective exposure

in total) of flight dispersion. The same information represented in a different fashion, has been included in Fig. 8b. Nevertheless, note that these numbers correspond to this particular case (route, day, and weather forecasts) and may or may not be representative of characteristic costs and benefits. Further studies should assess these quantities in a more systematic fashion.

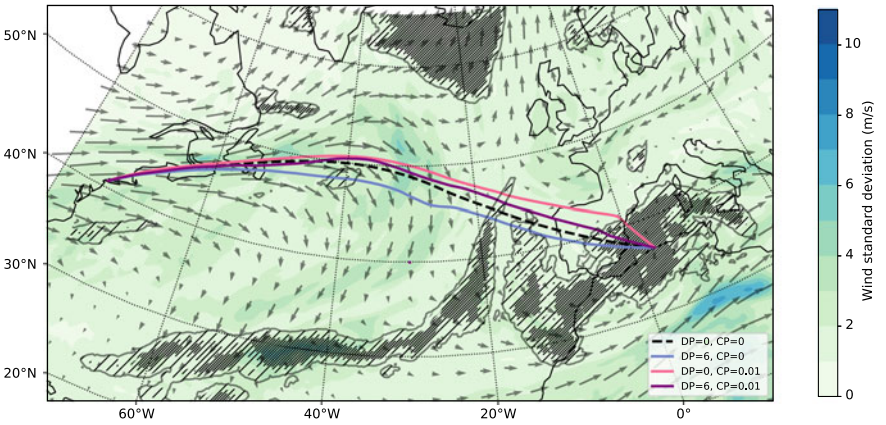


Fig. 6. Optimal paths for different DP-CP pairs

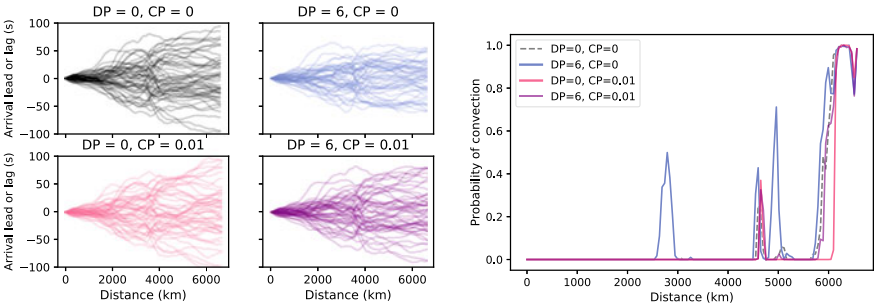
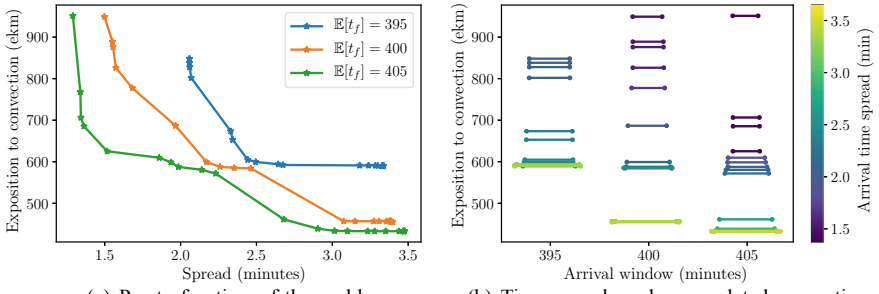


Fig. 7. Evolution of time spreads and convective exposure



(a) Pareto frontiers of the problem. (b) Time spreads and accumulated convection.

Fig. 8. Variable relations in the problem. Pareto frontiers (spread versus convection exposure) for three different arrival times

5 Conclusions and Future Work

A robust optimal control methodology has been used for computing efficient and predictable routes based on Ensemble Prediction Systems, including an approach to calculate risk of convection. This risk, a necessary though not sufficient condition for the formation of storms, has been included in the objective functional of the robust optimal control problem. We have demonstrated its utility in studying trade-offs between efficiency, predictability (measured in terms of dispersion in the final time) and exposure to convention. We can conclude that uncertainty (in this case due to wind) can not only be quantified, but also reduced by proposing alternative trajectories. In addition, we find that convective areas can be avoided at the cost of efficiency and predictability. This algorithm could prove useful for both flight dispatchers (demand side of the problem) and networks managers (capacity side of the problem) for the design of a more robust ATM system in order to improve ATM performance in terms of safety, capacity (delays), and environmental impact.

Acknowledgments. This work has been partially supported by project TBO-MET project (https://tbomet-h2020.com/), which has received funding from the SESAR JU under grant agreement No 699294 under European Union's Horizon 2020 research and innovation programme.

References

1. Casado E, Goodchild C, Vilaplana M. Sensitivity of trajectory prediction accuracy to aircraft performance uncertainty. In: Proceedings of the AIAA INFOTECH@AEROSPACE (I@A) CONFERENCE
2. Vazquez R, Rivas D (2013) Propagation of initial mass uncertainty in aircraft cruise flight. J Guid, Control, Dyn 36(2):415–429
3. Casado E, Vilaplana M, Goodchild C (2013) Sensitivity of continuous climb departure predictions to aircraft intent uncertainties. In: 3th international conference on application and theory of automation in command and control systems
4. Gonzalez-Arribas D, Soler M, Sanjurjo M (2016) Wind-based robust trajectory optimization using meteorological ensemble probabilistic forecasts. In: SESAR innovation days 2016 (SIDs 2016)
5. Gonzalez-Arribas D, Soler M, Sanjurjo-Rivo M (2018) Robust aircraft trajectory planning under wind uncertainty using optimal control. J Guid, Control, Dyn 41(3):673–688. https://doi.org/10.2514/1.G002928
6. Hacker JP, Krayenhoff ES, Stull RB (2003) Ensemble experiments on numerical weather prediction error and uncertainty for a north pacific forecast failure. Weather Forecast 18 (1):12–31
7. Bauer P, Thorpe A, Brunet G (2015) The quiet revolution of numerical weather prediction. Nature 525(7567):47–55
8. Cheung J, Brenguier J-L, Heijstek J, Marsman A, Wells H (2014) Sensitivity of flight durations to uncertainties in numerical weather prediction. In: SESAR innovation days
9. Cheung J, Hally A, Heijstek J, Marsman A, Brenguier J-L (2015) Recommendations on trajectory selection in flight planning based on weather uncertainty. In: Proceedings of 5th SESAR innovation days (SID2015), Bologna, Italy, pp 1–8

10. Vazquez R, Rivas D, Franco A (2017) Stochastic analysis of fuel consumption in aircraft cruise subject to along-track wind uncertainty. Aerosp Sci Technol 66:304–314
11. Legrand K, Puechmorel S, Delahaye D, Zhu Y (2016) Aircraft trajectory planning under wind uncertainties. In: Digital Avionics Systems Conference (DASC), 2016 IEEE/AIAA 35th. IEEE, pp 1–9
12. Carbone R, Tuttle J, Ahijevych D, Trier S (2002) Inferences of predictability associated with warm season precipitation episodes. J Atmos Sci 59(13):2033–2056
13. Ellrod G, Field G (1984) The characteristics and prediction of gulf stream thunderstorms. In: Conference on Weather Forecasting and Analysis, 10 th Clearwater Beach, FL, pp 15–21
14. Tiedtke M (1989) A comprehensive mass flux scheme for cumulus parameterization in large-scale models. Mon Weather Rev 117(8):1779–1800
15. Bechtold P, Köhler M, Jung T, Doblas-Reyes F, Leutbecher M, Rodwell MJ, Vitart F, Balsamo G (2008) Advances in simulating atmospheric variability with the ecmwf model: From synoptic to decadal time-scales. Q J R Meteorol Soc 134(634):1337–1351
16. Bechtold P, Semane N, Lopez P, Chaboureau J-P, Beljaars A, Bormann N (2014) Representing equilibrium and nonequilibrium convection in large-scale models. J Atmos Sci 71(2):734–753
17. Ross IM, Proulx RJ, Karpenko M, Gong Q (2015) Riemann–stieltjes optimal control problems for uncertain dynamic systems. J Guid, Control, Dyn 1–13
18. Li X, Nair PB, Zhang Z, Gao L, Gao C (2014) Aircraft robust trajectory optimization using nonintrusive polynomial chaos. J Aircr 51(5):1592–1603
19. Fisher J, Bhattacharya R (2011) Optimal trajectory generation with probabilistic system uncertainty using polynomial chaos. J Dyn Syst, Meas, Control 133(1):014501
20. Flanzer TC, Bower GC, Kroo IM (2012) Robust trajectory optimization for dynamic soaring. In: AIAA guidance, navigation, and control conference, p 4603
21. Nuic A (2010) User manual for the base of aircraft data (bada) rev 3.11. Atmosphere 2010:001
22. Betts JT (2010) Practical methods for optimal control and estimation using nonlinear programming, vol 19. Siam
23. Andersson J, Åkesson J, Diehl M (2012) Casadi: A symbolic package for automatic differentiation and optimal control. In: Recent advances in algorithmic differentiation. Springer, Berlin, pp 297–307
24. Hart WE, Watson J-P, Woodruff DL (2011) Pyomo: modeling and solving mathematical programs in python. Math Program Comput 3(3):219–260
25. Wächter A, Biegler LT (2006) On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math program 106(1):25–57

Optimal Location of Dynamic Military Areas Within Civil Aviation Traffic

N. Wang(✉), D. Delahaye, M. Mongeau, and A. Gondran

ENAC, Université de Toulouse, Toulouse, France
{ningwang, delahaye, mongeau, gondran}@recherche.enac.fr

Abstract. In this study, we focus on the problem of locating optimally dynamic military areas with the aim of minimizing the number of civil flight trajectories potentially impacted by the military activity, and the distance between the military area and the military base. We model the military areas by 2D geometry shapes with a vertical extension associated with given flight levels during the temporary area-activation time window. We propose a mathematical formulation of this problem as a constrained-optimization problem. We then introduce a global-optimization methodology based on a simulated annealing algorithm featuring tailored neighborhood-search strategies and an astute computational evaluation of the otherwise costly objective function. This is applied to 1-day of French traffic involving 8,836 civil flights. The results show that the proposed method is efficient to locate the military area that is nearest to the military base, while minimizing the potential impact on civil flight trajectories.

Keywords: Dynamic military area · Optimization · Location problem · Simulated annealing algorithm

1 Introduction

Along with the fast development of aviation technology, air traffic experienced its highest growth over the past 5 years [1]. Traditionally, prohibited airspace or no-fly areas, where no civil flights are allowed at any time, are established for security reasons or military activities. Figure 1 shows the real potential military areas (represented by the pink and orange polygons) in the lower French airspace. If we zoom in around the Charles de Gaulle airport (Paris), seven potential military areas are found, represented in Fig. 2 by yellow polygons of various shapes surrounded by fences, and whose dimensions range from 40 to 120 NM. These areas are placed so as to avoid the main civilian traffic flow (represented by different colors of lines). However potential conflicts still exist for a large number of trajectories. Therefore, military areas have become significant constraints for civilian traffic.

Electronic Navigation Research Institute, *Air Traffic Management and Systems III*, Lecture Notes in Electrical Engineering 555,
https://doi.org/10.1007/978-981-13-7086-1_7

Fig. 1. Potential military areas in lower French airspace

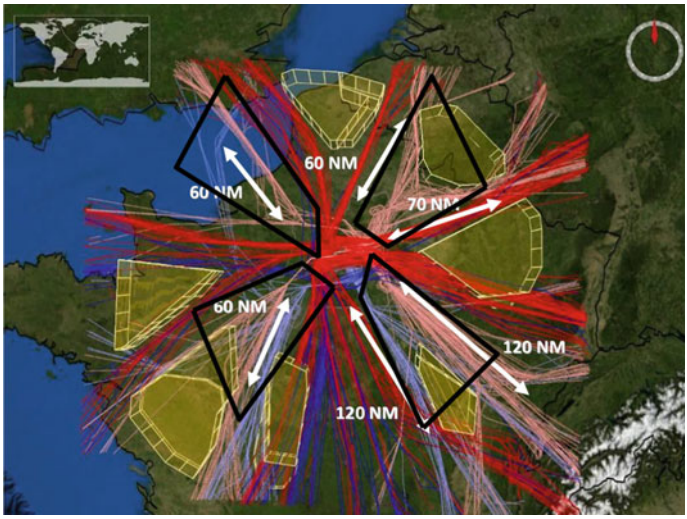


Fig. 2. Military areas (yellow polygons surrounded by fences) around Charles de Gaulle airport, Paris

To meet the increasing civilian traffic demand and to improve airspace utilization, implementing temporary/dynamic prohibited areas to perform military training or missions for short periods of time seems to be a better choice. A civil–military coordination program is thereby necessary for airspace planners to take the military activities into consideration. In a static airspace planning process, the military part provides the time and location information of the upcoming military missions online one or several days before the operation. Airspace planners take these constraints into account to come up with an operational airspace use plan and then publish it. Then, any changes after publication such as weather changes, cancelations of missions, etc., might not be captured, leading thereby to wastage of resources. To realize flexible airspace management, a dynamic airspace planning process is preferred: the coordination process continues after the airspace use plan is published, until a time much closer to the time of operations. This yields increased benefit and better use of resources through more accurate data while maintaining full flexibility for military operations. In this case, airspace management is performed at the tactical level.

Ideally, the airspace planning platform is online and in real time. It should allow civil and military parts for airspace bookings and coordinations. It can also carry out analyzing functionalities such as conflict detection, automation of tasks such as NOTAM (Notice to Airmen) requests for pilots, airspace use plan drafting, and visualization display for airspace planners. With such an airspace planning platform, common situation awareness is available for tactical civil–military coordination. Real-time activation and deactivation of airspace are viable based on planning and acknowledgments of the air traffic controllers in charge. Finally, airspace status can be displayed on different interfaces of the ATM system.

The left part of Fig. 3 displays a scenario in which civilian traffic flow (represented by white two-way arrows) overlaps a military area (represented by the blue polygon). To tackle this issue, various methods are proposed: one can divide the current military area into subsets, consider for example the three parts A, B, and C, as presented by Fig. 3 left. One can then activate only the A, C parts during the time (e.g., at night) during which civilian traffic is not active. One can then relocate the military area by airspace discretization, and find the grids of airspace that civilian traffic does not occupy (Fig. 3 center). One can also relocate the military area dynamically in time and space, taking the civilian traffic into consideration (Fig. 3 right). The first two methods are relatively easy to implement, while having recourse to the Dynamic Military Areas (DMA) constitutes a scientific challenge. In this study, we carry out a preliminary study on such a DMA approach. The objective of this study is to allow optimal military airspace reservations in time, flight levels, and/or geographical location without limiting mission effectiveness while enabling more efficient use of the airspace for civilian flights.

In this article, we concentrate on DMA for military missions that rely on designing an airspace volume that envelopes the whole military mission trajectories. It features user-defined lateral and vertical dimensions, along with user-defined activation-time duration. Our decision variables include the 3D geographical location of the military area related to the military base position. The activation time is also an optimization

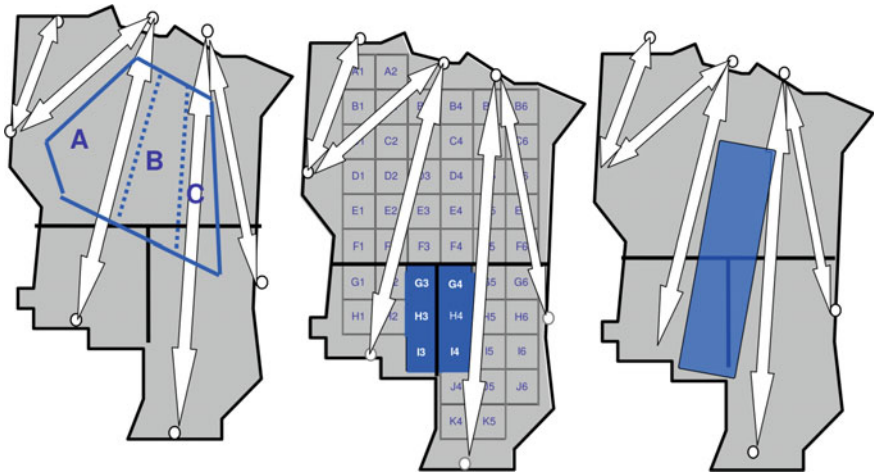


Fig. 3. Variable and dynamic military areas

variable, it can be chosen within a user-defined time window. These parameters will be computed so as to minimize both the amount of civilian flights overlapping the military area (at the strategic level) and the distance between the DMA and the military base.

To our knowledge, there is no previously published study on the military area location problem. Nevertheless, one is tempted to consider our problem as a special case of the single facility location problem. For this type of problem, "quick and dirty" methods are recommended [2]. For our problem, approximate models are used due to limited input data available for decision makers. Contrary to the classical facility problem, here the facility is considered "toxic", therefore one aims at minimizing the harmful effect (the impact on civil trajectories in our problem) inside the facility (the military area) coverage range. Another specific feature of our problem is the fact that it deals with *continuous* candidate locations. The first survey on continuous location problems was conducted by F. Plastria. In his work [3], exhaustive topics on continuous location theory are discussed. A recent survey on continuous location-allocation problems was performed by Brimberg [4] in 2008; his work confirms that heuristic methods contributed significantly to improving the methods addressing this type of problem.

The remaining of this article is organized as follows: in Sect. 2, a mathematical model for the military area optimal location problem is proposed. Section 3 introduces a simulated annealing algorithm to address this problem with dedicated neighborhood operators. Several test results on the French airspace are presented, and the performance of the algorithm is analyzed in Sect. 4. Finally, conclusions are drawn along with future perspectives for military area location problems.

2 Mathematical Model

In France, there are more than 8,000 civil flights in 1 day. Their trajectories cover almost the whole airspace above the French territory, as shown in Fig. 4. An example of a location of Military Area (noted as MA in the following sections) is represented by the green rectangle in this figure, and the position of the military base is the center of the blue circle. Planned flights passing through this rectangle during the activation time of the MA are affected and have to be modified. The MA is to be relocated inside this circle at different time windows so as to minimize the impacts on the civil flights planned during this same period. Moreover, from the fuel consumption saving point of view, minimizing the distance between the MA and its military base is also desirable.

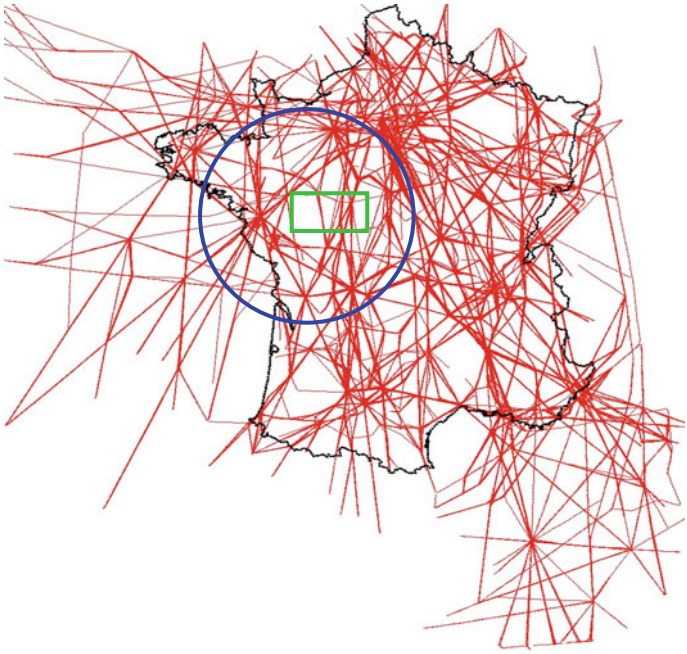


Fig. 4. Example of an MA location (green rectangle) near the military base (center of blue circle) within civil flight trajectories (red lines) over France

In this section, we model the MA by a 2D-geometry shape (a polygon) with a cylindrical vertical extension (i.e., a 3D cylinder with a polygonal section), and then we formulate our MA location problem as a constrained optimization problem. Remark that although one should properly speak of a military *volume*, we shall stick in the sequel to the more familiar MA terminology.

In this study, we aim to locate an MA for one given military base; the position of this military base is given as input data. To simplify the exposition, the shape of this MA is defined as a rectangle with a given vertical extension, i.e., a cuboid with given

length L, width W, and height H. It has also a given activation-time duration D. We also assume, as showed in Fig. 5, that it can then be rotated and/or moved around the military base, where (ρ, θ) are the polar coordinates of this MA (center of the rectangle), and ϕ is the rotation angle of the MA. We also denote the altitude of the lowest point of this MA as z, considering the military base as the origin of the space, and its starting activation time as t. To summarize, the decision variables are ρ, θ, ϕ, z, and t. In theory, all five variables are continuous variables. In practice, aircraft follows discretized Flight Levels (FL) in order to ensure safe vertical separation. Here, we, therefore, consider discrete elementary altitude shifts $\delta z = 5$ FL (1 FL = 100 ft), and we consider the discrete set $Z := \{z_{min} + \delta z, z_{min} + 2\delta z, \ldots, z_{min} + J\delta z\}$ for the possible values of z, where $J = \lfloor (z_{max} - z_{min})/\delta z \rfloor$ ($\lfloor z \rfloor$ denotes the largest integer greater than or equal to z), z_{min} and z_{max} are, respectively, the lower and upper bounds of altitude constraints for the MA. To simplify the time allocation, we also consider discrete time shifts $\delta t = 5$ min, and we define the discrete set $T := \{t_{min} + \delta t, t_{min} + 2\delta t, \ldots, t_{min} + K\delta z\}$ for the possible values of t, where $K = (\lfloor t_{max} - t_{min})/\delta t \rfloor$, t_{min} and t_{max} are respectively the lower and upper bounds of activation-time constraints for the MA. Let us denote ρ_{max} to be the maximum allowed distance between the center of the MA and the center of the military base. Let MA (x) denote the 4D (time + space) MA defined by an instantiation of our five-dimensional vector of decision variables $x = (\rho, \theta, \phi, z, t)$.

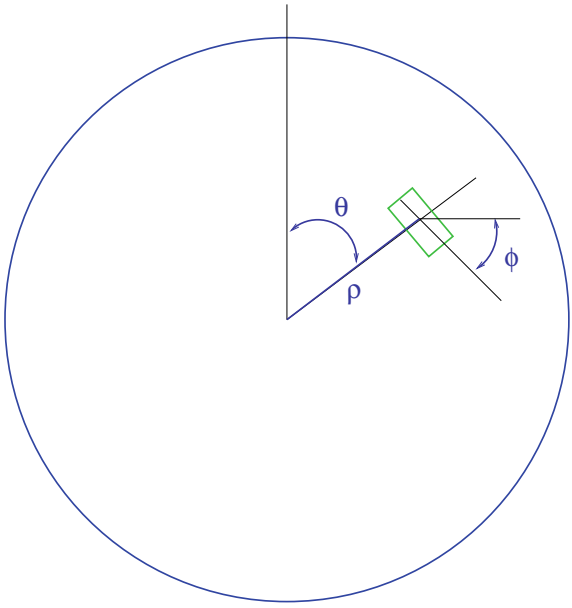


Fig. 5. MA location model: (ρ, θ) are the polar coordinates of the center of the MA (green rectangle), ϕ is the rotation angle of the MA

The civil flight (4D) trajectories are given under the form of discretized sample points $P_i \in \mathbb{R}^4, i = 1, 2, \ldots, M$, at intervals of 15 s, where M is the total (cumulated) number of sample points for all the trajectories. In order to minimize the impact of the military activity on the civil traffic, we aim at positioning the MA (in 4D) so as to minimize the number of trajectory sample points intersecting it. We also wish to minimize the distance between the MA and the military base. This is therefore a bi-objective optimization problem that we propose to model as

$$\underset{x=(\rho,\theta,\phi,z,t)}{\text{minimize}} \quad f(x) = \sum_{i=1}^{M} 1_{P_i \in \text{MA}(x)} + \mu.\rho \tag{1}$$

$$\text{subject to} \quad 0 \leq \rho \leq \rho_{max} \tag{2}$$

$$0 \leq \theta \leq 2\pi \tag{3}$$

$$0 \leq \phi \leq 2\pi \tag{4}$$

$$z \in Z \tag{5}$$

$$t \in T \tag{6}$$

where the indicator function $1_{P_i \in \text{MA}_{(x)}}$ is equal to 1 if $P_i \in \text{MA}(x)$ and is equal to zero otherwise. Hence, $\sum_{i=1}^{M} 1_{P_i \in \text{MA}_{(x)}}$ corresponds to the number of 4D sample trajectory points that lie inside the MA, and μ is a weighting parameter whose value is to be set by the decider according to his priority between the two criteria: distance ρ (expressed in NM) versus traffic impact. In our tests, we use $\mu = 1$ so as to prioritize the impact on civil traffic. Remark that this model could easily be adapted so as to evaluate the traffic impact by counting rather the number of *trajectories* intersecting the MA.

When we consider only minimizing traffic impact (the case where the user sets $\mu = 0$), this problem can also be considered as the dual problem of the classical maximum coverage problem involving only one set (one MA in our case). This one-set situation removes the combinatorial nature of the maximum coverage problem. Nevertheless, instead of the typical discrete candidate sets or a planar (2-dimensional) continuous location problem, we deal here with a 5-dimensional location problem involving 3 continuous location parameters. The difficulty of this continuous optimization problem comes here from the fact that the objective function is non-differentiable and costly to evaluate (a black-box optimization problem), as counting the number of intersecting points is time consuming and this computation must be performed at every iteration. This is why we propose here a heuristic resolution method rather than having recourse to an exact approach.

3 Simulated Annealing Algorithm

Simulated Annealing (SA) is a typical Monte Carlo global-optimization algorithm. It is a paradigm that simulates the relationship between the internal energy and the temperature T during the process of metal cooling [5]. Metal is in liquid state at high temperature; its internal energy is at its highest level but this state is highly unstable because of the high disorder of its atomic structure. If a fast cooling procedure is applied, then temper phenomena appears: internal energy quickly drops down but does not converge toward a globally minimal value; the metal has then an amorphous atomic structure with a high disorder. However, during a slow cooling process, at each temperature, the internal energy has some probability to reach a locally minimal value. When the temperature reaches room temperature, the internal energy of the metal stabilizes at a minimal value; metal has then a crystal atomic structure.

At the beginning of the cooling process, an initial state x_0 must be provided by the user. This initial state can be randomly generated or can be particularly chosen to improve the SA performance; for instance, using prior knowledge on the problem, or using any heuristic. The temperature decrease at each iteration i is dealt with some *cooling schedule* such as: $T_i = \alpha^i T_0$, where T_i is the temperature at iteration i, T_0 is the initial temperature chosen by the user, and $\alpha \in [0, 1]$ is a user-defined cooling parameter. At each iteration/temperature, N_t number of *transitions* (from one solution to another) are performed. At each transition, a neighbor state x' of the current state x is generated by some neighborhood operator $getNeighbor(x)$. This new state is then evaluated through the objective function f. If the new state is better than the previous state ($f(x') \leq f(x)$), it is accepted. When the new state is worse than the current state, it is accepted with some probability $I\!P(x, x', T)$ related to the current temperature and the objective-function degradation. The most-used acceptance probability for a minimization problem is

$$I\!P(x, x', T) = \exp \frac{(f(x') - f(x))}{T}$$

The idea is that a bad transition is more likely to be accepted at high temperatures.

After acceptance or rejection of the new state, the algorithm proceeds to the next transition. The algorithm stops when T_i goes below some predefined final temperature T_f (for instance, we use $T_f = 0.0001 T_0$ in our tests), or some predefined target objective-function value is reached, or when after performing N iterations, where N is the maximum number of iterations decided by the user. SA is given in Algorithm 1:

Algorithm 1: Simulated annealing algorithm for a minimization problem

$T \leftarrow T_0, x \leftarrow x_0$
while $T > T_f$ **do**
  for $i = 1, \ldots, N_t$ *transitions* **do**
    $x' \leftarrow getNeighbor(x)$
    if $f(x') \leq f(x)$ **then** $x \leftarrow x'$;
    else $x \leftarrow x'$ with probability $\exp \frac{f(x')-f(x)}{T_i}$;
  end
  $T \leftarrow \alpha T$;
end
return X;

The neighborhood operator $getNeighbor(x)$ we are proposing for our problem returns a neighbor state $x' = (\rho', \theta', \phi', z', t')$ of the current state $x = (\rho, \theta, \phi, z, t)$ through the following five functions, chosen randomly with, in our tests, an identical probability, 1/5:

- $Translate(x)$: return x', where ρ' is chosen randomly in $[0, \rho_{max}]$, and $\theta' := \theta$, $\phi' := \phi$, $z' := z$, $t' := t$.
- $Rot_\theta(x)$: Rotate the MA around the military base, return x', where θ' is chosen randomly in $[0, 2\pi]$, and $\rho' := \rho$, $\phi' =: \phi$, $z' := z$, $t' := t$.
- $Rot_\phi(x)$: Rotate the MA around its center, return x', where ϕ' is chosen randomly in $[0, 2\pi]$, and $\rho' := \rho$, $\theta' := \theta$, $z' := z$, $t' := t$.
- $ShiftAltitude(x)$: return x', where z' is chosen randomly in Z, and $\rho' := \rho$, $\theta' := \theta$, $\phi' := \phi$, $t' := t$.
- $ShiftActivationTime(x)$: return x', where t' is chosen randomly in T, and $\rho' := \rho$, $\theta' := \theta$, $\phi' := \phi$, $z' := z$.

All random choices above are assuming a uniform distribution.

To evaluate the new state, the MA is first located on its new geographical position, and then located in time. Once we have the current 4D position, x, of the MA, one can count the number of trajectory sample points lying inside MA (x). Again, in order to avoid checking each point which would involve excessive amount of calculation, we propose proceeding as follows. First, discretize the 3D airspace and time to construct a 4D grid. Then, extract the neighbor grid elements around the MA as illustrated in red on Fig. 6; these grid elements envelop the MA (green) completely. Trajectory sample points lying outside these grid elements are not inside the MA. It remains to consider the trajectory sample points lying within the (red) envelop. We apply the winding number algorithm to check if the points are inside the MA shape. The winding number algorithm is a computationally efficient procedure used to determine whether a given

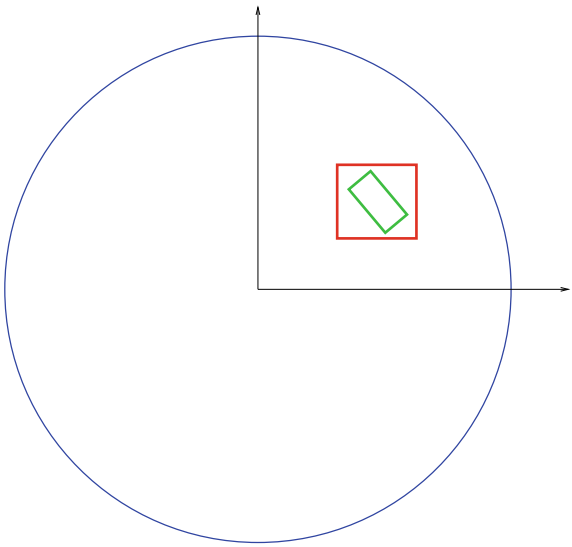


Fig. 6. Extract potential affected area (red) in the neighborhood of the current MA (green)

2D point lies inside a given 2D polygon [6], based on the sum of the angles subtended by each side of the polygon.

4 Simulation Results in the French Airspace

In this section, we apply the simulated annealing algorithm on 1 day of French traffic involving 8,836 civil flight trajectories, discretized into $M = 1,851,029$ trajectory sample points at intervals of 15 s. We first present the input MA parameters used in the numerical tests and the values we choose for the user-defined parameters involved in the SA algorithm. Then, numerical results obtained are presented and analyzed.

The chosen input MA parameter values are listed in Table 1:

The SA algorithm parameters are set as follows:

- Cooling parameter: $\alpha = 0.95$.
- Number of transitions for each temperature: $N_t = 200$.
- Initial temperature: T_0 calculated empirically to yield an 80% acceptance rate.
- Final temperature: $T_f = 0.0001T_0$.

We first apply the algorithm for three different values of the range ρ_{max}: 100 NM, 200 NM, and 400 NM. Table 2 indicates the number of trajectories (*nbTrajs*) and sample points (*nbPts*) that can potentially be impacted for each case. The SA algorithm was programmed in standard Java language. The results presented in this section are obtained on a computer with an Intel i7-4700MQ 2.40 GHz processor with 4 cores. We run the SA algorithm 10 times with the same parameter setting; average results will be computed over these 10 runs.

Table 1. Input MA parameter values

Parameter	Value	Unit
Length (L)	40	NM
Width (W)	20	NM
Height (H)	5,000	Feet
z_{min}	10,000	Feet
z_{max}	30,000	Feet
δz	1,000	Feet
J	20	–
D	120	Minutes
t_{min}	10 AM	–
t_{max}	12 AM	–
δt	5	Minutes
K	24	–

Table 2. Simulation results with length $L = 40$ NM for different values of ρ_{max}

	Case 1	Case 2	Case 3
ρ_{max} (NM)	100	200	400
nbTrajs	1,339	1,778	2,113
nbPts	100,741	171,331	226,316
$\overline{f(x^*)}$	53.2	51.1	50.8
$\overline{nbPts^*}$	0	0	0
$\Delta nbPt^*$	0	0	0
$\overline{\rho^*}$ (NM)	53.2	51.1	50.8
$\Delta\rho^*$ (NM)	2.3	3.8	1.4
CPU (s)	114	299	285

In each case, the algorithm finds a location of the MA with the minimal possible impact on the civil flights: zero impact on the civil flights. The maximum solving time of $\approx$ 2 min is viable in a real-time operational context. The results obtained are presented in Table 2, where $\overline{f(x^*)}$ is the average minimal objective function (cf. Eq. (1)), which is the sum of $\overline{nbPts^*}$, the average minimal number of impacted trajectory sample points (equal to $1_{P_i \in \mathrm{MA}_{(x^*)}}$) and of $\overline{\rho^*}$, the average minimal distance of the MA found by the SA algorithm. The average deviation of $nbPts^*$ for 10 runs is denoted $\Delta nbPt^*$, $\Delta\rho^*$ is the average deviation of ρ^*, and *CPU* represents the average SA algorithm run time in seconds.

Figures 7, 8, and 9 display the number of impacted trajectory sample points (first term of objective function (1) in red) and the distance between the MA (second term of objective function (1) in green) and the base at each iteration of the SA algorithm for three different ranges ρ_{max}. In the objective function, the second criterion (distance

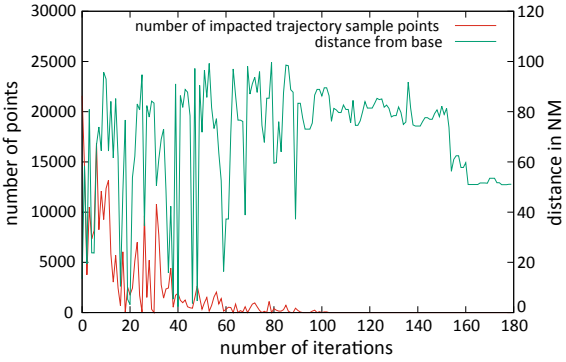


Fig. 7. Evolution of the two criteria for $\rho_{max} = 100$ NM

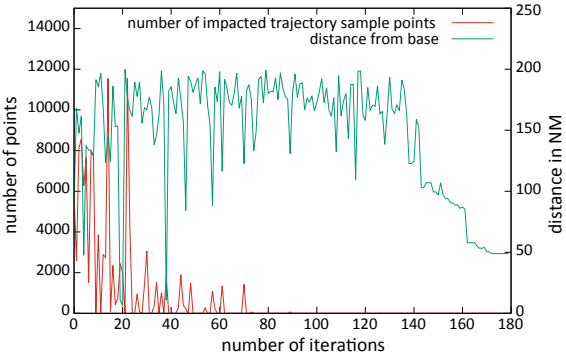


Fig. 8. Evolution of the two criteria for $\rho_{max} = 200$ NM

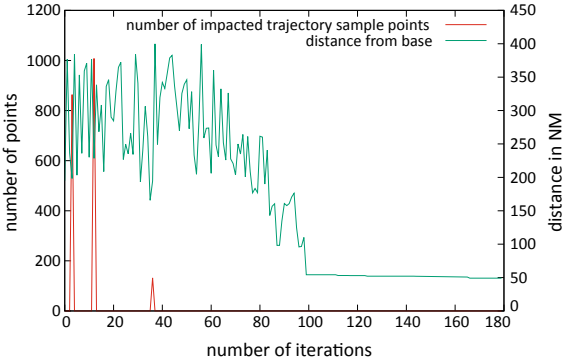


Fig. 9. Evolution of the two criteria for $\rho_{max} = 400$ NM

from the military base, expressed in NM) is a small value in comparison with the first criterion (number of impacted points), since we chose to set the weighting parameter to $\mu = 1$. As a consequence, we observe that the SA algorithm concentrates first on minimizing the first criterion (impact), and then pursues with minimizing the distance from the base while keeping the first criterion at its minimal value. The nearest MA found is 51 NM away from the base for the range 100 NM. The objective-function value converges faster to zero-impact solutions with the increasing range limits, and the nearest MA found is situated at 49 NM from the base for ranges 200 and 400 NM. Indeed, the SA algorithm often finds better locations because a larger range allows SA to converge rapidly to zero-impact solution, concentrating thereafter more time on minimizing the distance.

To analyze whether the algorithm is sensitive to the shape and size of the MA, we test different shapes and sizes of the MA by modifying the input parameters: length, width, height, and activation duration. The results obtained by SA for shape change and size reduction of the MA are similar to those obtained above. However, if we increase the size of the MA, SA may fail to find an MA with zero impact on civil trajectories under small range limitations. Table 3 displays the results obtained, for example, when doubling the length of the MA; one observes that $\Delta nbPt^*$ and $\Delta\rho^*$ are relatively large for the 100 NM range since for 3 runs out of 10, the algorithm finds an MA location near the base but impacting hundreds of trajectory sample points.

Table 3. Simulation results with length $L = 80$ NM for different values of ρ_{max}

	Case 1	Case 2	Case 3
ρ_{max} (NM)	100	200	400
nbTraj	1,561	1,932	2,117
nbPt	131,272	188,288	229,178
$\overline{f(x^*)}$	146.2	104	93.1
$\overline{nbPt^*}$	62.4	0	0
$\Delta nbPt^*$	88	0	0
$\overline{\rho^*}$ (NM)	83.8	104	93.1
$\Delta\rho^*$ (NM)	18.8	2.4	0.18
CPU (s)	281	843	596

5 Conclusion and Perspectives

As air traffic increases, military areas become a critical constraint for civilian traffic. More flexibility is requested for airspace management. In this article, we proposed a Dynamic Military Area (DMA) approach, concentrating on the optimal location problem of one DMA within civilian traffic dedicated to short-time military activities. The military area is modeled by a 2D geometry shape, with a cylindrical extension associated with given specified flight levels, that has the flexibility to move both in space and in time. We proposed a constrained-optimization formulation that aims at minimizing both the impact of military activities on civil flights, and the distance between the military area and a given military base. A simulated annealing algorithm was designed to address the obtained black-box optimization problem. Numerical experiments were conducted on instances involving one day of French traffic. The results indicate that the proposed methodology is a viable decision-aid tool to search for the nearest military area having no impact on the traffic, as in our tests none of the existing 8,836 civil flights were impacted under different range limitations. Various tests on different input parameters show that the SA algorithm is robust and not significantly sensitive to the shape and size changes of the military area.

In future work, one may address other *dynamic* versions of this location problem, in which the position of the center of the military area follows a given mission trajectory. Another promising track of research could envisage using a deterministic black-box optimization methods such as those proposed in [7, 8]. The military location problem may be combined with a real-time flight-trajectory conflict prediction and resolution problem. In this case, conflict resolution approaches will be applied to the trajectories impacted by the chosen military area. A further extension of this work would involve several types of military aircraft operating from different bases. Finally, the concepts introduced in this study can be applied for UAV (Unmanned Aerial Vehicle) planning in the civilian traffic, or to any other dynamic restricted airspace planning.

Acknowledgement. This work has been supported by the Civil Aviation University of China (Tianjin) and the program of China Scholarships Council.

References

1. Airbus global market forecast for 2017–2035, Airbus Group (2017). http://www.airbus.com/aircraft/market/global-market-forecast.html
2. Moradi E, Bidkhori M (2009) Single facility location problem. In: Zanjirani FR, Hekmatfar M (eds) Facility location: concepts, models, algorithms and case studies, Physica-Verlag HD, Heidelberg, pp 37–68
3. Plastria F (1995) Continuous location problems: research, results and questions. In: Drezner Z (ed) Facility location: a survey of applications and methods, Springer, Berlin, pp 85–127

4. Brimberg J, Hansen P, Mladenovic N, Salhi S (2008) A survey of solution methods for the continuous location allocation problem. Int J Oper Res 1–12
5. Simulated Annealing: Theory and Applications, (1987) In: Laarhoven PJM, Aarts EHL (eds) Kluwer Academic Publishers, Norwell, MA, USA
6. Hormann K, Agathos A (2001) The point in polygon problem for arbitrary polygons. Comput Geom 20(3):131–144
7. Conn Andrew R, Scheinberg K, Vicente Luis N (2009) Introduction to derivative-free optimization. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA
8. Audet C, Hare W (2018) Derivative-Free and Blackbox Optimization. Springer series in operations research and financial engineering, Springer International Publishing

Part III
Communication, Navigation and Surveillance

Coordinated Validation for SWIM Concept-Oriented Operation to Achieve Interoperability

Xiaodong Lu(✉), Tadashi Koga, and Yasuto Sumiya

Surveillance and Communications Department, ENRI, 7-42-23 Jindaiji-Higashi Machi, Chofu, Tokyo 182-0012, Japan
{luxd, koga, sumiya}@mpat.go.jp

Abstract. In order to achieve safe, secure, efficient, and environmentally sustainable air traffic management at global, regional and local levels, a collaborative environment for system-wide flight and flow information exchange is required. Therefore, the new provisions to enable a richer set of information exchange prior to departure between different aviation stakeholders with new flight planning and filing capabilities has been structured by the International Civil Aviation Organization (ICAO). This collaborative information exchange will enable a common operational picture for all related stakeholders in order to improve strategic planning. However, with the different conditions, it is difficult for all air traffic management service providers to transform from the current operation to the new operation and not all airspace users will adopt the changes at the same time. To validate the impact of changes for potential implementation of new provisions, the project of international validation has been conducted. In this paper, the observations and analysis of validation exercises consisting of Tabletop exercises and Lab exercises related to the regional implementation will be reported. Moreover, the operational processes, procedures, and automation changes required for new provisions implementation between related aviation stakeholders are clarified. Finally, the problems and challenges for constructing the collaborative operating environment to include interactions of related stakeholders using data, systems, and services through a system-wide information management environment are discussed.

Keywords: System Wide Information Management (SWIM) · Flight and Flow Information for a Collaborative Environment (FF-ICE) · ATM Service Providers (ASPs) · Airspace Users (AUs) · Interoperability · Validation

1 Introduction

With the rapid increase in local and global air traffic, a collaborative environment for system-wide flight and flow information exchange is required to improve safe, secure, efficient, and environmentally sustainable Air Traffic Management (ATM). In order to achieve system-wide operational information exchange and collaborative decision making, the System Wide Information Management (SWIM) concept has been proposed, and the practical implementation has been conducted in some regions [1].

To enable a richer set of information exchange between different aviation stakeholders, the provisions of Flight and Flow Information for a Collaborative Environment

Electronic Navigation Research Institute, *Air Traffic Management and Systems III*, Lecture Notes in Electrical Engineering 555,
https://doi.org/10.1007/978-981-13-7086-1_8

(FF-ICE) has been developed by the International Civil Aviation Organization (ICAO) [2]. Its implementation is based on the SWIM infrastructure and its concept enables to illustrate information for flow management, flight planning, and trajectory management associated with ATM operational components. Moreover, in order to ensure that the FF-ICE concept can be implemented globally and used by the ATM community as the basis, the ICAO Standards and Recommended Practices (SARPs) will be developed [3].

The FF-ICE implementation has been divided into two phases. The first phase is FF-ICE Planning (FF-ICE/1) that is focused on achieving the interoperability of ground-to-ground information exchanges by using standard information exchange models in the predeparture phase of flight. The second phase is FF-ICE Execution (FF-ICE/2) that will support Trajectory Based Operation (TBO) through ground-to-ground and air-to-ground information exchanges in the post-departure phase of flight.

The FF-ICE/1 provisions will provide guidance for new flight planning and filing capabilities structured to improve collaboration and coordination prior to departure. The provisions allow for airspace users to receive feedback on a planned flight as Air Navigation Service Providers (ANSPs) can provide constraints far in advance of departure. This information exchange will enable a common operational picture between aviation stakeholders in order to improve strategic planning. Therefore, ATM Service Providers (ASPs), Airspace Users (AUs), and related aviation stakeholders will need to determine the operational processes, procedures, and automation changes required for FF-ICE/1 provision implementation [4].

However, according to the different operational requirements, the system architecture, network and messaging infrastructure are different from each other. In addition, the required efficiency, reliability, safety, and environmental impact are also different. With these different conditions, it is difficult for all ATM Service Providers (ASPs) to transform from the current operation to the FF-ICE/1 based operation at the same time. Moreover, not all AUs will adopt the changes at the same time. Therefore, the impact of FF-ICE/1 changes for ASPs, particularly relative to adjacent Flight Information Regions (FIRs) and AUs, is unknown and unpredicted at this time.

To validate the ICAO provision changes for potential implementation, accounting for operational and technical interactions between different ATM systems within ANSP and AU domains, the International Interoperability Harmonization and Validation (IIH&V) project has been conducted by Federal Aviation Administration (FAA). Three Validation Exercises consisting of Tabletop and Lab exercises are planned in the 2016–2018 timeframe to provide recommendations to enhance implementation guidance material. The current status and observations of this project have been reported by FAA at ATM Requirements and Performance Panel (ATMRPP) meetings [5, 6].

In order to accelerate the FF-ICE/1 implementation and promote the SWIM concept-oriented operation in Asia-Pacific region, Japan Civil Aviation Bureau (JCAB) joined this project from January 2017. As a technical supporter of JCAB, Electronic Navigation Research Institute (ENRI) participated in Tabletop exercise and Lab exercise of Validation #1 that were conducted from February to August, 2017. Now, ENRI is cooperating with other members to develop the evaluation system for air-ground integration validation exercises.

In this paper, the observations and analysis of validation exercises consisting of Tabletop exercises and Lab exercises related to the regional SWIM implementation are reported. Not only the potential operational feasibility challenges between ASPs and AUs but also additional functional capabilities of communication and application required to support FF-ICE/1 are identified. Moreover, the operational processes, procedures, and automation changes required for FF-ICE/1 provision implementation between ASPs, AUs, and aviation stakeholders are clarified. Finally, the problems and challenges for constructing the FF-ICE/1 operating environment to include interactions of the ATM stakeholders using data, systems, and services through a SWIM environment are discussed.

The paper is structured as follows. In the next section, the FF-ICE concept and the overview of IIH&V are introduced. In Sect. 3, the discussion of vignette for Tabletop 1b exercise and the observations are described. The development of local system for Validation #1 Lab exercise and the lessons learned are presented and analyzed in Sect. 4. The problems and challenges to achieve interoperability are discussed in Sect. 5, and the paper is concluded in Sect. 6.

2 FF-ICE Concept and IIH&V

2.1 FF-ICE Concept

The present-day ICAO flight planning provisions were developed on the basis of a manual, paper-based, point-to-point, teletype communications system. A fundamental change is required to support the implementation of the Global ATM Operational Concept that has greater data requirements [7]. These include system-wide information sharing, providing early intent data, management by trajectory, coordinated decision making, and high automation support requiring machine readability and unambiguous information. The limitations of current flight planning provisions and how the FF-ICE concept addresses them are summarized in Table 1.

Table 1. Current provisions and FF-ICE concept

Items	Current provisions	FF-ICE concept
Information sharing	Multiple two-party exchange	All related stakeholders
Advance notification	Short term	Long term
Flight information	Local management	GUFI-based global management
Information distribution	Peer-to-peer communications	SWIM-based interoperability
Information security	Single policy	Multilayered governance
Information set	Local definitions; fixed data lengths	Standard models; flexible format
Derivable information	Independence; inconsistency	Interaction; consistency

In order to achieve a safe, secure, efficient, and environmentally sustainable air navigation system at global, regional and local levels, future ATM requires a collaborative environment with extensive information content. The FF-ICE concept will provide a globally harmonized process for planning and providing consistent flight information [1].

FF-ICE/1 is the first step towards achieving the FF-ICE concept and is primarily concerned with predeparture data and processes in a mixed-mode environment. This will involve the interoperability for flight plan coordination between partners that have SWIM flight plan filing capabilities and partners that are filing through existing systems.

2.2 Overview of IIH&V

As a main technical supporter of JCAB, ENRI is collaborating with FAA and other international aviation participants on the IIH&V to validate FF-ICE/1 concepts for potential implementation in the local system, as well as the interoperability between the local system and international ANSPs components. It is expected that this project will align current and future Collaborative Action for Renovation of Air Traffic Systems (CARATS) activities with the ICAO provisions to improve the accuracy and availability of flight information, and consistency of flight planning in different ATM environments and ANSP domains. This project consists of three Validation Exercises consisting of several Tabletop and Lab exercises to validate key ICAO provisions.

The goal of Validation #1 is to evaluate the viability of the implementation of FF-ICE/1 in the 2020 timeframe, that includes Flight Plan Submission, Monitoring, and Distribution which targets the predeparture coordination of flight plans between ASPs in a mixed-mode environment. Validation #2 and #3 expanded trajectory negotiations to the post-departure portion concerning with the A/G SWIM integration by applying Electronic Flight Bag (EFB) with single or bidirectional data link communications. And the considerations and recommendations that have been observed from the exercises of Validation #1 will be applied and improved in the lab exercises of Validation #2 and #3.

The Tabletop Exercises focus on operational, policy, and procedure questions. The development of the system capability in Lab Exercises is determined according to these discussions. Tabletop Exercises are performed for all vignettes. Based on defined down selection criteria, some scenarios of Tabletop Exercises are selected and conducted as Lab Exercises. Results from these validation exercises will be used to inform the development of any future implementation guidance, ASP procedures, AU procedures, and future revisions of ICAO Provisions (Fig. 1). Additional details on the Tabletop and Lab Exercises of Validation #1 are presented in Sects. 3, 4, respectively.

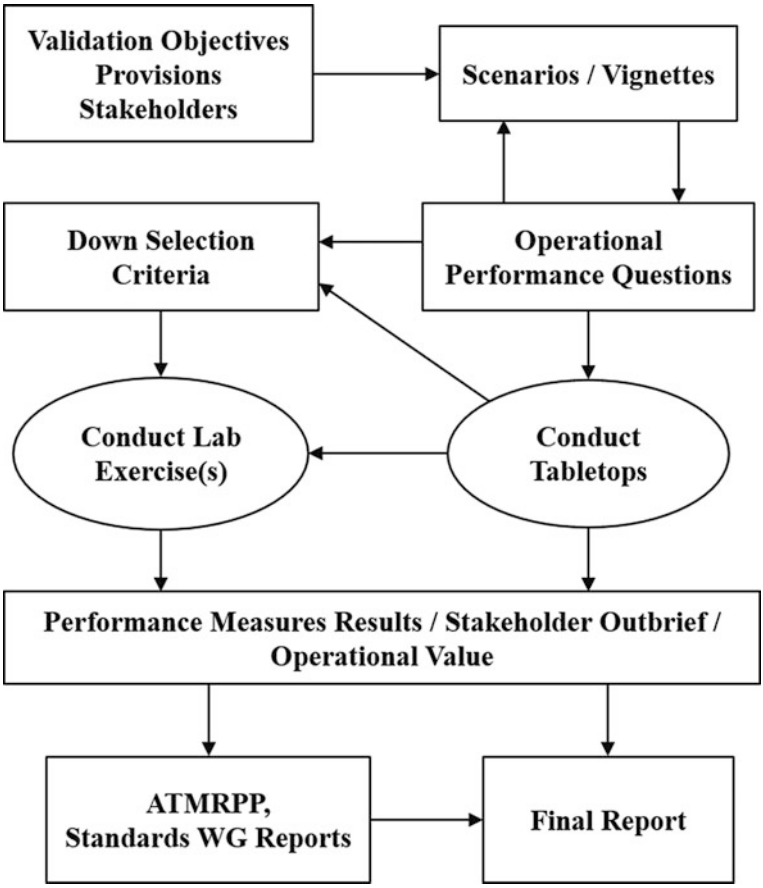


Fig. 1. Validation exercise approach

3 Tabletop Exercise

The Tabletop Exercises are discussion-based sessions where team members meet in an informal setting to discuss their roles during operation and their responses to particular situations. Tabletop Exercises give the user the ability to pose broad questions across numerous types of operational scenarios as well as allow for more detailed questions focused on a specific position or type of operation.

Validation #1 activities formally started with the Tabletop 1a exercise, which was held on September, 2016. The JCAB team jointed the IIH&V project from Tabletop 1b exercise and the objectives are as follows [8]:

1. Identify potential operational feasibility challenges with FF-ICE/1 between ASPs and AUs.
2. Identify additional functional capabilities of communication and application required to support FF-ICE/1.

3. Consider different implementations and FF-ICE/1 solutions across ASPs and understand how the differences affect AU operations.
4. Identify the message process of FF-ICE/1 based operation and understand varying feedback and constraint definitions according to different operation levels of ASPs.
5. Refine vignette and identify key topics to be validated in Lab exercise.

3.1 KJFK-RJAA Vignette

To validate the operational and technical aspects of global implementation of FF-ICE/1, the mixed-mode environment that includes participation by both FF-ICE/1 capable and FF-ICE/1 noncapable ASPs is considered. The FF-ICE/1 capable ASP and AU are referred to as eASP and eAU (enabled ASP and AU). There is no unique acronym to refer to noncapable ASP and AU.

In the Tabletop 1b exercise, several scenarios and vignettes are discussed. Each vignette is defined by the origin and destination city pairs and associated constraints. The Flight Plans are submitted by eAUs that contain additional information including 4 Dimension Trajectory (4DT), aircraft dynamics, weight, etc. This enables the eASPs to generate a more accurate model of the flight path.

The vignette proposed by JCAB is the flight planning for Japan Airlines (JAL) flight 5, a Boeing 777 aircraft with regularly scheduled service from John F. Kennedy International airport (KJFK) through the Canadian airspace to Narita International Airport (RJAA). The stakeholders of FAA, JCAB and NAV CANADA in this vignette are eASP. And the Japan Airline Air Operation Center (JAL) is an eAU. The operational views of this vignette are as follows:

1. 5 hours prior to the Estimated Off Blocks Time (EOBT), the JAL publishes a Preliminary Flight Plan (PFPL) message to all three relevant eASPs. The FAA replies to the JAL with Planning Status message that includes the Special User Airspace (SUA) constraint in Anchorage FIR.
2. The JAL publishes the PFPL Update to FAA/JCAB/NAVCANADA. The three eASPs evaluate the updated PFPL and respond with Planning Status Concur messages indicating operational acceptability of the PFPL.
3. 4 hours prior to EOBT, the JAL submits a Filed Flight Plan (FFPL) message to FAA/JCAB/NAVCANADA. The JAL receives Filing Status Accept messages from all eASPs.
4. 2 hours prior to EOBT, the JAL receives information from the FAA that the planned departure route has been heavily impacted with traffic during the EOBT for JAL5.
5. The JAL elects to use a Trial Request to determine whether a departure over another route would be acceptable. And Trial Response Concur message is returned from the FAA.
6. The JAL publishes a FFPL Update message for JAL5 and Filing Status Acceptance messages are received.
7. One and a half hours prior to EOBT, the continuous monitoring of JCAB reevaluates Updated FFPL and updates Filing Status message with airspace closure that is occurred in the Fukuoka FIR.

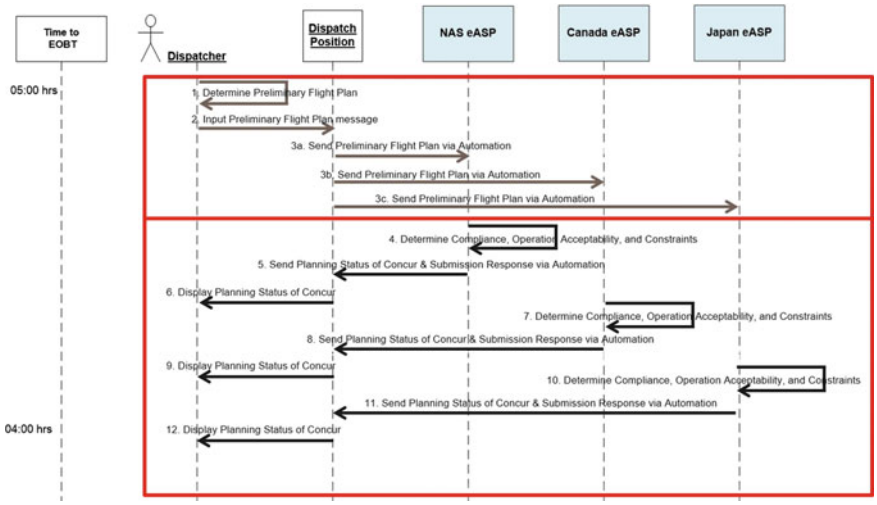


Fig. 2. Operational event trace

8. The JAL submits a FFPL Update message with revised air route and receives Filing Status Acceptance messages from all three eASPs.
9. In the Tabletop 1b, the Operational Event Trace (OET) is discussed and applied to represent the information exchanges between the operational stakeholders within the vignette. The example of PFPL submission and response by OET is shown in Fig. 2.

3.2 Observations

According to the Tabletop 1b exercise, some feedback to ICAO Implementation Guidance and other provisions for global aviation community have been reported by FAA [5, 6]. As an Asia-Pacific regional participator, some observations for improving regional SWIM implementation are provided below:

1. To implement FF-ICE/1 based operation, not only additional operational functions for supporting FF-ICE/1 message exchange has to be developed but also the regional common communication infrastructure and service providers for network connection are necessary.
2. For national ASPs and AUs, the benefit of FF-ICE/1 based operation depends on the accuracy of four-dimensional trajectory (4DT)/flight plan and operational levels of related FIRs. The priority implementation and test operation in some density areas are more efficient for regional SWIM development.
3. For achieving efficient FF-ICE/1 based operation, the understanding and cooperation between not only the departure and arrival pair of ASPs but also adjacent and nonadjacent ASPs involved in air route are required and important.

4 Lab Exercise

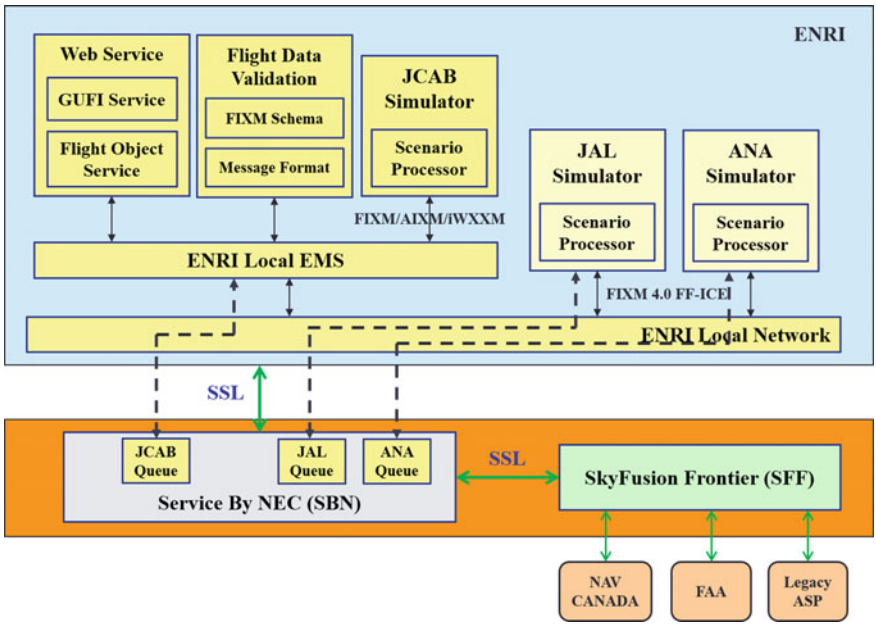


Fig. 3. System architecture

4.1 System Architecture

In the Lab exercise of Validation #1, there are two Global Enterprise Messaging Service (GEMS) providers that facilitate data sharing between a variety of partners and applications. As shown in Fig. 3, the FAA, NAV CANADA and legacy ASPs connect to SkyFusion Frontier (SFF), which is supported by Harris Corporation. NEC provides the GEMS connections (SBN) to local eASPs and eAUs.

The GEMS Providers are charged to enforce the use of the standardized aeronautical, flight and weather exchange models (AIXM, FIXM, and iWXXM) with the updated versions for each of their SWIM nodes to ensure the interoperability of information exchange via Cloud-based platform [9]. The communication between SFF and NEC is based on Secure Sockets Layer (SSL). And the SSL is also used for communication between NEC and local users. The communication standard for Publish/Subscribe messaging is Advanced Messaging Queuing Protocol (AMQP).

In this Lab exercise, the evaluation systems of JCAB, ANA, and JAL are constructed on the ENRI local network and EMS. There is a set of services and applications developed by ENRI that support both FF-ICE/1 based message process and local validation.

- The Globally Unique Flight Identifier (GUFI) is a key component of flight object management. The GUFI Service in local system provides the functionality of generating and tracking GUFIs.
- The Flight Object Service maintains an up-to-date version of all subscribed flight data. The flight data is organized into flight objects by GUFI and then stored in a database for continuous updates and queries.
- The Flight Data Validation provides validation and reporting on FIXM 4.0 based FF-ICE/1 messages conformance to schema and set of business rules. It also monitors FIXM messages and provides near real-time status (valid or invalid) to users on any findings related to nonconformance of schema and noncompliance to business rules.
- The Simulator of JCAB processes received FF-ICE/1 Flight Plan messages from eAUs. It also supports publishing constraints with AIXM and iWXXM messages and submitting the response that includes the constraints to eAUs.
- The Simulators of ANA and JAL generate FF-ICE/1 Flight Plan messages and submit Trail Requests and updated Flight Plans according to the constraints provided by eASPs.

4.2 Message Process

According to the definition in the ICAO FF-ICE/1 Provisions, the following message types in the FIXM 4.0 format shown in Table 2 are validated for the predeparture phase of flight.

Table 2. Message types

Message types	Descriptions
E_PFP	Preliminary flight plan
E_FPL	Filed flight plan
SUB_RESP	Submission response
FIL_STATUS	Filing status
PLAN_STATUS	Planning status
E_FPL_UPDATE	Flight plan update
E_ARRIVAL	Arrival
E_DEPARTURE	Departure
E_CANCEL	Cancel
E_TRIAL_REQ	Trial request
E_TRIAL_RESP	Trial response
E_INFO_REQ	Request flight information
E_INFO_RESP	Flight information response

There are two phases in the FF-ICE/1, Preliminary phase and Filed phase. In each phase, eASPs should reply to the eAU regarding operational acceptability of their flight plans. The decision tree of process flow for the message of Preliminary Flight Plan

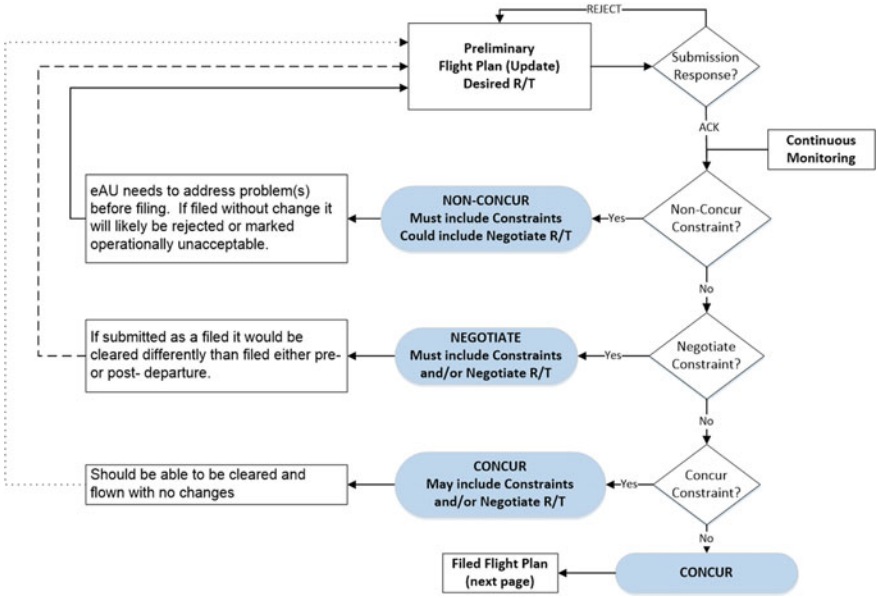


Fig. 4. Process flow for E_PFP

(E_PFP) is shown in Fig. 4. If the message can be processed and is free of syntax errors, a Submission Response of ACK will be responded by eASPs. And then, each eASP will check the constraints and publish Planning Status (PLAN_STATUS) to the eAU for the received E_PFP message. There are three status options in the

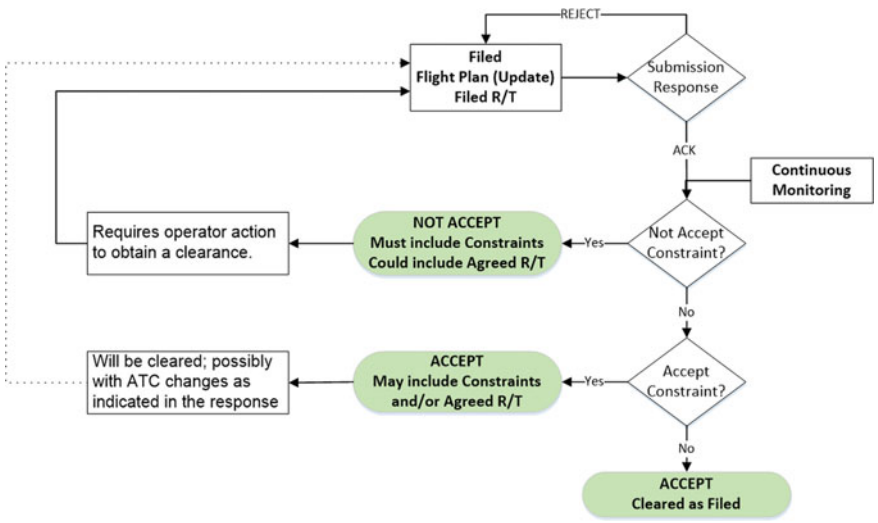


Fig. 5. Process flow for E_FPL

PLAN_STATUS message: NON-CONCUR, NEGOTIATE, and CONCUR. Due to the different operation situations, different eASPs will have different specific implementations. For example, one eASP may define a constraint type as Non-Concur while another eASP may define that same constraint type as Negotiating.

And the process flow for the Filed Flight Plan is shown in Fig. 5. After getting Concur for Preliminary Flight Plan from all three eASPs, it is available for eAU to publish Filed Flight Plan (E_FPL). According to the continuous monitoring, each eASP continuously checks the constraints that will affect the Filed Flight Plan in its own managed airspace and publish Filing Status (FIL_STATUS) to the eAU. There are two status options in the FIL_STATUS message: NOT ACCEPT and ACCEPT.

As shown in Fig. 6, the Trial Request is utilized for negotiation between eASP and eAU. In this scenario, when JAL received the message of Filing Status Update with departure fix constraint from the FAA, the Trial Request message was used to determine whether a departure over another route would be acceptable. And the Trial Response message with Concur is returned from the FAA.

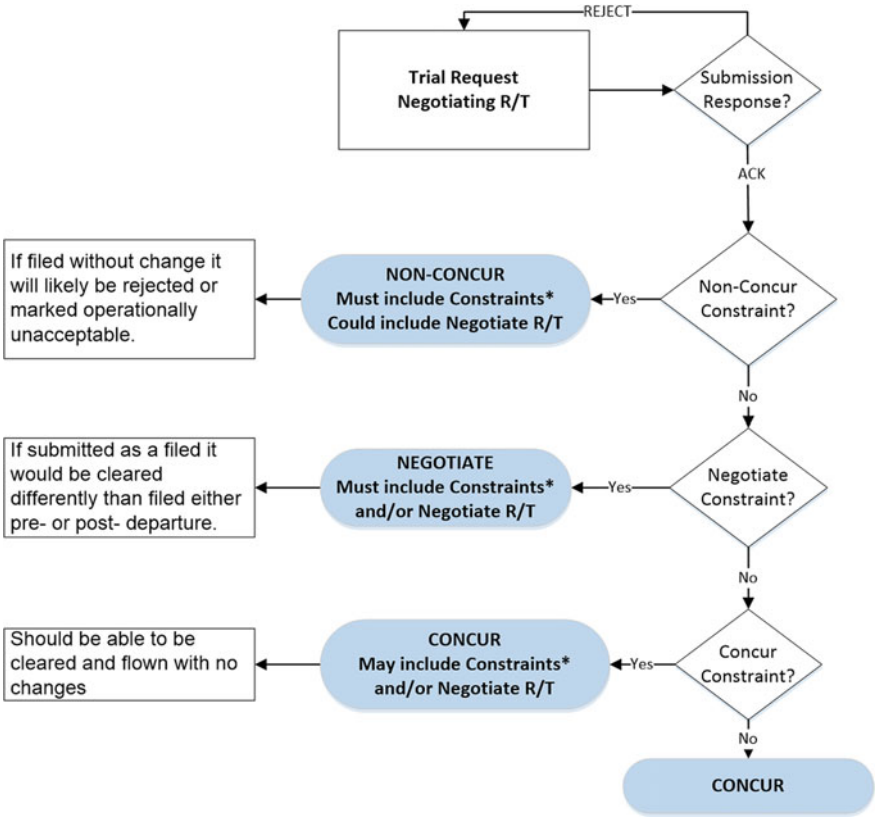


Fig. 6. Process flow for E_TRIAL_REQ

4.3 Validation and Analysis

The validation of flight information exchange model (FIXM 4.0) implementation is an essential component of the Lab exercise. The goal of the Lab exercise is not only to validate FF-ICE/1 messages but also evaluate all states and status options of the messages. The FF-ICE/1 message-based operations of JAL and JCAB for the KJFK-RJAA vignette are shown in Figs. 7, 8.

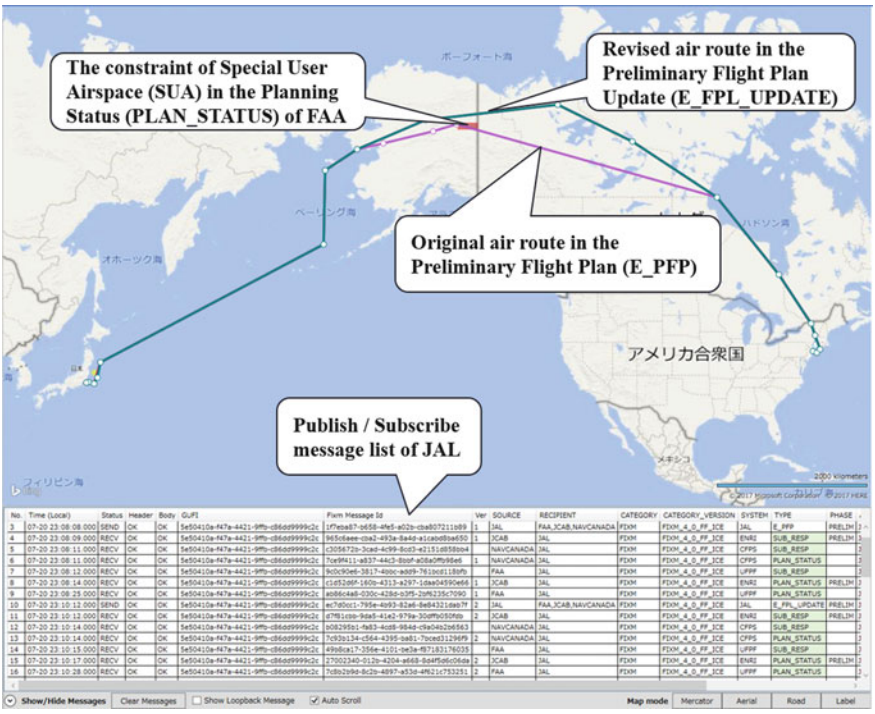


Fig. 7. JAL simulator

In the FF-ICE/1 messages, there are two parts (Fig. 8). One is message information part that includes contact information, flight plan version, and referenced constraints. The other is flight information part, such as aircraft information, 4D trajectory, and GUFI.

A key enabler to sharing flight data internationally is having the ability to unambiguously identify each flight. This is accomplished by having a GUFI assigned to each unique flight. A unique flight is defined as a single operation of an aircraft from takeoff to touchdown. The GUFI is intended to provide a unique reference to a specific flight. Its purpose is to assist in associating a message to the correct flight and to help in distinguishing between similar flights.

However, in the draft of FF-ICE/1 Implementation Guidance, it does not specify GUFI in Submission Response as a mandatory or optional field. Including the GUFI

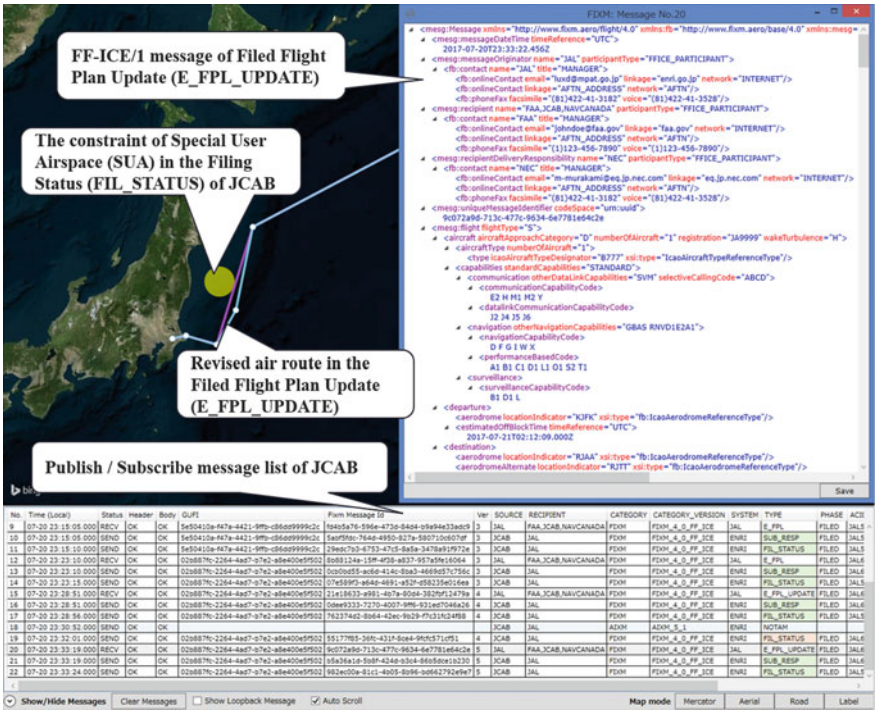


Fig. 8. JCAB simulator

allows participants to easily identify Submission Responses as part of their respective FF-ICE/1 message conversations (as opposed to relying solely on the reference message ID), and allows developers to be consistent in their implementations. It is therefore recommended that at least during initial implementation, the GUFI should be included in all FF-ICE/1 messages.

For consistent and automatic message process in a certain operation, each message should have different message identifiers. A message identifier allows identification of a message exchange between two parties. And a message identifier between two parties should be unique for a certain operation or a certain time period.

In addition, to assure all stakeholders are using the same flight plan information, the eAU is required to provide an indication of the flight plan version. The flight plan version should be incremented by the eAU each time a Flight Plan Update is submitted, i.e., at least one flight plan data element is changed. The flight plan version number is intended to provide both a reference to a particular version but also an indication of the sequence in which versions have been created. The last flight plan version provided by the eAU to an eASP is expected to be included by the eASP when providing feedback in the form of a Submission Response or a Planning or Filing Status.

However, in the FIXM v4.0, the Flight Plan Version field is a free text field without any mask. This can lead to incompatible data in this field. It is recommended that this

field should be defined as an INTEGER and more details for processing the incorrect flight plan version number should be given in the implementation guidance.

In the Lab exercise, there is an assumption that each eASP is able to provide constraints, such as aeronautical information, traffic flow management data, and severe weather conditions to the eAU. The eASP evaluation and continuous monitoring of Preliminary and Filed Flight Plans will check for and send message updates for changes to published constraints affecting the route. It not only assists the eAU in determining the optimal route/trajectory for a flight by identifying the operational environment and ATM constraints applicable to the flight as proposed but also enables eASPs to obtain an earlier, more detailed and more accurate assessment of the anticipated traffic demand.

However, how to decide the relevant ASPs for a certain operation is not clearly defined in the related documents. In the FF-ICE/1 Implementation Guidance, a relevant ASP is defined as any ASP who could potentially issue constraints on a flight. And the FF-ICE/1 Provisions states a relevant ASP is any ATM Service Provider whose airspace is along the flight plan route of flight or along the possible route of flight described in filed routing to a revised destination. The different interpretations exist within different participants to determine relevant ASPs who should receive the flight plan from the originator. Further examples would be helpful to clarify off-nominal cases and avoid different interpretations of the FF-ICE/1 Implementation Guidance.

5 Discussion

5.1 Communication

In order to evaluate the local communication performance, the message sending time and receiving time of JAL, JCAB, and NEC are recorded. Since there is an error for Windows operation system based synchronization among distributed computers, the following method is applied to calculate the communication time (T_C) between JAL and JCAB.

$$T_{JAL} = T_{JCAB} + \Delta T \tag{1}$$

$$T_{JCAB\_IN} - T_{JAL\_OUT} = T_C - \Delta T \tag{2}$$

$$T_{JAL\_IN} - T_{JCAB\_OUT} = T_C + \Delta T \tag{3}$$

$$T_R = T_{NEC\_IN} - T_{NEC\_OUT} \tag{4}$$

where ΔT is the synchronization error of current time between the JAL Simulator (T_{JAL}) and the JCAB Simulator (T_{JCAB}), and T_R is the message routing time in SBN. Then T_C can be calculated as:

$$T_C = [(2) + (3)]/2 \tag{5}$$

Based on the results of 100 tests during the different time frames, the average communication time between JAL and JCAB is 322 ms in which the average local message routing time in SBN is 151 ms.

In the Mini-Global Demonstration I, the IPsec VPN over the Internet was utilized to make connections between different users [10]. And the network delay in the demonstration was more than 700 ms that affected the performance of message exchange. Compared with the communication time of IPsec VPN over the Internet used in the Mini-Global Demonstration I, the network delay has been much reduced and the performance of message exchange has been improved by using SSL VPN over the Internet and Cloud-based platform. Even the communication time over the Internet cannot be assured, there is no significant delay during the tests and the communication time in this level is possible to satisfy the general requirement of eAUs and eASPs for FF-ICE/1 operation.

5.2 Interoperability

From the view of technical aspects, to achieve the interoperability of SWIM services, the technical interoperability, the semantic interoperability, and the process interoperability are required.

The technical interoperability is the basis of SWIM concept-oriented operation. Based on acceptable technology standards, the common and secure infrastructure for network communication and message exchange should be constructed at local, regional and global levels.

The implementation of semantic interoperability requires the defined and precise meaning of exchanged information that can be preserved and understood by all related stakeholders. And it needs to bundle of information into meaningful messages based on the different information exchange models that have been defined by ICAO. As a result, it enables new interoperable services to be identified, designed and implemented based on the different information exchange services.

To assure the consistent operation, the process interoperability is necessary to perform actual information exchange based on appropriate process alignment. Coordinated and standardized processes enable SWIM responsible authorities as well as stakeholders to work together based on sufficiently aligned processes. Therefore, the metadata, format, and process of different messages should be defined in advance to ensure that eAUs and eASPs implement the global operation.

5.3 Challenges

The SWIM-based FF-ICE operation will provide related information in greater detail and allow the eAU and the eASP to share their expectations in an unambiguous manner via the exchange of trajectory information. However, as shown in Table 3, there are still many problems and challenges we should face to achieve the FF-ICE-oriented operation in local, reginal and global areas.

For standards-based interoperability, the routing standards between different systems to ensure message delivery should be addressed. For seamless information sharing, additional guidance is needed on translating between ATS and FF-ICE

Table 3. Problems and challenges

SWIM	FF-ICE	Problems and challenges
Infrastructure	Standards-based interoperability	Messaging infrastructure for FF-ICE operation
Exchange models	Seamless information sharing	FIXM-based definition for FF-ICE messages
Exchange services	Situation-awareness service cooperation	Heterogeneous services provision and utilization
Governance	Life cycle management	Definition for quality, security and business rules

messages to avoid ambiguity and data loss. It was observed that not all elements map one-to-one between ATS and FF-ICE messages and this can result in misinterpretation between translators and message consumers. For example, translating 4D points to ATS route text. For situation-awareness service cooperation, it is required to establish a common format for referencing constraints in AIXM and iWXXM over the different systems.

Moreover, to facilitate interoperability and harmonization and avoid integration issues, it was necessary for participants to share a common, agreed-upon set of business/data rules which were derived from the FF-ICE/1 Provisions and Implementation Guidance. Participants could validate messages using methods compatible with their needs and resources, as long as the validation method used the common set of business rules. This helped avoid issues where participants may have interpreted the FF-ICE/1 Implementation Guidance in different ways.

And for the safety-critical information exchange, the life cycle management is required to provide timely, relevant, accurate, authorized, and quality-assured information to related aviation stakeholders. Currently, the blockchain is a potential option to achieve life cycle management for collaborative information exchange in ATM field.

6 Conclusion

In this paper, the overview of IIH&V international project for validating FF-ICE/1-oriented operation is introduced. And the observations and analysis of validation exercises consisting of Tabletop exercises and Lab exercises related to the regional implementation are reported. In addition, according to a certain scenario, the operational processes, procedures, and automation changes required for FF-ICE/1 provision implementation between ASPs, AUs, and aviation stakeholders are clarified. Finally, the problems and challenges for constructing the FF-ICE/1 operating environment to achieve interoperability are discussed.

References

1. ICAO (2015) Manual on system wide information management (SWIM) concept. ICAO Doc 10039
2. ICAO (2012) Manual on flight and flow—information for a collaborative environment. ICAO Doc 9965
3. ICAO (2016) 2016–2030 Global air navigation plan. ICAO Doc 9750, 5th edn
4. ICAO (2007) Procedures for air navigation services—air traffic management. ICAO Doc 4444, 15th edn
5. FAA (2016) Validating ICAO provisions via tabletop exercises. ATMRPP/2-IP/01, ICAO
6. FAA (2017) Validating ICAO provisions via tabletop exercises—update. ATMRPP-WG/32-IP/08, ICAO
7. ICAO (2005) Global air traffic management (ATM) operational concept. ICAO Doc. 9854, 1st edn
8. ENRI (2017) International and regional coordination for FF-ICE/1 validation. SWIM-TF/1-IP/02
9. Lu XD, Koga T (2017) System wide information management for heterogeneous information sharing and interoperability. In: IEEE 13th international symposium on autonomous decentralized systems, pp 199–204
10. Lu X, Koga T (2015) Real-time oriented system wide information management for service assurance. In: IEEE 12th international symposium on autonomous decentralized systems, pp 175–180

Development of a Ground Subsystem Prototype for Ground-Based Augmentation System (GBAS) Approach Service Type D (GAST-D) and the Evaluation of Its Performance in a Low Magnetic Latitude Region

T. Yoshihara(✉), S. Saito, A. Kezuka, K. Hoshinoo, S. Fukushima, and S. Saitoh

Navigation Systems Department, Electronic Navigation Research Institute (ENRI), Tokyo, Japan
{yosihara, susaito, kezuka, hoshinoo, fks442, saitoh}
@mpat.go.jp

Abstract. A ground-based augmentation system (GBAS) is a navigation system using global navigation satellite system (GNSS) that enables precision approaches and landing for aircraft. In May 2010, the International Civil Aviation Organization (ICAO) Navigation Systems Panel (NSP) working group completed development baseline standards and recommended practices (SARPs) for GBAS ground subsystems to support GBAS approach service type D (GAST-D), which refers to Category III precision approach services using the single-frequency L1-C/A signal. The Electronic Navigation Research Institute (ENRI) developed a prototype of the GAST-D ground subsystem to operationally validate the development baseline SARPs. Owing to the fact that ionospheric delays with large spatial gradients represent one of the most significant risks to the integrity of the GAST-D operation, the system was installed in a low magnetic latitude region where plasma bubble causes steep spatial gradients in the ionospheric delay. Preliminary results were reported to the NSP working group before the development baseline SARPs were approved in December 2016 with an expectation that they would go into effect in 2018. Here, we report the development of a prototype for a GAST-D ground subsystem to validate the development baseline SARPs and preliminarily evaluate the system's performance.

Keywords: GNSS · Landing system · Safety system design · Ionospheric delay · Precision approach

1 Introduction

A ground-based augmentation system (GBAS) is a navigation system using global navigation satellite system (GNSS), and it is designed to support precision approaches and landing for aircraft. There are three categories of GBAS operation defined on the

Electronic Navigation Research Institute, *Air Traffic Management and Systems III*, Lecture Notes in Electrical Engineering 555,
https://doi.org/10.1007/978-981-13-7086-1_9

basis of the capability to serve the different approach phases: Category I (CAT-I) GBAS can provide guidance for precision approaches with a decision height of 200 feet, whereas Category III (CAT-III) GBAS can support precision approaches including the final phase below 200 feet as well as landing and rollout from the runway. Because GBAS is designed to support the most critical flight phase, the International Civil Aviation Organization (ICAO) standards and recommended practices (SARPs) specify a high level of safety requirements. The system performance requirements include accuracy, integrity, and continuity along with service availability.

The fundamental concept of GBAS is based on a local differential GNSS technique to provide aircraft with range correction values and integrity information for each ranging signal of GNSS satellites. Figure 1 shows the basic configuration of the GBAS. The GBAS ground subsystem comprises three major parts: GNSS reference stations, a data processor, and a VHF data broadcast (VDB) transmitter. GNSS reference stations typically comprise four sets of an L1-single-frequency GNSS antenna coupled with a receiver. The data processor generates range correction values and integrity information and converts them into GBAS messages. These messages are finally transmitted to the aircraft via the VDB transmitter.

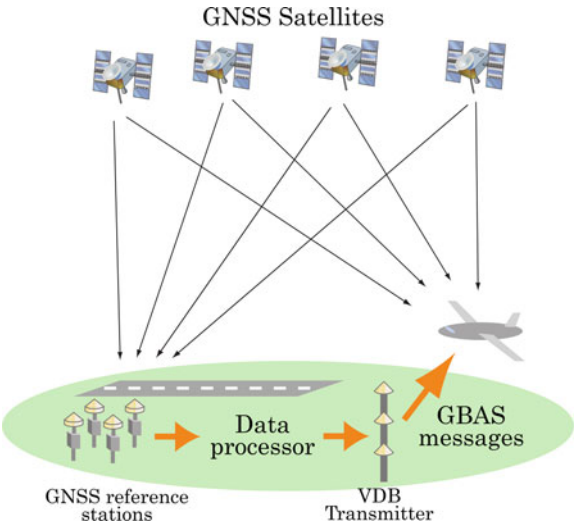


Fig. 1. GBAS configuration

ICAO SARPs for CAT-I GBAS operations were published in 2001, and the first operational systems have been implemented in airports since 2012. One of the most notable reasons for the 10-year delay between the publication and the first implementation of the CAT-I GBAS SARPs was that the integrity risk associated with ionospheric anomalies was found to be more significant than expected before 2001. It is well known that the spatial decorrelation of ionospheric delays causes ranging errors in the affected satellite signals with additional components owing to the carrier smoothing process [1], which smooths a code range with carrier range measurement for each

ranging source. Through this process, the ionospheric range errors almost double because code and carrier ranges are delayed and advanced by ionospheric effects, respectively. In fact, this effect was considered for "nominal" ionospheric conditions in the integrity parameters in GBAS messages. However, it was found that large spatial gradients were caused by the ionospheric disturbances than initially assumed, resulting in significant ranging and positioning errors in the GBAS [2, 3]. Such events were observed during the ionospheric disturbances associated with the storm-enhanced density around and after the solar maximum of 2001.

Figure 2 shows an example situation that is considered the worst-case scenario, in which the ionospheric delays between the GNSS signals received by the GNSS reference stations and the aircraft differ significantly. Therefore, CAT-I GBAS operations should always assume the possibility of large errors because the ground subsystem cannot detect ionospheric anomalies around the aircraft [4–6]. Against the integrity risk due to ionospheric anomalies, it is important to specify its threat space, which describes an assumed range to consider in GBAS safety design to protect the users. The threat space can be defined by ranges of parameters characterizing the anomalies, such as the gradient, width, depth, and moving speed of the special change of the ionospheric delays [7, 8]. Figure 2 also shows the integrity monitor of the ionosphere field monitor (IFM), which can be used to reduce the potential errors associated with ionospheric anomalies by monitoring the ionospheric delay differences between the GNSS reference stations and the IFM station [9].

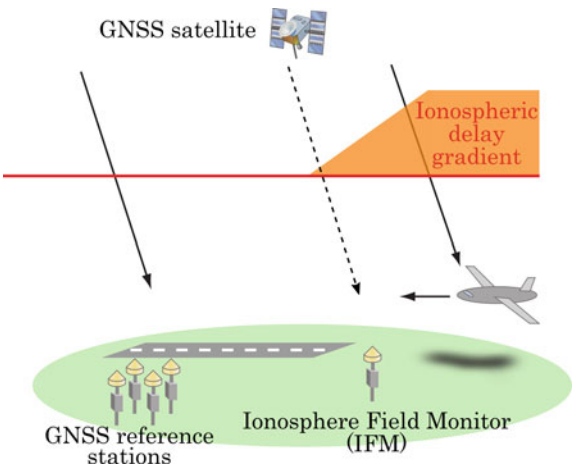


Fig. 2. Integrity risk associated with ionospheric anomalies with large spatial gradients in the ionospheric delays

GBAS approach service type D (GAST-D) has been proposed as a service type to support CAT-III precision approaches with a single-frequency L1-C/A signal. GAST-D requires extensive measures to ensure safety and is consequently designed to mitigate ionospheric threats using both ground and airborne integrity monitors. The ICAO

Navigation Systems Panel (NSP) Working Group 1 (WG1) CSG (Category II/III Subgroup) developed the development baseline SARPs for GAST-D, completing their technical validation in May 2010 [10]. After the operational validation in the NSP working group, which took about six and a half years, the final SARPs were approved by the NSP in December 2016. After consultation with ICAO member states, these SARPs are expected to go into effect in 2018.

The GAST-D development baseline SARPs define the requirements for CAT-III GBAS using ranging sources with single-frequency L1-C/A signals. In the GAST-D SARPs, a new classification of GBAS service types is introduced, and CAT-I GBAS services are classified as GAST-C. The development baseline SARPs specify that the requirements of GAST-D operation should include all of the requirements of GAST-C along with additional requirements that are higher than those for GAST-C to support the final approach phase and landing below 200 ft. For example, the integrity requirement for the hardware of a GAST-D ground subsystem is $1–1 \times 10^{-9}$.

The Electronic Navigation Research Institute (ENRI) participated in these operational validation activities as part of a working group under the ICAO NSP. A total of four projects were conducted all over the world to develop a prototype of the GAST-D ground subsystem. Prototypes were developed and installed in airports in Atlantic City (USA), Frankfurt (Germany), Toulouse (France), and Ishigaki (Japan). The first three sites are located in the magnetic midlatitude region, whereas the fourth airport in Ishigaki is located in a low magnetic latitude region, where the ionospheric conditions differ from those in the midlatitude region and ionospheric disturbances occur frequently. This is one of the unique features of this study site because the characteristics of ionospheric disturbances differ at different magnetic latitudes. The ENRI conducted its own development and validation of a GAST-D ground subsystem at the airport in Ishigaki.

Plasma bubbles are depletion of the ionospheric density; this phenomenon is frequently observed at nighttime in the low and equatorial ionosphere, during periods of high solar activity, and particularly during the equinox seasons around the longitudes near Japan. Thus, the impacts of plasma bubbles on the system performance were considered in the evaluation. Hence, ENRI's GAST-D subsystem prototype was installed in February 2014, when the solar activity was declining but still moderately high.

In this study, we report the initial results of a system performance evaluation of ENRI's GAST-D subsystem prototype. The former report [11] described the outline of the program including an airborne evaluation equipment. This paper describes details of the development and validation of the prototype of the GAST-D ground subsystem. Although the operational validation of the development baseline SARPs has been completed, it is important to evaluate its system performance and identify hazardous events over a yearly period. Therefore, we further discuss the future steps toward implementation of GAST-D operations in terms of long-term data that has been collected since February 2014.

2 Development of a Prototype of GAST-D Ground Subsystem

The objectives of ENRI's GAST-D program included an operational validation of the GAST-D development baseline SARPs, identification of the major technical subjects to achieve the required safety levels, and fundamental solutions to the challenges hindering future GAST-D implementation in Japan. The research program was conducted to last for four years from April 2011 to March 2015. The project comprised three major parts: development of a research prototype of a GAST-D ground subsystem, development of experimental airborne GAST-D equipment (including the major airborne integrity monitors), and validation of the GAST-D ionospheric threat model for the low magnetic latitude region. Again, the ionospheric threat model defines a range for each parameter including the gradient, width, depth, and speed of the ionospheric disturbances; these ranges should cover all of the possible events sufficient to design the GAST-D ground and airborne subsystems with the required safety levels. Figure 3 shows the actual schedule of ENRI's GAST-D program.

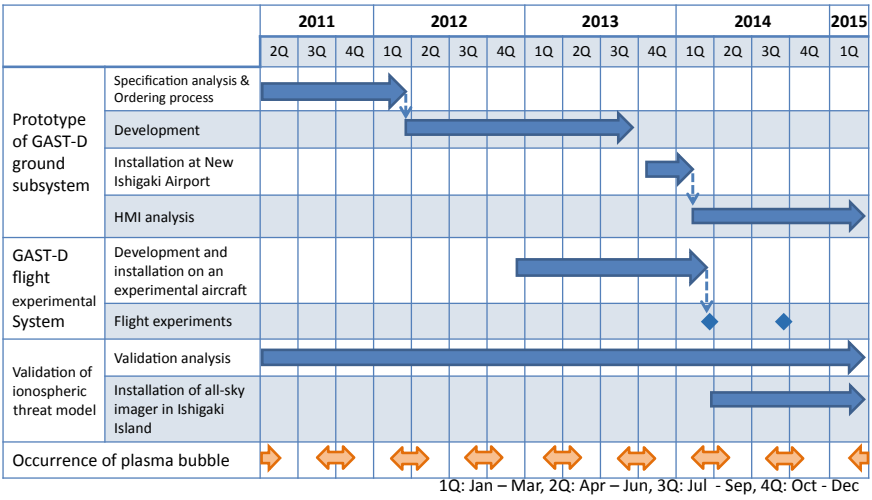


Fig. 3. Actual schedule of ENRI's GAST-D validation program

The ENRI contracted NEC Corp. in March 2012 to develop a prototype of the GAST-D ground subsystem according to the GAST-D development baseline SARPs, RTCA DO-246D [12], and DO-253C [13]. The safety design and validation processes were conducted according to the guidelines outlined in SAE ARP4754 [14] and ARP4761 [15] with the exception of the software integrity assurances and redundant hardware. In total, 23 safety design review meetings were held between the ENRI and the manufacturer at intervals of almost every three weeks over a period of one and a half years until the prototype of the GAST-D ground subsystem was delivered in September 2013.

Dual-frequency (L1/L2) GNSS receivers (Novatel Euro-3) were used as the reference receivers in the prototype to make it possible to validate the system's performance against ionospheric disturbances, including the performance of the integrity monitors, by dual-frequency measurements. The prototype was also designed to broadcast GBAS messages to allow SBAS ranging sources to be used. The following major GAST-D integrity monitors were implemented in the prototype:

- Ionospheric spatial gradient monitor (ISGM)
- Signal deformation monitor (SDM)
- Code-carrier-divergence monitor (CCDM)
- Excessive acceleration monitor (EAM)
- Ephemeris monitor (EPHM)
- Radio-frequency interference monitor (RFIM).

Additionally, the prototype had all of the integrity monitors required for CAT-I capabilities based on ENRI's CAT-I prototype developed in a previous project, including IFM [9]. The prototype had additional novel features for advanced studies including the functionality to use a reference signal with a chip-scale atomic clock (CSAC) [16] and the implementation of an integrity monitor to detect multiple receiver faults (MRFM). Table 1 shows the different integrity monitors and the corresponding integrity risks that are mitigated through the use of these monitors as well as the requirements specified in the GAST-D development baseline SARPs.

Table 1. Integrity monitors and the associated risks

Integrity monitor	Integrity risks for the ground				Development baseline SARPs
	Ionospheric anomaly	Satellite faults	RFI	Receiver faults	
ISGM	X				Required
SDM		X			Required
CCDM	X	X			Required
EAM		X			Required
EPHM		X			Required
RFIM			X		_
MRFM				X	_

Figure 4 shows the schedule of the GAST-D ground subsystem development. The fundamental hardware design was followed by a software design and coding. In general, a preliminary system safety assessment and an algorithm description document are important in the design and validation of safe systems. The system safety design was conducted as follows:

- Extraction of risks and fault tree analysis
- Quantitative risk analysis and identification of key risks
- Risk allocation

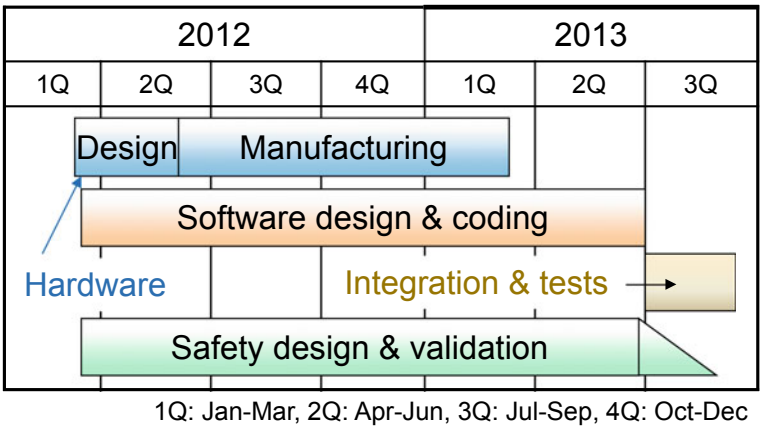


Fig. 4. Development schedule for the GAST-D ground subsystem

- Development of mitigation methods such as integrity monitors
- Evaluation of the remaining risks.

Snow accumulation is one of the important safety risks associated with GAST-D ground subsystems in Japan, especially in the northern areas; therefore, the ENRI has performed winter experiments and evaluated the effects of snow stuck to GPS antennas and multipath signals reflected from the snow surface [17]. The results showed that these effects could negatively impact the performance of the integrity monitors but would not significantly affect the performance of the GAST-D ground subsystem.

3 Installation and Continuous Data Collection at New Ishigaki Airport

The GAST-D development baseline SARPs are based on an operational concept of mitigating the ionospheric threat through the use of integrity monitors to detect ionospheric anomalies at both the ground and the airborne subsystems [18]. Therefore, it is necessary to demonstrate and validate the requirements defined in the SARPs under various ionospheric conditions, including actual disturbed conditions. As mentioned in Sect. 1, the ionospheric characteristics depend on the magnetic latitude. While other validation programs have been conducted in the magnetic midlatitude regions, in this study, the development baseline SARPs were validated in the low magnetic latitude region. In the southern part of Japan, the equatorial ionization anomaly (EIA) is considered to be part of the nominal conditions in spring and autumn. Further, plasma bubbles frequently occur at nighttime in spring and autumn, causing steep spatial gradients in the ionospheric delay. Figure 5 shows an example of global distribution of the ionospheric vertical delay at 20JST (Japan Standard Time) in March, which is derived from the NeQuick model [19, 20].

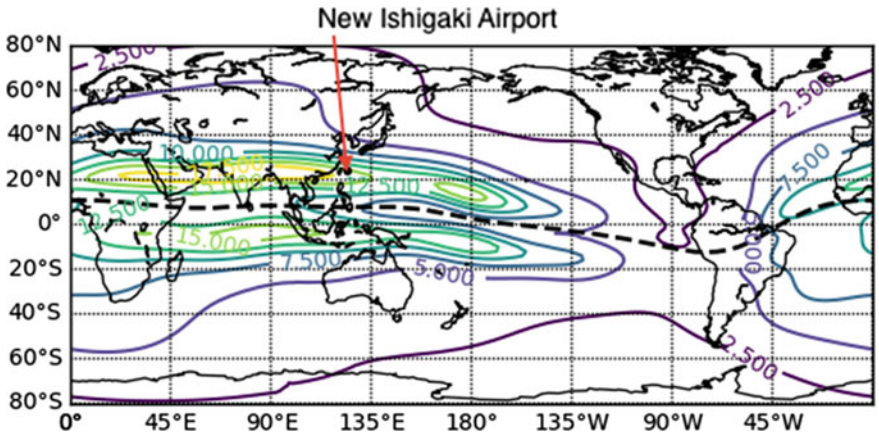


Fig. 5. An example of global distribution of the ionospheric vertical delay at 20JST (Japan Standard Time) in March; numerical values are shown in meters and the red arrow indicates the location of New Ishigaki Airport

The prototype of the GAST-D ground subsystem was installed at New Ishigaki Airport (24.4°N, 124.2°E, 19.6° magnetic latitude; see Fig. 5). This airport was selected because it is the southernmost modern airport in Japan that is compatible with the activities of this project, including long-term data collection under real airport conditions and flight trials with both ground and airborne integrity monitors to validate the ability mitigate the ionospheric threat under actual disturbed ionosphere conditions.

It is important to optimize the parameters to set for each integrity monitor by long-term data collection. Hazardous misleading information (HMI) analysis is also important to monitor the occurrence of unexpected events that threaten the users. The fundamental concept of GAST-D was evaluated and validated in terms of the performance of the integrity monitors in detecting ionospheric anomalies from the data obtained during the flight trials [21].

Figure 6 shows the location of each component of the prototype in New Ishigaki Airport. The prototype has four GNSS reference stations to generate differential corrections. It has one additional station for IFM, which is a component of the GAST-C function of ENRI's prototype to monitor ionospheric delay differences from the GBAS reference stations and reduce the potential ranging errors due to ionospheric spatial gradients in geometry screening processes. The GNSS reference stations and the IFM station each have a GNSS antenna (Novatel ANT-C2GA-TW-N) and a receiver (Novatel Euro-3). The observed GNSS data is transmitted via optical fibers to the data processor located in a building at the airport. GBAS messages are generated by the data processor and sent to the VDB transmitter via optical fibers. A VDB antenna with three elements is used to reduce the nulls, which correspond to gain drops in the radiation pattern due to ground reflection and to maintain the electromagnetic intensity on the approach path and runway surface as specified by the GAST-D development baseline SARPs.

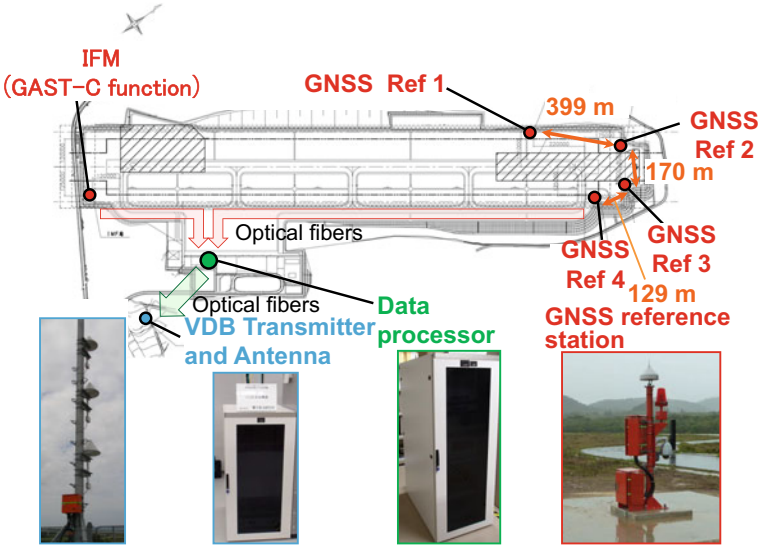


Fig. 6. Location of each component in the prototype installed at New Ishigaki Airport

To monitor the ionospheric conditions independently, an all-sky airglow imager and a short-baseline GNSS-based ionospheric measurement system were installed near the prototype as shown in Fig. 7. The short-baseline GNSS-based ionospheric

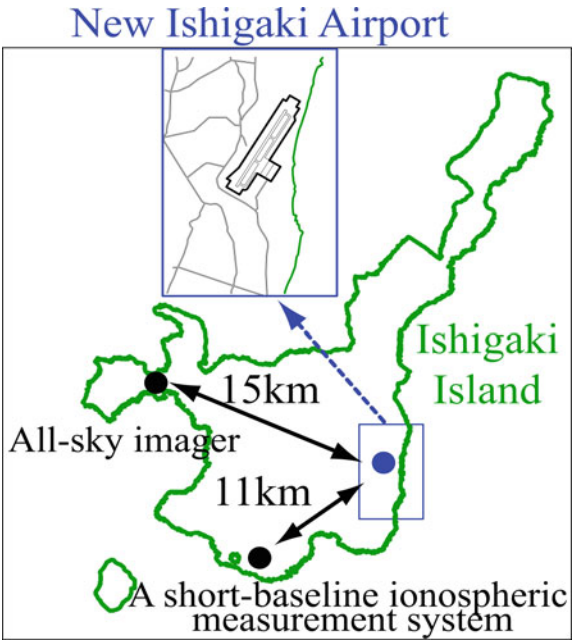


Fig. 7. Locations of the ionospheric observational instruments

measurement system observes ionospheric gradients and scintillation. The all-sky imager detects airglow emissions from the ionosphere at wavelengths of 630.0 and 777.4 nm, which typically have intensities that are proportional to the electron densities at 250 and 350 km, respectively. Thus, the two-dimensional shapes of plasma bubbles can be determined from the all-sky airglow images. These observational systems were also utilized to make go/no-go decisions for the flight experiments based on the plasma bubble conditions.

4 Preliminary Performance Evaluation

As mentioned in the previous section, one of the purposes of continuous operation of the prototype in a real airport is to determine an appropriate set of thresholds for the integrity monitors. For example, a set of parameters for the SDM were calculated from the observational data collected on the roof of the building and used in the initial validation stage before the prototype was delivered. Therefore, it is necessary to reanalyze the monitoring performance and validate the requirements using actual data collected in a real airport. HMI analysis, which is part of the safety monitoring and recording procedure against occurred events that threaten the users, is another important component of long-term data collection.

The GAST-D ground subsystem prototype continuously operated after it was installed in February 2014. We selected two data sets with a duration of a week in spring and summer, which were corresponding to seasons with active and quiet ionospheric conditions, respectively. Table 2 shows the positioning accuracy (95%) for GAST-C and GAST-D operations based on GBAS messages resulted from the prototype and the GNSS observational data obtained by treating the IFM station as a pseudo user [11]. The horizontal and vertical results were, respectively, well within the tolerances of 16 m and 4 m for GAST-C (although no range was given for GAST-D), and there were no significant differences between the two seasons. It should be noted that the accuracy of GAST-D was slightly worse than that of GAST-C in both the horizontal and the vertical components. This is because the position solution is based on carrier smoothing with a time constant of 30 s in GAST-D and a time constant of 100 s in GAST-C. Therefore, the GAST-D solution is susceptible to observational errors such as multipath effects.

Table 2. Positioning accuracy (95%) of the prototype

2014	GAST-C		GAST-D	
	Horizontal (m)	Vertical (m)	Horizontal (m)	Vertical (m)
March 21–27	0.1455	0.3848	0.2010	0.5212
August 7–13	0.1541	0.3705	0.2060	0.4983

In addition, fundamental performance evaluation was conducted on the ISGM and reported in [22]. The initial results with a baseline distance of 399 m between the Ref1 and Ref2 GNSS reference stations showed that the ISGM could feasibly detect

gradients of more than 300 mm/km, which was a typical requirement for the minimum detectable value in the GAST-D development baseline SARPs. In the prototype, a threshold for test statistic of the ISGM was determined with the number of missed detections (P_{md}) of 10^{-9} and the number of false alarms (P_{ffd}) of 2×10^{-7}. Regarding the SDM, it was found to be difficult to satisfy the requirements at low elevation angles of 5–10° because of the measurement noise caused by the ground multipath and the low intensity of the received signals. There are two candidate methods to solve this issue. The first approach is to set a long smoothing time constant for the SDM test statistics to reduce the noise; however, this would slow the response time of the monitor. The other approach is to use a multipath-limiting antenna, which is designed to effectively reject multipath signals from low elevation angles [23], instead of the choke-ring antennas that were employed in the prototype system. The CCDM is designed to detect rapid temporary changes in the ionospheric delay due to ionospheric disturbances or satellite failure. It was found that the CCDM test statistic could be used after removing the contribution of nominal ionospheric delay gradients from the original test statistic and generating a probability distribution of the data to determine a threshold because the divergence due to the ionospheric delay associated with the EIA behaves in a bias-like manner rather than being randomly distributed. Thus, if a CCDM test statistic associated with the EIA exceeds the threshold, even though the EIA always exists and more or less affects the test statistic, the corresponding ranging source can be excluded as an anomalous data from GBAS messages.

The availability of GAST-D service is specified in the development baseline SARPs as a ratio of an available time duration to a required one, and it depends not only on the GAST-D ground subsystem but also on the airborne subsystem because the integrity monitors at both sides are used to mitigate the ionospheric threat. Therefore, the GAST-D positioning analysis was carried out offline using software to evaluate the GAST-D airborne positioning as originally developed for real-time on-board analysis during flight experiments. Figure 8 shows an example of a vertical integrity-availability chart based on GNSS observational data collected on April 9, 2014, when no ionospheric disturbances were observed. The GNSS data was collected by the IFM station as a pseudo user at a sampling rate of 2 Hz. Although final solutions were obtained for only 86.100% of the total of 172,800 epochs, the availability of 99.973% of the final solutions was calculated without HMI events, as shown in Fig. 8. We predict that the low rate of the obtained solutions can be improved by optimizing the parameters for both the ground and the airborne integrity monitors. Further analysis is needed to check whether there are any HMI events that have the actual positioning errors exceeding VPLs, which would not be tolerable for safety reasons.

5 Summary and Future Work

Here, the ENRI developed a prototype of a GAST-D ground subsystem and installed it in a low magnetic latitude region to operationally validate the GAST-D development baseline SARPs. The system was operated continuously to collect a long-term dataset. This dataset was used to analyze the overall system performance and identify HMI events to solve issues that may hinder future GAST-D implementation in Japan and

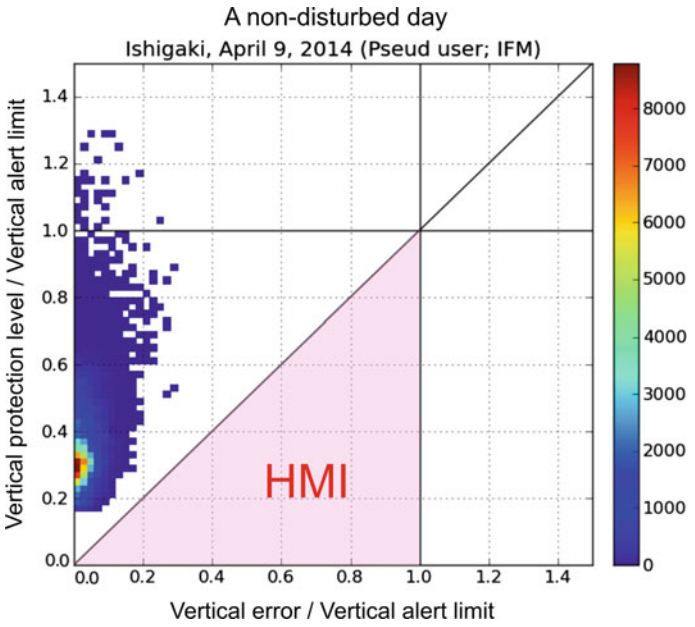


Fig. 8. An example of a result from an availability analysis using GNSS observational data collected on April 9, 2014, when no ionospheric disturbances were observed. The vertical error and vertical protection level (VPL) are normalized to the vertical alert limit (VAL). Color means number of data for each grid

other locations with similar ionospheric conditions. Moreover, an initial fundamental performance evaluation was conducted with regard to the system's accuracy and some integrity monitors with a short-term dataset. In the future, a long-term dataset will be used to (i) validate the parameters for the integrity monitors that were derived from our initial analysis with a short-term dataset, (ii) evaluate the performance of the integrity monitors including missed detection and false alarm rates, and (iii) measure the overall system performance including the system integrity, accuracy, and availability.

Moreover, it will be important to address the issues that became apparent during the long-term operational validation of the GAST-D development baseline SARPs in the NSP working group. Although the validation testing revealed that the SARPs fundamentally covered ionospheric conditions in the magnetic midlatitude region, it also indicated that further studies are needed to enhance the availability of GAST-D in the low magnetic latitude region, where steep spatial gradients in the ionospheric delay and scintillations occur frequently owing to plasma bubbles. The data obtained by the GAST-D ground subsystem can be utilized to investigate the impacts of ionospheric disturbances on the GAST-D ground subsystem in future efforts to achieve the best performance of the GAST-D system in low magnetic latitude regions like Japan.

Acknowledgements. Authors deeply thank Japan Civil Aviation Bureau, Okinawa prefecture, Ishigaki city for their cooperation for the installation and operation of ENRI's GAST-D prototype in New Ishigaki airport.

References

1. Christie J, Ko§ P, Pervan B, Enge P, Powell JD, Parkinson B, Shan C (1998) Analytical and experimental observations of ionospheric and tropospheric decorrelation effects for differential satellite navigation during precision approach. In: Proceedings of the ION GPS 1998 Nashville, TN, pp 739–747
2. Luo M, Pullen S, Walter T, Enge P (2004) Ionospheric spatial gradient threat for LAAS: mitigation and tolerable threat space. In: Proceedings of the ION NTM 2004, San Diego, CA, pp 490–501
3. Luo M, Pullen S, Ene A, Qiu D, Walter T, Enge P (2004) Ionosphere threat to LAAS: updated model, user impact, and mitigations. In: Proceedings of the ION GNSS 2004, Long Beach, CA, pp 2771–2785
4. Lee J, Luo M, Pullen S, Park YS, Enge P, Brenner M (2006) Position-domain geometry screening to maximize LAAS availability in the presence of ionosphere anomalies. In: Proceedings of ION GNSS 2006. Fort Worth, TX, pp 393–408
5. Ramakrishnan S, Lee J, Pullen S, Enge P (2008) Targeted ephemeris decorrelation parameter inflation for improved LAAS availability during severe ionosphere anomalies. In: Proceedings of ION NTM 2008, San Diego, CA, pp 354–366
6. Lee J, Seo J, Park Y, Pullen S, Enge P (2011) Ionospheric threat mitigation by geometry screening in ground-based augmentation systems. J Aircraft 48(4), 1422–1433
7. Ene A, Qiu D, Luo M, Pullen S, Enge P (2005) A comprehensive ionosheric storm data analysis method to support LAAS threat model development. In: Proceedings of ION NTM 2005. San Diego, CA, pp 110–130
8. Pullen S, Park YS, Enge P (2009) Impact and mitigation of ionospheric anomalies on ground-based augmentation of GNSS. Radio Sci 44, RSOA21. https://doi.org/10.1029/2008rs004084
9. Fukushima S, Kudo M, Saitoh S, Yoshihara T, Saito S, Fujita S, Fujii N (2011) Development and Safety assurance of satellite-based precision approach and landing system. IEICE J94–B(7), 802–811 (Japanese)
10. ICAO/NSP WGW May 2010 Report—Attachment G, NSP May 10/Flimsy 29, "Development Baseline SARPs Proposal", Presented by Tim Murphy and Stefan Naerlich
11. Yoshihara T, Saito S, Kezuka A, Hoshinoo K, Fukushima S, Saitoh S (2015) Development and validation of a CAT-III GBAS (Ground-based Augmentation System) prototype for magnetic low latitude region. In: Proceedings of the 56th aircraft symposium, 2G02 (JSASS-2015–5145), Matsuyama-shi (Japanese)
12. GNSS-Based Precision Approach Local Area Augmentation System (LAAS) Signal-in-Space Interface Control Document (ICD), RTCA/ DO-246D, December 16, 2008
13. Minimum operational performance standards for GPS local area augmentation system airborne equipment, RTCA/DO-253C, December 16, 2008
14. Certification considerations for highly-integrated or complex aircraft systems, SAE ARP4754, November 1996
15. Guidelines and methods for conducting the safety assessment process on civil airborne systems and equipment, SAE ARP4761, December 1996
16. Yoshihara T, Murashi F, Saito S, Hoshinoo K (2015) A study on integrity improvement of GBAS ground subsystem using CSAC (Chip Scale Atomic Clock). In: Proceedings of ION ITM 2015, Dana point, CA, pp 593–599
17. Yoshihara T, Motoyoshi H, Sato T, Yamaguchi S, Saito S (2013) GAST-D integrity risks of snow accumulation on GBAS reference antennas and multipath effects due to snow-surface reflection. In: Proceedings of ION ITM 2013, San Diego, CA, pp 112–120

18. ICAO/NSP WGW Nov 2009 Report—Attachment H, NSP Nov 09/Flimsy 6, "Conceptual Framework for the Proposal for GBAS to Support CAT III Operations", Presented by CSG Rapporteur
19. Di Giovanni G, Radicella S (1990) An analytical model of the electron density profile in the ionosphere. Adv Space Res 10(11):27–30
20. Radicella S, Zhang M (1995) The improved DGR analytical model of electron density height profile and total electron content in the ionosphere. Annali di Geofisica, 38(1):35–41
21. Saito S, Yoshihara T, Kezuka A, Saitoh S, Fukushima S, Otsuka Y (2015) GAST-D flight experiment results with disturbed and quiet ionospheric conditions. In: Proceedings of ION GNSS + 2015, Tampa, FL, pp 1494–1499
22. Saito S, Yoshihara T, Nakahara H (2015) Performance of GAST-D ionospheric gradient monitor studied with low latitude ionospheric disturbance data obtained in a real airport environment. In: Proceedings of ION Pacific PNT 2015, Honolulu, HI, pp 815–820
23. Lopez A (2008) LAAS/GBAS ground reference antenna with enhanced mitigation of ground multipath. In: Proceedings of ION NTM 2008, San Diego, CA, pp 389–393

Interoperability of the ENRI GAST-D Ground-Station Prototype with Different Airborne Software Implementations

S. Saito[1(✉)], M. Stanisak[2], T. Yoshihara[1], T. Feuerle[2], and A. Lipp[3]

[1] Navigation Systems Department, Electronic Navigation Research Institute (ENRI), Tokyo, Japan
{susaito, yosihara}@mpat.go.jp
[2] Institute of Flight Guidance, Technische Universität Braunschweig (TU-BS), Braunschweig, Germany
{m.stanisak, t.feuerle}@tu-braunschweig.de
[3] EUROCONTROL Experimental Center, Bretigny-sur-Orge, France
andreas.lipp@eurocontrol.int

Abstract. Ground-based augmentation system (GBAS) interoperability trials were conducted in Ishigaki, Japan in June 2016. The interoperability of different implementations of GBAS airborne equipment/software including the TriPos from TU Braunschweig, a software from ENRI, and the Pegasus software developed by EUROCONTROL were tested with an experimental GAST-D ground prototype developed by ENRI and manufactured by NEC. Some differences were observed in protection levels, likely because of differences in the software tools' satellite selection strategies. Even so, the position solutions and course deviations were in good agreement between the different implementations, so that interoperability was demonstrated successfully.

Keywords: Ground-Based augmentation system (GBAS) · GBAS approach service Type-D (GAST-D) · Interoperability · Low latitude ionosphere

1 Introduction

The Ground-based augmentation system (GBAS) is now used globally for Category-I precision approach operations. For Category-II and III operations with GBAS, new international standards (Standards and Recommended Practices or SARPs) [1] for GBAS Approach Service Type-D (GAST-D) that use GPS and GLONASS L1 signals were finalized by the Navigation Systems Panel (NSP) of International Civil Aviation Organization (ICAO) in December 2016 and will take effect in November 2018.

To contribute to the operational validation of the GAST-D SARPs, the Electronic Navigation Research Institute (ENRI) has developed a prototype GAST-D ground subsystem and installed it at New Ishigaki Airport, Japan [2]. The prototype was designed with the development baseline SARPs for GAST-D [3], which were defined for the sake of operational validation and differ slightly from the final version of the GAST-D SARPs. Since the integrity requirements are assured not only by a ground subsystem but also jointly by an airborne subsystem, ENRI also developed a GAST-D

Electronic Navigation Research Institute, *Air Traffic Management and Systems III*, Lecture Notes in Electrical Engineering 555,
https://doi.org/10.1007/978-981-13-7086-1_10

experimental airborne system and conducted flight trials successfully in different ionospheric conditions [4].

To ensure that various ground- and airborne systems manufacturers interpret the SARPs uniformly, interoperability between different combinations of ground and airborne subsystems is a vital aspect of the operational validation of the GAST-D SARPs. Interoperability between a ground facility mockup of Multi-Constellation/Multi-Frequency (MC/MF) GBAS, which is a further generation GBAS and includes GAST-D capability, developed by TU Braunschweig (TUBS), and different airborne software/receivers including ENRI's one has been successfully demonstrated in trials conducted at Toulouse Airport, France in May 2016 [5, 6]. The trials were conducted under the frame of SESAR 15.3.7.

To demonstrate interoperability between the ENRI's GAST-D ground prototype and airborne software/receivers, another joint interoperability trial was conducted in June 2016 at Ishigaki, Japan in the same framework as the interoperability trials conducted at Toulouse. This paper presents the results of this interoperability trial.

The rest of the paper is organized as follows. The experimental setup, including the software, hardware, and the observation geometry is introduced in Sect. 2. Section 3 explains the tools used for analysis and their relevant parameters. The results are presented and discussed in Sect. 4. Section 5 concludes the paper.

2 Experimental Setup

The Interoperability trials between ENRI's GAST-D ground prototype and airborne software/receivers were conducted from 21 to 24 June 2016 at Ishigaki, Japan. Measurements were taken in daytime only, because of limitations on access to the test site. The airborne software and receivers included a GAST-D-capable Multi-Mode Receiver (MMR) prototype developed by Thales, the TriPos airborne software from TUBS, and ENRI's airborne experimental system (ENRI-Airborne). The Pegasus tool developed by EUROCONTROL (http://www.eurocontrol.int/pegasus) were used to later analyze offline the raw data recorded by GNSS and VDB receivers. The PEGASUS software provides a software-based emulation of a minimum-performance receiver to provide manufacturer-independent references.

All the airborne subsystems were connected to a fixed stationary GNSS antenna on the ground. The GNSS antenna (Trimble GNSS Choke Ring Antenna) was located at the Japan Civil Aviation Bureau's aeronautical communications transmitting site near the New Ishigaki Airport. It is about 2 km southwest from the GBAS reference point for the GAST-D ground prototype.

Figure 1 shows the geometry of the broadcast approach paths and the position of the user's GNSS antenna. This user position is outside the Precision Approach Region (PAR) for the two approaches of the runway of the airport that are usually served by GBAS final approach segments (FASs) in the GBAS message type-4. To address this limitation, an additional virtual approach was created so that the user position was inside the PAR of this approach. The virtual approach path was set up to be parallel to the approach to the RWY-04 of the New Ishigaki Airport and was offset horizontally and vertically so that the vertical and lateral deviations from the virtual approach path at

the user's GNSS antenna would be finite and nonzero. The Landing Threshold Point (LTP)/Fictitious Threshold Point (FTP) and the Flight Path Alignment Point (FPAP) for the virtual path are shown in Fig. 1, along with the PAR associated with the virtual path.

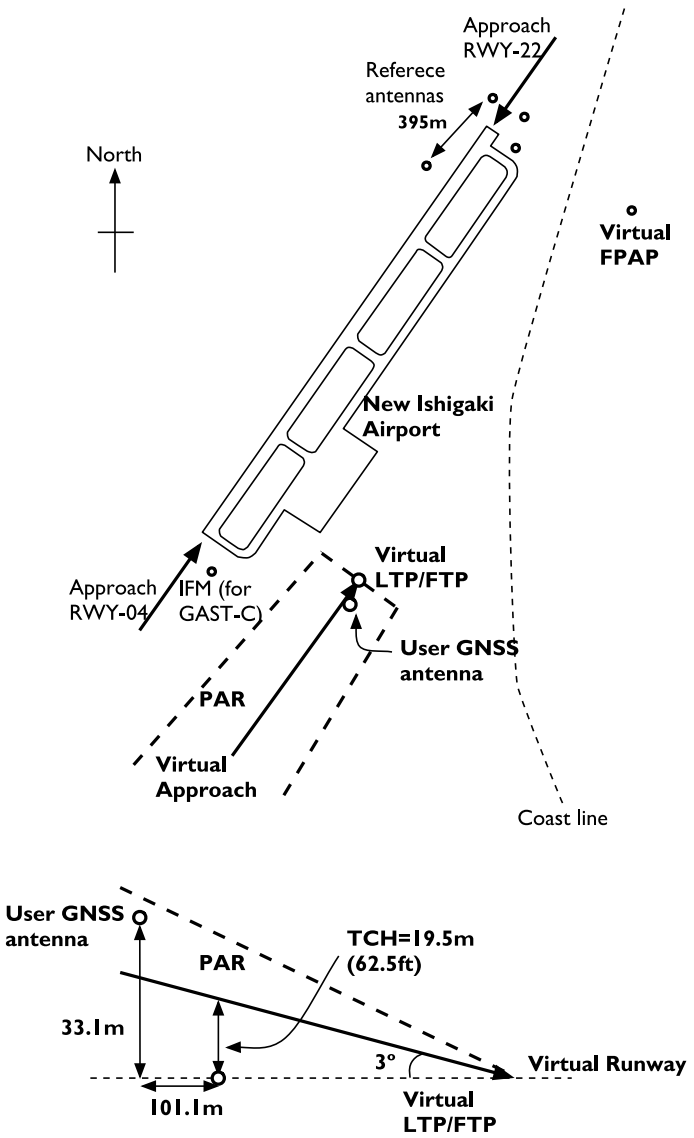


Fig. 1. Test setup at new Ishigaki airport. Horizontal and vertical layouts are shown in the top and bottom panels, respectively. Vertical dimensions are not drawn to scale

No severe ionospheric disturbances, such as plasma bubbles, affected the experiment, because the measurements were made in daytime and geomagnetic conditions were quiet throughout the period of the trial.

3 Analysis

TriPos, ENRI-Airborne, and Pegasus software were used to analyze the same GNSS and VDB data recorded by a Javad Delta GNSS receiver and a Telerad VDB receiver, respectively. To keep the analysis conditions as uniform as possible, all the three GAST-D airborne processing tools used the same set of parameters listed in Table 1. The Aircraft Accuracy Designator (AAD) value is set as "B," which is a requirement for airborne subsystems that support GAST-D [7]. The σD_V and σD_L values were estimated roughly from the D_V and D_L values obtained from the analysis discussed in this paper. However, these values are common among the tested software, and will not affect the results of the interoperability analysis. The limitSvert, limitSlat, and limitSvert2 values depend on the aircraft integration [8]. The value of 3.5 was used here for limitSlat and limitSvert, both are within the range of values used in the conceptual validation of GAST-D [8]. And the same value was adopted for the limitSlat.

Table 1. Parameters of airborne data analysis

Parameter	Meaning	Value
AAD	Aircraft Accuracy Designator	B
$CN_{0,min}$	Minimum Carrier-to-Noise density	33 dBHz
θ_{min}	Minimum satellite elevation angle	5°
σD_V	Standard deviation of D_V	0.3 m
σD_L	Standard deviation of D_L	0.3 m
limitSvert	Largest allowed value of the local vertical component of the projection matrix	3.5
limitSlat	Largest allowed value of the local lateral component of the projection matrix	3.5
limitSvert2	Maximum sum of the two maximum absolute values of the local vertical component of the projection matrix	4.5
maxVAL	Largest allowed vertical alert limit	10
maxLAL	Largest allowed lateral alert limit	40

According to the latest standards for GBAS airborne subsystems (DO-253D) [7], limitSvert2 in Table 1 is computed from the maximum undetected pseudorange error caused by an ionospheric gradient (E_{IG}). E_{IG} is computed by the airborne receiver based on the broadcast parameters (Y_{EIG} and M_{EIG}) in the additional data block 3 of GBAS message type-2. However, this value was explicitly defined, because ENRI's experimental GAST-D ground subsystem was designed according to the development baseline SARPs [2] before the E_{IG} parameter was introduced. The value of limitSvert2

was chosen to be smaller than the conservative values with a worst-case assumption appearing in the literature [8], to avoid excessive screening of the satellite geometries.

For reference, the position of the user's GNSS antenna was determined through a kinematic analysis with accuracy of about 1 cm. This allows the direct calculation of the resulting GBAS position errors.

4 Results and Discussion

4.1 Deviation Output

Figure 2 compares the vertical and lateral guidance outputs calculated by the three different software tools. In order to be comparable with legacy systems, all the deviations are given as Differences in Depth of Modulation (DDM), as used in the Instrument Landing System (ILS). Aside from some minor differences, the results are in good agreement with each other. In addition, they are in good agreement with the DDM values computed using the position of the reference antenna and the FAS parameters. This means that all the receiver software tools processed the GBAS corrections as intended, obtained very similar GBAS solutions, and computed the DDM values from the correct interpretations of the FAS information contained in the GBAS message type-4. Although the experimental GAST-D MMR by Thales suffered technical problems and functioned only intermittently, visual inspection of the control software screen during the trials confirmed that the system output was consistent during the experiment, when outputs were available.

4.2 Position Solutions with Different Smoothing Time Constants

As one of GAST-D specific outputs, Fig. 3 compares the vertical components of position solution differences with two different smoothing-time constants (100 and 30 s, D_V), which are used to add additional margins to the vertical protection level in GAST-D.

The results show occasional spikes caused probably by ground multipath and other external disturbances because the position solution depends on the geometry of the satellites used for the position computation as well as the timing of the smoothing. Some of the differences of ENRI-Airborne results from others may be explained by its unique satellite selection strategy, and that it also uses SBAS GEO [9]. Nevertheless, the D_V values delivered by the three receiver software tools were in good agreement.

4.3 Protection Levels

Figure 4 compares the vertical protection levels (VPL) derived by each of three-receiver software tools for a selected 30-minute interval. In contrast to the vertical and lateral DDM results, noticeable differences appear, though the trends are similar. In addition to bias-type differences in VPL, the ENRI-Airborne output shows spikes in the VPL.

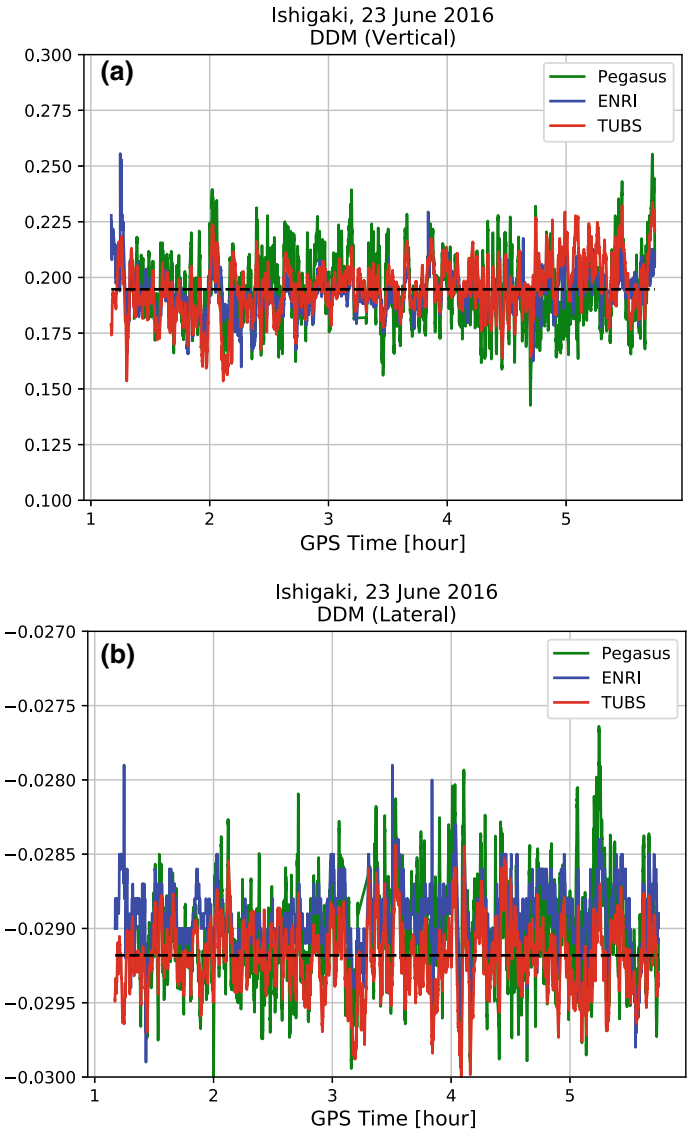


Fig. 2. **a** Vertical and **b** lateral DDM values, respectively derived by ENRI-Airborne, TriPos (TUBS), and Pegasus software respectively from the data collected on 23 June 2016. Dashed lines show the vertical and lateral DDM values computed from the true positions and the FAS information

Figure 5 shows the differences in the VPL values by ENRI-Airborne and TriPos with respect to those from Pegasus for the entire period of the experiment. The VPL values estimated by TriPos and Pegasus were mostly similar. The mean and the standard deviation of differences were 0.10 m and 0.37 m, respectively. On the other

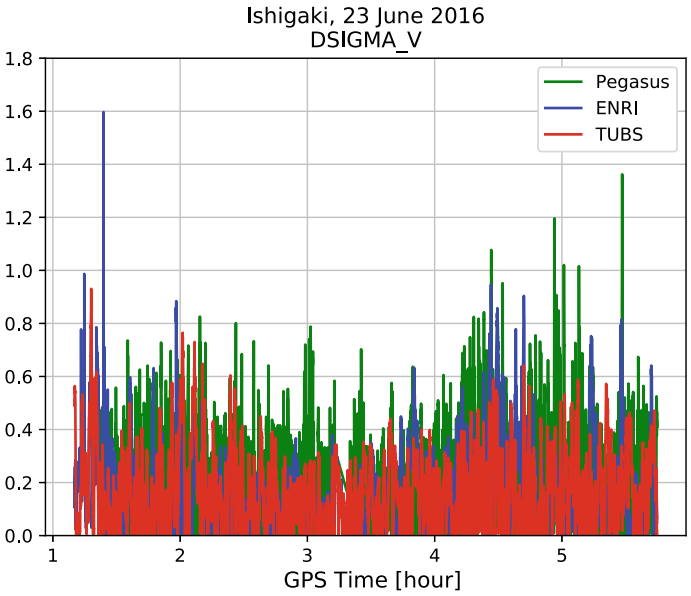


Fig. 3. D_V values generated by derived by ENRI-Airborne, TriPos (TUBS), and Pegasus software for the data collected from 03:30 to 04:30 GPST on 23 June 2016

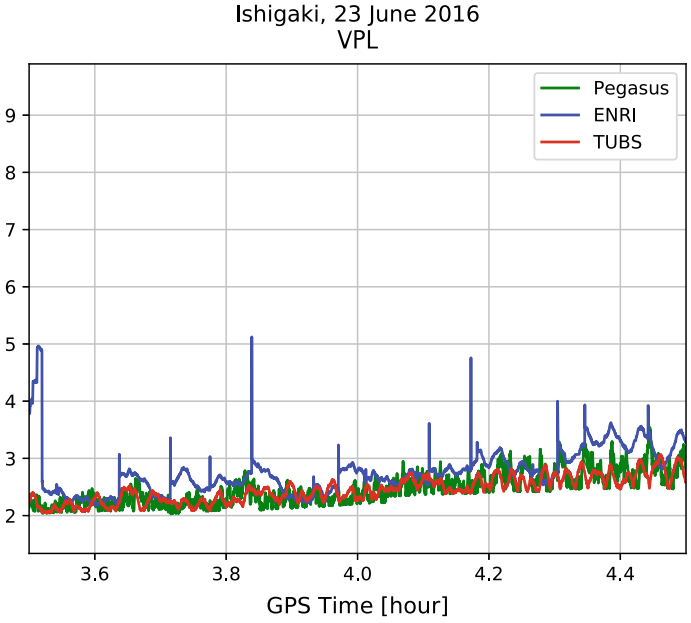


Fig. 4. VPL values generated by derived by ENRI-Airborne, TriPos, and Pegasus software for the data collected from 03:30 to 04:30 GPST on 23 June 2016

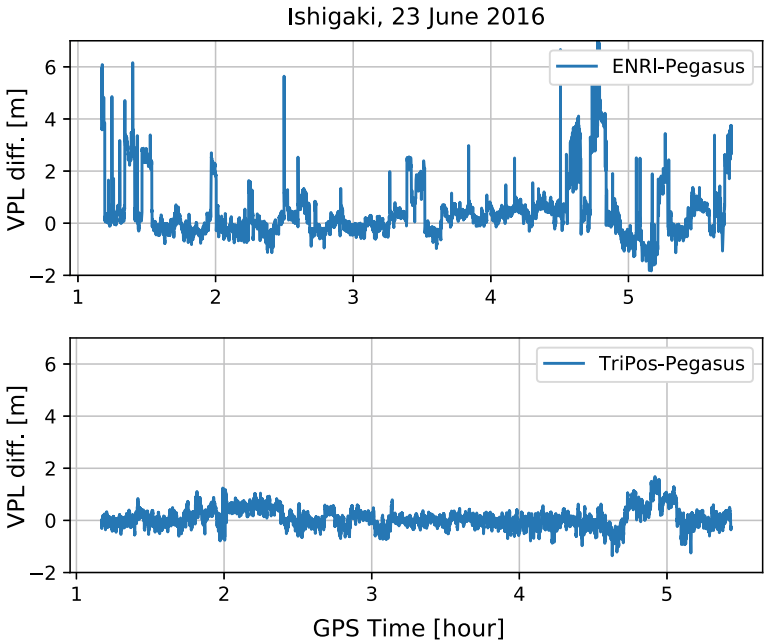


Fig. 5. Differences in VPL values estimated by (top) ENRI-Airborne and Pegasus and (bottom) TriPos and Pegasus

hand, the VPL values estimated by ENRI-Airborne were sometimes noticeably different from those estimated by Pegasus, with as much as 6.5 m disagreement, though the values still agreed overall. The mean and the standard deviation of differences of VPL values estimated by ENRI-Airborne and Pegasus were 0.45 m and 1.1 m, respectively.

Some differences in VPL are expected, because unlike the other outputs, VPL depends not only on the GAST-D airborne parameters but also on algorithm and thresholds of the integrity monitors that may exclude satellites and satellite geometry selection strategy including re-admittance after exclusion by integrity monitors. Additionally, ENRI's experimental ground and airborne subsystems support the geosynchronous equatorial orbit satellite of satellite-based augmentation system (SBAS GEO) as a ranging source, while the other tools do not support this. Indeed, the SBAS GEO of MTSAT Satellite-based Augmentation System (MTSAT-2) was tracked and used throughout the trials.

The differences in the VPL values are within expectations and are acceptable as long as the integrity is maintained, though larger VPL values can potentially degrade availability.

4.4 Integrity and Availability

Figure 6 shows the normalized vertical integrity diagrams or Stanford charts derived by the three receiver software tools. TriPos returned the smallest distribution with no integrity failure and no availability loss. Pegasus returned a similar distribution as TriPos, which was just slightly wider. Pegasus suffered no integrity failure or availability loss. ENRI-Airborne delivered a vertically elongated distribution of data points, which correspond to the spiky VPL results. However, the core of the distribution is similar to those of TriPos and Pegasus. Like the other software tools, ENRI-Airborne also suffered no integrity failure or availability loss.

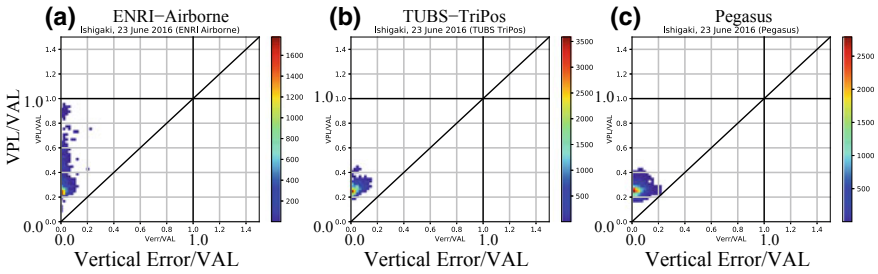


Fig. 6. Vertical integrity diagrams derived by **a** ENRI-Airborne, **b** TriPos, and **c** Pegasus software for the data collected on 23 June 2016

5 Concluding Remarks

Good overall agreement was observed between the different airborne software implementations. The deviations and position solution differences with different smoothing time constants (D_V) were in very good agreement. The Thales MMR prototype performed consistently. The VPL values varied somewhat, likely because of different implementations of the satellite selection strategy, the designs of integrity monitors, and the use of SBAS GEOs. However, all the software tools delivered very good overall integrity and availability. All the three receiver implementations, ENRI-Airborne, TriPos, and Pegasus, proved to be interoperable with ENRI's GAST-D experimental ground subsystem.

Acknowledgements. This work presented in this paper was funded by EUROCONTROL in scope of contracts between EUROCONTROL and TU Braunschweig. The experiment in Ishigaki, Japan was supported technically by Japan Civil Aviation Bureau.

References

1. ICAO NSP Flimsy 17 (2016) Proposed Amendment to International Standards and Recommended Practices, Aeronautical Telecommunications, Annex 10 to the Convention on International Civil Aviation. Volume I (Radio Navigation Aids), Montreal, Canada, November-December, 2016
2. ICAO (2010) GBAS CAT II/III Development Baseline SARPs
3. Yoshihara T, Saito S, Kezuka A, Fukushima S, Saitoh S (2017) Development of a CAT-III GBAS (GAST-D) ground subsystem prototype and its performance evaluation with a long term-data set. In: Proceedings of the 5th international workshop on ATM/CNS (EIWAC2017), November 2017
4. Saito S, Yoshihara T, Kezuka A, Saitoh S, Fukushima S, and Otsuka Y (2015) GAST-D flight experiment results with disturbed and quiet ionospheric conditions. Proc. ION GNSS + 2015, 1494–1499
5. Feuele T, Stanisak M. Saito S, Yoshihara T (2016) GBAS interoperability trials and multi-constellation/multi-frequency ground mockup evaluation. In: Proceedings of the 6th SESAR innovation days, EUROCONTROL. ISSN 0770-1268
6. Feuele T, Stanisak M, Saito S, Yoshihara T (2017) GBAS interoperability and multi-constellation/multi-frequency trials. In: Proceedings of the 5th international workshop on ATM/CNS (EIWAC2017), November 2017
7. RTCA/DO-253D (2017) Minimum operational performance standards for GPS local area augmentation system airborne equipment. July 13, 2017
8. ICAO NSP Flimsy 10 (2016) Conceptual framework for the proposal for GBAS to support CAT III operations. Montreal, Canada, November–December, 2016
9. Saito S, Vernez K, Yoshihara T, and Kezuka A (2014) Performance improvement of GAST-D Airborne monitor algorithms under disturbed ionospheric conditions. In: Proceedings of ION GNSS + 2014, pp 815–821

GBAS Interoperability and Multi-constellation/Multi-frequency Trials

T. Feuerle[1(✉)], M. Stanisak[1], S. Saito[2], T. Yoshihara[2], and A. Lipp[3]

[1] Institute of Flight Guidance, Technische Universität Braunschweig, Braunschweig, Germany
{t.feuerle,m.stanisak}@tu-braunschweig.de
[2] Electronic Navigation Research Institute, Navigation Systems Department, Tokyo, Japan
{susaito,yosihara}@enri.go.jp
[3] Eurocontrol Experimental Centre, Bretigny-sur-Orge, France
andreas.lipp@eurocontrol.int

Abstract. This paper summarizes GBAS tests conducted by the Technische Universität Braunschweig (TUBS), the Electronic Navigation Research Institute (ENRI) and European Organisation for the Safety of Air Navigation (EUROCONTROL) in the frame of the Single European Sky ATM Research Programme (SESAR, subproject 15.3.7). For these tests, an experimental GBAS ground facility transmitted VDB data for current and future approach services. On the one hand, this paper summarizes the efforts to ensure backward compatibility and interoperability. To this end, existing GBAS hardware receivers were tested and compared with different software solutions. These tests confirmed that the proposed VDB formatting does not interfere with existing GBAS implementations. On the other hand, these tests were intended to support the development of future Multi-Constellation and Multi-Frequency GBAS services. GNSS data recorded at the ground reference receivers and the research aircraft D-IBUF was used to calculate an experimental ionosphere-free GBAS solution not threatened by ionospheric gradients. This paper presents initial results for the proposed GAST-F service, demonstrating that (despite the low number of L1 and L5 capable GNSS satellites currently available) this kind of processing could be a candidate for a future MC/MF CAT-II/III GBAS service.

Keywords: Ground-Based Augmentation System (GBAS) · Multi-Constellation (MC) · Multi-Frequency (MF) · GAST-F

Electronic Navigation Research Institute, *Air Traffic Management and Systems III*, Lecture Notes in Electrical Engineering 555,
https://doi.org/10.1007/978-981-13-7086-1_11

1 Introduction

Recent publications [1, 2] demonstrated that the use of different Global Navigation Satellite Systems (GNSS) such as the Global Positioning System (GPS, USA), Glonass (Russia) or Galileo (Europe) can improve the Ground-Based Augmentation System (GBAS) significantly. Even if not being included in the relevant standards yet, the use of these additional ranging sources in a Multi-Constellation (MC) GBAS offers tremendous potential for improving GBAS availability, continuity and especially integrity.

The current planning of GBAS manufacturers and certification agencies foresees a type certificate for GBAS Approach Service Type (GAST) D ground stations soon. GAST-D has been designed to offer CAT-III precision approach capabilities based on GPS L1 C/A as the sole usable GNSS signal. The main challenges for CAT-II/III GBAS precision approach services are ionospheric anomalies threatening the overall integrity. GAST-D will thus require a combination of extensive airborne and ground monitoring algorithms in order to mitigate all threats. However, with the increasing availability of modern GNSS signals on different frequencies, the ionospheric threat can be mitigated almost completely by Multi-Frequency (MF) GNSS techniques.

Currently, four main GBAS approach service types (GAST) are defined to cover different operational requirements (see [3, 4]):

- GAST-A: APV-I approach service (Approach Procedure with Vertical guidance)
- GAST-B: APV-II approach service
- GAST-C: CAT-I approach service (using L1 signals only)
- GAST-D: CAT-II/III approach service (using L1 signals only).

The trials described in this paper focused on the GBAS approach service types C and D for legacy and interoperability purposes. In addition, experimental MC/MF services were addressed. To this end, two additional GBAS approach services have been defined in the frame of subproject 15.3.7 of the Single European Sky ATM Research Programme (SESAR 15.3.7):

- GAST-E: CAT-I approach service (using L5 signals only)
- GAST-F: CAT-II/III approach service (using L1 and L5 signals).

All named services can be provided technically both using one GNSS constellation (usually GPS) or in combination with another one. However, next to GPS, only the Russian Glonass system is included in the current ICAO SARPs, all other systems have to be considered experimental at the moment.

For the proposed GAST-F dual-frequency GBAS service, two different modes of operation are foreseen at the moment. This paper focuses on the ionosphere-free (I/Free) mode in which the L1 and L5 corrections transmitted by a ground facility are combined into an I/Free correction which, in turn, is applied to the I/Free combination of the L1 and L5 measurements of the airborne receiver. Thus, whenever we talk about GAST-F in this paper, we address the proposed I/Free MC/MF GBAS processing.

This paper is divided into three major sections. First, Sect. 2 describes the trials used for collecting MC/MF GBAS data. The processing and evaluation of this data is then shown in Sect. 3. Finally, Sect. 4 concludes this paper. Despite the low number of available L5 capable satellites, this paper demonstrates that the proposed processing could meet the requirements for a future MC/MF CAT-II/III GBAS service.

2 Equipment and Trial Setup

This section describes the overall setup and equipment used for the experiments. This includes flight trials conducted in Toulouse (Sect. 2.1) as well as static trials conducted in Ishigaki (Sect. 2.2). During both campaigns, all data was recorded for further post processing.

2.1 Flight Trials in Europe

All flights presented in this paper were conducted in 2016 by the research aircraft D-IBUF of the Technische Universität Braunschweig (TUBS) [5]. This Dornier Do 128-6 (as shown in Fig. 1) is operated by the Institute of Flight Guidance and is generally used for airborne measurement campaigns in various research areas.

Fig. 1. Research aircraft D-IBUF at Toulouse Airport

For this GBAS measurement campaign, different GBAS-related equipment was integrated (see Table 1). This included a Rockwell–Collins GAST-C Multi-Mode Receiver (MMR), an experimental GAST-D MMR prototype developed by Thales in the frame of SESAR, an MC/MF Javad GNSS receiver as well as a Telerad VDB receiver. All raw data of these devices was time-stamped and recorded using the

aircraft's real-time data acquisition and recording system for later post processing. In addition, the data of the GNSS and the VDB receiver was also used online for two GBAS airborne simulations running in parallel. The first package was a GAST-D airborne simulation developed by ENRI (Japan), the second was an experimental GBAS simulation developed by TUBS focusing on MC/MF GBAS.

Table 1. Equipment of research aircraft D-IBUF

	Equipment	Notes
Hardware	Rockwell–Collins MMR GLU-925-430	Test of interoperability and compatibility. Baseline for trials
	Thales MMR TLS-2060	Test of interoperability with SESAR prototype
	Javad delta GNSS receiver	GNSS reception (GPS L1 and L5, Galileo E1 and E5a, Glonass L1)
	Telerad RE 9009A	VDB reception
Software	ENRI-Airborne	Test of interoperability. Independent GAST-D software implementation developed by ENRI (Japan)
	TriPos	Test of interoperability. Experimental MC/MF processing along with GAST-C/-D processing. Software developed by Technische Universität Braunschweig (Germany)

After having received the approval of airworthiness, initial tests of the installation were performed at the research airport Braunschweig (being the home base of the research aircraft). Here, the installed equipment was tested successfully with an experimental GAST-D ground facility prototype (manufactured by Thales Air Systems and operated by the German Aerospace Center DLR) and an experimental MC/MF ground-station software developed by TUBS [2]. GAST-C approaches were conducted in Bremen (Germany) where an approved GAST-C station (Honeywell SLS4000) is operated by the German air navigation service provider (Deutsche Flugsicherung—DFS). At Frankfurt airport, one approach on runway 25 L was performed using the SESAR GAST-D ground station prototype of Indra Navia installed there. These flights verified that all equipment had been integrated correctly and worked flawlessly with the different ground stations.

Finally, the TUBS/ENRI flight experiment team participated in the evaluation flight trials of SESAR 15.3.7, which took place from May 17–20th 2016 at Toulouse airport. In total, 32 approaches were conducted throughout three research flights with the research aircraft D-IBUF (see Fig. 2). All approaches were flown with a 6–12 NM final approach segment.

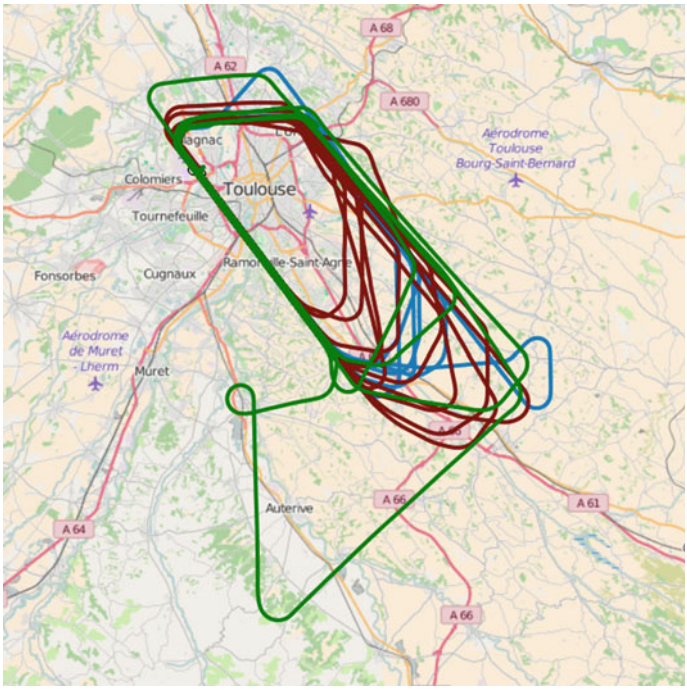


Fig. 2. Flight Patterns in Toulouse

The approaches in Toulouse were flown using the VDB broadcast of the experimental ground station mock-up developed by TUBS. This broadcast used an experimental VDB message formatting developed by TUBS as well (see [1]). Existing hardware of the French air navigation service provider (Direction des Services de la navigation aérienne—DSNA) at Toulouse airport was used as intensively as possible. This way representative hardware with corresponding performance could be used for the trials. During the campaign, Honeywell used a Dassault Falcon to verify their airborne developments independently.

In addition, TUBS performed interoperability trials in order to prove that ENRI's software package was also compatible with the SESAR prototypes. These interoperability tests were needed for validating the documents drafted on ICAO level as a baseline for worldwide GAST-D developments.

2.2 Static Trials in Japan

In Japan, GBAS ground station equipment (hard- and software) as well as airborne components have been developed in recent years independently. ENRI developed together with NEC a GAST-C ground station prototype, which was installed at Kansai International Airport (KIX). This installation was already tested in different campaigns

with ENRI's experimental aircraft (with a Rockwell–Collins GLU-925 MMR installed), an experimental aircraft of Japan's Aerospace Exploration Agency (JAXA, equipped with a Rockwell–Collins GNLU-930 MMR), as well as Boeing 787 aircraft operated by All Nippon Airlines (ANA) and Japan Airlines (JAL).

Based on this GAST-C ground-station prototype, a follow-on development of a GAST-D prototype was developed and installed at Ishigaki New International Airport (ISG). This airport is located at the Island of Ishigaki in the southwestern part of Japan (24.4°N, 124.1°E, see Fig. 3). Due to its location in low magnetic latitudes, this installation is heavily affected by ionospheric disturbances called plasma bubbles. More details on the GAST-D development of ENRI can be found in [6].

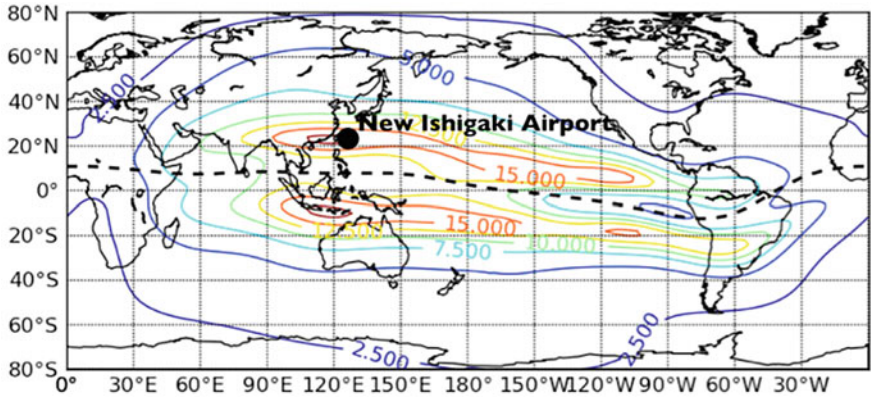


Fig. 3. Location of the New Ishigaki Airport. Typical distribution of the ionospheric delay at the L1 frequency (1.57542 GHz) at 03 UT is also plotted

3 Data Evaluation

3.1 Interoperability

The main purpose of the conducted test campaigns was to prove the feasibility of MC/MF GBAS according to the proposed VDB formatting and processing, especially with regards to the backward compatibility with the existing GBAS services and the VHF Data broadcast.

It can be summarized that the different pieces of equipment worked as expected during all approaches conducted at the different airports in Europe and the static trials in Japan. All GBAS solutions were able to provide valid GBAS approach guidance for their respective approach service all the time. Of course, a performance assessment is not meaningful due to the experimental ground mock-up. Nevertheless, a comparison of the different GBAS services allowed validating the correct overall operation.

Additionally, in order to qualify the correct operations, one specific approach was used to compare and demonstrate the guidance information from different independent GBAS solutions as an example. Four distinct solutions have been excerpted or calculated from the recorded onboard data:

- Rockwell–Collins GLU-925: This Multi-Mode Receiver (MMR) is capable of GBAS CAT-I (GAST-C) and was analyzed by its ARINC 429 output.
- ENRI software: This experimental GBAS software developed by ENRI (Japan) implements the GBAS Approach Services C and D and allows live processing of the received GNSS and VDB data. A Javad GNSS receiver and a Telerad VDB receiver was used here.
- PEGASUS: The navigation toolbox PEGASUS was developed by EUROCONTROL and is one of the main references for GBAS processing worldwide. Even though a MC/MF implementation is being developed, GAST-D is the highest service used here.
- TriPos: This navigation framework was developed by the Institute of Flight Guidance at TU Braunschweig and includes experimental GBAS processing for GAST-C and GAST-D, as well as GAST-E and GAST-F for MC/MF GBAS.

The interoperability trials have been analyzed in [5] in detail, demonstrating only minor differences between the guidance information of all GBAS solutions and services. These interoperability trials proved that the draft ICAO SARPs are mature and stable enough to allow independent developments worldwide for the upcoming GAST-D GBAS CAT-II/III service. In addition, a first attempt of calculating an MC/MF GBAS solution was presented in that paper. The achieved performance of this MC/MF GBAS service will be further assessed in the next section.

3.2 MC/MF Concept Validation

The main objective of the tests in Toulouse in May 2016 was to demonstrate the feasibility of a MC/MF VDB broadcast using live GPS and Galileo data only. Nevertheless all elements for calculating an MC/MF ionosphere-free GAST-F solution were present in principle.

In order to calculate an ionosphere-free GBAS solution, GNSS satellites capable of providing L1 and L5 signals are required for both ground and airborne processing. In addition, the used navigation data has to cover the signals in both frequency bands. For instance, the legacy navigation data provided by L1 C/A (LNAV) does not contain any parameters describing the L5 signal at all. This is why L5-capable GPS satellites provide new CNAV navigation data as part of their L5 signal. CNAV not only contains parameters for L5, but also for the legacy L1 C/A signals. However, in order to ensure backward compatibility with existing equipment, GBAS L1 corrections still have to use LNAV even when the L5 corrections use CNAV. This is why an airborne user needs to receive and process both navigation data sets. For Galileo, the situation is similar.

Galileo I/NAV navigation data is transmitted via the E1 signal, while F/NAV is transmitted via E5a. In order to allow single-frequency processing in both bands, the ground facility also uses I/NAV for calculating the L1 differential corrections and F/NAV for calculating the L5 corrections. An airborne user thus also has to receive and process both data sets for an ionosphere-free GBAS service.

During the tests in May 2016, 12 GPS satellites and 9 Galileo satellites broadcast L1 and L5 signals as well as valid navigation messages. As not all satellites are in view continuously, the number of usable L5 satellites varied significantly over the duration of the flight trials. This is shown in Fig. 4 for the duration of the overall test campaign. Three flights have been conducted by the research aircraft D-IBUF within this campaign; these periods of time are marked in orange.

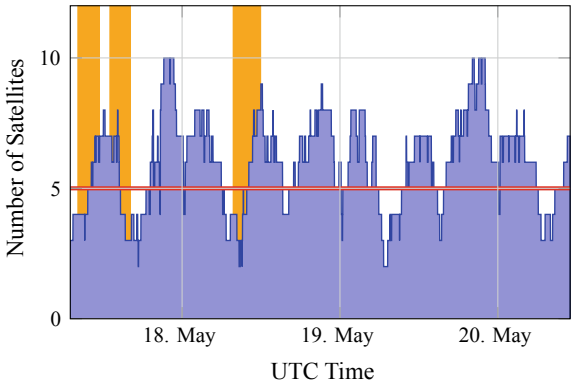


Fig. 4. Number of usable L5 GPS and Galileo satellites above an elevation of 5° for the test campaign in Toulouse (blue). Orange markings: flights of the research aircraft D-IBUF. Red line: Minimum of 5 satellites for a combined GPS/Galileo solution

In order to assess the behavior and initial performance of GAST-F I/Free processing, at least five L5 satellites have to be visible with an elevation of more than 5° over the horizon for a combined GPS/Galileo solution. Thus, a period of approximately 1:45 h was selected during the second half of the third flight for this analysis. During this period of time, the research aircraft conducted 6 approaches to runway 32R of Toulouse airport. The trajectory of this portion of the flights is shown in Fig. 5.

Even though usable VDB data has been broadcast in real time and recorded on board of the research aircraft, VDB data was recomputed due to some configuration optimizations and bug fixes in the ground facility software developed after the trials. The ground facility simulation is based on the recorded data of the two GBAS reference receivers (Septentrio PolaRx MC/MF GNSS receivers). On the airborne side, the GNSS data recorded by a Javad Delta MC/MF GNSS receiver with a sampling rate of 10 Hz is used. The skyplot shown in Fig. 6 visualizes the usable L5 GNSS satellites during the selected period of time.

Fig. 5. Analyzed GAST-F approaches into Toulouse airport (LFBO, runway 32R)

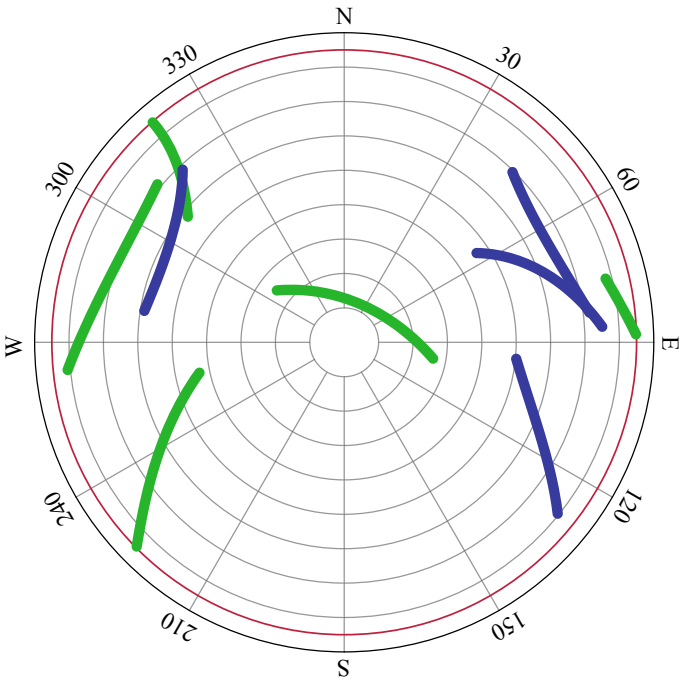


Fig. 6. Skyplot of usable L5 satellites for the analyzed period of time. Green: GPS satellites. Blue: Galileo satellites

It is obvious that the number and combinations of usable satellites is far away from being optimal at the moment so that the performance is not expected to be on par with an L1 solution. With at least five satellites in view, a dual-constellation GAST-F solution could be calculated nevertheless.

In agreement with the VDB format proposal developed within SESAR 15.3.7, the L5 corrections are transmitted in a new VDB message (message type 42), based on the L5 navigation data (i.e., GPS CNAV, Galileo F/NAV). On board of the aircraft, the differential L1 and L5 corrections are combined into an I/Free L1/L5 correction which is in turn applied to the I/Free L1/L5 combination of airborne measurements. This kind of processing (which is only one of the two proposed processing schemes) eliminates most of the ionospheric effects and thus mitigates ionospheric threats very effectively, but suffers from an increased noise level. The other proposed processing scheme uses differentially corrected pseudorange on a single frequency for positioning only, but incorporates dual-frequency ionospheric monitoring. This processing is not analyzed in this paper.

For the assessment of accuracy, a reference trajectory for the whole flight was calculated using Precise Point Positioning (PPP). The calculated used position is then compared to the reference position in order to calculate the lateral and vertical position errors.

The results of the GAST-F I/Free processing are shown in Fig. 7. This plot shows the position errors (PE) and the corresponding GBAS protection levels (PL) for the lateral and vertical case.

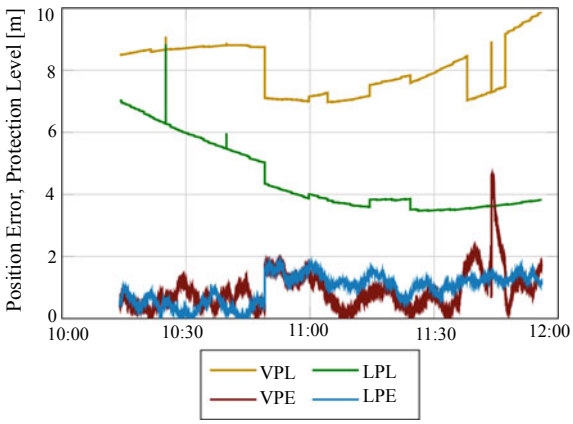


Fig. 7. Position errors and protection levels over time for GAST-F (I/Free)

As the (lateral/vertical) position errors are bound by the respective protection levels continuously, the requirements on integrity could be fulfilled for the whole analyzed period of time. In addition, the respective alert limits (not shown in Fig. 7) are above the protection levels for all times, resulting in 100% availability of GAST-F I/Free here.

In order to compare these results with legacy GBAS CAT-I, a GPS-only GAST-C solution has been processed in parallel, using the same input data than for the GAST-F solution. Of course, more GPS satellites are usable for this L1 only service. The position errors and protection levels for GAST-C (shown in Fig. 8) are significantly

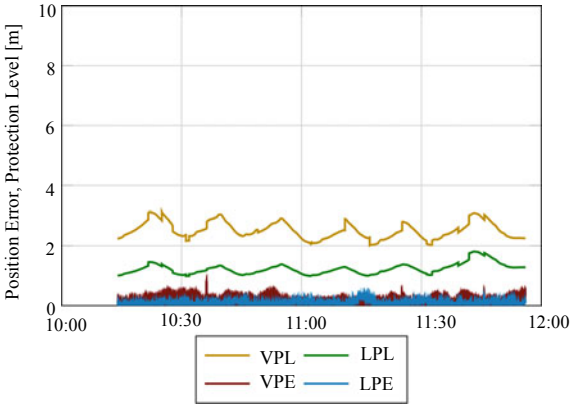


Fig. 8. Position errors and protection levels over time for GAST-C (GPS only)

lower than the respective values for GAST-F I/Free (shown in Fig. 7). The position error of GAST-F is larger as each ionosphere-free combination suffers from the amplification of noise. In addition, the protection levels are larger due to the lower number of satellites and different sigma values used.

However, despite these disadvantages and limitation, these results demonstrate that an I/Free GAST-F service is doable already today and has the potential to meet all requirements for precision approaches. Due to the ionosphere-free processing, all ionospheric effects are mitigated so that the ionosphere no longer poses a threat for the overall integrity. This way, even ionospheric gradients exceeding the current GBAS design threat space would not endanger a future I/Free GAST-F service. In the future, with more L5-enabled GNSS satellites, the overall performance will improve significantly.

However, some challenges still need to be addressed for a reliable GAST-F service which would comprise MC as well as MF techniques. At the moment, GBAS is limited to the legacy L1 signals of GPS and Glonass. In order to ensure compatibility, these signals need to be supported in the future as well. All new services thus have to be designed as an addition to the existing components. This might lead to challenges regarding the used navigation data, inter-system or interfrequency offsets for example.

This paper demonstrates the general usability of GAST-F I/Free processing using real data for the first time. However, as the whole processing algorithms are experimental, this work should be considered as a proof of concept only. Much more work has to be conducted prior to an MC/MF GBAS operational implementation.

4 Conclusions

The measurement campaigns presented in this paper were conducted in 2016 in order to pave way for a future MC/MF GBAS service. For this, the interoperability of existing legacy equipment with the proposed VDB formatting as well as the general feasibility of an ionosphere-free MC/MF GBAS service has been demonstrated.

For the assessment of interoperability, various GBAS hard- and software was involved in these trials. All legacy airborne equipment worked flawlessly with the proposed VDB message scheme, demonstrating the compatibility both to GAST-C and GAST-D. In addition, tests together with Japan's Electronic Navigation Research Institute (ENRI) showed that their independent GAST-D developments are interoperable with other developments. This shows that the draft ICAO SARPs [7] are mature and stable enough to allow independent developments for GAST-D.

For the demonstration of an experimental MC/MF GBAS service, an ionosphere-free GAST-F solution was calculated and tested successfully. The number of satellites usable for this service was limited significantly during the trials, but a valid ionosphere-free GAST-F MC/MF solution could be calculated throughout the analyzed approaches nevertheless. Even though some challenges will have to be addressed for a possible future operational introduction, the results show that all performance requirements could be met. In the future, with more L5 capable satellite in orbit, the performance is expected to improve further. This, of course, has to be assessed carefully prior to any standardization effort.

These results have been achieved by joint research of a multi-national research from Germany (TU Braunschweig), France (DSNA/DTI), and Japan (ENRI) with the support of EUROCONTROL.

Acknowledgement. Parts of the work presented in this paper have received funding by EUROCONTROL in the frame of SESAR 15.3.7. The trials in Toulouse have been supported by the *Direction des Services de la Navigation Aérienne* (DSNA). The trials in Ishigaki were technically supported by the *Japan Civil Aviation Bureau* (JCAB). The authors would like to thank all persons involved in the preparation and conduction of the tests for their help and support.

References

1. Stanisak M, Lipp A, Feuerle T (2015) Possible VDB formatting for multi-constellation/multi-frequency GBAS services. In: Proceedings of the 28th international technical meeting of the satellite division of the institute of navigation (ION GNSS+ 2015). Tampa, FL, pp 1507–1518
2. Feuerle T, Stanisak M, Hecker P (2016) GBAS flight trials for multi-constellation/multi-frequency GBAS concept validation. In: Proceedings of the 2016 international technical meeting of the institute of navigation. Monterey, CA, pp 380–388

3. RTCA SC-159. DO-246D: (2008) GNSS-based precision approach local area augmentation system (LAAS) signal-In-space interface control document (ICD). Washington, DC
4. RTCA SC-159. DO-253C: (2008) Minimum operational performance standards for GPS local area augmentation system airborne equipment. Washington, DC
5. Feuerle T, Stanisak M, Saito S, Yoshihara T, Lipp A (2016) GBAS interoperability trials and multi-constellation/multi-frequency ground mockup evaluation. In: Proceedings of the 6th SESAR innovation days.
6. Saito S, Stanisak M, Yoshihara T, Feuerle T, Lipp A (2017) Interoperability of ENRI GAST-D prototype with different airborne software implementations. In: Proceedings of the 5th ENRI international workshop on ATM/CNS (EIWAC2017)
7. ICAO. (2016) Annex 10: standards and recommended practices, radio navigation aids. vol I, Draft version

ADS-B Latency Estimation Technique for Surveillance Performance Assessment

Yasuyuki Kakubari(✉), Yoshio Kosuge, and Tadashi Koga

Electronic Navigation Research Institute (ENRI), Tokyo, Japan
{kakubari,kosuge,koga}@mpat.go.jp

Abstract. ADS-B is one of the key available technologies for aircraft surveillance. ADS-B provides high accurate GNSS positions with high update rate for air traffic control and airline operations. This feature will improve the safety and efficiency of air traffic operations. However, ADS-B user cannot benefit from the accuracy capability because the ADS-B message does not include information of GNSS measured time and because of the latency between the GNSS measured time and ADS-B message arrival time. In this paper, we propose a technique to estimate the variable element of ADS-B latency. Then, the proposed technique is applied to real data obtained by ENRI experimental system. The results of the estimations are presented in support of performance assessment.

Keywords: ADS-B · Latency · Surveillance performance

1 Introduction

Automatic Dependent Surveillance-Broadcast (ADS-B) is one of the key available technologies for aircraft surveillance. Aircraft positions with high update rate and GNSS accuracy will improve the safety and efficiency of air traffic operations. With ADS-B, aircraft broadcasts its position with meter-order accuracy by GNSS positioning. However, ADS-B user can not benefit from the accurate capability because the ADS-B position report does not contain GNSS measured time. As a consequence, an ADS-B ground station can obtain only ADS-B report arrival time measured by own clock.

The total latency of position information is the delay between the time of applicability of the position measurement and the time of arrival of the ADS-B message for that position. It is one of the important factors for the performance of ADS-B systems, because the total latency has a direct influence on its positions. Large latency becomes a cause of serious position error.

In [1, 2], the total latency has been estimated by summing up the values regulated in international standards or regulations of ADS-B systems. However, it is difficult to measure and evaluate the actual value of the total latency.

In this paper, we propose a technique to estimate the variable element of the ADS-B total latency. Then, we apply the proposed technique to real data obtained by ENRI experimental system and show results of the estimations that indicate the effectiveness of the proposed technique in support of surveillance performance assessment.

Electronic Navigation Research Institute, *Air Traffic Management and Systems III*, Lecture Notes in Electrical Engineering 555,
https://doi.org/10.1007/978-981-13-7086-1_12

The structure of this paper is as follows. In Sect. 2, we discuss the latency of ADS-B. In Sect. 3, we propose a technique to estimate the standard deviation of the total latency. In Sect. 4, we apply the proposed technique to real ADS-B data observed in ENRI experimental system. In Sect. 5, we conclude this paper with some results.

Extended squitter ADS-B has two different timing modes, one is coupled to GNSS timing and another is uncoupled with it. In this paper, we only discuss uncoupled mode because most of the aircraft are using uncoupled mode.

2 ADS-B Latency

The latency of ADS-B is discussed in [1, 2]. The functional architecture of a transmitting aircraft and a ground ADS-B receiver is shown in Fig. 1. This figure is a combination diagram of the functional architecture and general timing.

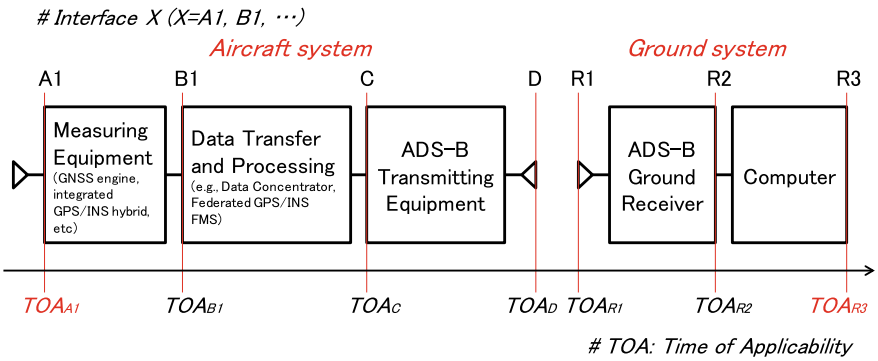


Fig. 1. Functional architecture diagram with the general timing

In Fig. 1, each block shows components of ADS-B systems. The capital letters at the top show the interfaces between components. For example, A1 indicates Input to the Measuring equipment. The considerations in [2, 3] do not include the delay at the ground systems. We add two blocks (ADS-B receiver and computer) to the functional architecture and define their interfaces as R1, R2, and R3. Even if the latencies in the ground systems are negligible, it is important to understand the relation between onboard measured time and ground measured time.

TOA_X is the true time of applicability of the data that crosses interface X. The detailed relations between time and delay in subsystems and in total system is discussed in [1–3].

The total latency (TL) of position information is the delay between the time of applicability of the position measurement and the time of arrival of the decoded report on that position.

TL is shown as follows

$$TL = TOA_{R3} - TOA_{A1}. \quad (1)$$

3 Latency Estimation Technique

In this section, we propose a technique to estimate the variable element of ADS-B latency. This technique can estimate the standard deviation of the total latency. We focus on the dynamic part of total latency which leads to noisy position errors.

3.1 Requirements and Assumptions

The requirements and assumptions for applying this technique are the following:

- The onboard GNSS receiver measures accurate positions and time.
- The velocities (ground speed) of aircraft are computed from the positions and time.
- The clock at a ground receiver keeps accurate time.

If the ADS-B systems do not satisfy these requirements, it is not possible to estimate the deviation of the total latency of ADS-B.

3.2 Consideration of the Total Latency

To simplify the discussion, we define two types of the latency, the constant latency (*CoL*) and the inconstant latency (*InL*) (in other words, variable latency) in ADS-B system.

CoL is constant and specific to each ADS-B equipment. It is a fixed value and does not change by each event. The main cause of the *CoL* is processing and transmission delays in the system.

InL is variable and changes in every event. The main cause of the *InL* is the transmitting jitter at the transmitting phase. Mode S transponders intentionally shift signal transmission time [4].

The relation among *TL*, *CoL,* and *InL* at event i is shown as follows:

$$TL_i = CoL + InL_i (i = 1, 2, \ldots, n). \quad (2)$$

From this definition, *CoL* means mathematically offset element of TL_i. InL_i can be both plus and minus value and its average is zero.

3.3 Computation

We consider that an ADS-B onboard system transmits the ADS-B position reports at the time event 1 and 2. ADS-B ground equipment receives the reports containing GNSS positions at each time of $TOA_{A1\_i} (i = 1, 2)$. These reception times for the event 1 and 2 are shown as follows:

$$TOA_{R3\_1} = TOA_{A1\_1} + CoL + InL_1, \tag{3}$$

$$TOA_{R3\_2} = TOA_{A1\_2} + CoL + InL_2. \tag{4}$$

The ground speed (v_{mes}) that the ground receiver computes from ADS-B position reports and ground clock measured time is shown as follows:

$$v_{mes\_12} = \frac{R_{12}}{TOA_{R3\_2} - TOA_{R3\_1}}, \tag{5}$$

where R_{12} is the distance between ADS-B positions at the event 1 and 2.

The ground speed (v_{true}) that the GNSS receiver computes from GNSS position and GNSS clock measured time is shown as follows:

$$v_{true\_12} = \frac{R_{12}}{TOA_{A1\_2} - TOA_{A1\_1}}. \tag{6}$$

The time difference of TOA_{R3_1} and TOA_{R3_2} is shown as follows:

$$TOA_{R3\_2} - TOA_{R3\_1} = \frac{R_{12}}{v_{mes\_12}}. \tag{7}$$

By the Eqs. (3), (4), the Eq. (7) is modified as follows:

$$(TOA_{A1\_2} + CoL + InL_2) - (TOA_{A1\_1} + CoL + InL_1) = \frac{R_{12}}{v_{mes\_12}}. \tag{8}$$

Then, the constant latency is canceled, and the difference of the inconstant latency remains in the equation.

$$(TOA_{A1\_2} - TOA_{A1\_1}) + (InL_2 - InL_1) = \frac{R_{12}}{v_{mes\_12}}. \tag{9}$$

By the Eq. (6), the Eq. (9) is modified as follows:

$$\frac{R_{12}}{v_{true\_12}} + (InL_2 - InL_1) = \frac{R_{12}}{v_{mes\_12}}. \tag{10}$$

The difference of the inconstant latencies (ΔInL_{12}) is shown as

$$\Delta InL_{12} = InL_2 - InL_1 = \frac{R_{12}}{v_{mes\_12}} - \frac{R_{12}}{v_{true\_12}}. \tag{11}$$

In the same way, the difference of the inconstant latencies at the event i and $i+1$ is shown as follows:

$$\Delta InL_{i,i+1} = (InL_{i+1} - InL_i) = \frac{R_{i,i+1}}{v_{mes\_i,i+1}} - \frac{R_{i,i+1}}{v_{true\_i,i+1}}. \tag{12}$$

The mean of inconstant latency (*InL*) is zero, because *InL* is randomly changing value due to the definition, and the *InL* at each event has no correlation. The mean and standard deviation of $\Delta InL_{i,i+1}$ is computed as follows:

$$Mean = \frac{1}{n}\sum_{i=1}^{n} \Delta InL_{i,i+1} = \frac{1}{n}\sum_{i=1}^{n} (InL_{i+1} - InL_i) = 0. \tag{13}$$

$$\begin{aligned} Std &= \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left\{\Delta InL_{i,i+1} - Mean\right\}^2} \\ &= \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left\{\Delta InL_{i,i+1}\right\}^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left\{InL_{i+1} - InL_i\right\}^2} \\ &= \sqrt{\frac{1}{n}\sum_{i=1}^{n} \left\{(InL_{i+1})^2 - 2 \cdot InL_{i+1} \cdot InL_i + (InL_i)^2\right\}} \\ &= \sqrt{\frac{1}{n}\sum_{i=1}^{n} (InL_{i+1})^2 - \frac{1}{n}\sum_{i=1}^{n} (2 \cdot InL_{i+1} \cdot InL_i) + \frac{1}{n}\sum_{i=1}^{n} (InL_i)^2} \end{aligned} \tag{14}$$

Here, if *InL* behaves as random noise and *n* (observation count) is large enough, the second member in the last square root of the Eq. (14) is negligible as follows:

$$\frac{1}{n}\sum_{i=1}^{n} (2 \cdot InL_{i+1} \cdot InL_i) \approx 0. \tag{15}$$

By the Eq. (15), the standard deviation of $\Delta InL_{i,i+1}$ is approximated as follows:

$$\begin{aligned} Std &\approx \sqrt{\frac{1}{n}\sum_{i=1}^{n} (InL_{i+1})^2 + \frac{1}{n}\sum_{i=1}^{n} (InL_i)^2} \\ &\approx \sqrt{\frac{2}{n}\sum_{i=1}^{n} (InL_{i+1})^2}. \end{aligned} \tag{16}$$

This equation means that the standard deviation of the inconstant latency (*InL*) is $1/\sqrt{2}$ of the standard deviation of $\Delta InL_{i,i+1}$ that we can obtain from observations. This value also means the standard deviation of the total latency of ADS-B.

4 Experiments

4.1 Experimental ADS-B Ground System

In experiments, we use the OCTPASS system which is developed as airport surface multilateration [5]. The system involves an ADS-B signal processor based on DO-260A. By bench tests, we confirmed that the system has a highly accurate time stamp capability. The timestamp error is less than 1ms. This is sufficiently small compared to a hundred milliseconds latency at airborne systems [2, 3].

4.2 Ground Speeds

Three types of ground speed (GS) is available from obtained data.

The first GS is "ADS-B GS" which contained in ADS-B velocity report. This ADS-B GS is corresponding to v_{true} in Eq. (6).

The second GS is "measured GS" that computed from positions in consecutive ADS-B position reports and time at which the ground equipment receives the ADS-B position reports. This measured GS is corresponding to v_{mes} in Eq. (5).

The third GS is "smoothed GS" that computed from positions in ADS-B position reports that separated for certain amount of time (ex. 20s) and the time that the ground equipment receives the corresponding ADS-B position reports.

Figure 2 shows that three ground speed of an aircraft measured trajectory. The blue, magenta and cyan lines indicate the measured GS, the ADS-B GS, and the smoothed GS, respectively.

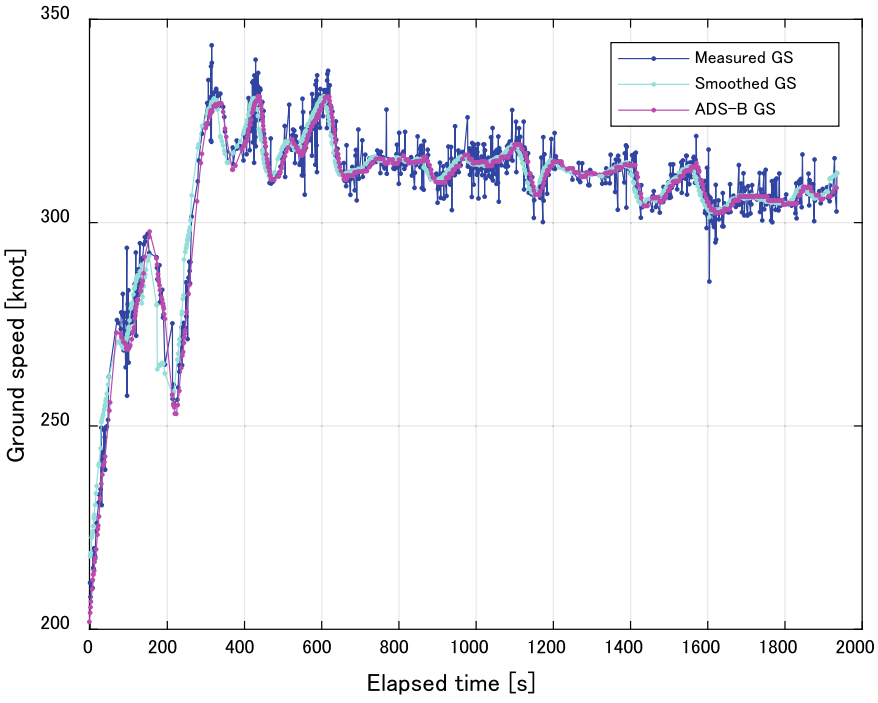


Fig. 2. The GS of an aircraft

The smoothed GS and ADS-B GS vary smoothly and both values are matching very well. On the other hand, measured GS is noisy and contains errors. In this experiment, we adopt the smoothed GS as v_{true} for latency computation.

4.3 Estimated Latency

Then, we compute the difference of the inconstant latency $(\Delta InL_{i,i+1})$ from ADS-B position report and time stamp on the ground.

Figure 3 shows computed $\Delta InL_{i,i+1}$ from the measured aircraft trajectory corresponding to the GS of Fig. 2. In this aircraft track, although short fluctuation is observed at the beginning of the climb phase, the mean of the $\Delta InL_{i,i+1}$ is 0.76 ms, and the standard deviation of $\Delta InL_{i,i+1}$ is 37.4 ms. Then, the estimated standard deviation of InL is 26.5 ms $(= 37.4/\sqrt{2})$. If aircraft is flying at the speed of 300 knots, 26.5 ms deviation is equal to 4.1 m position error. In this case, the estimated position error of this aircraft is small.

Figure 4 shows $\Delta InL_{i,i+1}$ from other aircraft trajectories. In this aircraft track, large values of $\Delta InL_{i,i+1}$ are frequently observed. The standard deviation of $\Delta InL_{i,i+1}$ is 226 ms and the estimated standard deviation of InL is 160 ms. Therefore, quite a large position error of over 24.7 m is contained in ADS-B data. In this case, although the Eqs. (13), (16) are not established, such large error data indicate some irregular situation.

We also compute the standard deviations in the same way for different 138 aircraft as shown in Fig. 5. The x-axis shows the number of aircraft, and the y-axis shows estimated standard deviations of the inconstant latency.

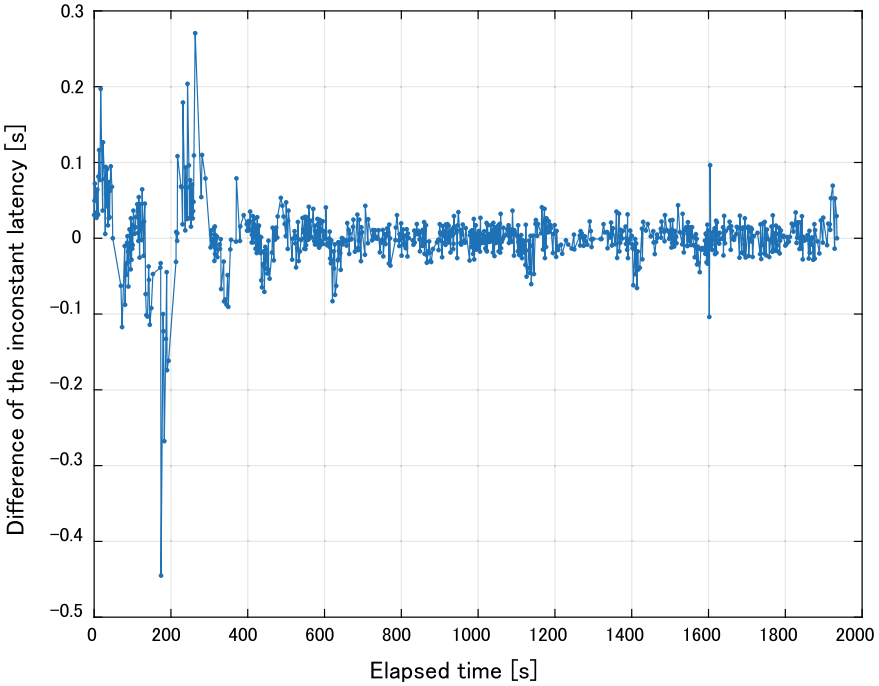


Fig. 3. The difference of the inconstant latency $\Delta InL_{i,i+1}(1)$

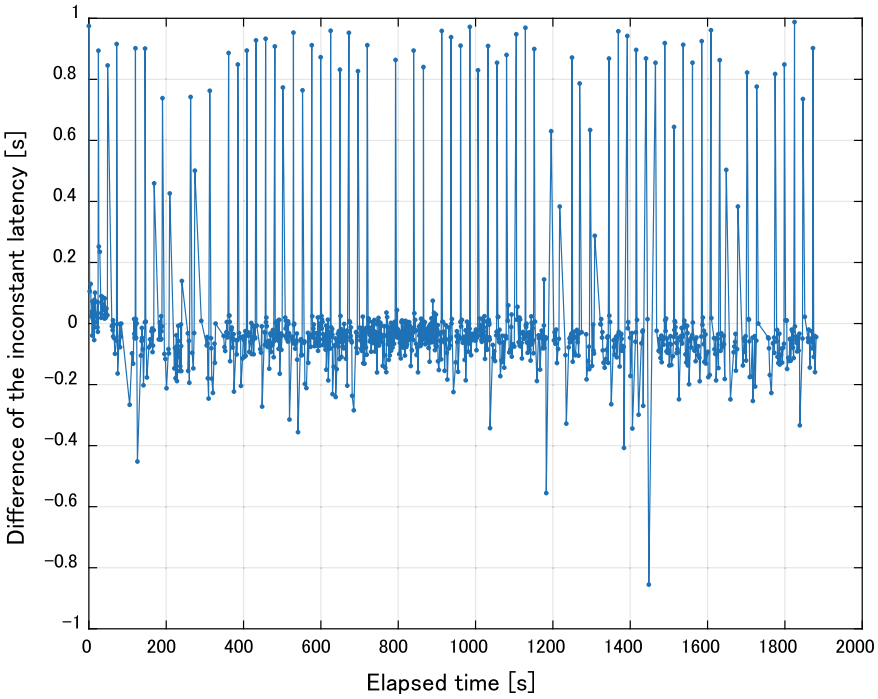


Fig. 4. The difference of the inconstant latency $\Delta InL_{i,i+1}(2)$

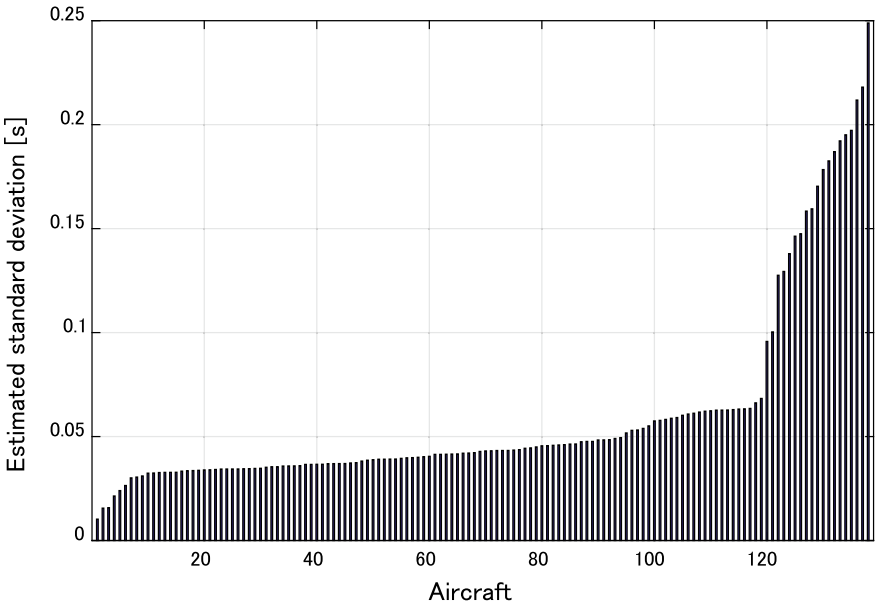


Fig. 5. The estimated standard deviation of 138 aircraft

120 out of 138 aircraft were within 0.1 second that is within 15.4 m longitudinal positioning error at 300 knots flight speed. These small deviations are mainly caused by the transmitting jitter at the transmitting phase, which is a regular operation for transponders.

12 aircraft were more than 0.15s, that is also equal to more than 23.2 m positioning errors. These large deviations may be caused by the irregular operations in onboard ADS-B systems or inaccuracy of broadcasted GNSS positions.

5 Conclusion

In this paper, we propose a technique to estimate the variable element of ADS-B latency and apply this technique to real data obtained by ENRI experimental system.

We have two conclusions. First, thanks to the proposed technique, we found the inconstant latency of the majority of aircraft is within 0.1 second. The standard deviation of ADS-B positioning random error is within 15.4m at the speed of 300 knots. These results could be useful to support performance assessment. Second, we found that some aircraft are transmitting ADS-B positions with a large error. At the present time, we speculate that this error is due to the irregular operations of onboard systems or the GNSS position inaccuracy.

We have two remaining works. First, we are going to investigate the detail of the second result. We are going to apply the proposed technique to a large amount of real data. Then, we are going to investigate whether some relations exist between "larger error" and "transponder types (or aircraft types)". Aircraft with a large error should be eliminated from ADS-B surveillance and should be fixed in the future.

The second one is to estimate the constant latency and then to estimate the total latency. For example, the constant latency may be calculated by comparing broadcasted GNSS positions and multilateration positions corresponding to the ADS-B signal.

The analysis of the ADS-B latency will contribute to the improvement of safety and efficiency of air traffic operations.

References

1. RTCA SC-186 (2009) Minimum Operational Performance Standards (MOPS) for 1090 MHz Extended Squitter Automatic Dependent Surveillance-Broadcast (ADS-B) and Traffic Information Services-Broadcast (TIS-B). RTCA/DO-260B, Washington DC, December 2, 2009. Appendix U, "Total and Uncompensated Latency 1090MHz Extended squitter ADS-B"
2. Thedford WA (2011) ADS-B IN/OUT Technical issues. In: CNS/ATM Conference, Florida
3. Levitt I et al (2012) Modeling ADS-B out system latency. In: IEEE/AIAA 31st digital avionics systems conference (DASC), Virginia
4. International Civil Aviation Organization (2007) Surveillance and Collision Avoidance Systems. Annex 10 to the Convention on International Civil Aviation, Vol 4, 4th edn, Montreal, July, 2007
5. Kakubari Y et al (2014) Enhancement of Passive Surveillance System for Airport Surface Movement. Electron Commun Jpn 97(3):24–30

Jamming and Spoofing Protection for ADS-B Mode S Receiver Through Array Signal Processing

Junichi Naganawa[1(✉)], Camille Chomel[2], Tadashi Koga[1], Hiromi Miyazaki[1], and Yasuyuki Kakubari[1]

[1] Surveillance and Communications Department, Electronic Navigation Research Institute (ENRI), Tokyo, Japan
{naganawa, koga, hiro, kakubari}@mpat.go.jp
[2] French Civil Aviation School (ENAC), Toulouse, France
camille.chomel@recherche.enac.fr

Abstract. Automatic-dependent surveillance-broadcast (ADS-B) receivers are vulnerable to jamming and spoofing attacks. Air traffic management is expected to heavily rely on ADS-B in the future so it is important to provide protection against these attacks. One promising solution is array signal processing. Jamming signals can be suppressed by beamforming and ADS-B positions can be verified by comparing them to the measured angles of arrival (AoAs). Many studies have been done on this topic but they generally consider either jamming or spoofing. However, in a real environment, anti-jamming and anti-spoofing functions must be provided together. It is important to understand the performance when these two functions are implemented simultaneously in order to guide future research and development strategies and implementation plans. Thus, we conducted an experimental evaluation of anti-jamming and anti-spoofing functions. For jamming protection, a power-inversion beamformer was used to suppress the jamming signal. For spoofing protection, a Capon's beamformer was used to measure the AoA. The array used was a uniform linear array of four elements. The experiments were conducted in an anechoic chamber using a four-channel oscilloscope and the signal processing was conducted offline. The results of the anti-jamming experiment showed that the jammed signals were successfully recovered and the signal-to-jamming-plus-noise (SJN) ratio was increased to above the decodable level. The effects of various parameters on the overall performance and the lower bound performance were also investigated. Moreover, the anti-spoofing experiment showed that the AoA of the ADS-B signal could be measured successfully. The AoA accuracy was evaluated in terms of the standard deviation, which was further converted into the probability of spoofing detection. These findings will be important in the implementation of effective infrastructure to ensure safe air travel.

Keywords: Aeronautical surveillance · Automatic-dependent surveillance-broadcast (ADS-B) · Jamming · Spoofing · Array signal processing

Electronic Navigation Research Institute, *Air Traffic Management and Systems III*, Lecture Notes in Electrical Engineering 555,
https://doi.org/10.1007/978-981-13-7086-1_13

1 Introduction

Aeronautical surveillance systems provide aircraft positions to the air traffic controller and, thus, represent an essential infrastructure for safe air travel. Secondary surveillance radar (SSR) is a typical aeronautical surveillance system. Although the current systems have successfully supported the worldwide air traffic control (ATC) services for many years, the ever-growing demand for air transportation is intensifying the need for high-performance surveillance systems to support the increasing ATC capacity.

As a new means of aeronautical surveillance, the deployment of automatic-dependent surveillance-broadcast (ADS-B) has been either planned or initiated worldwide. In the ADS-B regime, aircraft periodically measure their own positions and broadcast them to the ground stations, thereby enabling accurate and fast-update-rate surveillance. However, ADS-B relies on the reports from aircraft, which introduces to significant security risks: jamming and spoofing [1–9]. Jamming is the emission of radio-frequency (RF) signals that compromise a receiver's operational capabilities. Spoofing is the emission of false ADS-B messages to introduce false "ghost" targets in the system. Because air traffic management will likely rely heavily on ADS-B in the future, it is important to implement protections against both of these types of attacks.

One promising solution is array signal processing [10–12] in which an antenna array is used to receive a set of signals having phase and amplitude variations. By summing these signals, the directional pattern of the array can be controlled dynamically to achieve a variety of functions. For protection against jamming, the antenna pattern can be controlled such that null signals are created in the directions of the jamming signals. Moreover, for protection against spoofing, the angle of arrival (AoA) can be measured to verify the reported aircraft position.

So far, numerous anti-jamming or anti-spoofing techniques based on array signal processing have been proposed [13–19]. However, these approaches were only designed to address either jamming or spoofing. In real environments, however, anti-jamming and anti-spoofing functions must be provided at the same time. Thus, it is important to study the performance when these two functions are provided simultaneously to gain information to guide future research and development strategies and implementation plans.

Therefore, we conducted an experimental evaluation of simultaneous anti-jamming and anti-spoofing functions with a uniform linear array (ULA) of four elements. The anti-jamming function was provided by using a power-inversion (PI) beamformer to suppress the jamming signal. Spoofing protection was provided by using a Capon's beamformer (CB) to measure the AoA and verifying the ADS-B position based on this AoA. The experiment was conducted in an anechoic chamber. The measurement system down-converted the RF signals to an intermediate frequency (IF) and recorded them using a four-channel digital sampling oscilloscope. The recorded signals were analyzed offline.

This paper is organized as follows. In Sect. 2, the anti-jamming and anti-spoofing functions provided by array signal processing are explained. In Sect. 3, the measurement system is described. In Sects. 4 and 5, the anti-jamming and anti-spoofing experiments, respectively, are discussed. In Sect. 6, this paper is concluded.

2 Array Signal Processing

2.1 Assumptions on Attacks and Signals

Jamming and spoofing can be carried out by various techniques. For example, the jamming signal can be incorporated into various waveform such as continuous waves, pulses, Mode S pulses, etc. A spoofed trajectory can be either random, artificial, or a replay. Therefore, it is necessary to define what kinds of attacks are to be prevented. In this paper, we focus on the attacks that are easiest to conduct but can incur significant damage to operations. Specifically, the following assumptions are made.

1. The jamming signal is a high-power continuous wave.
2. The spoofed signals report random positions.

Furthermore, it is assumed that the jamming signal, ADS-B signals, and noise are uncorrelated.

2.2 Overview of the Signal Processing

This study is the first time that simultaneous jamming and spoofing protection has been evaluated. Therefore, basic techniques for anti-jamming and anti-spoofing functions were selected separately and combined. Figure 1 shows an overview of the proposed signal processing. First, signals received by the antenna array are processed by the anti-jamming function, yielding a jamming-free signal. Then, the jamming-free signal is decoded to yield the ADS-B position. Next, the anti-spoofing function is executed to obtain the actual AoA based on the raw signals as well as the expected AoA from the ADS-B position and compare them to verify the reported ADS-B position based on binary hypothesis testing as the decision criterion [19].

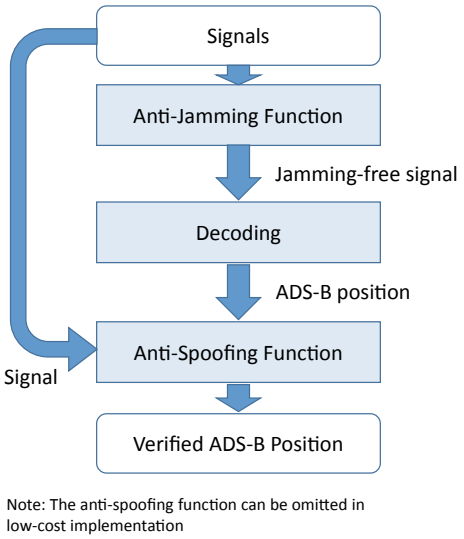


Fig. 1. Overview of the signal processing

2.3 Algorithm Selection

To provide the anti-jamming function, the signals are combined such that the jamming signals are suppressed while the ADS-B signals remain intact. There are five algorithms described in [12]: the MMSE (minimum mean square error) algorithm, the MSN (maximum signal-to-noise ratio) algorithm, DCMP (directional constrained minimization of power), PI, and the CMA (constant modulus algorithm). Among them, the PI and CMA do not require prior information such as a replica or the direction of arrival. However, the CMA is not applicable to this study because the ADS-B signal does not have a constant envelope. Therefore, the PI algorithm was selected for this study. PI is advantageous because of its simplicity; however, the adaptive beamforming function is not as powerful as CMA or MMSE and, thus, may ultimately limit the coverage range.

Six algorithms for estimating the AoA are listed in [12]: beamformer, CB, LP, Min-Norm, MUSIC (MUltiple SIgnal Classification), and ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques). These methods are categorized as either beam-steering or null-steering methods. Beam-steering methods are generally simpler to implement than null-steering methods. Therefore, we selected a beam-steering method, CB, because it provides a better AoA resolution than the beamformer method. Moreover, both CB and PI are based on the inversion of a spatial covariance matrix and use the same mathematical operation, which can reduce the implementation cost. However, CB provides a lower AoA resolution than null-steering methods and, ultimately, a lower chance of detecting a spoofing signal.

2.4 PI: Anti-jamming Signal Processing

The PI signal processing is described in this subsection. The number of array elements, sampling index, baseband signal vector, and weight vector are denoted as N, m, $\boldsymbol{x}(m) \in C^{N\times 1}$, and $\boldsymbol{w} \in C^{N\times 1}$, respectively. After applying the weight vector, the baseband signals becomes $y(m) \in C^{1\times 1}$:

$$y(m) = \boldsymbol{w}^{\mathrm{H}}\boldsymbol{x}(m). \tag{1}$$

Here, the weight vector is defined as

$$\boldsymbol{w} = \frac{\boldsymbol{R}^{-1}\boldsymbol{u}}{\boldsymbol{u}^{\mathrm{H}}\boldsymbol{R}^{-1}\boldsymbol{u}}, \tag{2}$$

where $\boldsymbol{R}$ is a spatial covariance matrix and $\boldsymbol{u}$ is a vector used to apply a fixed unitary gain to one of the elements (for example, when the first element is selected, $\boldsymbol{u} = [\,1 \quad 0 \quad \cdots \quad 0\,]^{\mathrm{T}}$.)

The standard approach to estimate $\boldsymbol{R}$ is based on the signal to be beamformed, $\boldsymbol{x}(m)$. However, the ADS-B signal can be suppressed if the jamming signal is not sufficiently stronger than the ADS-B signal or does not exist. Therefore, our proposed technique uses different windows: the estimation window has a length M' and is in front of the beamforming window, which has a length of M, with a transition period of ΔM

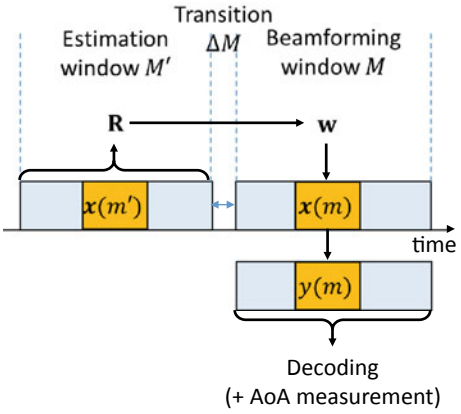


Fig. 2. Estimation and beamforming windows

(see Fig. 2). The estimation window is used for estimating $\boldsymbol{R}$ and obtaining the beamformer, $\boldsymbol{w}$, while the beamforming widow is used for applying $\boldsymbol{w}$ and recovering the ADS-B signal. Thus, $\boldsymbol{R}$ is estimated as

$$\boldsymbol{R} = \frac{1}{M'} \sum_{m'} \boldsymbol{x}(m') \boldsymbol{x}^{\mathrm{H}}(m'). \tag{3}$$

The mechanism of the proposed technique can be explained as follows. For simplicity, assume there is only one desired signal. Three cases are considered: a strong jamming case, a no jamming case, and a weak jamming case. When a strong jamming attack is carried out, the estimation window contains the jamming signal while the beamforming window contains both the jamming and ADS-B signals. The spatial covariance matrix in the estimation window is

$$\boldsymbol{R} = \boldsymbol{R}_{\mathrm{JJ}} + \mathrm{diag}\left(\sigma^2, \ldots, \sigma^2\right), \tag{4}$$

where $\boldsymbol{R}_{\mathrm{JJ}}$ is the autocorrelation matrix for the jamming signal, diag represents a diagonal matrix, and σ^2 is the variance due to noise. The resulting beamformer suppresses only the jamming signal when it is applied to the beamforming window.

When there is no jamming, the estimation window does not contain any signals and the beamforming window contains the ADS-B signal. For the noise-only estimation window, the spatial covariance matrix reduces to

$$\boldsymbol{R} = \mathrm{diag}\left(\sigma^2, \ldots, \sigma^2\right) \text{and} \tag{5}$$

$$\boldsymbol{R}^{-1} = \frac{1}{\sigma^2} \mathrm{diag}(1, \ldots, 1). \tag{6}$$

This yields the following weight vector:

$$\boldsymbol{w} = [\,1 \quad 0 \quad \cdots \quad 0\,]^{\mathrm{T}}, \tag{7}$$

which represents reception by only one array element. By applying this weight vector to the beamforming window, the ADS-B signal can be obtained.

Finally, when there is a weak jamming signal, the resulting $\boldsymbol{R}$ is between those for the other two cases. Thus, it can be concluded that the proposed technique functions appropriately.

2.5 CB: Anti-spoofing Signal Processing

CB generates an angular spectrum representing the signal power arriving from each direction as follows:

$$P(\phi) = \frac{1}{\left|\boldsymbol{a}^{\mathrm{H}}(\phi)\boldsymbol{R}^{-1}\boldsymbol{a}(\phi)\right|^2} \tag{8}$$

where $\boldsymbol{a}(\phi) \in C^{N\times 1}$ is called a mode vector and represents the directional characteristic of the array. For a ULA with half-wavelength spacing, $\boldsymbol{a}(\phi)$ is given by

$$\begin{aligned} \boldsymbol{a}(\phi) &= [\,a_1(\phi) \quad \cdots \quad a_N(\phi)\,]^{\mathrm{T}} \\ a_k(\phi) &= \exp(-j\pi(k-1)\sin\phi). \end{aligned} \tag{9}$$

The actual mode vector can be different due to manufacturing error and mutual coupling, and a measured phase pattern is used in this work. The spatial covariance matrix, $\boldsymbol{R}$, is estimated from the signals in the beamforming window as follows:

$$\boldsymbol{R} = \frac{1}{M}\sum_{m} \boldsymbol{x}(m)\boldsymbol{x}^{\mathrm{H}}(m). \tag{10}$$

Equations (2) and (7) show that both PI and CB are based on the inversion of the spatial covariance matrix, $\boldsymbol{R}$. The use of the same computational operation is advantageous because it can reduce the implementation cost. Finally, by identifying the peak of the angular spectrum, the AoA can be estimated as follows:

$$\hat{\phi} = \max_{\phi} P(\phi). \tag{11}$$

2.6 Window Size

The window sizes were determined as follows. First, M was selected such that the value of MT, where T is the sampling period, is 122 μs, which is equal to the ADS-B signal length (120 us) with a 2 μs (two-pulse) margin. Second, M' was selected to be equal to M to facilitate efficient implementation. The spatial covariance matrix, $\boldsymbol{R}$, and the

inverse, $\boldsymbol{R}^{-1}$, are then computed with both the estimation and beamforming windows. Therefore, using the same size for both windows allows a single computational operation to be used twice. Finally, a ΔM value was arbitrary selected such that $\Delta MT = 1$ us (one pulse) because larger ΔM values require more memory.

3 Measurement System

3.1 Hardware

The measurement system consists of an antenna array, a receiver, an ADS-B transmitter, and a jamming transmitter as shown in Fig. 3. The antenna array is a ULA with four elements separated by a half wavelength that was built of off-the-shelf A3 ADS-B antennas. The array element is omnidirectional in the horizontal plane. Figure 4 shows the measured phase pattern which is used to construct the mode vector.

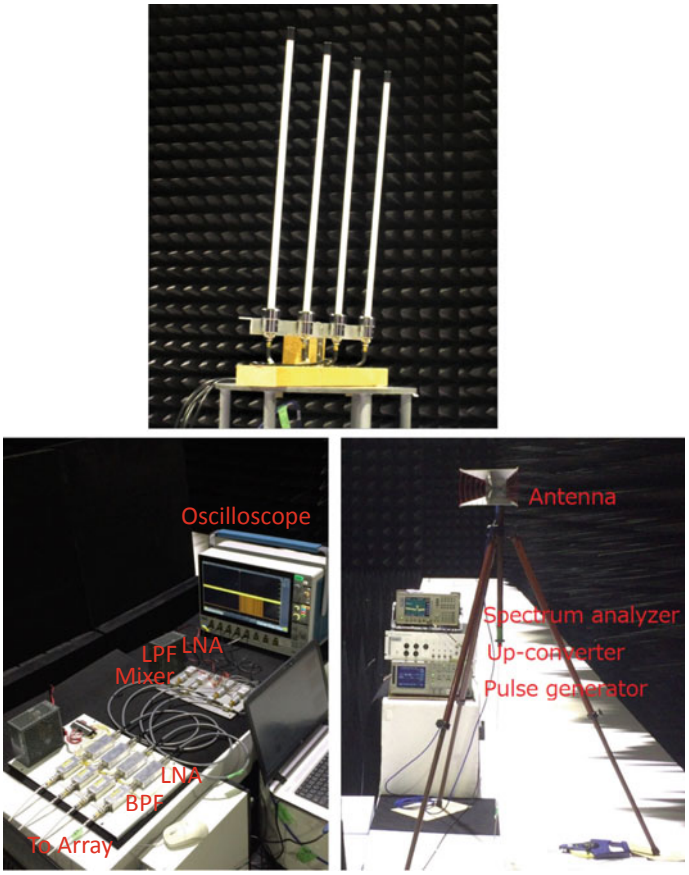


Fig. 3. Photographs of the array (top), receiver (left), and ADS-B transmitter (right)

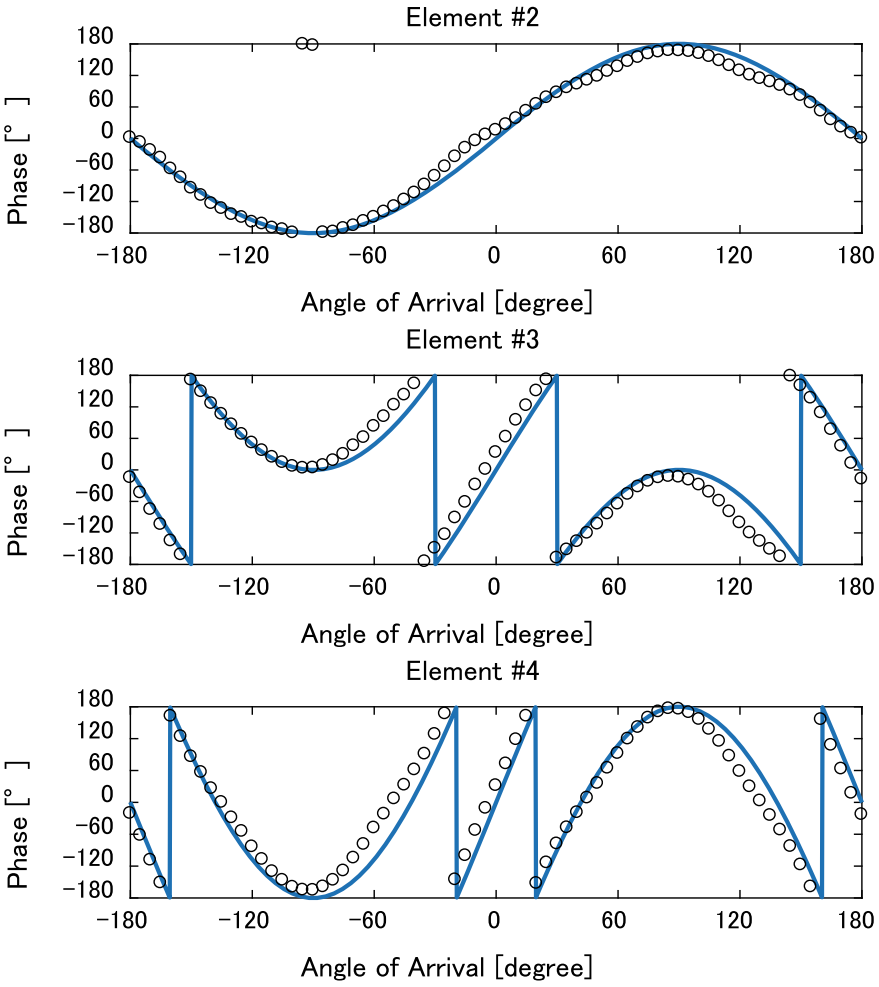


Fig. 4. Measured phase pattern for each element. Element #1 is omitted because the phase is always zero

Figure 5 shows a schematic of the receiver. The receiver consists of low-noise amplifiers (LNAs; R&K LA130 and Mini-Circuits ZKL 1R5+), band-pass filters (BPFs), mixers (Mini-Circuits ZX05-10L-S+), a local oscillator (LO; an Agilent E4438C signal generator), low-pass filters (LPFs; SPL 100+), and a multi-channel digital sampling oscilloscope (DSO; Tektronix MSO58). The radio-frequency (RF) signal received at each antenna element was down-converted to an intermediate-frequency (IF) signal of 10 MHz. A 10 MHz reference signal was generated by a rubidium oscillator (Stanford Research Systems FS 725). The sampling and triggering parameters are shown in Table 1.

The ADS-B transmitter consists of an antenna (an ETS-Lindgren Model 3115 double-ridged waveguide horn), a commercial pulse generator (Tektronix AWG410

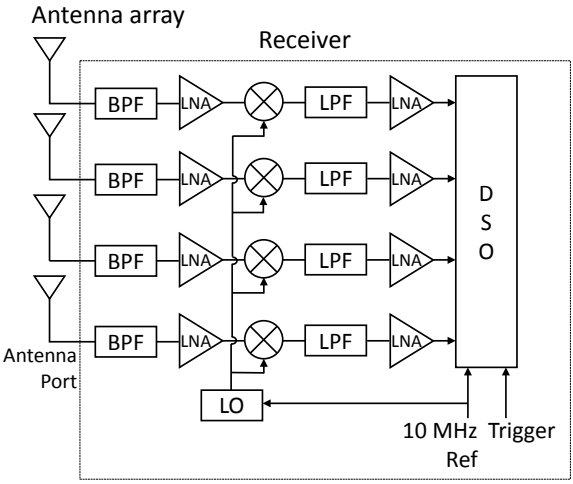


Fig. 5. Schematic of the receiver

Table 1. Parameters for sampling and triggering

Sampling speed	4 ns(250 MHz)
Sampling length	100 kSamples
Trigger	External (ADS-B)
Resolution	15 bit

Arbitrary Waveform Generator), a custom-built up-converter, and a spectrum analyzer for monitoring the transmission frequency. The pulse generator outputs a monitor signal that triggers the DSO in the anti-jamming experiment. The up-converter is equipped with an internal attenuator to control the power. The jammer transmitter consists of a signal generator (Agilent E4438C) and an antenna (the same model that was used for the ADS-B transmitter).

3.2 Data Processing

The recorded IF signals were further down-converted to baseband signals by a quadrature demodulation process involving multiplication with sine and cosine waves, low-pass filtering, and decimation. As a result, an in-phase component, $x_{\mathrm{I}}(m)$, and a quadrature component, $x_{\mathrm{Q}}(m)$, were obtained to form the baseband signal vector:

$$\boldsymbol{x}(m) = [\, x_1(m) \quad \cdots \quad x_N(m) \,], \tag{12}$$

where $x_n(m) = \mathrm{x}_{\mathrm{I},n}(m) + jx_{\mathrm{Q},n}(m)$ for every nth channel index. The sampling frequency of the baseband signals is 10 MHz, which corresponds to using 10 samples per symbol, which is consistent with the advanced decoding technique in DO-260 B [20]. Finally, $\boldsymbol{x}$ was processed by the anti-jamming and anti-spoofing functions as described in the previous section.

3.3 Received Power Calibration

To prepare for the experiments, the receiver is calibrated to output the received power at the antenna ports in dBm. To accomplish this function, the nth antenna is removed and the receiver was connected to an SG and a continuous wave with a frequency of 1090 MHz and a power of $P_{\mathrm{t,CW}}$ is transmitted. In this way, a set of coefficients, $C_{\mathrm{cal},n}$, can be obtained as follows:

$$C_{\mathrm{cal},n} = P_{\mathrm{t,CW}} - 10\log_{10}\left(\mathrm{E}\left[|x_n(m)|^2\right]\right). \tag{13}$$

These coefficients can be used to obtain the received powers, $P_{\mathrm{r},n}$, as follows:

$$P_{\mathrm{r},n} = 10\log_{10}\left(\mathrm{E}\left[|x_n(m)|^2\right]\right) + C_{\mathrm{cal},n}. \tag{14}$$

This process is repeated for all the ports.

3.4 Transmit Power Setting

The ADS-B transmitter power is controlled to simulate an imaginary aircraft-receiver distance of r. Therefore, the ADS-B transmitter power is set such that the received power is as follows:

$$P_{\mathrm{r}} = P_{\mathrm{t}} - L_{\mathrm{a}} + G_{\mathrm{t}} + G_{\mathrm{r}} - 20\log_{10}\left(\frac{r}{4\pi\lambda}\right), \tag{15}$$

where $P_{\mathrm{t}} = 54\,\mathrm{dBm}$, $L_{\mathrm{a}} = 3\,\mathrm{dB}$, $G_{\mathrm{t}} = 0\,\mathrm{dB}$, and λ are the nominal values of aircraft transmission power, aircraft cable loss, airborne antenna gain, and wavelength, respectively, and $G_{\mathrm{r}} = 4.94\,\mathrm{dB}$ is the average of the measured array element gains. The measured values of r were 50, 100, and 150 NM ($1\,\mathrm{NM} = 1.852\,\mathrm{km}$).

The jamming power was controlled to maintain a fixed jamming-to-signal ratio (JSR). The measured JSR values were −20, −10, 0, 10, 20, and 30 dB.

4 Anti-jamming Experiment

4.1 Setup

The anti-jamming experiment was conducted in an anechoic chamber. The jamming and ADS-B signals were simultaneously transmitted and measured. The received signal was processed offline to evaluate the anti-jamming performance. The evaluation was repeated with different array layouts, jammers, and ADS-B transmitters and the different transmission powers for the jamming and ADS-B signals. The signal-to-jamming-plus-noise (SJN) ratio after PI was calculated as a performance measure as follows:

$$\begin{aligned} P_{\mathrm{r,PI}} &= \mathrm{E}\left[|y(m)|^2\right] \text{for } m \in \text{pulse} \\ P_{\mathrm{jn,PI}} &= \mathrm{E}\left[|y(m')|^2\right] \\ \text{SJN Ratio} &= 10\log_{10} P_{\mathrm{r,PI}} - 10\log_{10} P_{\mathrm{jn,PI}} \end{aligned} \tag{16}$$

The signal recording was repeated for ten times for each configuration and the obtained SJNs were averaged.

The layout of the experiment is shown in Fig. 6. The array was mounted on a rotator positioned at the origin; the Cartesian coordinates spanned laterally and longitudinally parallel to the walls of the chamber. The rotation angle (i.e., the array baseline angle measured clockwise from the lateral axis) was denoted as ϕ_{rot}. The two-dimensional positions of the jammer and ADS-B transmitter were denoted as $(x_{\mathrm{j}}, y_{\mathrm{j}})$ and $(x_{\mathrm{a}}, y_{\mathrm{a}})$, respectively, and their directional angles measured counterclockwise from

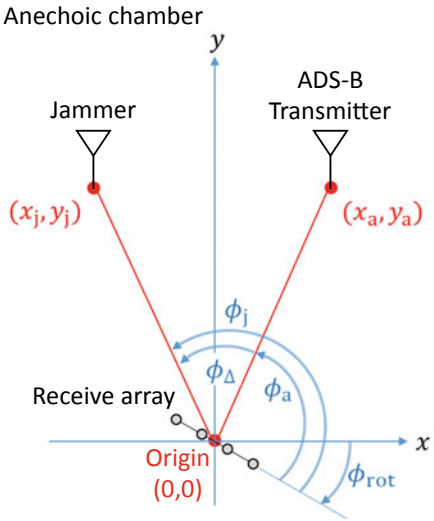


Fig. 6. Measurement layout

the array baseline were denoted as ϕ_{j} and ϕ_{a}, respectively. The separation angle, denoted as ϕ_{Δ}, was defined as the difference between ϕ_{j} and ϕ_{a}. The range of ϕ_{a} was [30, 150], assuming that the 360° in the horizontal plane is divided into three sectors. In the following experiments, the distances from the jammer and ADS-B transmitter to the receiver array are greater than the far-field distance, d_{f}, which is given by [21]:

$$d_{\mathrm{f}} = \frac{2d^2}{\lambda} = \frac{9\lambda}{2} = 1.2\,\mathrm{m}, \tag{17}$$

where the array dimension, d, is $\frac{3\lambda}{2}$ for the ULA used so the incident waves can be assumed to be far field.

4.2 Result

4.2.1 Initial Test

An initial test of the anti-jamming function was conducted with the layout shown in Fig. 7a. Figure 7b shows that the signal before beamforming was completely corrupted by the jamming signal. Figure 7c shows the directional pattern given by the PI beamformer:

$$D(\phi) = \boldsymbol{w}^{\boldsymbol{H}}\boldsymbol{a}(\phi). \tag{18}$$

A null was created toward the direction of the jamming signal. Figure 7d shows the signal after beamforming, showing that the pulses were clearly recovered. The SJN ratio was 14.9 dB. To interpret this SJN ratio, a threshold of 6 dB was employed as recommended in the decoding technique described in RTCA DO-260B [20]; in other words, SJN ratios greater than 6 dB were considered to be acceptable. Because the proposed function provided a sufficient SJN ratio, the initial test was considered to be successful.

4.2.2 Effects of Aircraft Distance and Jamming Power

Using the same layout as in the initial test, the measurement was repeated with different ADS-B powers (imaginary aircraft-receiver distances) and jammer powers. Figure 8 shows the resulting beamformed SJN ratio as a function of the JS ratio for different aircraft-receiver distances. These results show that stable anti-jamming performance is achieved. Moreover, the SJN ratio decreases as the aircraft-receiver distance increases; this means that aircraft that are farther from the receiver are more susceptible to jamming.

4.2.3 Effect of AoA

The measurement was repeated as the receiver array was rotated to examine the effect of the AoA with the same measurement layout as the initial test. Three ϕ_a angles $(90^\circ, 120^\circ, \text{and } 150^\circ)$, which correspond to ϕ_{rot} angles of $0^\circ, 30^\circ, \text{and } 60^\circ$, were

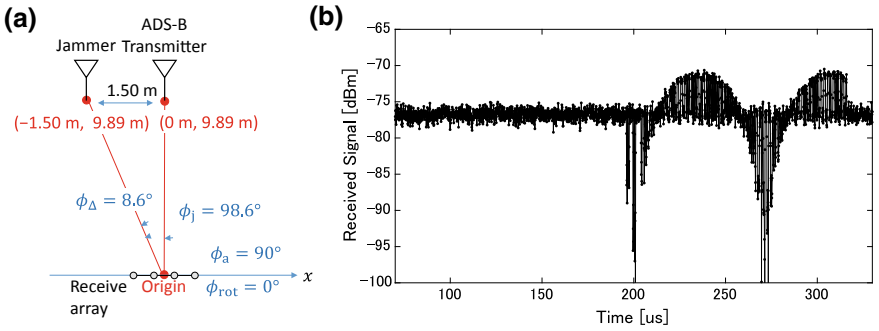


Fig. 7. Initial test of the anti-jamming function (**a**) Measurement layout. (**b**) Signal before the beamforming: one of the four channels. (**c**) Directional pattern of the beamforming. (**d**) Signal after the beamforming

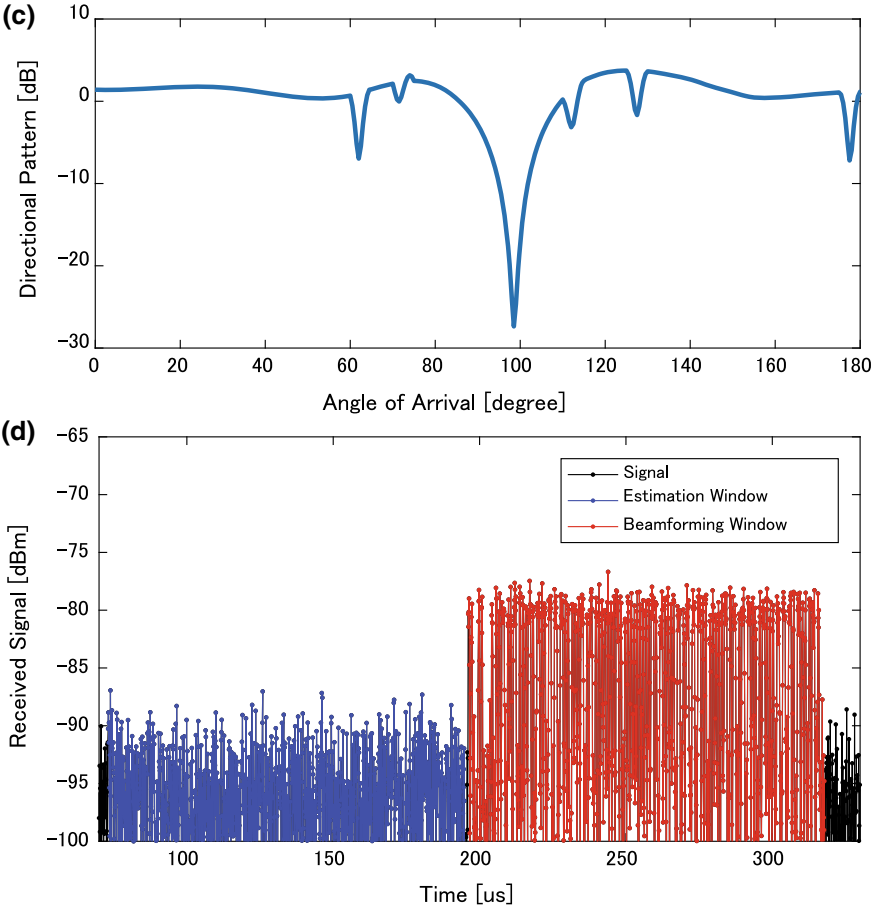


Fig. 7. (*continued*)

measured as shown in Fig. 9a. The resulting beamformed SJN ratio as a function of the AoA for different aircraft-receiver distances are shown in Fig. 9b. The results for $r =$ 150 NM are omitted because a good SJN ratio was not fully achieved. The SJN ratio decreases as ϕ_a increases; this means that the anti-jamming performance decreases when the ADS-B signal and jamming signal arrive obliquely at the array. This is because the sharpness of the null is degraded as shown in the directional pattern in Fig. 9c.

4.2.4 Effect of Separation Angle

The measurement was repeated as the separation angle between the jammer and ADS-B transmitter is changed. Three configurations were tested with three ϕ_Δ angles $\left(4.7^\circ, 8.6^\circ, \text{and } 16.1^\circ\right)$ as shown in Fig. 10a. Figure 10b shows the resulting beamformed SJN ratio as a function of the separation angle, ϕ_Δ. The results for $r = 150$ NM are omitted because a good SJN ratio was not fully achieved. The SJN ratio decreases as ϕ_Δ decreases. This is because the null in the directional pattern suppresses the ADS-B

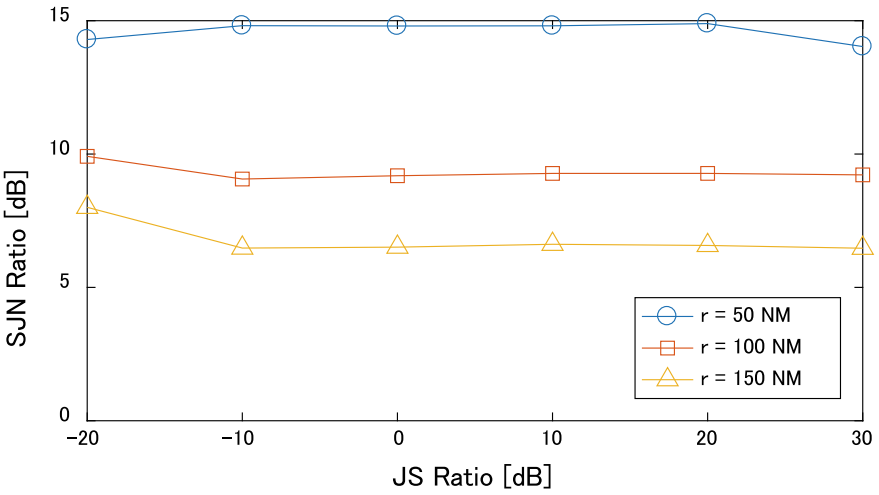


Fig. 8. Effects of the jamming power (JS Ratio) and the aircraft-receiver distance, r, on the anti-jamming performance

signal, when the ADS-B and jamming signals are close. As shown in the directional pattern in Fig. 10c, the null moves away from the ADS-B signal (150°) as ϕ_Δ increases.

4.2.5 Lower Bound Performance

Finally, the lower bound performance was examined. Based on the previous discussions in Sects. 4.2.2–4.2.4, the result for $\phi_a = 150^\circ \left(\phi_{\text{rot}} = 60^\circ\right)$ and $\phi_\Delta = \mathbf{4.7}^\circ$ were selected as the lower bound performances as shown in Fig. 11. The SJN ratio was above 6 dB when the JS ratio was less than or equal to 30 dB with $r = 50$ NM. Therefore, it was concluded that the decodable SJN ratio level, 6 dB, is guaranteed when the following conditions are satisfied: $r \leq 50$ NM, JS ratio ≤ 30 dB, $\phi_\Delta \geq 4.7^\circ$, and $30 \leq \phi_a \leq 150^\circ$.

5 Anti-spoofing Experiment

5.1 Setup

The anti-spoofing experiments were conducted in the same anechoic chamber as those described in Sect. 4. The ADS-B signals were transmitted and measured and the received signal was processed offline to evaluate the AoA measurement accuracy. The AoA accuracy was further translated into the expected anti-spoofing performance. The decoding process and the detection process (the binary hypothesis testing) were skipped because the focus of this experiment was the AoA measurement accuracy. The AoA error was computed as a performance measure as follows:

$$\epsilon_\phi = \hat{\phi}_a - \phi_a. \tag{19}$$

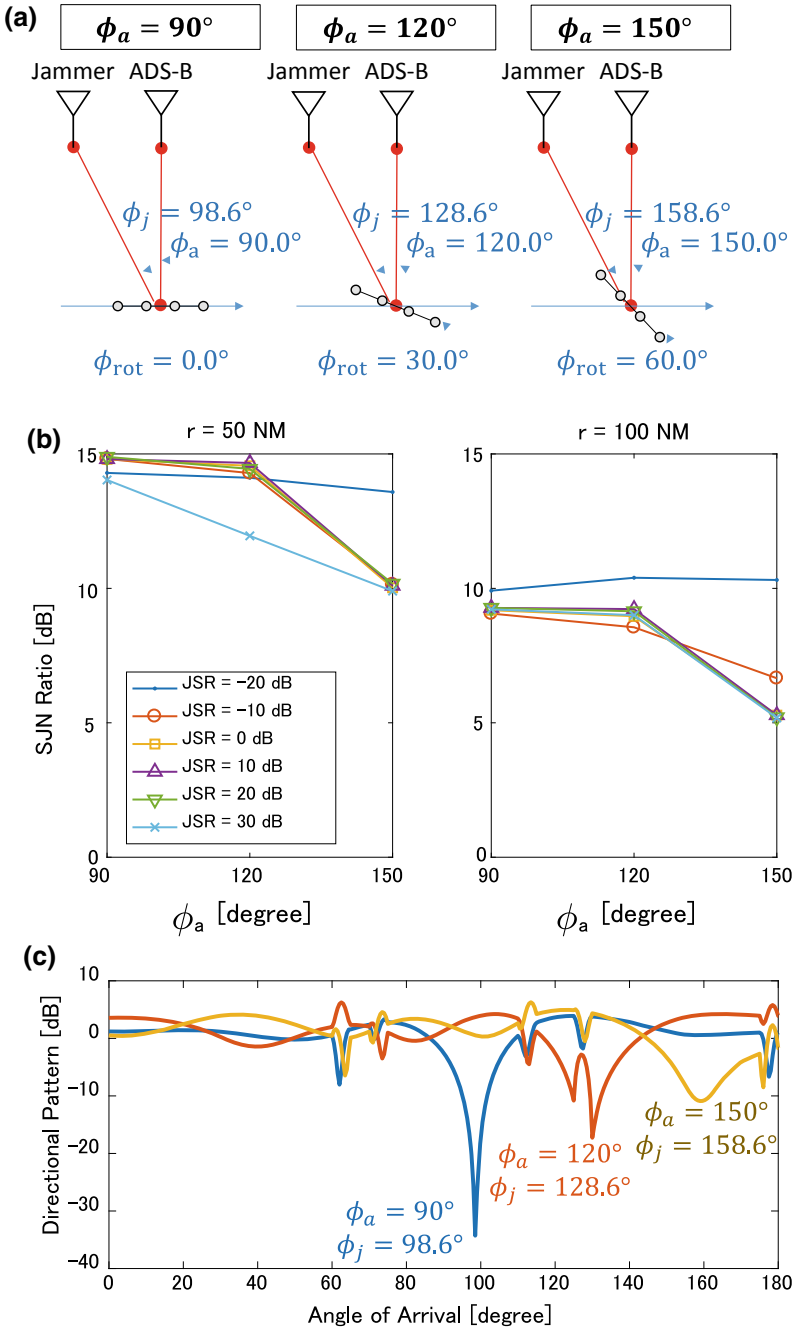


Fig. 9. Effect of AoA on the anti-jamming performance (**a**) Three measured AoA configurations. (**b**) Result. (**c**) Directional pattern

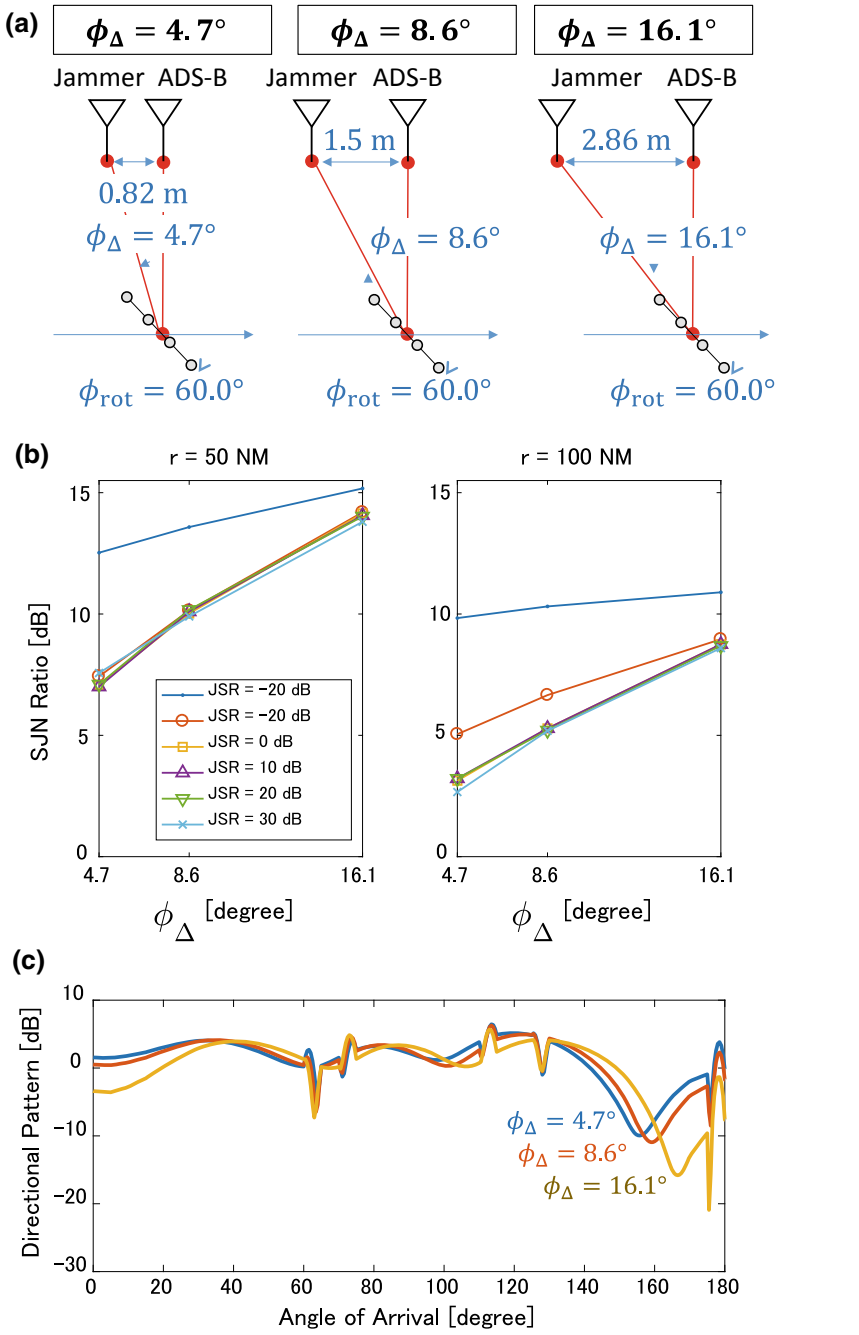


Fig. 10. Effect of separation angle on the anti-jamming performance (**a**) Three measured separation angles. (**b**) Result. (**c**) Directional pattern

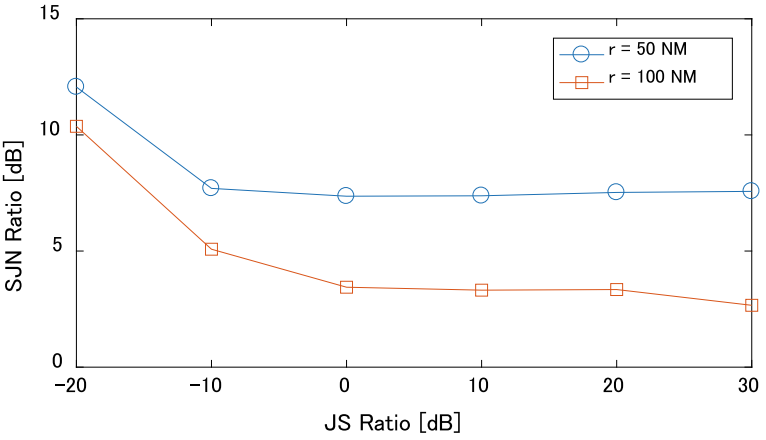


Fig. 11. Result of tests of lower bound performances: $\boldsymbol{\phi_a = 30^\circ}$ $\left(\boldsymbol{\phi_{rot} = 60^\circ}\right)$ and $\boldsymbol{\phi_\Lambda} = 4.7^\circ$

The mean and standard deviation (SD) of the ϵ_ϕ values measured in 30 trials were calculated; the SD is equivalent to the root mean square error (RMSE). Only samples that showed good synchronization among the oscilloscope channels were used. The measurement layout was the same as the anti-jamming experiment and $\phi_a = 30^\circ, 35^\circ, \ldots, 150^\circ$ $\left(\phi_{rot} = -60^\circ, -55^\circ, \ldots, 60^\circ\right)$. The ADS-B transmission power was set to simulate an aircraft-receiver distance of 50 NM; the results for other distances are omitted because the anti-jamming function was effective only for this value as described in the previous section.

5.2 Result

5.2.1 Initial Test

The initial results Fig. 12 shows the angular spectra as a function of the AoA for $\phi_a = 90^\circ, 120^\circ,$ and 150° $\left(\phi_{rot} = 0^\circ, 30^\circ,\right.$ and $\left.60^\circ\right)$. The spectra exhibit peaks corresponding to the directions of arrival of the ADS-B signal. The peaks were located at

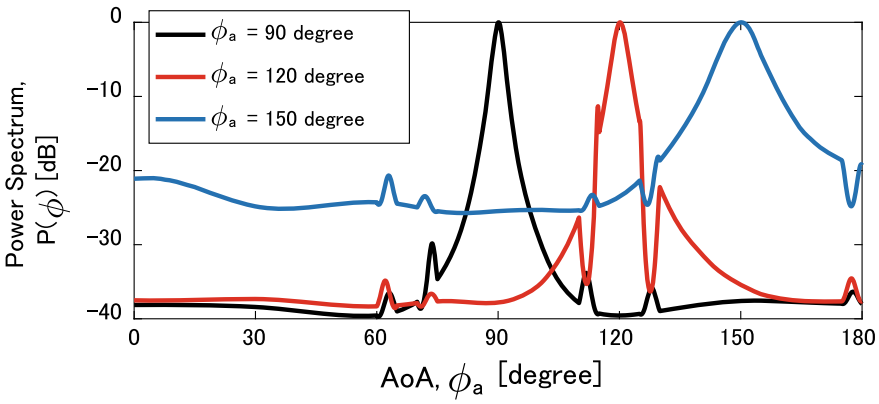


Fig. 12. Angular spectrum

$\hat{\phi}_{\mathrm{a}} = 90.1^{\circ}, 120.2^{\circ}$, and 150.0° and $\epsilon_{\phi} = 0.1^{\circ}, 0.2^{\circ}$, and 0.0° for $\phi_{\mathrm{a}} = 90^{\circ}, 120^{\circ}$, and 150°, respectively. Thus, the initial test was successful.

5.2.2 Effect of AoA

The mean and standard deviation of the ϵ_{ϕ} values that were evaluated for different AoA values are shown in Fig. 13(a) and (b), respectively. The mean is between -0.37° and 0.30° and the highest observed standard deviation was 0.13° at $\phi_{\mathrm{a}} = 30^{\circ}$. Thus, good AoA accuracy was obtained. It should be noted that these performances are for an aircraft-receiver distance less than or equal to 50 NM.

Finally, the obtained standard deviation was further translated into an anti-spoofing performance by applying an anti-spoofing framework based on binary hypothesis testing [19]. The assumptions are listed in Table 2. It was assumed that the mean was

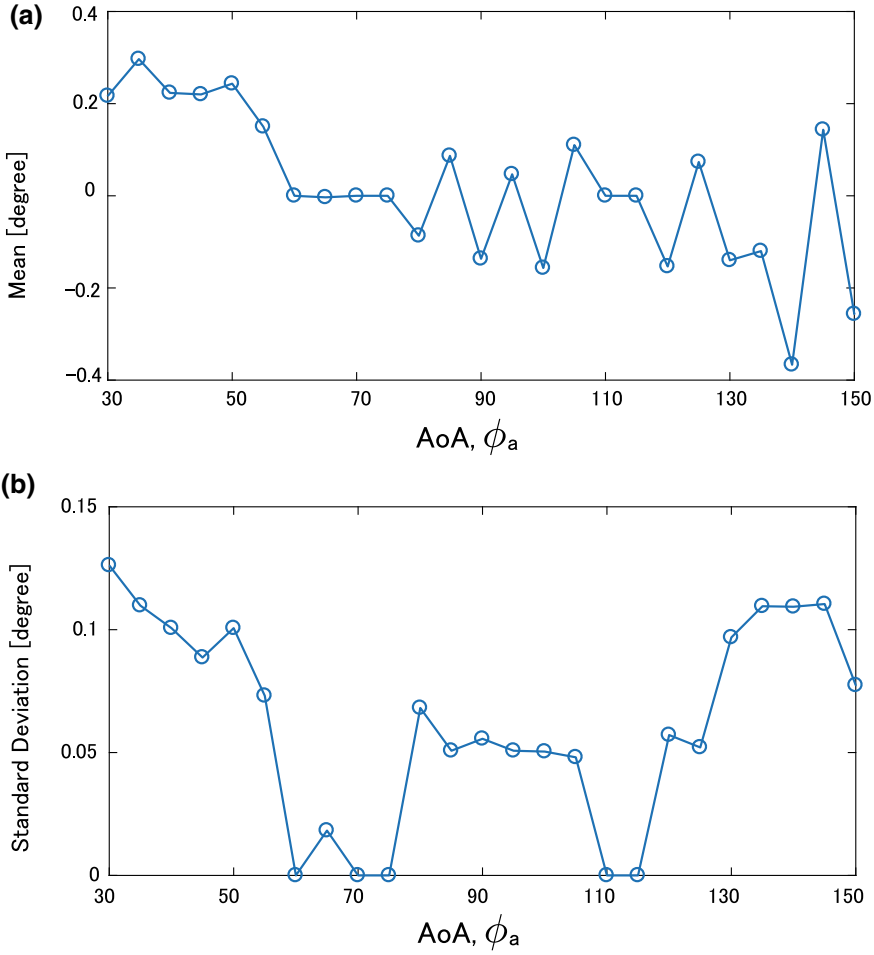


Fig. 13. Results of the AoA measurements (**a**) Mean. (**b**) Standard deviation

Table 2. Assumptions for estimating the anti-spoofing performance

ADS-B Position Error (EPU: Estimated Positon Uncertainty)	Not considered
Probability of false alarm	0.05
Number of samples	1

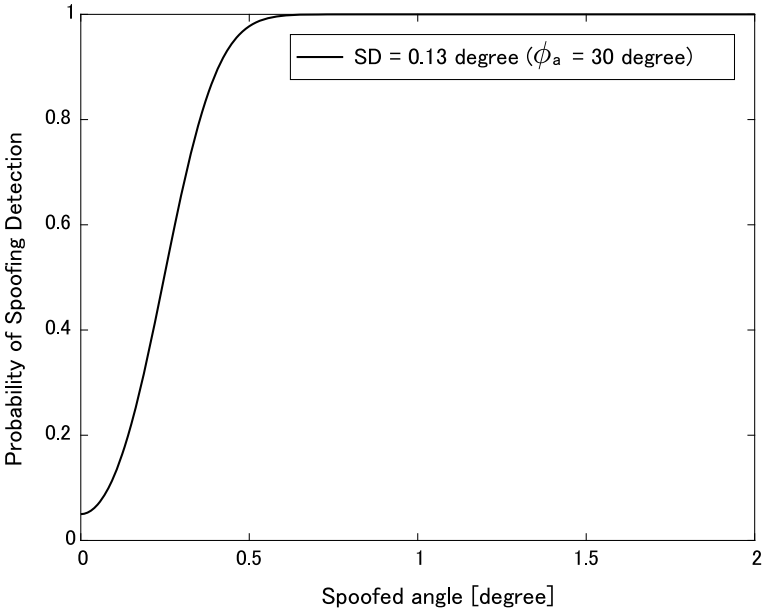


Fig. 14. Probability of spoofing detection

further calibrated to zero. Figure 14 shows the probability of spoofing detection as a function of the spoofed angle (i.e., the difference between the AoA from the ADS-B position and the AoA of the attacker position). The detection probability was above 90% at spoofed angles of 0.41°. This value is considered to be effective for initial anti-spoofing protection, especially when the spoofed trajectories are randomly distributed as in a flooding attack [19].

6 Conclusion

In this study, we conducted an experimental evaluation of anti-jamming and anti-spoofing array signal processing for securing ADS-B reception. For the jamming protection, a PI beamformer was used to suppress the jamming signal. For the spoofing protection, a Capon's beamformer was used to measure the AoA, which is compared to that corresponding to the reported position to verify the ADS-B position.

The anti-jamming experiment showed that the jammed ADS-B signals were successfully recovered and SJN ratio was improved to be above the decodable level.

The effect of the jamming power, received ADS-B power (used to simulate different aircraft-receiver distances), aircraft-receiver distance, AoA, and separation angle on the anti-jamming performance were evaluated. Thus, the lower bound performance was measured and found to meet the decodable for the SJN ratio, 6 dB, when the following conditions are satisfied: $r \leq 50\,\text{NM}$, $\text{JS Ratio} \leq 30\,\text{dB}$, $\phi_\Delta \geq 4.7^\circ$, and $30 \leq \phi_a \leq 150^\circ$.

The anti-spoofing experiment showed that the AoA of the ADS-B signal was measured successfully. The AoA accuracy was evaluated in terms of the worst standard deviation and was calculated to be 0.13° for an aircraft-receiver distance of 50 NM. Based on this standard deviation, the probability of spoofing detection was above 90% at a spoofed angle of 0.41°.

Future studies should be done to further improving the anti-jamming and anti-spoofing performance by employing advanced algorithms and by implementing these two functions with a single beamforming method.

Acknowledgements. The authors would like to thank Dr. Yonemoto and Mr. Soga for their support to the experiments in the anechoic chamber.

References

1. Strohmeier M, Lenders V, Martinovic I (2015) On the security of the automatic dependent surveillance-broadcast protocol. IEEE Commun Surv Tutor 17(2):1066–1086
2. Strohmeier M, Schafer M, Lenders V, Martinovic I (2014) Realities and challenges of NextGen air traffic management: the case of ADS-B. IEEE Commun Mag 52(5):111–118
3. Laonardi M, Piracci E, Galati G (2014) ADS-B vulnerability to low cost jammers: risk assessment and possible solutions. In: Proceedings of the 2014 Tyrrhenian international workshop on digital communications-enhanced surveillance of aircraft & vehicles, September 2014
4. SESAR JU (2016) Final project report improved 1090 MHz ADS-B ground station capacity and security
5. Manesh MR, Kaabouch N (2017) Analysis of vulnerabilities, attacks, countermeasures and overall risk of the automatic dependent surveillance-broadcast (ADS-B) system. Int J Crit Infrastruct Prot 19:16–31
6. McCallie D, Butts J, Mills R (2011) Security analysis of the ADS-B implementation in the next generation air transportation system. Int J Crit Infrastruct Prot 4:78–87
7. Purton L, Abbass H, Alam S (2010) Identification of ADS-B system vulnerabilities and threats. In: Proceedings of the Australasian transport research forum, Canberra, Australia, October 2010
8. Sampigethaya K, Poovendran R, Shetty S, Davis T, Royalty C (2011) Future e-enabled aircraft communications and security: the next 20 years and beyond. Proc IEEE 99 (11):2040–2055
9. Costin A, Francillon A (2012) Ghost in the air (Traffic): on insecurity of ADS-B protocol and practical attacks on ADS-B devices. Paper presented at Black Hat USA 2012, Las Vegas, NV, USA, July 2012
10. Pillai SU, Burrus CS (1989) Array signal processing. Springer, Berlin
11. Chen Z, Gokeda G, Yu Y (2010) Introduction to direction-of-arrival estimation. Artech House

12. Kikuma N (1999) Adaptive signal processing with array antenna. Science and Technology Publishing Company, Inc. (in Japanese)
13. Reck C, Reuther MS, Jasch A, Schmidt L-P (2012) Verification of ADS-B positioning by direction of arrival estimation. Int J Microw Wireless Technol 4(2):181–186
14. Wang W, Chen G, Wu R, Lu D, Wang L (2015) A low-complexity spoofing detection and suppression approach for ADS-B. In: Proceedings of the 2015 integrated communications, navigation, surveillance conference, Herdon, VA, USA, April 2015, pp (K2-1)–(K2-8)
15. Wu R, Chen G, Wang W, Lu D, Wang L (2015) Jamming supression for ADS-B based on a cross-antenna array. In: Proceedings of the integrated communications, navigation, surveillance conference, April 2015, pp (K3-1)–(K3-9)
16. Piracci EG, Galati G, Pagnini M (2014) ADS-B signals reception: a software defined radio approach. In: Proceedings of the 2014 IEEE metrology for AeroSpace, Benevento, Italy, May 2014, pp 543–548
17. Burrows ML (1996) Six-sector antenna for the GPS-squitter en-route ground station. Project report ATC-248, MIT Lincoln Laboratory, May 1996
18. Leonardi M, Piracci E, Galati G (2017) (2017) ADS-B jamming mitigation: a solution based on a multichannel receiver. IEEE Aerosp Electron Syst Mag 32(11):44–51
19. Naganawa J, Tajima H, Miyazaki H, Koga T, Chomel C (2017) ADS-B anti-spoofing performance of monopulse technique with sector antennas. In: Proceedings of the 2017 IEEE conference on antenna measurement and applications, Tsukuba, Japan, December 2017, pp 87–90
20. RTCA (2009) Minimum operational performance standards for 1090 MHz extended squitter automatic dependent surveillance - broadcast (ADS-B) and traffic information services - broadcast (TIS-B). RTCA DO-260B, December 2009
21. Balanis CA (2005) Antenna theory, 3rd edn. Wiley, New York, pp 34–35

Automatic Detection of Airport Runway Area Based on Super-Pixel PolSAR Image Classification

Ping Han(✉), Zeshan Lin, Xiaoguang Lu, Qingyan Shi, and Zhe Zhang

Tianjin Key Lab for Advanced Signal Processing, Civil Aviation University of China, Tianjin, China
hanpingcauc@163.com

Abstract. This paper proposes an unsupervised algorithm for airport runway area detection based on super-pixel PolSAR image classification. First, the simple linear iterative clustering (SLIC) algorithm are used to obtain super-pixel image by segmenting the PauliRGB image in order to reduce computational complexity and save computing time. Then, VAT-DBE algorithm is used to estimate and obtain the number of clusters of the image automatically. Combing the polarization information, the super-pixel image is classified by the method of spectral clustering. After that, the suspected airport runway area is extracted according to the scattering characteristics of the runway and classification result. Finally, the airport runway area is detected by using structural and topological characteristics of the runways. The experimental results show that the proposed algorithm can detect the airport runway area effectively with a clear outline, complete structure, and low false alarm rate. It also needs less time and a priori information compared with other methods.

Keywords: PolSAR image · **S**uper-pixel image · Airport runway detection · Spectral clustering · Region of interest

1 Introduction

Airport targets have an extremely important strategic significance both in the military and national economy fields. Its automatic detection is very necessary in the military strike, emergency rescue, aircraft navigation, and other domains. As a typical and stable symbol of an airport, runways are often used as a key feature to identify the airport area in an image. Synthetic-Aperture Radar (SAR) is widely used in remote sensing field because of the characteristics of its all-weather, all-time, strong penetrability, and long-distance imaging. Polarimetric SAR (PolSAR) image contains more polarization information of targets and can better characterize their features and details than that of single-polarization SAR, which show a great potential in classification, target detection, and recognition [1–4].

The existing airport runway detection algorithms based on PolSAR images normally adopt the image classification method to extract the region of interest (ROI) first, and then perform further identification. For the pixel-based image classification, the

Electronic Navigation Research Institute, *Air Traffic Management and Systems III*, Lecture Notes in Electrical Engineering 555,
https://doi.org/10.1007/978-981-13-7086-1_14

computation complexity is very high when the image size is large. Wishart classifier is often used with a good classification performance, but it is sensitive to initial clustering center and easy to fall into a local optimum.

To solve the above problems, this paper proposes a new algorithm for automatic detection of airport runway area based on super-pixel PolSAR image and spectral clustering classification. First, we construct a super-pixel image from the original PolSAR image. Then, unsupervised classification based on spectral clustering is applied to extract ROI which includes the runway. The airport runway area is detected finally by using structural and topological characteristics of the runways in the ROI. Since SAR is an active imaging system, this technology can be used to monitor disaster situations such as earthquakes and mudslides, when other facilities are damaged and it can also be used to assist aircraft navigation, detect airport runway, and road damage or some other fields.

2 Basic Theory and Methods

2.1 Polarimetric SAR Data and Pauli Decomposition

The polarized SAR data records the polarization scattering information of each pixel. For PolSAR, the complex relationship between the incident wave and the scattered wave is usually expressed by the Sinclair matrix [5].

$$S = \begin{bmatrix} S_{HH} & S_{HV} \\ S_{VH} & S_{VV} \end{bmatrix} \tag{1}$$

where H and V, respectively, represent the horizontal and vertical polarization states, which shows the complex scattering amplitude of the horizontal polarization emission and the vertical polarization reception. For single station backward scattering, the objective of satisfying the reciprocity theorem has $S_{HV} = S_{VH}$, so the 3D Pauli eigenvector can be expressed as follows:

$$\underline{k} = \frac{1}{\sqrt{2}}[S_{HH} + S_{VV}, S_{HH} - S_{VV}, 2S_{HV}] \tag{2}$$

In practical use, the polarization coherency matrix T is commonly used to represent the scattering process in order to clearly represent the physical meaning:

$$\begin{aligned} T = <\underline{k} \cdot \overset{*T}{\underline{k}}> &= \left\langle \begin{bmatrix} |k_1|^2 & k_1 k_2^* & k_1 k_3^* \\ k_2 k_1^* & |k_2|^2 & k_2 k_3^* \\ k_3 k_1^* & k_3 k_2^* & |k_3|^2 \end{bmatrix} \right\rangle \\ &= \begin{bmatrix} \left\langle |X|^2 \right\rangle & \langle XY^* \rangle & \langle XZ^* \rangle \\ \langle X^* Y \rangle & \left\langle |Y|^2 \right\rangle & \langle YZ^* \rangle \\ \langle X^* Z \rangle & \langle Y^* Z \rangle & \left\langle |Z|^2 \right\rangle \end{bmatrix} \end{aligned} \tag{3}$$

where $X = S_{HH} + S_{VV}$, $Y = S_{HH} - S_{VV}$, $Z = 2S_{HV}$.

2.2 SLIC Super-Pixel Segmentation

Super-pixel [6] refers to a small area in which pixels of similarity between pixels are merged according to gray scale, color, texture, etc. The super-pixel segmentation of images with similar pixels, can both reduce redundant information while preserving image details and useful information, greatly reduce the subsequent processing task of computational complexity and stability, and improve the efficiency of the algorithm. Therefore, the super-pixel is widely used in various tasks such as image classification, image segmentation, and so on. Recently, SLIC super-pixel-segmentation method is proposed by some scholars [7]. Because it is proved to be very effective and time saving in the image segmentation, the method is not only used for the optical image but also for the SAR image. SLIC is evolved by improving the k-means clustering method. The basic idea is to describe the RGB space using CLELAB color space and pixel location information X, Y, which are five-dimensional feature vectors. The similarity metric and local clustering of eigenvector is used to form super-pixel through iterative convergence. The detailed steps are as follows:

First, we set the number of super-pixels K, for an image with N pixels, the size of each super-pixel is therefore N/K pixels. Then, N seed points are evenly distributed on the image and labels are assigned to each seed point, so the interval between each seed point is $S = \sqrt{N/K}$. In order to avoid placing seed points at an edge and choosing a noisy pixel, we move them to seed locations corresponding to the lowest gradient position in a 3×3 neighborhood. Then, according to the Eqs. (4)–(6), the similarity between the seed points within $2S \times 2S$ area around the super-pixel center (seed point) is obtained. Finally, we complete SLIC super-pixel segmentation by calculating the similarity between the seed point and the adjacent pixel, merging the pixels with similar similarity and assign the label of the nearest seed point repeatedly.

$$d_{lab} = \sqrt{(l_i - l_j)^2 + (a_i - a_j)^2 + (b_i + b_j)^2} \tag{4}$$

$$d_{xy} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \tag{5}$$

$$D_i = d_{lab} + \frac{m}{S} d_{xy} \tag{6}$$

where d_{lab} is the color distance between the pixels in LAB color space, d_{xy} is the plane distance between pixels, and D_i is the sum of the lab distance and the xy plane distance normalized by the grid interval S. A variable m is introduced to control the compactness of a super-pixel.

2.3 VAT and DBE Algorithm

In the image classification problem, it often involves the determination of the number of classes, a reasonable class number directly affect the complexity and accuracy of classification results. Most of the existing classification algorithms are either according to the experience of manual settings or some characteristics to set a fixed number and

cannot adaptively determine the number of data categories, which affects the classification results [8]. To solve the problems above, Bezdek et al. [9] proposed a VAT algorithm (visual assessment of cluster tendency). VAT is a graph-based algorithm that can estimate and visualize the potential clustering information of data, and combined with dark block extraction (DBE) algorithm [10], which is used to postprocess the estimation results, the estimated class number, and the clustering center is automatically extracted. Liu Bin [11] and others applied it to SAR image adaptive classification processing and achieved good results, so we choose the algorithm before clustering to estimate the class number. The basic steps of the algorithm are as follows:

First, we use a measure to calculate the distance of data to construct the dissimilarity matrix and sort the data with minimum dissimilarity along the main diagonal of the matrix to form a dissimilarity image. Then, we construct a new dissimilarity matrix to form a reordering dissimilarity image, the RDI graph is binarized with Ostu automatic threshold segmentation, and perform a distance transformation (DT) after directional morphological filtering. In the end, we project all pixel values of the DT image onto the main diagonal axis to obtain a projection signal and calculate the first-order derivative of the projection signal. So, the number and position of "zero-crossing" points are obtained, the number of "zero-crossing" points represents the estimated class number of data, and the location of "zero-crossing" points corresponds to the cluster center. The main steps of the algorithm are shown in Fig. 1:

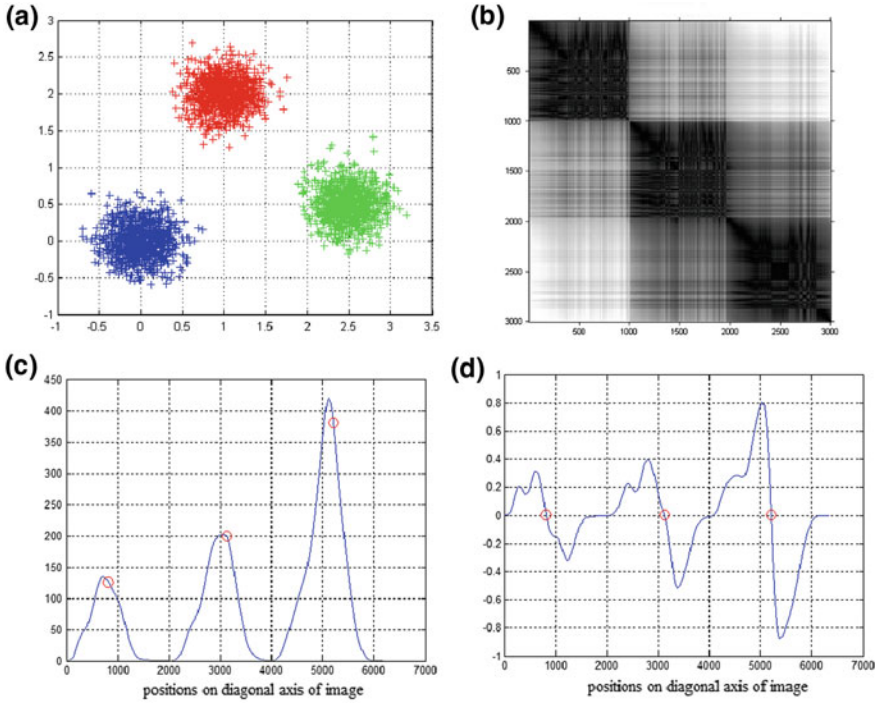


Fig. 1. Main steps of an example of VAT and DBE algorithm. **a** Scatter plot of 3,000 points. **b** Reordered VAT image. **c** Projection signals. **d** First-order derivative signal

Figure 1a shows the randomly generated three types of data points scatter plot, each type of data contains 1000 points, and is represented with different colors, respectively; Fig. 1b shows the reordered VAT image after calculating the Euclidean distance dissimilarity of the three types of data; Fig. 1c shows the projection signal, the ordinate represents the projection value, and the abscissa represents the position of the main diagonal axis; Fig. 1d shows the first-order derivative of the projection signal, where the number and position of the circle are the estimated class number and the position of the cluster center.

2.4 Spectral Clustering Algorithm

Wishart classifier, which is based on ideas of k-means, is often used with a good classification performance in polarimetric SAR image unsupervised classification, but it is sensitive to initial clustering center and easy to fall into a local optimum [12, 13]. Spectral clustering [12] is a clustering method based on spectral graph theory. Compared with the traditional clustering method, it has the characteristics of clustering in any shape data space and converging to the global optimal. The core idea is to consider each sample as the vertex of the graph, and the dissimilarity between the samples is treated as the edge of the weighted graph. The method of spectral graph segmentation is used to find the optimal method for grouping and the final result is that the weights of the edges of the same group of samples are as high as possible. The weights of the edges of the different groups are as low as possible, that is, the dissimilarity in the group is high, the dissimilarity between groups is low, so as to achieve the purpose of clustering. It is important to apply spectral clustering method to SAR image processing, and some good results have been obtained. Such as Ersahin [14] and others applied spectral clustering method to the classification of polarized SAR images for the first time and achieved a good result. Anfinsen [15] and others proposed a SAR image classification algorithm based on the polarization coherence matrix combined with the Wishart classifier and spectral clustering method, and it has high accuracy. Therefore, we choose spectral clustering for the proposed algorithm. Nyström method [16] is used in order to reduce the time and space cost. The detailed steps are as follows:

Step1 Construct affinity matrix $W_{n\times n}$ with a measure.

Step2 Calculate the Laplacian matrix, $L_{sym} = D^{-\frac{1}{2}}(D - W)D^{-\frac{1}{2}}$ where D is the diagonal matrix, whose diagonal elements are $D(j,j) = \sum_{i=1}^{n} w_{i,j}$, that is the sum of all the column elements of the matrix $W_{n\times n}$.

Step3 Calculate the eigenvalues of the Laplacian matrix L_{sym} and arranged in order from small to large, then feature vectors corresponding to the first k eigenvalues to construct the matrix $V_{n\times k} = \{v_1, v_2, \ldots v_k\}$.

Step4 Cluster the row vectors of $V_{n\times k}$ using k-means algorithm. Complete spectral clustering by dividing the n k-dimensional data into is k class finally.

3 Proposed Method in This Paper

The specific flow of the algorithm is shown in Fig. 2. First PauliRGB image is obtained by Pauli decomposition of the fully polarized SAR image and it is segmented by SLIC algorithm to construct a super-pixel image. Then, the classification step is completed and suspected region of interest (ROI) of airport area is extracted by using spectral clustering after that VAT and DBE algorithms are used to estimate the class number of super-pixel image. Finally, according to the airport runway polarization characteristics, weak echo characteristics and geometric characteristics, further identification is executed for the suspected ROI extraction to determine the airport area.

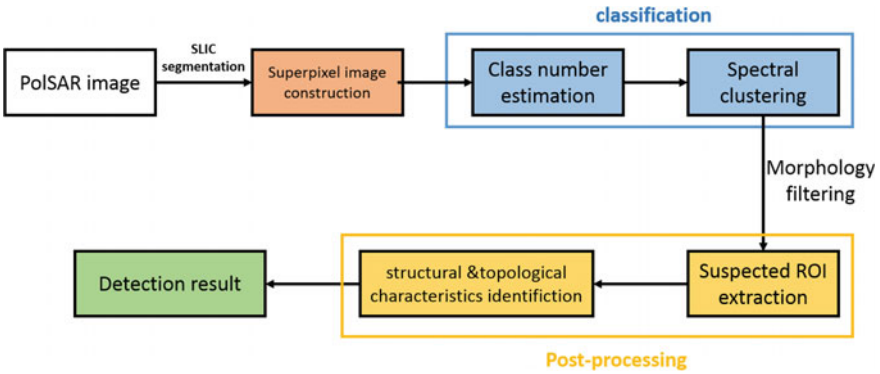


Fig. 2. The flowchart of the proposed method

3.1 Super-Pixel Image Construction

After PauliRGB image is formed by Pauli decomposition of polarized SAR image, we first set the number of super-pixels to be generated as K. The number of super-pixels determines the complexity and execution effect of the latter classification. The more the number of super-pixels is, the higher the dimension of affinity matrix will be in the process of spectral clustering, which costs lots of the memory and calculating time; and if the number of super-pixels is too small, it is easy to form under-segmentation, resulting in poor segmentation effect. Therefore, all the SAR images used in the experiment are set different super-pixels number to segment and count the average boundary recall [6] and computing time. Figure 3a shows that as the number of super-pixels increases, the boundary recall continues to increase; Fig. 3b shows that time-consuming is increasing with the increase of the super-pixel number. When the number of super-pixels is greater than 10000, the increasing trend of boundary recall slows down, so we choose 10000 super-pixels as a standard for the experiment.

According to the number of super-pixels to be set, the PauliRGB image is segmented by the SLIC algorithm mentioned in Sect. 2.2 to form the super-pixel image I, then the mean polarization information of the pixels contained in each super-pixel is computed and used as the polarization information of the super-pixel. When the

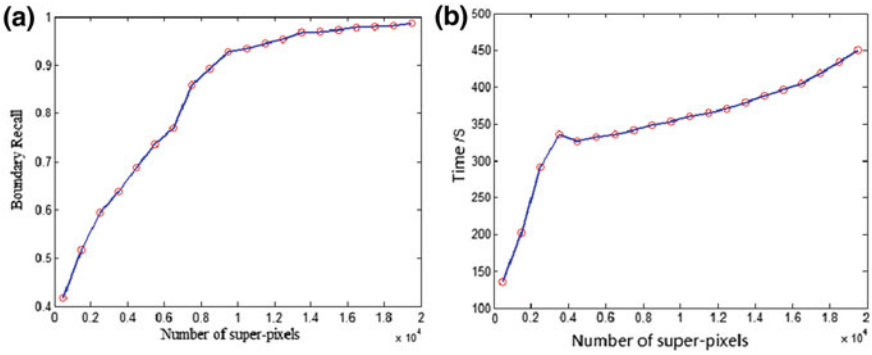


Fig. 3. The impact of the super-pixels number on boundary recall and algorithm computing time. **a** Impact on boundary recall. **b** Impact on computing time

polarization coherence matrix T is chosen as the feature, the average coherence matrix $\overline{T}_s$ of each super-pixel is shown in Eq. (7). So, the number of super-pixels is much smaller than the pixels of the original SAR image, and not only is the complexity of the later process reduced but also the polarization information of the original SAR image is preserved.

$$\overline{T}_s = \frac{1}{n}\sum_{i=1}^{n} T_i \tag{7}$$

where $n = N/K$ represents the pixels number in each super-pixel, that is, the size of the super-pixel, N is the pixels number included in the PauliRGB image.

3.2 Classification Processing

After the super-pixel image I is obtained, the gray scale information contained in super-pixel image I is used as the feature to measure the pairwise dissimilarity by Euclidean distance of pixels value, then the VAT and the DBE algorithm in Sect. 2.3 are used to estimate the class number c of the I for the further clustering. When we get the class number c, we construct the dissimilarity $w(i,j)$ between super-pixels with distance of $\overline{T}_s$ by the Gaussian kernel function, as is shown in Eq. (8).

$$w(i,j) = \exp\left(\frac{-\left\|\overline{T_{si}} - \overline{T_{sj}}\right\|^2}{2\sigma^2}\right) \tag{8}$$

where σ is the scaling parameter and $\left\|\overline{T_{si}} - \overline{T_{sj}}\right\|$ is the Euclidian distance between super-pixels i and j.

In the end, the affinity matrix is formed by calculating the dissimilarity between all super-pixels and the spectral clustering algorithm mentioned in Sect. 2.4 is implemented to complete classification step and get the classification result I1.

3.3 Suspected ROI Extraction and Identification

Corresponding to the classification result in classification step, we calculate the average power of each class in I1. As the runway appears weak scattering echo characteristics in PolSAR image, the power of pixels belonging to runway is very small. So, the pixels in a class with minimum average power are extracted as suspected airport runway areas (ROI). The pixels inside ROI are labeled as "1" and others are labeled as "0" to form a binary image. Some small isolated areas labeled as "1" are removed with morphologic processing method. Finally, runway structural features such as parallels, topologic property are utilized to identify which region is the true runway area just like the method used in the literature [3].

4 Experiment Results

In order to verify the validity of the proposed algorithm, we use the four-look real PolSAR image data of San Andreas Fault area collected by NASA/JPT UAVSAR system at L band in 2009. The azimuth and range resolution are 7.2 and 4.9 m, respectively. Figure 4 shows PauliRGB and the optical image about this area. The image size is 1051 × 1151 pixels. There are airport, river, road, grass, farmlands, and buildings, etc., in the image. Parameters in the experiment are set as follows: morphological filter threshold Th0 = 450, the number of super-pixel is 10000, the number of Euler E and parallel line spacing D is are 0 and 10, respectively.

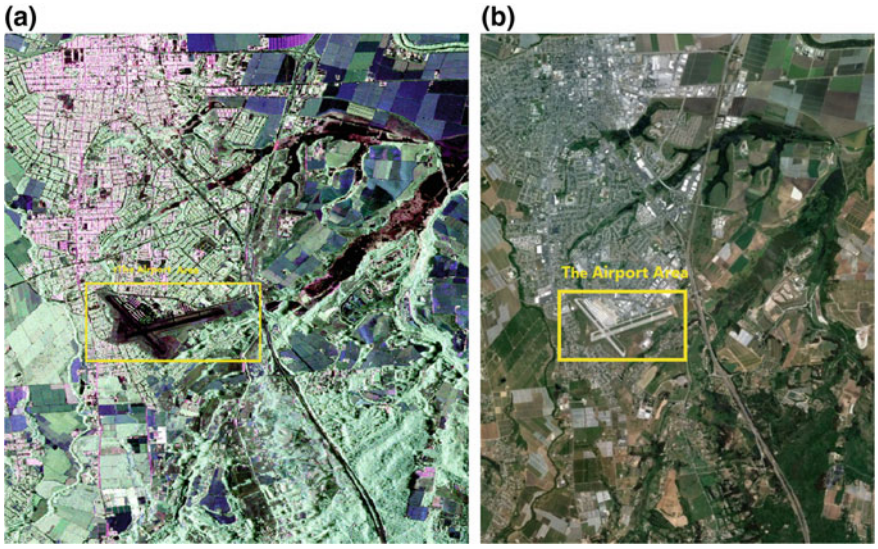


Fig. 4. The images of San Andreas fault area. **a** PauliRGB image. **b** Optical image. *Source* https://www.asf.alaska.edu/

Figure 5a shows the super-pixel image obtained by SLIC segmentation algorithm with the uniform size and clear edge and the super-pixels number is about one percent

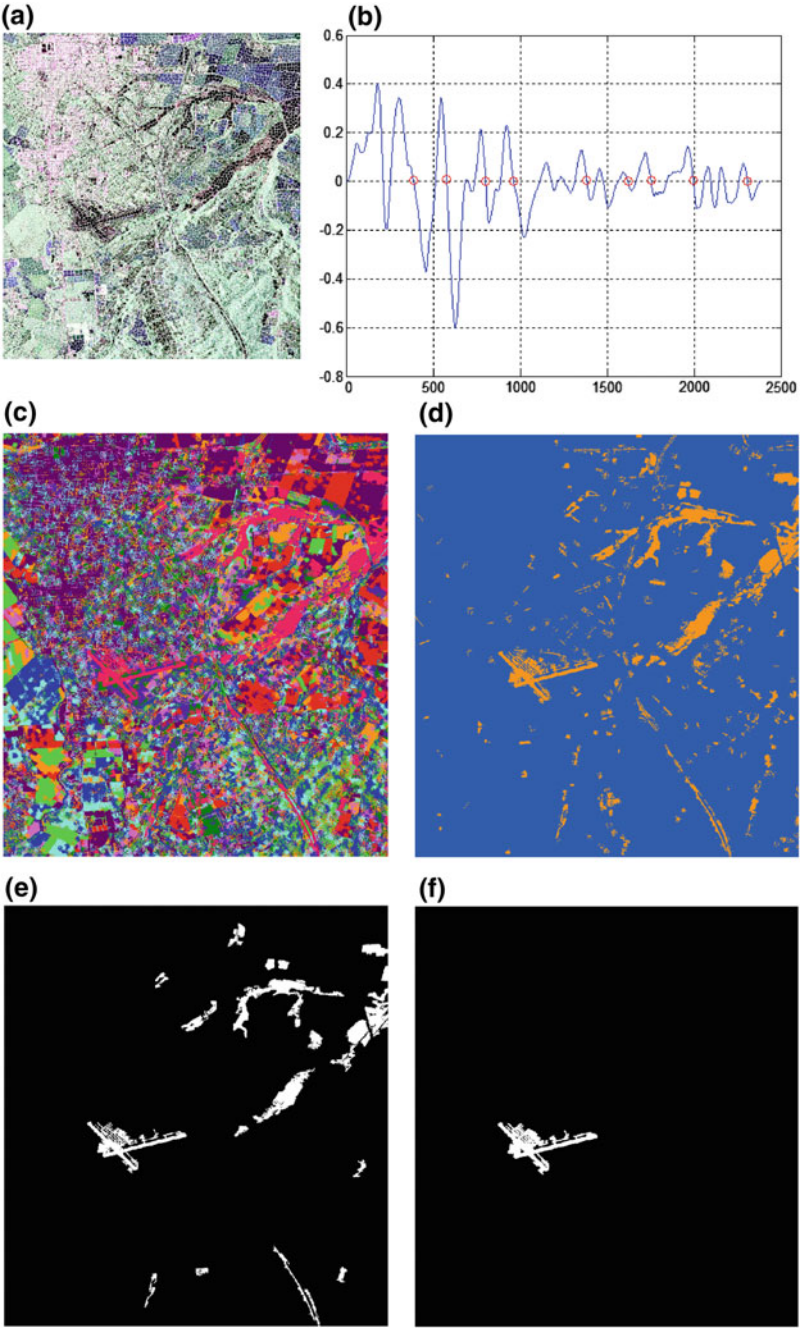


Fig. 5. The experiment result of the proposed method. **a** Super-pixel image. **b** Class number estimation result. **c** The classification result. **d** Class with power minimum extraction. **e** Extraction of ROI. **f** Detection result

of the original pixels number. This means that the SLIC segment step not only reduces the number of pixels to process, improves the efficiency of the algorithm, but also retains the boundary information of the image. Figure 5b shows the estimated class number by using VAT and DBE algorithm with the gray scale information of the super-pixels and the class number. According to the estimated number, the spectral clustering method is used to classify the super-pixel images, and the results are shown in Fig. 5c, it can be seen clearly that the light red runway area and its various colors represent the different types of ground objects, which indicates that the classification result is good. Figure 5d shows the result of extracted class with minimum power according to the weak echo scattering characteristics of the runway, and a small amount of other interfering objects can be clearly distinguished in the image. Figure 5e shows the extracted ROI of the suspected airport runway area by morphological processing for the binary image of Fig. 5d to remove some small areas. Figure 5f shows the final detection result of the proposed method. Compared with the method shown in Figs. 6 and 7, the three methods can effectively detect the airport area all. However, there are false alarms in the final results of the method used in the literature [3], and under the same experimental conditions, the proposed method takes less time, about 4 percent of the method in [3], 30 percent of the method in [17]. From the experiment, it proves that the proposed algorithm can effectively extract the airport runway area in the polarized SAR image without supervision and takes less time.

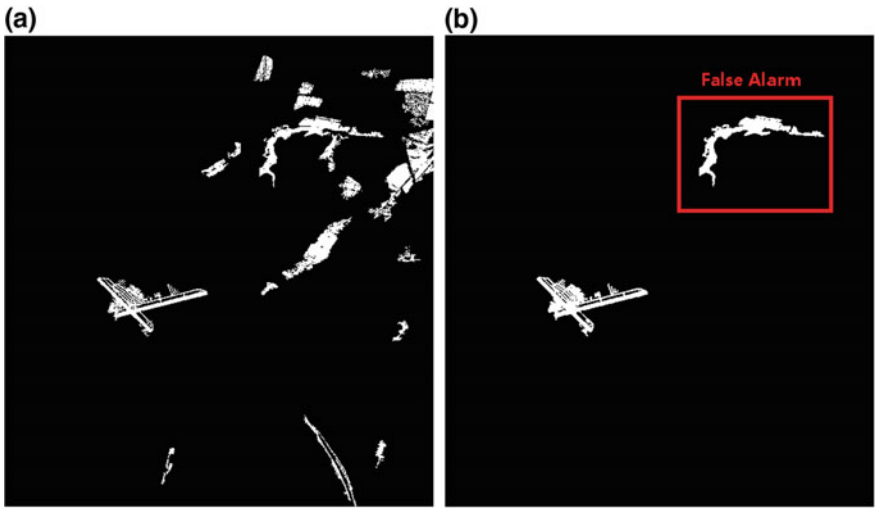


Fig. 6. The experiment result in [3]. **a** Extraction of ROI. **b** Detection result

In addition, we have applied the proposed method to other 12 different PolSAR images collected by UAVSAR system at L band. Detection results show that the proposed method is of high detection accuracy and low false rate.

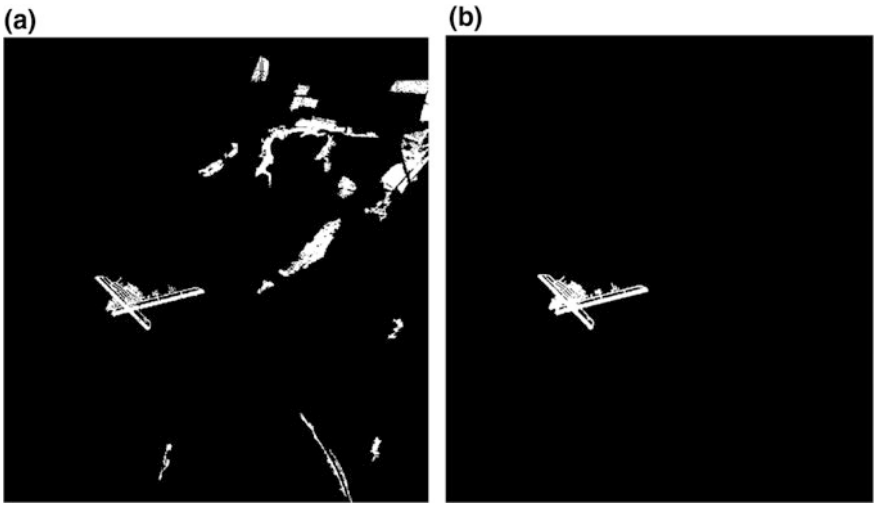


Fig. 7. The experiment result in [17]. **a** Extraction of ROI. **b** Detection result

5 Conclusions

In this paper, we propose an unsupervised method based on super-pixel image and spectral clustering to detect the airport runway area. The SLIC segmentation algorithm is used to obtain the super-pixel image first in order to reduce the number of pixels and the complexity of post-algorithm processing. Then the VAT and DBE algorithms are used to implement the adaptive determination of the number of classification and the super-pixel image is classified using the spectral clustering method according to the estimated number. Finally, runway area is detected by using target scattering properties, structural, and topological characteristics of the runways. Through the experiment of the multiple sets of fully polarized SAR data and the comparison with the algorithm in [3, 17], the results show that the proposed algorithm can detect the airport runway area more accurately with less calculation time, and the false alarm rate is low. The algorithm needs less priori information and need not set the class number artificially before classification, also, it is insensitive to speckle noise, which means more robust and higher efficiency.

Acknowledgements. The authors also would like to thank the support of the National Natural Science Foundation of China (No.61571442, No. 61471365), the National Key Research and Development Program of China (No.2016YFB0502405), and National University's Basic Research Foundation of China (No. 3122014C004).

References

1. Liping Z, Hong Z, Chao et al (2010) A fast method of airport detection in large-scale SAR image with high resolution. J Image Graph Beijing, J Image Graph China 15(7):1112–1120
2. Yong H, Xin X, Hong S et al (2004) Detection of airport runways in airborne SAR images. J Wuhan Univ China 50(3):393–396
3. Ping H, Ling C, Cheng Z, QingYan S (2016) Runways detection based on H/Q decomposition and iterative Bayesian classification. J Syst Eng Electron China 38(19):2048–2054
4. Ruijin J, Wei Z, Yang J (2014) Airport automatic detection in large-scale polarimetric SAR images. J Tsinghua Univ China 54(12):1588–1593
5. Lee JS, Pottier E (2009) Polarimetric radar imaging: from basics to applications. CRC Press, Florida, pp 66–240
6. Achanta R, Shaji A, Smith K et al (2012) SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans Pattern Anal Mach Intell USA 34(11):2274–2282
7. Qin F, Guo J, Lang F (2015) Superpixel segmentation for polarimetric SAR imagery using local iterative clustering. Geosci Remote Sens Lett IEEE USA 12(1):13–17
8. Shibao W (2014) Research on unsupervised polarimetric SAR image classification. Xidian University, China
9. Bezdek JC, Hathaway RJ (2002) VAT: a tool for visual assessment of (cluster) tendency. In: International joint conference on neural networks, February 2002. IEEE, pp 2225–2230
10. Wang L, Leckie C, Ramamohanarao K et al (2009) Automatically determining the number of clusters in unlabeled data sets. IEEE Trans knowl Data Eng USA 21(3):335–350
11. Liu B, Hu H, Wang H et al (2013) Superpixel-based classification with an adaptive number of classes for polarimetric SAR images. IEEE Trans Geosci Remote Sens USA 51(2):907–924
12. Weiying G (2013) Classification of polarimetric SAR image based on spectral clustering. Xidian University, China, p 2013
13. Ming C, Jie Y, Yanbing W et al (2015) Superpixel based PolSAR image spectral clustering. Sci Surv Mapp China 40(3):80–84
14. Ersahin K, Cumming IG, Ward RK (2010) Segmentation and classification of polarimetric SAR data using spectral graph partitioning. IEEE Trans Geosci Remote Sens USA 48 (1):164–174
15. Anfinsen SN, Jenssen R, Eltoft T (2007) Spectral clustering of polarimetric SAR data with the wishart-derived distance measures, vol 644
16. Fowlkes C, Belongie S, Fan C et al (2004) Spectral grouping using the Nyström method. IEEE Trans Pattern Anal Mach Intell USA 26(2):214
17. Ping H, Cheng, Z, Ling C (2016) Automatic runway detection based on unsupervised classification in PolSAR image. In: Integrated communications navigation and surveillance (ICNS), April 2016. IEEE, pp 6E3-1–6E3-8

Part IV
Future Technologies

Wireless-Optical Signal Converter Utilizing Stacked-Patch Antennas and Electro-Optic Substrates

Hiroshi Murata[1(✉)], Hironori Aya[2], Toshiyuki Inoue[2], Atsushi Sanada[2], Yusuf Nur Wijayanto[3], Atsushi Kanno[4], and Tetsuya Kawanishi[4,5]

[1] Graduate School of Engineering, Mie University, 1577 Kurima-Machiya-Cho, Tsu, Mie 514-8507, Japan
murata@elec.mie-u.ac.jp
[2] Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-8531, Japan
[3] Indonesian Institute of Sciences, Jl. Sangkuriang, Cisitu, Bandung 40135, Indonesia
[4] National Institute of Information and Communications Technology, 4-2-1 Nukuikitamachi, Koganei, Tokyo 184-8795, Japan
[5] Waseda University, 3-4-1 Ohkubo, Shinjuku, Tokyo 169-8555, Japan

Abstract. We have proposed and developed a new wireless-optical signal converter utilizing stacked-patch antennas with a narrow gap and electro-optic (EO) crystal films for millimeter-wave radio-over-fiber systems. This converting device utilizes a structure of vertically stacked-patch antennas like the Yagi–Uda antenna, with a hybrid substrate composed of EO crystal films and low-dielectric-constant material plates. The electric field for wireless-optical signal conversion and the size of the patch antenna are increased by introducing the antenna and substrate stacked structure. Therefore, the conversion efficiency from wireless to optical signals can be improved. In the analysis by use of HFSS, the operational frequency was shifted to the higher range by 7 GHz, and the optical modulation index was improved by 3.2 dB with the double-stacked structure. In the experiments, the operational frequency shift by 8 GHz and the modulation index improvement by 3.0 dB were verified successfully, and the measured frequency response and directivity were in good agreement with the designed ones. The conversion of wireless 60 GHz band signals with QPSK modulation and the transfer of the converted signal over 1 km long fiber were also demonstrated.

Keywords: Planar antenna · Electro-optic effect · Yagi–Uda antenna · Radio over fiber · Optical modulator

Electronic Navigation Research Institute, *Air Traffic Management and Systems III*, Lecture Notes in Electrical Engineering 555,
https://doi.org/10.1007/978-981-13-7086-1_15

1 Introduction

Millimeter-wave (MMW) wireless technology is promising for the applications to high-resolution radar, remote sensing/imaging, molecular spectroscopy, and navigation systems [1–3]. MMW is also important for fifth-generation (5G) mobile communication systems owing to its outstanding characteristics for broadband data transfer and massive connection capability by use of time, frequency, and free-space-division multiplexing techniques [4–6]. Therefore, studies on MMW components, devices, modules, and subsystems have recently been attracting a lot of interest in the world.

However, MMW is not suitable for long-distance (>10 km) free-space nor wire transmission since it has a large propagation loss in both air and metal cables. Radio-over-fiber (RoF) technology is a promising solution to connect many stations/cells of a relatively small coverage area by use of extremely low-loss silica optical fibers ($\sim$0.2 dB/km at the light wavelength of 1550 nm), and is applicable in many wireless MMW application systems of radar, communication, remote sensing, and so on [7–10]. In MMW RoF systems, a signal conversion device from MMW to lightwave (LW) is significant. We have proposed several new devices for converting MMW signals to LW signals directly using planner antennas and electro-optic (EO) crystals [11–14]. In particular, these MMW-LW signal conversion devices are attractive for compact MMW signal receivers in MMW distributed radar systems to detect foreign object debris on an airport runway and other avionics systems [15, 16].

In this paper, a new wireless MMW-LW signal converter with an improved conversion efficiency by use of the patch antenna and substrate stacked structure is discussed [17]. Utilizing a double-stacked structure, the antenna gain, antenna aperture, and electric field induced across the narrow gap for optical modulation are all increased. Therefore, the conversion efficiency from MMW to LW is expected to be improved. In the experiment, a 3.0 dB increase in the optical modulation index was successfully demonstrated in the operational frequency band of 60 GHz.

2 Device Structure and Operation

Figure 1 shows the structure of the proposed EO MMW-LW signal converter, which utilizes the structure of stacked-patch antennas of multiple substrates combined with thin EO crystal (z-cut $LiNbO_3$) films and low-dielectric-constant material (SiO_2) plates. In the lower level patch antenna, a narrow ($\sim$5 μm) gap is introduced at the center of the patch metal film along the *y*-axis as shown in Fig. 1a, and a single-mode straight optical waveguide is also fabricated along the gap in the $LiNbO_3$ crystal film. This gap is used as the optical modulation electrode for MMW-LW signal conversion by use of a displacement current and the first-order EO effect (Pockels effect) in the $LiNbO_3$ film.

When a wireless MMW signal with *x*-polarization is irradiated to this converter from above, a strong MMW electric field (displacement current) is induced across the gap owing to the continuity of the electric current. This electric field is rather strong compared to the field irradiated from above (about 100 times). Therefore, by introducing an optical waveguide just close to the gap, an LW propagating in the optical

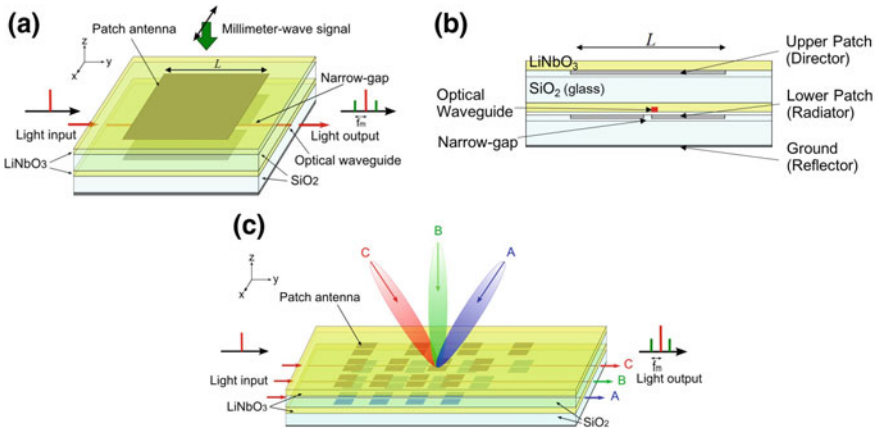


Fig. 1. Structure of the proposed EO MMW-LW signal converter utilizing stacked-patch antennas embedded with a ~5 μm gap. **a** The whole view. **b** The cross-sectional view. **c** Array configuration with multiple directivity. © 2017 IEEE, Reprinted, with permission, from [17]

waveguide is modulated by the wireless MMW signal through the Pockels effect, and signal conversion from MMW to LW is obtainable [12, 14].

Utilizing the stacked structure as shown in Fig. 1b, the antenna gain and electric field across the gap for optical modulation are enhanced [13, 14], since this structure can be regarded as a kind of Yagi–Uda antennas along the normal direction of the substrate; the gap-embedded patch antenna can be considered as a radiator composed of a printed dipole antenna of a half-wavelength, and the ground and upper patch antenna can be considered as a reflector and a director, respectively. Therefore, the antenna gain is expected to be enhanced like a Yagi–Uda antenna, and the signal conversion efficiency can be improved. The directivity in MMW-LW signal conversion is also expected to be narrower than that of the single stack device by using the stacked structure like a Yagi–Uda antenna.

In addition, utilizing the array structure along the *y*-axis as shown in Fig. 1c, the directivity in MMW-LW conversion can be also designed using the successive optical modulation in each patch antenna like an array antenna [12, 14]. Therefore, several MMW space-division multiplexing (SDM) signals coming from different directions can be discriminated and converted to different optical signals simultaneously as shown in Fig. 1c.

3 Analysis and Design

The proposed MMW-LW signal converter was designed using three-dimensional electromagnetic field analysis software, HFSS ver. 16. In the device design with an operational frequency in the 60 GHz band, we decided the thickness values of the $LiNbO_3$ crystal film and the SiO_2 glass plate as 50 μm and 250 μm, respectively, through the repetitive calculation considering the dielectric constants of $LiNbO_3$ crystal ($\varepsilon_{rx} = \varepsilon_{ry} = 43$, $\varepsilon_{rz} = 28$) and SiO_2 glass ($\varepsilon_r^{glass} = 4$) in the 60 GHz band.

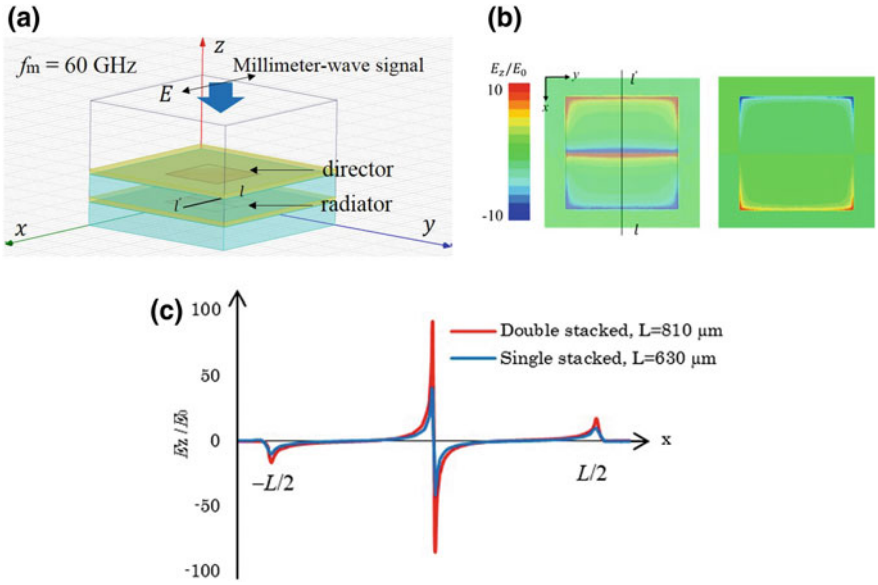


Fig. 2. **a** Analysis model. **b** Electric field distributions on the surface of the lower patch metal (Radiator) and the upper patch metal (Director). **c** Electric field distribution along the line of *l-l'* on the lower patch metal (Radiator). Electric field distribution for the single-stacked case with the same operational frequency is also plotted for comparison. © 2017 IEEE, Reprinted, with permission, from [17]

Figure 2b, c shows the calculated distributions of the normalized surface electric field of E_z/E_0 on the patch metals with the proposed stacked structure when an *x*-polarized 60 GHz plane-wave signal is irradiated to the device. We can see that the electric field distributions are mutually out of phase between the upper and lower patches. This is the same as the behavior of the Yagi–Uda antenna. Figure 2c shows the electric field distribution along the line of *l-l'* on the lower patch metal film shown in Fig. 2a, b. We can also see that a strong electric field is induced across the gap owing to a displacement current to satisfy the current continuity; this gap can be considered as a small capacitor for MMW signals and the surface electric current on the patch metal can pass through this capacitor as the displacement current. The peak value of the induced electric field is about 100 times of the irradiated field from above.

In the analysis, we found an interesting characteristic in the stacked structure; both the $LiNbO_3$ film and the SiO_2 base plate should be stacked to obtain a stronger electric field. If there is no $LiNbO_3$ layer on the top of the upper layer in the patch metal stacked structure, the electric field induced across the gap for optical modulation is less enhanced. We also found another interesting result on the electric field enhancement by introducing a gap at the center of the patch metal; the electric field induced across the gap for optical modulation become stronger when the gap is set at both the upper and lower patch metal.

Through the repeated calculations, we decided the patch metal size as $L = W =$ 810 μm with a 5 μm gap at the center along the *y*-axis for the operation at 60 GHz.

Therefore, utilizing the stacked structure, the surface electric field on the patch surface is increased, and patch length becomes longer for the same resonance frequency. As a result, the conversion efficiency from wireless MMW to LW signals is expected to be improved.

The proposed MMW-LW converter can be considered as an optical phase modulator driven by the wireless MMW signal from the photonic device point of view. Therefore, the optical modulation index is an important figure of merit to discuss its performance, which is calculated by using the following equation [14].

$$\Delta\theta = \frac{\pi n_e^3 r_{33}}{\lambda} \Gamma \int_0^L E_0 \sin\left(n_g k_m y + \phi\right) dy \tag{1}$$

where r_{33} is the Pockels coefficient of $LiNbO_3$, n_e is the refractive index for the extraordinary ray in $LiNbO_3$, Γ is the overlapping factor between the MMW electric field distribution across the gap and the LW electric field distribution in the cross section of the optical waveguide in the x–z plane, L is the patch metal length along the optical waveguide (the y-axis), E_z is the MMW electric field amplitude, n_g is the LW group index in the optical waveguide, k_m is the MMW wave number in vacuum, and ϕ is the initial phase of the MMW field when the LW is input at the end of patch antenna.

The calculated frequency responses of the optical modulation index in the single/double-stacked structures with the same patch size (L = 810 μm) are shown in Fig. 3. This optical modulation index value corresponds to the MMW-LW conversion efficiency and its square is proportional to the optical sideband power level in the converted LW signal [14]. We can see that by using the double-stacked structure, the peak frequency is shifted to the higher range by 7 GHz, and the peak optical modulation index value is improved by 3.2 dB. This means that optical modulation sideband power level will be enhanced by 6.4 dB with the double-stacked structure.

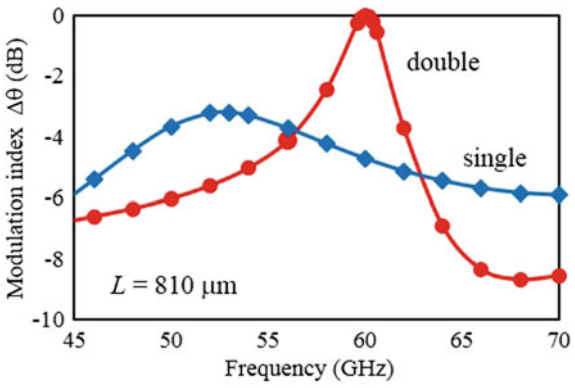


Fig. 3. Calculated frequency responses of MMW-LW conversion in the single and double-stacked converters with the same patch size. © 2017 IEEE, Reprinted, with permission, from [17]

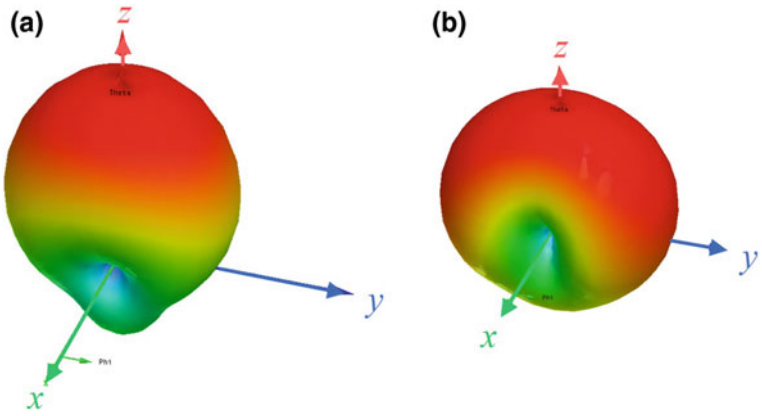


Fig. 4. Calculated antenna radiation patterns (f_m= 60 GHz). **a** Double and **b** single-stacked structures

Figure 4 shows the calculated antenna patterns of the single and double-stacked structures as MMW antennas. By using the double-stacked structure, the antenna gain is enhanced by ∼3 dB compared to the single-stacked case. The antenna patterns of the single and double-stacked structures in the *x*–*z* plane are also plotted in Fig. 5. We can see the antenna gain enhancement clearly.

In addition, the antenna pattern in the *y*–*z* plane can be controlled by introducing the array structure as shown in Fig. 1c. The angular dependence of the MMW-LW signal conversion in the *y*–*z* plane in the arrayed case (array factor) can be calculated by using the following equation assuming the linear array with the element number of *N* and the same distance of *d* between the adjoining antenna elements along the *y*-axis.

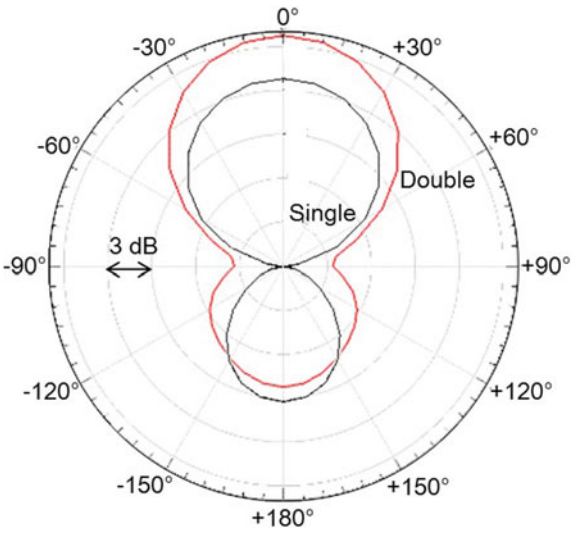


Fig. 5. Calculated antenna radiation patterns in the *x*–*z* plane (f_m= 60 GHz)

$$\Delta\theta_{array} = \frac{\pi n_e^3 r_{33}}{\lambda}\Gamma \times \sum_{h=0}^{N-1}\int_{hd}^{hd+L} E_0 \sin\{n_g k_m y - n_0 k_m h\, d \sin\theta + \phi\}dy \tag{2}$$

Examples of the calculated angular dependence are indicated in Fig. 6, where the MWW frequency f_m = 60 GHz, the element number N = 4, and the antenna distance d = 2.04, 2.28, and 2.59 mm. We can see that by changing the antenna distance, the peak conversion angle can be controlled. For a prototype device, we selected the antenna distance value as d = 2.28 mm.

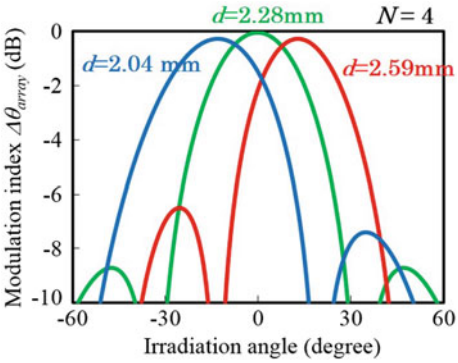


Fig. 6. Calculated angular dependence in the y–z plane of the MMW-LW signal conversion in the array of the double-stacked-patch antennas. (f_m = 60 GHz, N = 4, d = 2.04, 2.28, 2.59 mm)

The parameters of the designed MMW-LW signal converter are summarized in Table 1.

4 Device Fabrication

Based on the analysis results, the proposed device was designed and fabricated. The operational MMW frequency was set at 60 GHz and the operational LW wavelength was set at 1550 nm. A 250-μm-thick crystal plate of z-cut congruent $LiNbO_3$ was adopted as an initial EO material substrate, and this crystal plate was polished to 50 μm in the fabrication process by use of the MCP technique [14]. A 250-μm-thick low-dielectric-constant plate of fused SiO_2 glass was adopted as a base substrate. The device fabrication processes are essentially the same as our previous device described in [14].

Figure 7 shows the photograph of the fabricated devices just before the final Al metal deposition on the reverse side of the substrate for the ground electrode.

Table 1. The parameters of the designed MMW-LW converter

Operational frequency, f_m	60 GHz
EO crystal film	z-cut $LiNbO_3$
Thickness	50 μm
Dielectric constant components	(43, 43, 27)
Low-dielectric-constant material	fused SiO_2 glass
Thickness	250 μm
Dielectric constant	4
Square patch antenna	Al
Thickness	1 μm
Size	810 μm × 810 μm
Gap width at the center	5 μm
Number of vertical stacking	2 (double stack)
Bonding method for stacking	UV adhesive
Thickness of bonding layer	1.2 μm
Dielectric constant	4
Array antenna structure	Linear array
Number of elements	4
Distance between elements, d	2.28 mm
Peak conversion angle at 60 GHz	0 degree (normal)
Operational light wavelength, λ	∼ 1550 nm
Optical waveguide (APE waveguide)	
Core width	∼ 3 μm
Core depth	∼ 2 μm
Optical buffer layer (SiO_2 sputtered)	0.2 μm

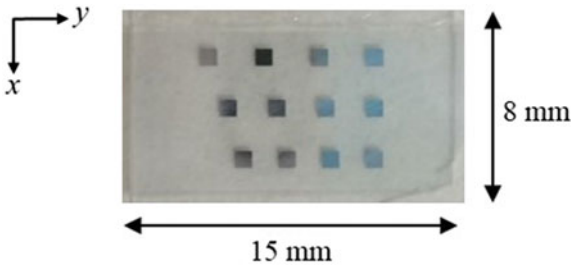


Fig. 7. Photograph of the fabricated double-stacked converter. © 2017 IEEE, Reprinted, with permission, from [17]

5 Basic Experiments

The performance of the fabricated EO MMW-LW converter was tested using the experimental set up shown in Fig. 8. An MMW signal in the 60 GHz band was amplified and irradiated to the fabricated device through a standard gain horn antenna of 19 dBi. The MMW power fed to the horn antenna was 10 mW. An LW from a DFB laser was coupled to the optical waveguide as a TM fundamental mode through a microscope objective lens, and the optical spectrum of the output LW was measured by use of a standard optical spectrum analyzer. Typical examples of the measured optical spectra from the fabricated devices with the single/double-stacked structures are shown in Fig. 9. Clear modulation sidebands were observed with the radiation of merely 10 mW MMW signals.

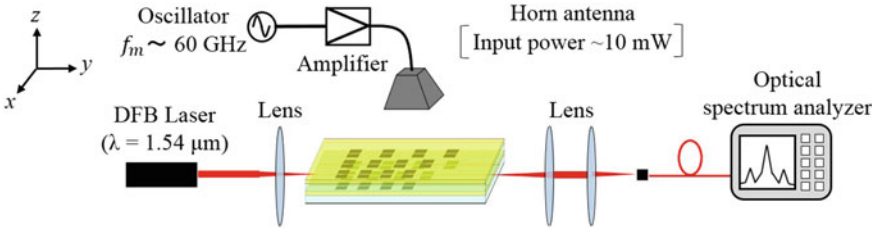


Fig. 8. Experimental set up. © 2017 IEEE, Reprinted, with permission, from [17]

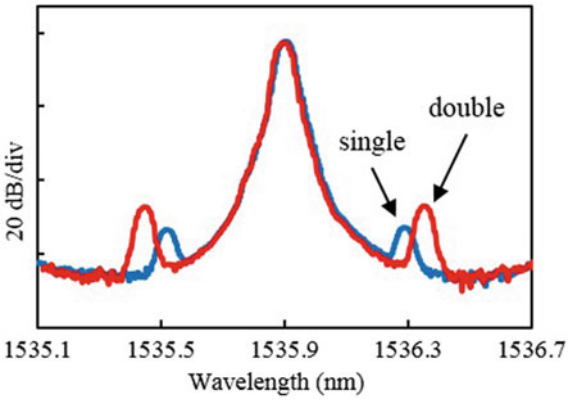


Fig. 9. Examples of the measured optical modulation spectra from the single/double-stacked converters. © 2017 IEEE, Reprinted, with permission, from [17]

Figure 10 shows the measured frequency response of the single and double-stacked converters. By using the double-stacked structure, the peak conversion frequency was shifted to the higher range by ~8 GHz, and the ratio between the optical carrier and first modulation sideband components was increased by 6 dB, which corresponds to

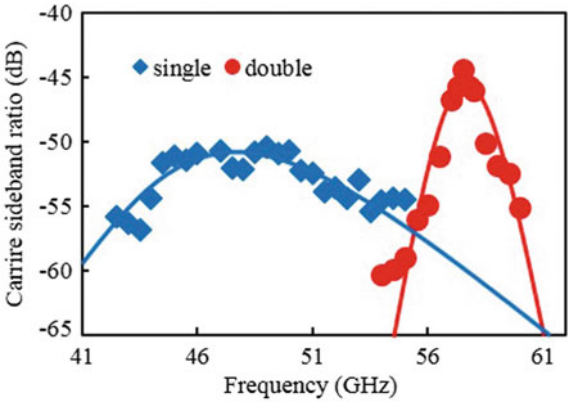


Fig. 10. Measured frequency responses of the single and double-stacked MMW-LW converters. © 2017 IEEE, Reprinted, with permission, from [17]

the increase in the optical modulation index by 3 dB. These results were in good agreement with the calculated characteristics shown in Fig. 3.

Figure 11 shows the measured directivities in the MMW-LW signal conversion of the single and double-stacked converters. The measured frequency was at f_m = 57.5 GHz for the double-stacked case and at f_m = 50 GHz for the single-stacked case. The peak conversion angle of the double-stacked converter was in good agreement with the theoretical result shown in Fig. 6.

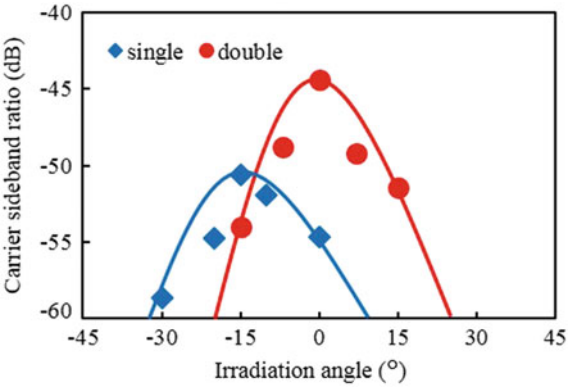


Fig. 11. Measured directivities in the MMW-LW signal conversion of the single and double-stacked converters. (N = 4 and d = 2.28 mm) The measured frequency was at f_m = 57.5 GHz for the double-stacked structure and at f_m = 50 GHz for the single-stacked structure. © 2017 IEEE, Reprinted, with permission, from [17]

6 Data Transfer Experiment

By using the fabricated double-stacked MMW-LW converter, wireless data transfer experiments over an optical fiber were also carried out. The experimental set up for the data transfer is shown in Fig. 12. A wireless 60 GHz band QPSK signal with a repetition frequency of 10 MHz was generated by use of the photonic two-tone technique [18] and was irradiated to the double-stacked converter from above. The irradiated MMW signal power level was about +3 dBm. Then, the MMW signal was converted to an LW signal by using the fabricated MMW-LW signal converter. The converted LW signal was detected by use of a high-speed photodiode after being transferred over 1-km-long-standard silica fiber. Figure 13 shows the measured MMW spectrum and constellation diagram of the detected signal by use of the MMW spectrum/signal analyzer, Rohde and Schwarz FSW67. A QPSK constellation was successfully identified from the detected signal. The measured EVM value was 28.7%. Therefore, the signal conversion and transfer over fiber by using the double-stacked MMW-LW converter were successfully demonstrated.

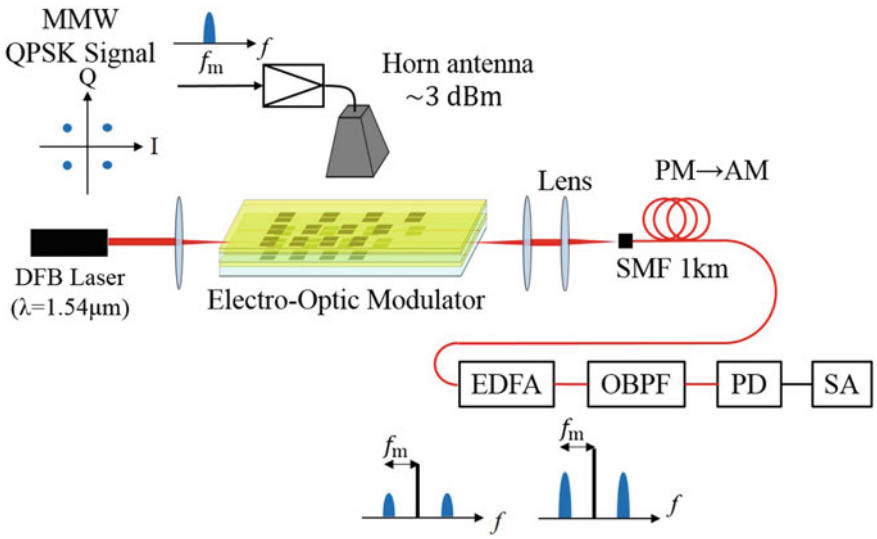


Fig. 12. Experimental set up for the data transfer over a 1-km-long-standard single-mode silica optical fiber

7 Conclusion

We have proposed and developed the new wireless MMW-LW signal converter based on an EO modulator using stacked planar patch antennas embedded with a narrow gap. By stacking patch antennas and substrates with the same size vertically, the length of the patch metal and the antenna gain become larger when the resonance frequencies

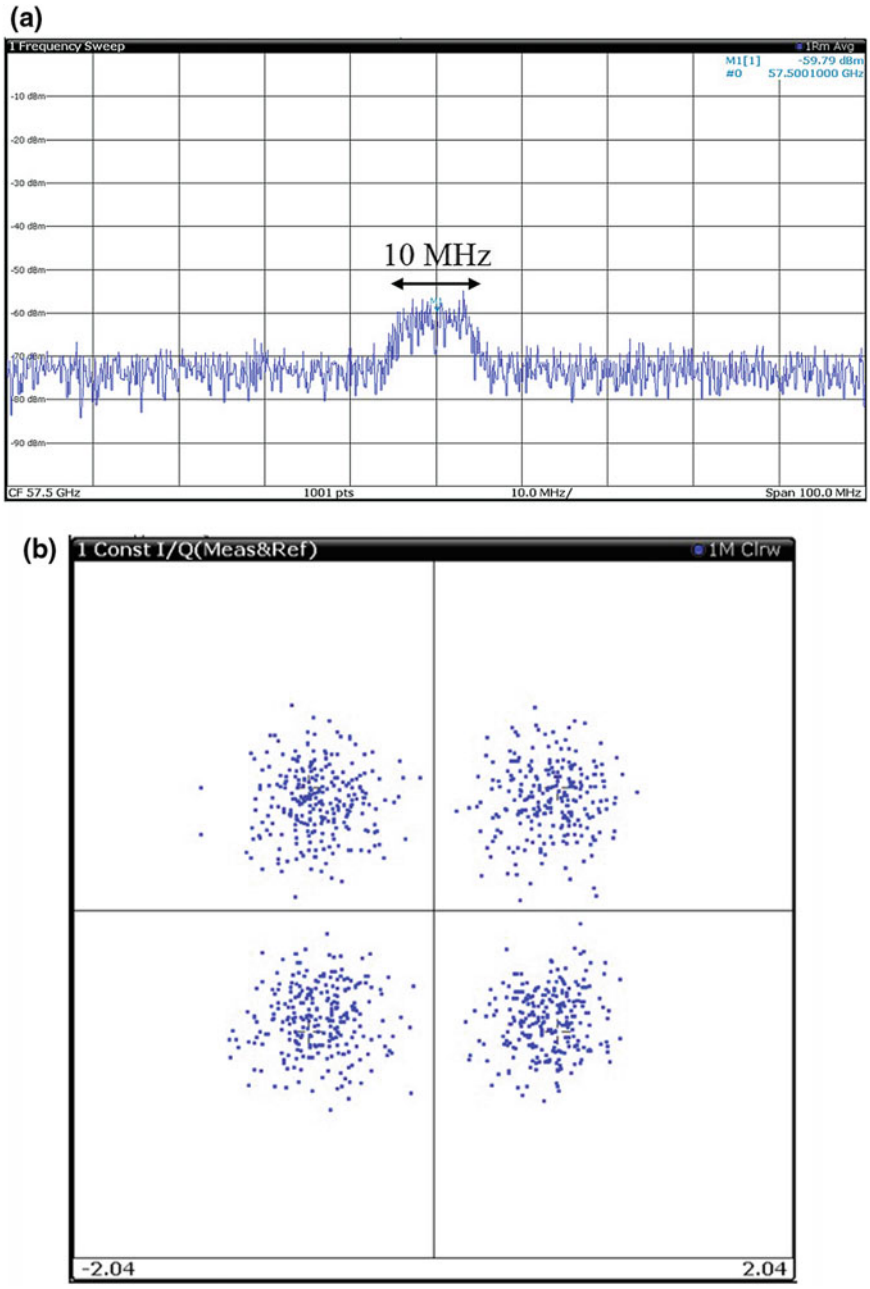


Fig. 13. Measured MMW spectrum (**a**) and constellation diagram (**b**) of the detected QPSK signal after the LW signal conversion and 1 km fiber transmission. The signal repetition frequency was 10 MHz. The EVM value of the detected signal was 28.7%

are tuned to the same value. As a result, the induced electric field across the narrow gap for optical modulation is considerably enhanced and the conversion efficiency from MMW to LW is improved, which was verified in the experiment. By stacking multiple (~ 10) patch antennas with multiple substrates vertically [19, 20] and by increasing the number of antenna elements in the array, further improvement in the conversion efficiency as the carrier/sideband ratio over 10 dB or enlargement of the operational frequency bandwidth over 10 times is expected [19, 20]. Therefore, the proposed MMW-LW converter will be able to apply high-resolution MMW distributed radar systems combined with radio-over-fiber links to monitor foreign debris on an airport runway.

Acknowledgements. The authors thank Dr. Hidehisa Shiomi and Prof. Yasuyuki Okamura of Osaka University, Japan for their helpful advice and discussion.

This research results were achieved in part by the research project of "Radio technologies for 5G using Advanced Photonic Infrastructure for Dense user environments" (RAPID), the Commissioned Research of National Institute of Information and Communications Technology (NICT), Japan.

References

1. Appleby R, Anderton RN (2007) Millimeter-wave and submillimeter-wave imaging for security and surveillance. Proc IEEE 95(8):1683–1690
2. Steinhauer M, Ruob H-O, Irion H, Menzel W (2008) Millimeter-wave-radar sensor based on a transceiver array for automotive applications. IEEE Trans Microw Theor Tech 56(2): 261–269
3. Jose E, Adams M, Mullane JS, Patrikalakis NM (2010) Predicting millimeter wave radar spectra for autonomous navigation. IEEE Sens J 10(5):960–971
4. Chin-Lin I, Rowell C, Han S, Xu Z, Li G, Pan Z (2014) Toward green and soft: a 5G perspective. IEEE Com Mag 52(2):66–73
5. Agyapong PK, Iwamura M, Staehle D, Kiess W, Benjebbour A (2014) Design considerations for a 5G network architecture. IEEE Com Mag 52(11):65–75
6. Dahlman E, Mildh G, Parkvall S, Peisa J, Sachs J, Selén Y, Sköld J (2014) 5G wireless access: requirements and realization. IEEE Com Mag 52(12):42–47
7. Capmany J, Novak D (2007) Microwave photonics combines two worlds. Nat Photon 1:319–330
8. Seeds AJ, Williams KJ (2006) Microwave photonics. IEEE/OSA J Lightwave Technol 24:4628–4641
9. Gomes NJ, Monteiro PP, Gameiro A (2012) Next generation wireless communications using radio over fiber. Wiley
10. Wake D, Nkansah A, Gomes NJ (2010) Radio over fiber link design for next generation wireless systems. IEEE/OSA J Lightwave Technol 28(16):2456–2464
11. Murata H, Miyanaka R, Okamura Y (2012) Wireless space-division-multiplexed signal discrimination device using electro-optic modulator with antenna-coupled electrodes and polarization-reversed structures. Int J Microwaves Wireless Technol 4(3):399–405
12. Wijayanto YN, Murata H, Okamura Y (2012) Electro-optic microwave-lightwave converters utilizing a quasi-phase-matching array of patch antennas with a gap. Electron Lett 48(1): 36–38

13. Kohmu N, Murata H, Okamura Y (2013) Electro-optic modulator using $LiNbO_3$ waveguide suspended to double-antenna-coupled electrodes on low-k substrate for millimeter-wave-lightwave signal conversion. PA-6, APMP, Gwangju, Korea
14. Wijayanto YN, Murata H, Okamura Y (2013) Electro-optic millimeter-wave-lightwave signal converters suspended to gap-embedded patch antennas on low-k dielectric materials. IEEE J Selected Top Quant Electron 19(6):3400709
15. Kanno A, Umezawa T, Kuri T, Yamamoto N, Kawanishi T, Wijayanto YN (2016) Key technologies for millimeter-wave distributed RADAR system over a radio over fiber network. In: 2016 international conference on radar, antenna, microwave, electronics, and telecommunications (ICRAMET), Jakarta, Indonesia
16. Wijayanto YN, Kanno A, Murata H, Kawanishi T, Adhi P (2017) W-band millimeter-wave patch antennas on optical modulator for runway security systems. 2017 IEEE conference on antenna measurements & applications (CAMA), MB2.1. Tsukuba, Japan, pp 79–82
17. Murata H, Aya H, Inoue T, Sanada A, Wijayanto YN, Kanno A, Kawanishi T (2017) Millimeter-wave electro-optic modulator using stacked patch-antennas embedded with micrometer-gap for radio-over-fiber system. In: 2017 47th European microwave conference (EuMC), EuMC30–4, Oct. 2017, Nuremberg, Germany
18. Inoue T, Ikeda K, Kakubari Y, Yonemoto N, Shibagaki N, Toda H, Murata H (2016) Millimeter-wave wireless signal generation and detection using photonic technique for mobile communication systems. AVFOP/MWP 2016, TuM1.8, Oct. 2016, CA, USA
19. Kramer O, Djerafi T, Wu K (2011) Very small footprint 60 GHz stacked yagi antenna array. IEEE Trans Antennas Propag 59(9):3204–3210
20. Kramer O, Djerafi T, Wu K (2010) Vertically multilayer-stacked yagi antenna with single and dual polarizations. IEEE Trans Antennas Propag 58(4):1022–1030

Multiuser 5G Hot-Spots Based on 60 GHz Leaky-Wave Antennas and WDM Radio-over-Fiber

M. Steeg[1(✉)], M. Szczęsny[2], and A. Stöhr[1]

[1] Center for Semiconductor Technology and Optoelectronics (ZHO), University of Duisburg-Essen, Duisburg, Germany
{matthias.steeg, andreas.stoehr}@uni-due.de
[2] Office of Architecture and Network Planning, Exatel, Warsaw, Poland
michal.szczesny@exatel.pl

Abstract. This paper reports on photonic multiuser 5G hot-spots, that provide multiple beams by employing frequency steerable leaky-wave antennas. Therefore, a 60 GHz periodic leaky-wave antenna (LWA) has been developed, based on low loss substrate-integrated waveguide (SIW) technology, which is fabricated through standard PCB processes. The developed LWA operates in the V-band between 50 and 70 GHz and provides over 40° beam steering in the H-plane via frequency scanning. The capability for beam steering and multiple simultaneous beams from only one feeding port is combined with Radio-over-Fiber (RoF) techniques to provide a simple, compact and low-cost system for new applications in mm-wave communications. The proposed system enables centralized photonic beam steering in fiber-wireless transmission links, where up to 6 Gbit/s data rates are demonstrated using 64 QAM with IF-OFDM and simple receiver architectures. The multibeam capabilities provide a strong addition to RoF links by utilizing dense WDM channels to support multiple wireless users. Thereby, multiple low latency and high data rate wireless services can be provided via a single fiber-fed antenna. Finally, this concept is demonstrated by lab experiments, where three 1 Gbit/s OOK data signals were simultaneously transmitted and by a week-long field trial in a shopping mall, where two 1.5 Gbit/s real-time SDI video streams were transmitted.

Keywords: 5G mobile communications · Antennas · Beam steering · Millimeter-wave radio · Radio-over-Fiber

1 Introduction

The advent of the next generation of mobile communications 5G is expected to bring a revolutionary 1000-fold increase in mobile data traffic and a substantially larger number of connected users per cell [1]. In order to sort the various use cases to be enabled by the new technology, the ITU-R has proposed three use case families for 5G [2]. One of them is “enhanced mobile broadband” (eMBB), which is most closely

Electronic Navigation Research Institute, *Air Traffic Management and Systems III*, Lecture Notes in Electrical Engineering 555,
https://doi.org/10.1007/978-981-13-7086-1_16

associated the overall increase in mobile data throughput. The eMBB family also covers the "5G Hot-Spot" scenario, which embraces all those use cases, where a large number of users in a limited area demand high data rate services [2, 3]. Thereby, applications include scenarios, where users wish to quickly download or access multimedia content while waiting or traveling, such as in an airport, a train station or on an airplane, a train. In the modern networking society, the use cases can further be expanded to office spaces, cafes or popular meeting places.

The EU-Japan project RAPID is researching technical solutions to support such high-capacity hot-spot scenarios especially for dense user areas, where the requirement for mobility is low but the user data rates are expected to be extremely high [4]. RAPID especially considers the use case scenarios of crowded infrastructure such as airplanes, train stations, motorways, etc., as well dense user areas in office buildings, shopping areas and malls and sports stadiums, where dedicated services are expected. For achieving this goal RAPID technology employs photonic techniques to exploit the large bandwidth at mm-wave frequencies. In order to demonstrate its 5G technology field trials in natural dense user environments in a football stadium and a shopping mall are designated.

The cornerstones to support the projected, significant increase in data traffic for 5G are seen in larger bandwidth, higher spectral efficiency and more, smaller cells [2, 5, 6]. Thereby, the large available spectrum at mm-wave frequencies, e.g., in the 60 GHz band can provide the necessary bandwidth for 5G communications and is thus considered as an enabling technology [3, 5, 6]. While the utilization of the mm-wave spectrum also brings new challenges, such as high free space path loss, innovative solutions are sought. One of the key demands for providing mm-wave mobile access is an antenna, which has a high gain to maintain a manageable link budget, while being steerable to support user mobility [5, 6]. In RAPID we further require an antenna, that supports multiple beams to provide uninterrupted, low-latency service to multiple users at the same time in densely populated environments.

In view of the large available bandwidth, but also the challenges that mm-waves provide for conventional phased array antenna techniques [6], we propose the utilization of leaky-wave antennas as a low-cost technology for 5G hot-spots. While the efforts to implement beam forming and beam switching solutions at 60 GHz have led to technological solutions, they either only provide a single steerable beam [7] or require multiple antenna feeds [8]. In contrast to that LWAs provide a simple mechanism for beam steering by frequency scanning [9], which requires only a single feeding port, can be implemented inexpensively and is compatible to FDMA techniques. Furthermore, they allow for new direction-of-arrival (DoA) estimation techniques, which are important for user localization in multiuser scenarios [10]. While LWAs are well known, e.g., in avionics and for radar applications, we want to illustrate their application for fiber-wireless communications. We will show that they can complement Radio-over-Fiber (RoF) techniques, which are often proposed because of their capability for centralizing network features. This architecture is depicted in Fig. 1, where a SIW LWA antenna in implemented in a WDM fiber-wireless communication system for multiuser support.

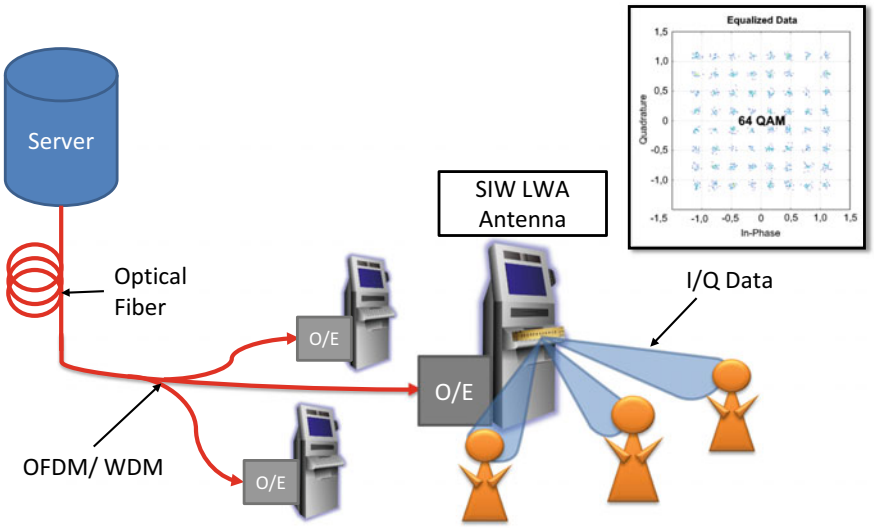


Fig. 1. Schematic architecture of an optical distribution network connecting a server to multiple 5G hot-spots, which can each deliver data streams to multiple users employing the presented SIW LWA

In this paper, the implementation of 5G hot-spots by the means of frequency steerable 60 GHz multibeam antennas is outlined. Therefore, leaky-wave antennas are employed in conjunction with dense WDM optical networks, which are extended via RoF wireless links. At first the utilized antenna structure and its performance is illustrated. After that the setup of the fiber-wireless communication link is described and RoF data transmission experiments are presented. Then the option to serve multiple wireless users via the connected optical WDM network explained and demonstrated by measurements. Finally, the demonstration of a 5G hot-spot is presented, which was employed to deliver two uncompressed video streams for a field trial in a shopping mall, before the paper is concluded.

2 Utilized SIW LWA Antenna Structure

2.1 Leaky-Wave Antenna Principle

In this work, periodic leaky-wave antennas (LWAs) are utilized, which are based on substrate-integrated waveguides (SIWs). In general, LWAs are waveguide structures that are modified to leak power for antenna operation. For periodic LWAs the waveguide is divided into unit cells, which are arrayed and periodically modified. Thus, they radiate similar to a linear array, with a beam angle that depends on the phase difference between the radiating elements of the arrayed unit cells. This phase shift between adjacent unit cells depends on the complex propagation constant of the modified waveguide [9]. Thereby, good power acceptance, simple feeding via the waveguide interface and a high directivity due to the array factor can be achieved.

Furthermore, as the phase difference between the periodic radiating elements of the LWA depends on the frequency according to their dispersion characteristic, a LWA scans its beam angle with changing the frequency. This means, that an LWA will radiate signals at different radio frequencies (RFs) into different directions, while only requiring a single RF feeding port [9–11].

2.2 Implemented SIW LWA

The specific structure of the used LWA has been proposed in [11]. It has been manufactured from a RO5880 laminate by standard PCB processes. The LWA employs longitudinal half-wavelength microstrip lines inset to the SIW as radiating elements, which are off-centered to improve broadside radiation [11]. The fabricated laminate of such a LWA is depicted in Fig. 2.

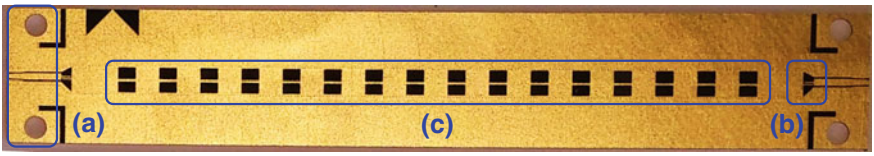


Fig. 2. Fabricated laminate of the SIW LWA containing the grounded CPW interface for the ELCs (**a**), the coupled line transition (**b**) and the periodic LWA array (**c**)

Apart from the periodic LWA array, the laminate contains a coupled line transition and holes and markers for the fixtures of the coaxial end launch connectors (ELCs) that are used to contact the antenna. Thereby, the coupled lines provide a transition from the SIW of the LWA to the grounded coplanar waveguide interface of the ELCs, while also providing impedance transformation to yield good matching. Since the coupled lines transition acts as a DC block they already provide one part of a bias-tee for the integration of photodiodes [12]. This gives the SIW LWA the potential to provide an integration platform for radio-over-fiber (RoF) photonic transmitters.

The presented SIW LWA operates in the V-band between 50 and 70 GHz, where it provides beam steering in excess of 40° in the H-plane. In Fig. 3 the radiation patterns at radio frequencies of 50 GHz and in the 57–66 GHz band are shown. It can be seen that the measured gain is around 14 dBi with ±1 dB deviation across the bandwidth with a sidelobe suppression of 13 dB. The SIW LWA is a linearly polarized antenna with a cross polarization discrimination (XPD) of 24.83 dB in the steering plane. The frequency scanning behavior of the SIW LWA can be observed from Fig. 3, demonstrating beam steering from –9.8° at 57 GHz through broadside around 61.5 GHz to 8.0° at 66 GHz.

Thus, the antenna beam can be steered, e.g., to follow a mobile user by changing the radio frequency. The four radiation patterns at different frequencies in Fig. 3 also illustrate that the SIW LWA can generate multiple beams to support multiple users simultaneously, which is demonstrated in the following sections. Therefore, different data signals are radiated at different RFs to be received by multiple users at different positions.

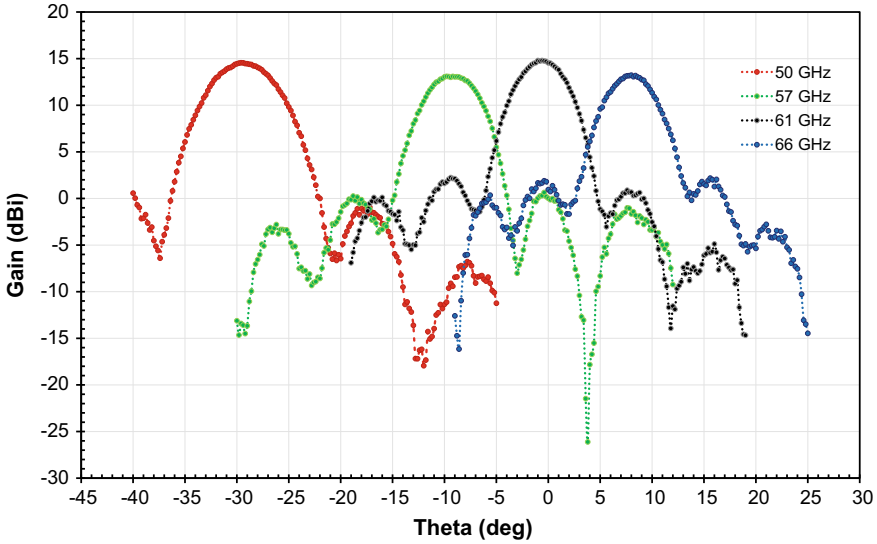


Fig. 3. Radiation pattern of the SIW LWA as the measured gain over the beam angle theta at radio frequencies of 50 GHz (red), 57 GHz (green), 61 GHz (black) and 66 GHz (blue)

2.3 Application of LWAs for 60 GHz Communications

Due to the large overall bandwidth and lack of necessary electrical control signals, the SIW LWA synergizes well with RoF systems, which utilize the vast optical bandwidth. In this paper it is employed for fiber-wireless communications, where it creates unique possibilities for centralized photonic beam steering to a coherent RoF link. With this approach, which is detailed in the next section, the RF of the radiated data signals is obtained from the optical frequencies of the LO and data signals. A transparent link is established from the data source to the data sink and the beam direction can be controlled via the laser frequency. It should be mentioned that for a full 5G communication system, besides beam steering it is also necessary to localize the mobile users. In this regard LWAs also provide new solutions for DoA estimation for multiple users [10]. While we focus on RoF communication with the presented SIW LWAs in this paper, we have also developed a simple technique for estimating the DoA of a user requesting service from the hot-spot. Therefore, a frequency upchirp signal send by the user, and often used for the preamble of OFDM waveforms, can be employed to choose the correct beam angle based on the LWA response [13].

3 Fiber-Wireless Communications

3.1 Employed Radio-Over-Fiber Approach

Radio-over-Fiber is a straightforward approach to connect remote antenna station to an optical fiber network by directly transmitting the wireless signals. Usually three

different analogue RoF schemes are differentiated based on where the baseband data signals are upconverted to the RF [14]. Thereby, it is assumed that always electrical mixers are used, which are located either at the central unit (CU) or at the radio access unit (RAU). In contrast to that, we have used a coherent RoF approach, where the baseband or intermediate frequency (IF) signal, generated at the CU, is photonically upconverted at the RAU by employing an optical local oscillator (LO) [15]. This approach is employed in conjunction with the introduced SIW LWA antennas for fiber-wireless data transmission with the setup illustrated in Fig. 4. The combination of optical upconversion with the LWA enables unique photonic beam steering capabilities, as outlined in the following. At the CU a laser generates an optical carrier, which is modulated by the data signal by means of a Mach–Zehnder modulator (MZM). Then the signal is amplified by an optical amplifier and transmitted over 10 km standard single mode fiber (SMF), emulating the optical distribution network. After the transmission, the signal arrives at the RAU, where an optical local oscillator (LO) from a tunable laser diode is added. Then a high-speed photo diode generates the RF signal, containing the data, through heterodyne detection of the arriving optical carrier from the data signal and the LO. As the RF corresponds to the beat frequency of the two optical waves, the frequency can be changed by tuning either the laser frequency of the optical carrier or of the LO. The RF signal is then amplified and radiated by the presented SIW LWA. Since the beam angle of the LWA can be changed through the RF, the radiated signal can be steered by tuning either the signal or the LO laser. This means that the system provides photonic beam steering functionality either locally at the RAU or centralized from the CU. After the wireless transmission, the RF signal is received by the mobile unit (MU) by means of a horn antenna. The received signal is then amplified and downconverted by a zero-biased Schottky barrier diode detector (SBD). As the SBD works as an envelope detector, it does not depend on the RF of the

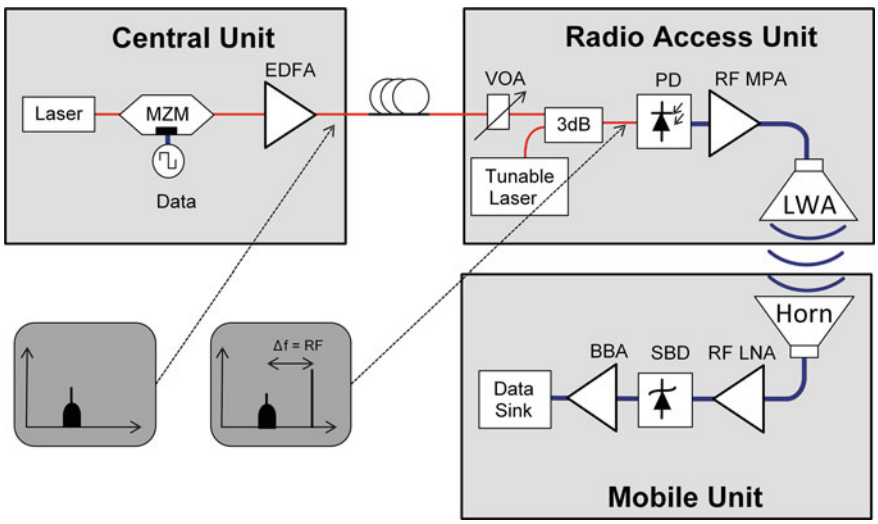


Fig. 4. Schematic architecture of the utilized fiber-wireless communication system

received signal and no electrical LOs are required at the MU. Furthermore, it is insensitive to RF drift and phase noise, which means that phase-locking of the lasers is not necessary. Finally, the downconverted baseband signals are amplified and then analyzed at the designated data sink.

3.2 Single Link OOK and IF-OFDM Measurements

At first single link transmission experiments using the outlined setup from Fig. 4 have been carried out. Therefore, a PRBS-9 (pseudo random binary sequence) baseband data signal is generated by a pulse pattern generator (PPG) at the CU and modulated onto the optical carrier. To produce bit error rate (BER) curves, an error detector acts as the data sink to analyze the received signal. The optical power is changed by a variable optical attenuator (VOA) to yield different signal-to-noise (SNR) values at the receiver. The measured BERs of 2 Gbit/s OOK data signals after wireless transmission over 2 m are presented in Fig. 5. As can be seen quasi error-free transmission is achieved and the difference in required optical power is low across the different radio frequencies from 57 GHz over 60 to 63 GHz. This means that the link budget is flat across the 60 GHz band, even though different beam angles are used.

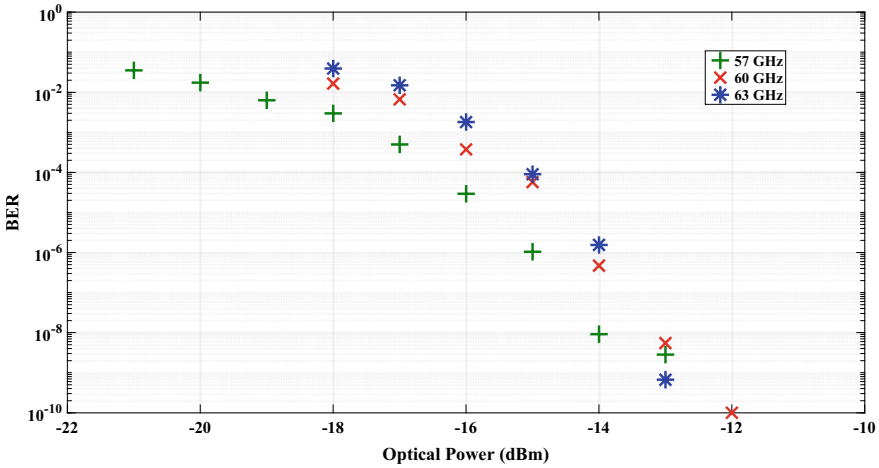


Fig. 5. Measured bit error rate over optical power of 2 Gbit/s OOK signals, which are transmit over 2 m wireless distance at 57 GHz (green), 60 GHz (red) and 63 GHz (blue)

The data rate of the RoF link can be further increased by employing higher order modulation techniques. In order to employ QAM modulation, while retaining the simple SBD receiver, IF-OFDM techniques can be employed [16, 17]. Therefore, QAM modulated OFDM channels are upconverted to an IF before being modulated onto the optical carrier. This way the SBD downconverts the received RF signal to the IF, so that the I/Q information is conserved. By using a digital sampling oscilloscope to analyze the data, 0.5 m wireless transmission with a throughput of 6 Gbit/s could be achieved with 64 QAM modulation at a bandwidth of 1 GHz. An IF of 1.5 GHz was

used to avoid signal-signal beat interference (SSBI) and an average SNR per subcarrier of 22.33 dB was achieved [17]. The constellation with an EVM of 7.65% can be seen in the inset of Fig. 1.

3.3 Optical WDM for Wireless Multiuser Support

In the previous measurements the data streams at different RFs and beam angles have been transmitted and analyzed one after the other. Now we want to serve multiple users simultaneously. Therefore, optical wavelength division multiplexing is used, i.e., three optical channels are used to provide data for three users. In this case, three signal lasers operating at different optical frequencies are each modulated by a MZM with individual data from different PRBS sources. These three signals each carry 1 Gbit/s OOK data and are generated with 3 GHz ultra-dense optical channel spacing. By utilizing optical heterodyning to generate the RF channels, all three wireless signals can be generated by one PD. The 3 GHz difference between the laser frequencies translates into an equal difference between the generated RFs and so the channel spacing is preserved. This means that in contrast to a conventional RoF system no optical filters are necessary [15]. Because of the ultra-dense spacing, the optical signals occupy only a small bandwidth of 8 GHz in total. Furthermore, the RAU can remain unchanged as the PD and the LWA antenna have sufficient bandwidth to generate and radiate all three signals. The LO laser is set to a wavelength 62 GHz away from the median channel, so that the wireless signals are generated at RFs of 59, 62, and 65 GHz, all within the 60 GHz band. By means of the SIW LWA the signals are radiated at different beam angles of –3.5°, 2°, and 7.5°. They can be received after 2 m wireless transmission at distinct positions by the mobile units. Again, a simple SBD detector is used for downconversion. In contrast to an electronic mixer no additional local oscillator is required, and the mobile unit does not need to be aware of the received signal RF to detect its data. Thus, all RF signals are downconverted to baseband and can be received by simply moving into the respective beam angles without further changes at the mobile unit. So, when moving a mobile unit in the steering dimension, the received RF signal changes according to the relative angle to the RAU. The best signal quality is achieved at –3.5° for the 59 GHz signal, 2° for 62 GHz and 7.5° for 65 GHz, respectively. Within ±2° around these angles the signal quality gracefully decreases. For larger angles the signal quality drops rapidly, as the gain drops according to the radiation pattern and eventually the power of interfering signals increases. As the beam width in the E-plane is around 75°, the directivity is limited in the dimension orthogonal to the steering direction. Thus, the mobile unit can be moved up to 0.5 m orthogonal to the beam steering direction and the signals can still be received, when angling the horn antenna accordingly. The successful reception of three simultaneously transmitted 1 Gbit/s OOK signals, yielding a total data rate of 3 Gbit/s, is confirmed by the BER measurements. Like for the single link a VOA is utilized to change the signal power of the combined optical channels. The measured BER curves are plotted against the received RF powers of the detected signals at the mobile unit and are presented in Fig. 6. BERs of 3.805 E-3, 6.654 E-4 and 2.779 E-4 have been achieved at –37.79 dBm and 65 GHz, at –39.38 dBm and 62 GHz and at –40.95 dBm and 59 GHz, respectively. When the signal powers were further increased an error floor due to the

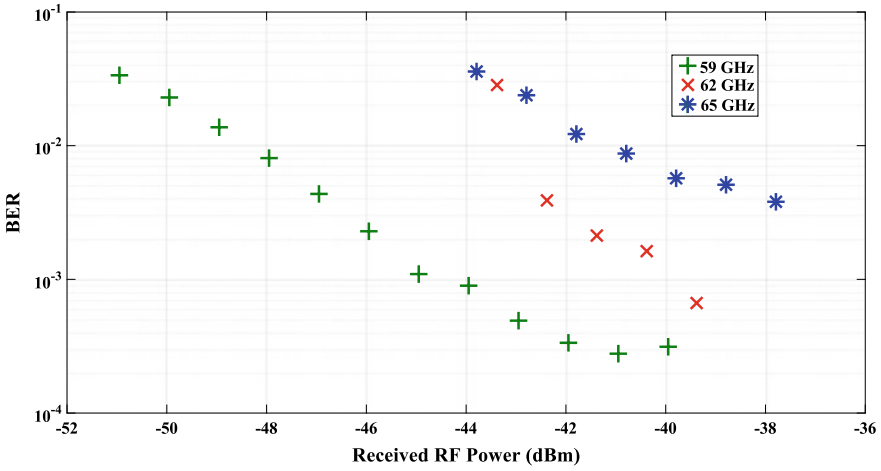


Fig. 6. Bit error rate over received electrical power of the three 1 Gbit/s OOK signals at 59 GHz (green), 62 GHz (red) and 65 GHz (blue) after 2 m simultaneous wireless transmission

interfering signals has been observed. As no filters are utilized and the SBD down-converts all incoming 60 GHz band RF signals to baseband, the other two signals interfere with the desired signal. Thereby, the system solely relies on the directivity of the SIW LWA to reduce undesired RF signals and maintain acceptable signal-to-interference-ratios (SIR).

4 5G Hot-Spot Demonstration for Video Streaming in a Shopping Mall

The system, described in the previous system has been utilized to implement a 5G hot-spot for a field trial demonstration in a shopping mall in Warsaw, Poland. Thereby, the 5G hot-spot supplies multiple users wirelessly with high data rate video streams. The field trial architecture contains transmission across four different locations as illustrated in Fig. 7: from a data center, where the optical carriers are modulated, over the shopping mall server room, where the optical LO is added, to the RAU, where the RF

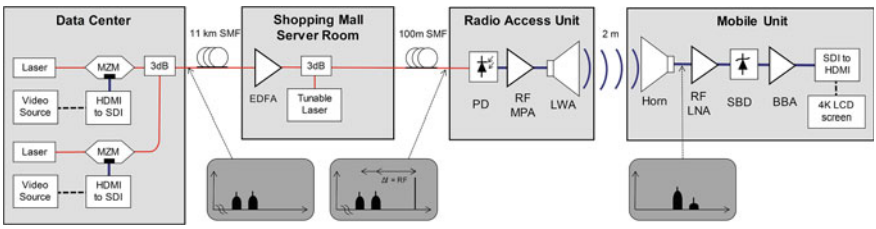


Fig. 7. Installed field trial architecture for the 5G video streaming demonstration in a shopping mall

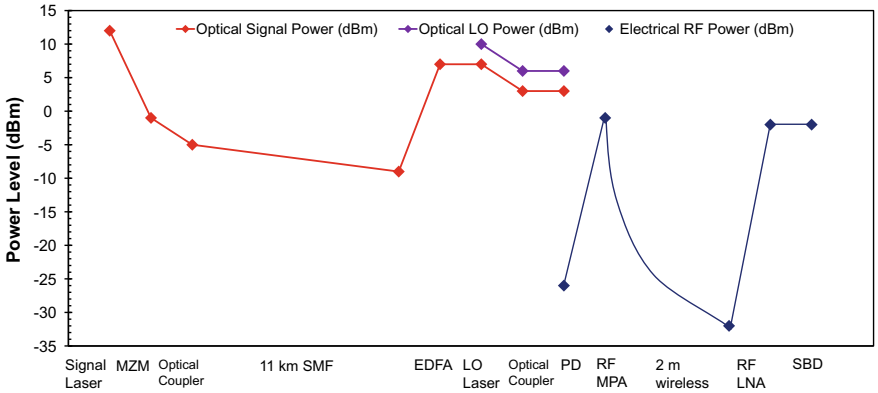


Fig. 8. Level diagram of the optical and RF link of the field trial system with indicated components

signals are generated and radiated to the mobile unit. An optical and RF level diagram is provided in Fig. 8. To demonstrate multiple data streams two optical carriers have been modulated with two different 1.5 Gbit/s SDI video signals by MZMs. SDI stands for serial digital interface and is a broadcasting standard for real time, uncompressed video transmission. The SDI signals are OOK baseband data streams with an amplitude level of 800 mV peak-to-peak. The SDI signals were generated via HDMI-to-SDI scalerboxes, which are commercial broadcast converters that generate the SDI signals from video stream input via an HDMI interface. The scalerboxes do not support buffering, an uninterrupted video stream confirms that the data stream is continuously available with a good signal quality. The central unit, which generates the two optical video signals was installed in the data center. From there the signals were transmitted within one WDM channel over 11 km SMF, installed in the field, to the server room of the shopping mall as depicted in Fig. 9. At the shopping mall the signals are extracted via a wavelength filter, amplified and combined with an optical LO signal. After that they are transmit over 100 m SMF to the hybrid integrated photonic RAU, depicted in Fig. 10. This RAU contains the high-speed PD, an RF amplifier (Gain: 25 dB; P_{RF} = –1 dBm) and the LWA (Gain: 14 dBi) and is ceiling mounted in the field trial area of the shopping mall. There the RAU establishes two wireless video streams by simultaneously radiating each of the two video signals at a distinct beam angle of –30° for the 50 GHz signal and –1.5° for the 60 GHz signal. After wireless transmission over 2 m they were received by the mobile unit. In this case the mobile unit is the demo vehicle, shown in Fig. 11. It contains a horn antenna (Gain: 23 dBi), RF LNA (Gain: 30 dB; NF: 4.5 dB), SBD for downconversion and baseband amplifier (BBA) as depicted in Fig. 7. The data sink of the mobile unit is another scalerbox, which converts the video signals from SDI back to HDMI to be displayed by a large 4 K LCD screen. Both video signals at 50 and 60 GHz could be successfully received by moving the demo vehicle to their respective beam angles. This was confirmed by the screen showing the high-quality video streams on the mobile unit. Therefore, an SDI amplitude of 250 mV peak-to-peak was received at the demo vehicle. The system was

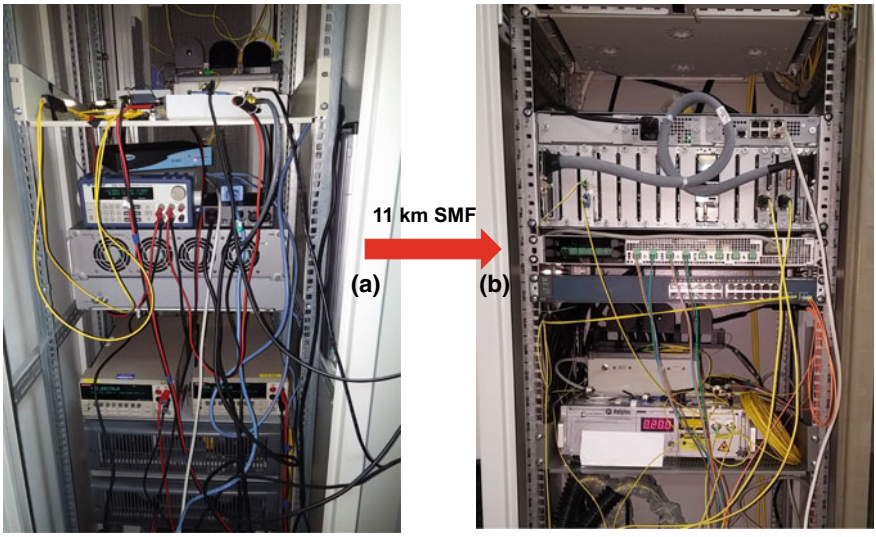


Fig. 9. Field trial installation in the data center (**a**) and in the shopping mall server room (**b**), which are connected by a 11 km SMF

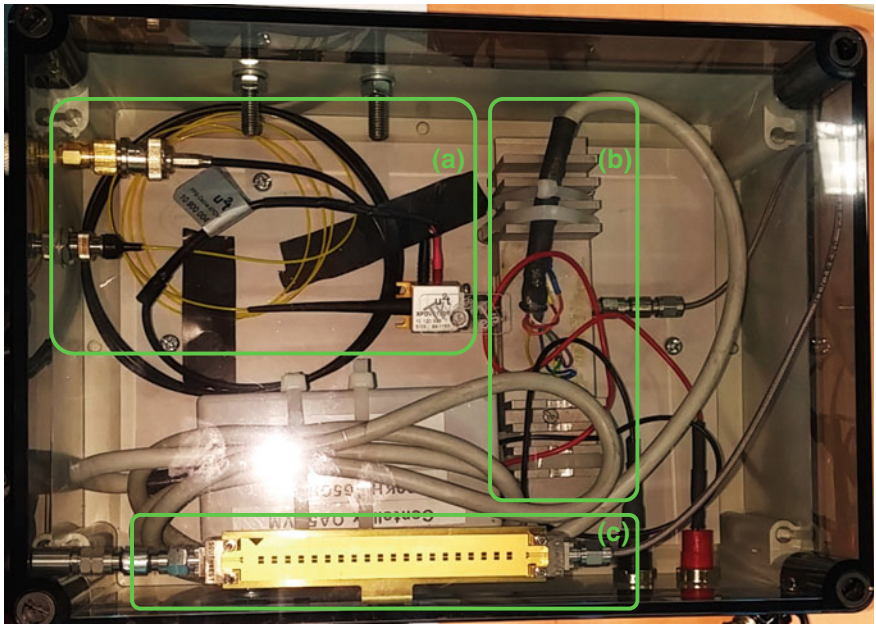


Fig. 10. Fabricated hybrid integrated RAU for the shopping mall ceiling installation consisting of a high-speed PD (**a**), a RF amplifier (**b**) and a SIW LWA antenna (**c**)

Fig. 11. RAPID5G demo vehicle used as mobile unit in the shopping mall, containing a receiver horn antenna, a RF amplifier, a SBD detector, a baseband amplifier a SDI-to-HDMI scalerbox and a 4 K high resolution screen to display the received video stream

operational over the course of a full week, thus demonstrating the implementation of such a 5G hot-spot under real-life conditions.

With this demonstration we provided an example for the generation of multiple simultaneous antenna beams providing a real-time high data rate service. One example, where such high data rates are required in real time is virtual reality (VR). We consider connecting multiple VR users wirelessly to an access point as a strong use case for 5G hot-spots for which multiple, steerable beams are essential.

5 Conclusion

In this paper, the implementation of 5G hot-spots via a combination of leaky-wave antennas and Radio-over-Fiber techniques has been demonstrated. The radiation patterns of the utilized PCB based LWA structure show gains around 14 dBi between 50 and 70 GHz. Beam steering and multiuser support are provided by the frequency scanning behavior of the LWA. It has been outlined, how to employ RoF techniques to

use the LWA for beam steering by changing the wavelength of the optical carrier. This setup has been employed to perform 2 Gbit/s OOK data transmission and up to 6 Gbit/s IF-OFDM-QAM data transmission, while utilizing a simple SBD receiver. Furthermore, it has been demonstrated to extend the system to support multiple users with multiple RF beams by employing dense WDM channels by transmitting three 1 Gbit/s OOK signals with 3 GHz spacing. Finally, the implementation of 5G hot-spots for a week-long operation during a field trial in a shopping mall has been illustrated. Therefore, two 1.5 Gbit/s uncompressed real-time video streams have been transmitted simultaneously over 11 km SMF and 2 m wireless to a demo vehicle, which acts as the mobile unit to show the received video stream.

Acknowledgements. This work was supported in part by the Horizon2020 EU-JP research project RAPID5G (www.rapid-5g.eu or www.rapid-5g.jp) under the grant no. 643297. The authors want to thank Naruto Yonemoto from ENRI for the support regarding the antenna measurements and Boris Shih from Corning for the support for the IF-OFDM measurements. Furthermore, the authors want to thank Michał Kościesza, Mason Lange from UDE, Andrzej Szczypawka and Maciej Buczkowski from Exatel and Adrian Juarez from Corning, who made the shopping mall field trial possible.

References

1. CISCO (2015) Cisco visual networking index: global mobile data traffic forecast update, 2014–2019, pp 1–42
2. Jiang D, Liu G (2016) An overview of 5G requirements. 5G mobile communications, 1st edn. Springer Publishing, New York
3. Chandra K, Venkatesha Prasad R, Niemegeers I (2015) An architectural framework for 5G indoor communications. In: 2015 international wireless communications and mobile computing conference (IWCMC), Dubrovnik, pp 1144–1149. https://doi.org/10.1109/iwcmc.2015.7289244
4. Stöhr A et al (2017) Radio technologies for 5G using advanced photonic infrastructure for dense user environments (RAPID). http://www.rapid-5g.eu/
5. Rappaport TS et al (2013) Millimeter wave mobile communications for 5G cellular: it will work! IEEE Access 1:335–349. https://doi.org/10.1109/access.2013.2260813
6. Weiler RJ et al (2014) Enabling 5G backhaul and access with millimeter-waves. In: 2014 European conference on networks and communications (EuCNC), Bologna, pp 1–5. https://doi.org/10.1109/eucnc.2014.6882644
7. Marnat L et al (2017) V-band transceiver modules with integrated antennas and phased arrays for mmWave access in 5G mobile networks. In: 2017 11th European conference on antennas and propagation (EUCAP), Paris, pp 2786–2790. https://doi.org/10.23919/eucap.2017.7928489
8. Zhu J, Peng B, Li S (2017) Cavity-backed high-gain switch beam antenna array for 60-GHz applications. IET Microw Antennas Propag 11(12):1776–1781.https://doi.org/10.1049/iet-map.2016.1129
9. Oliner AA, Jackson DR (2007) Leaky-wave antennas. Antenna engineering handbook, 4th edn. McGraw-Hill, New York

10. Paaso H et al (2017) DoA estimation using compact CRLH leaky-wave antennas: novel algorithms and measured performance. IEEE Trans Antennas Propag 65(9):4836–4849. https://doi.org/10.1109/tap.2017.2724584
11. Steeg M, Khani B, Rymanov V, Stöhr A (2016) Novel 50–70 GHz compact PCB leaky-wave antenna with high broadside efficiency and low return loss. In: 2016 41st international conference on infrared, millimeter, and terahertz waves (IRMMW-THz), Copenhagen, pp 1–2.https://doi.org/10.1109/irmmw-thz.2016.7758489
12. Khani B, Rymanov V, Steeg M, Buck A, Dülme S, Stöhr A (2015) Compact e-band (71-86 GHz) bias-tee module for external biasing of millimeter wave photodiodes. In: 2015 international topical meeting on microwave photonics (MWP), Paphos, pp 1–4. https://doi.org/10.1109/mwp.2015.7356673
13. Husain B, Steeg M, Stöhr A (2017) Estimating direction-of-arrival in a 5G hot-spot scenario using a 60 GHz leaky-wave antenna. In: 2017 IEEE international conference on microwaves, antennas, communications and electronic systems (COMCAS), Tel-Aviv, pp 1–4. https://doi.org/10.1109/comcas.2017.8244845
14. Novak D, Nirmalathas A, Lim C, Waterhouse R (2009) Fiber radio technology. Microwave photonics: devices and applications, 1st edn. Wiley, Chichester
15. Steeg M, Babiel S, Chuenchom R, Stöhr A (2015) 10 GHz channel spacing ultra-dense WDM networks transparently extended by mm-wave coherent RoF links. In: 2015 IEEE international conference on communications workshops (ICCW), London, pp 324–328. https://doi.org/10.1109/iccw.2015.7247199
16. Stöhr A, Shih B, Abraha ST, Steffan AG, Ng'oma A (2016) High spectral-efficient 512-QAM-OFDM 60 GHz CRoF system using a coherent photonic mixer (CPX) and an RF envelope detector. In: 2016 optical fiber communication conference and exposition (OFC), Anaheim, CA, pp 1–3. https://doi.org/10.1364/ofc.2016.tu3b.4
17. Steeg M, Stöhr A (2017) High data rate 6 Gbit/s steerable multibeam 60 GHz antennas for 5G hot-spot use cases. In: 2017 IEEE photonics society summer topical meeting series (SUM), San Juan, pp 141–142. https://doi.org/10.1109/phosst.2017.8012690

Millimeter-Wave Imaging for FOD Detection

F. Nsengiyumva, C. Migliaccio, and Ch. Pichot(✉)

Université Côte d'Azur, CNRS, LEAT, Sophia Antipolis, France
christian.pichot@univ-cotedazur.fr

Abstract. We present an efficient approach for reconstructing images of small objects and Foreign Object Debris (FODs) from scattered field synthetic data and measurements in W-band (75−110 GHz), taking advantages of spatial and polarization diversities, and process them with either qualitative or quantitative inverse methods. The target application is FOD detection for airport runways. We investigate FOD detection and identification capabilities through two approaches: a quantitative reconstruction under 2D assumption and qualitative reconstruction using a Backpropagation algorithm (BP). Most existing systems are monostatic but rapid advances in millimeter-wave technologies make it realistic to imagine the deployment of multistatic radars on either side of airport runways. In the following, we investigate the contribution of multistatic measurements, on the reconstruction of typical debris, courtesy of DGAC (French Civil Aviation).

Keywords: Millimeter-wave imaging · Foreign object debris

1 Introduction

Millimeter-wave radar systems in W-band (75–110 GHz) are booming, due to advances in integrated circuits, and fabrication of low-cost and high-resolution compact systems, thanks to the short wavelength. Application domains are various: automotive sector, pilot assistance for helicopters, etc. We aim to design a W-Band radar system for imaging Foreign Object Debris (FODs) on runways [1]. Detecting objects is important but assessing if the object is harmful to air navigation is fundamental and crucial in order to take the decision to stop or not the traffic on the runway. Most of mm-wave radar systems are already monostatic. Some also include multiview configurations obtained either by using electronic/mechanical scanning or SAR processing. Multistatic scenarios offer significant advantages for identification and detection purposes [2]. First, multistatic measurements appear to be an attractive solution for enhancing the detection, especially for nonmetallic objects. The reflectivity of nonmetallic objects is weak whereas the transmitted wave through dielectric objects brings significant information that a multistatic system can process. Second, millimeter-wave images are good candidates for identification. But building high-resolution 3D images also require measurements including multistatic configurations. In addition, polarization diversity also contributes to a best knowledge of targets and it will be investigated as well. FODs cost a considerable amount of money to companies and airport operators. A FOD is defined as an unwanted object, animate or inanimate, located in an inappropriate place in the aerodrome environment (runways, taxiways, etc.) that is

Electronic Navigation Research Institute, *Air Traffic Management and Systems III*, Lecture Notes in Electrical Engineering 555,
https://doi.org/10.1007/978-981-13-7086-1_17

hazardous to the airport and for personal security and integrity of an aircraft. Neither detection nor identification is easy to proceed because even small objects such as nuts, screws, debris of runway lights or tires, thin metallic parts of airplanes (like the one that caused the crash of the Concorde), are sources of severe not to say fatal damages.

All existing radar systems are based on monostatic configurations. Despite their performance, identification and detection are not 100% guaranteed. For that purpose, a quantitative reconstruction image will provide information not only on the shape but also on the properties of the material. The radar sensors are supposed to be located along the runway [3]. In order to estimate the investigation domain, a qualitative imaging is used by means of a back propagation algorithm. A simplified assumption for fast reconstructions is to use algorithms based on two-dimensional approaches (2D-TM and 2D-TE polarizations) of the electromagnetic imaging problem. Using 2D assumption and the polarization used for the incident field (horizontal or vertical), one can reconstruct different cross-sections of 3D objects. Recently, some results of quantitative imaging were presented reconstructions from both 2D-TM and 2D-TE synthetic data with additive noise [4]. We present results obtained from both synthetic and experimental 3D data for some typical objects or FODs that can be found on runways.

2 Direct Problem

Thanks to the short wavelengths in the W-band, many objects are large compared to the wavelength and can therefore be considered as 2D. We separate the direct problem into 2D-TE and 2D-TM cases. Considering 2D cylinder with infinite dimension along Oz, the 2D-TM case is, per definition, the case where the electric field parallel to Oz. On the contrary, for the 2D-TE case, the electric field is parallel to xOy plane. Electric fields are orthogonal in 2D-TE and 2D-TM cases and form an orthogonal basis that can be used to decompose any polarization of the electric field.

Considering two-dimensional inhomogeneities in free-space, the following integral representation is used for the total field

$$E(r) = E^{inc}(r) + \left[k_0^2 + \nabla\nabla.\right][G(r) * \chi(r)E(r)], \tag{1}$$

where

- $\chi(r) = \varepsilon_{rc}(r) - 1$: Contrast function.
- $\varepsilon_{rc}(r)$: Relative complex permittivity of the inhomogeneities.
- k_0: Free-space wavenumber.

Moreover, the relative complex permittivity is defined as:

$\varepsilon_{rc}(r) = \varepsilon_{rc}(r) - j\sigma/(\omega\varepsilon_0)$, with ε_0 the electric permittivity of the vacuum.

The Green's function expression is

$$G(r, r') = \frac{-j}{4} H_0^{(2)}(k_0|r - r'|), \tag{2}$$

For objects lying on runways, the Green's function has to be changed to take into account the presence of the ground.

The operator $\nabla\nabla.$ is null for the 2D-TM case. The 2D-TE case is more complex due to the second-order derivative in Eq. (1) and extreme care should be used for the basis functions when applying the method of moment. We use piecewise sinusoidal 1D functions

$$f_n(x) = \begin{cases} \frac{\sin[k_0(x-x_{n-1})]}{\sin[k_0(x_n-x_{n-1})]} & x_{n-1} \le x \le x_n \\ \frac{\sin[k_0(x_{n+1}-x)]}{\sin[k_0(x_{n+1}-x_n)]} & x_n \le x \le x_{n+1} \\ 0 & elsewhere \end{cases}, \tag{3}$$

Finally, and 2D basis functions are the product of two 1D functions

$$f_n(x,y) = f_n(x) f_n(y), \tag{4}$$

For cylinder with circular cross section, we also use an analytical model defined in [5], that we call ESCALE.

3 Inverse Scattering Problem

3.1 Backpropagation

An example case for applying Backpropagation (BP) algorithm is shown in Fig. 1. Using Green's object-receiver matrix G^R, scattered fields are backpropagated to obtain the distribution *S(x,y)* proportional to the induced currents in the target

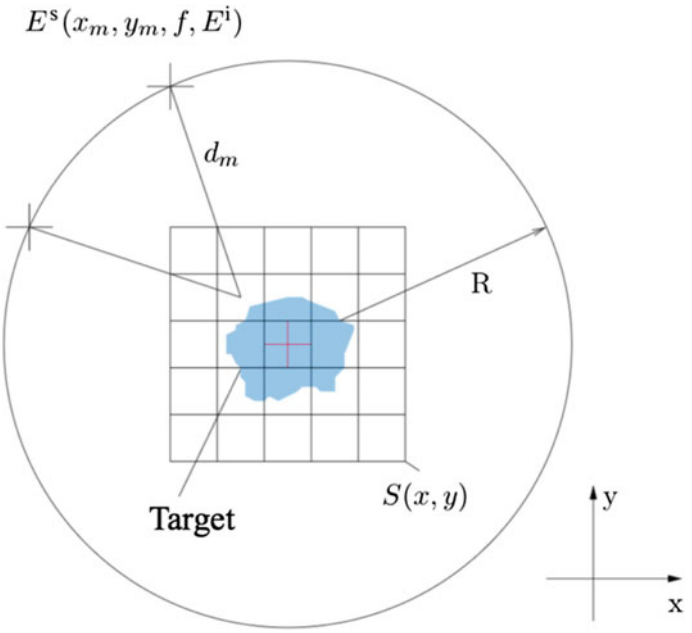


Fig. 1. Example case of backpropagation

$$S(x,y) = G^{R*}E^S = G^{R*}G^R\chi E^S, \tag{5}$$

The relation Eq. (5) is approximated with the following coherent summation over M receivers, L transmitters and F frequencies and normalized with respect to the number of observation points, including the number of frequencies

$$S(x,y) = \frac{1}{F}\frac{1}{L}\frac{1}{M}\sum_{f=1}^{F}\sum_{l=1}^{L}\left|\sum_{m=1}^{M}\frac{1}{d_m}E^S\left(d_m,f,E^i\right)e^{j\Phi(d_m,f)}\right| \tag{6}$$

with

- F: Number of frequencies.
- L: Number of incidences.
- M: Number of receivers.
- E^s: Scattered field measured on the mth receiver.
- d_m: Distance between the mth receiver (x_m, y_m) and the cell (x,y) of the object
- Φ (d_m, f): Phase shift linking the cell to the receiver position.

We have to distinguish two cases

- Receivers are apart from incidences, which is the multistatic case and we call Φ_{multi} the corresponding phase. It is simply calculated as the product between the free-space wave number and the distance between the cell and the receiver Eq. (7).
- Receivers and incidences are collocated, which is the monostatic case and we call Φ_{mono} the corresponding phase (Eq. 8). In this case, we have to take into account the round trip and calculate the product between the free-space wave number and twice the distance to the cell Eq. (8).

$$\Phi_{multi}(d_m,f) = k_0 d_m = 2\pi\frac{d_m}{\lambda}, \tag{7}$$

$$\Phi_{mono}(d_m,f) = 2k_0 d_m = 4\pi\frac{d_m}{\lambda}, \tag{8}$$

The Backpropagation scheme can allow at same time to locate targets and estimate the investigation domain for the quantitative algorithm. But it can also be used as an initial guess for the quantitative imaging algorithm and reconstruct a qualitative image.

3.2 Quantitative Inverse Algorithm

We assume that two-dimensional objects are embedded in the investigation domain D and illuminated successively by L different known incident fields. For each excitation l ($l = 1...L$), the scattered far-field is measured on receivers located on a circle of radius R. The integral representation of the scattered field on the measurement points, and using moment method, yields the following nonlinear matrix system:

$$E_l^S = G^R \chi \left(I - G^0 \chi\right)^{-1} E_l^{inc} \tag{9}$$

where

- χ: Diagonal matrix of the contrast
- E_l^i: Incident field vector.
- E_l^s: Scattered field vector.
- G^R and G^O: matrices of integrated Green's functions.

For reconstructing separately real and imaginary parts of the contrast χ, we use a bi-conjugate gradient method [6]. The inverse problem solution is obtained by minimizing a cost functional $J(\varepsilon_r,\sigma)$ representing the discrepancy between measured scattered field and simulated one for a given contrast $\chi(\varepsilon_r,\sigma)$. We consider the real and imaginary part of the contrast χ as independent variables.

$$J(\varepsilon_e, \sigma) = \sum_{f=1}^{F} \sum_{l=1}^{L} \left\| E_{meas}^S - E_{comp}^S \right\|^2, \tag{10}$$

with

- E_{meas}^S: Scattered field vector obtained from measurement.
- E_{comp}^S: Scattered field vector obtained from computation.

In the next section, we investigate 2D quantitative reconstructions of selected FODs.

4 2D-Quantitative Reconstructions

4.1 2D-Hypothesis

We have explained in Sect. 2 that in W-band many objects are large compared to the wavelength at least in one dimension. It is therefore possible to consider these objects as 2D, which significantly simplifies the theoretical problem. Nevertheless, this hypothesis must be discussed and verified.

For this, we consider a Perfect Electrical Cylinder (PEC) with 8 mm-diameter and variable height. Results of 3D full-wave simulations are compared with those from the 2D model. 3D simulations are conducted using Ansoft-HFSS [7] whereas the 2D model is computed with ESCALE. We have deliberately chosen ESCALE instead of the method of moment developed for the direct problem for getting rid of any numerical errors, since ESCALE uses analytical solutions. We consider a fixed incidence (at $\phi = 180°$) with receivers placed along a circle in the far field. Hence, we perform a multistatic study where the monostatic case corresponds to the particular point $\phi = 180°$ (receiver is collocated with the transmitter).

Figures 2, 3, 4 and 5 show results in amplitude and phase for 2D-TM and 2D-TE cases and heights from 40 to 100 mm. The frequency of simulation is 94 GHz. From these plots, we draw the following conclusions:

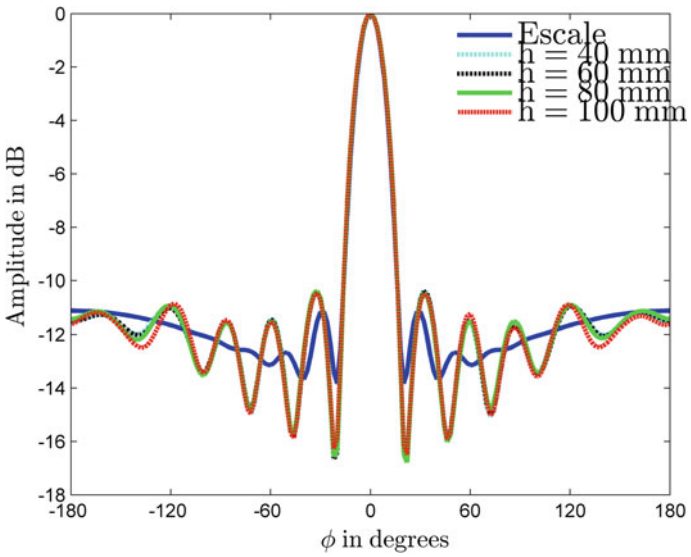


Fig. 2. PEC cylinder—Amplitude of the scattered field for TM case—Comparison between 2D case and various heights for 3D case

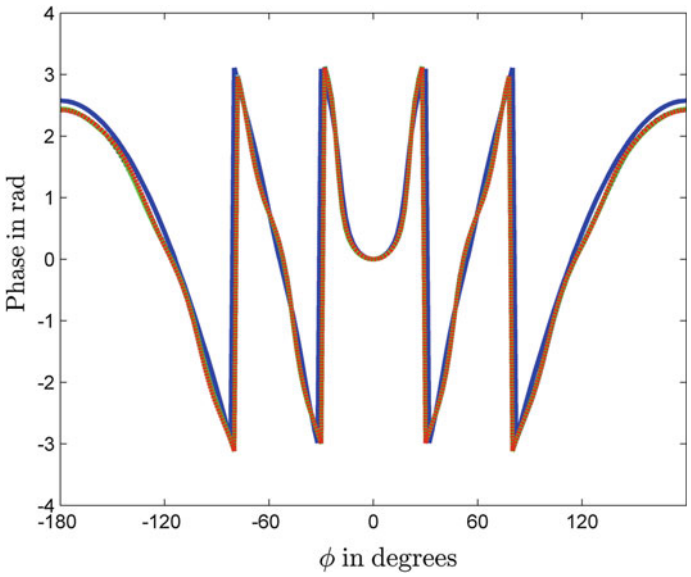


Fig. 3. PEC cylinder—Phase of the scattered field for TM case—Comparison between 2D case and various heights for 3D case (definition of the curves given in Fig. 2)

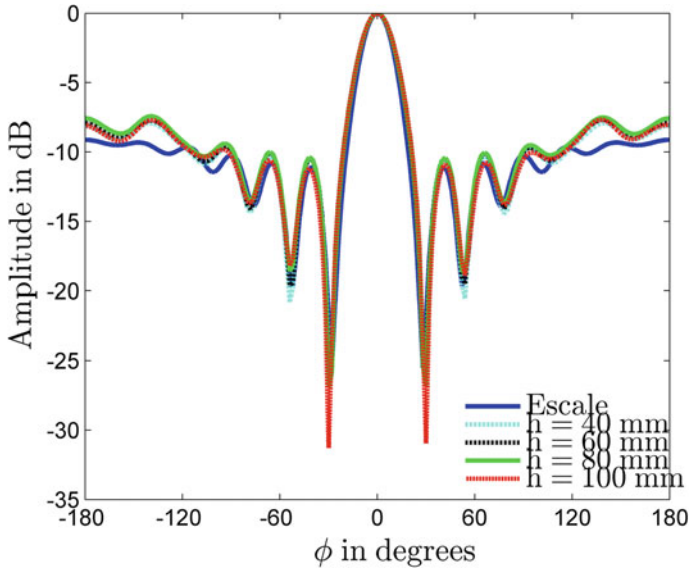


Fig. 4. PEC cylinder—Amplitude of the scattered field for TE case—Comparison between the 2D case and various heights for the 3D case

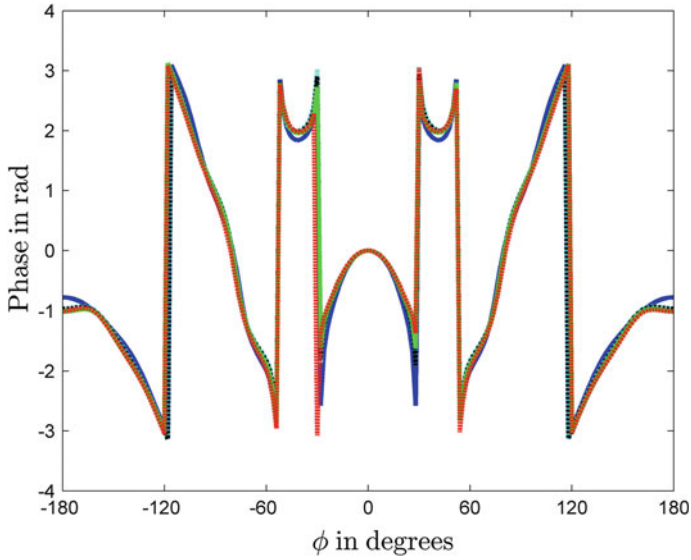


Fig. 5. PEC cylinder—Phase of the scattered field for TE case—Comparison between 2D case and various heights for 3D case (definition of the curves given in Fig. 4)

- There is no visible difference for 3D simulated radiation patterns for heights ranging from 40 to 100 mm.
- Phase patterns are less sensitive than amplitude ones to the 2D hypothesis with almost no differences between 3D and 2D results.
- 2D and 3D amplitude patterns are similar in $\phi = 0°$ (direction opposite to the incidence), and begin to differ for larger ϕ. Therefore, the error is maximal in the monostatic case ($\phi = 180°$). This means that we cannot use the 2D hypothesis for reconstructing objects in the monostatic case.
- The 2D hypothesis is valid for a total angular range of 120° in ϕ centered at 0° for the 2D-TE case. It is of the order of 90° for the 2D-TM one. Quantitative reconstructions based on incomplete field of view reconstructions will be part of further studies.

In addition, we define a quantitative error criterion (e). For this, we compute the difference (computed as the *L2* norm) between the 3D simulation and 2D model at every height.

$$e = \frac{\left\langle E^S_{3D} - E^S_{2D}, E^S_{3D} - E^S_{2D} \right\rangle}{\left\langle E^S_{3D}, E^S_{3D} \right\rangle}, \tag{11}$$

where

- $< >$: Hermitian scalar product.
- E^S_{3D}: Scattered field vector obtained from 3D simulation. It is taken as reference in the scalar product.
- E^S_{2D}: Scattered field vector obtained from 2D model.

Results are shown in Table 1. The error (e) is expressed in dB as 10*log(e). For this study, we have extended the range of the height from 2 to 100 mm.

Table 1. Error (L2 norm) between 3D simulation and 2D model

h (mm)	Error (dB)	
	TE-case	TM-case
2	−8.4	−7.4
3.2	−10.3	−11.7
6	−13.1	−12.6
10	−15.2	−13.7
20	−16	−14.3
40	−17.5	−16.2
60	−17	−16.2
80	−16.5	−16.1
100	−16.6	−16.2

The error decreases when the height of the cylinder increases but tends to stabilize from 40 mm. This perfectly matches the low differences observed in 3D simulations in

Figs. 2, 3, 4 and 5. Moreover, even for the cylinder of 3.2 mm height, which is twice as small as its diameter, the error remains less than -10 dB.

Note that, assuming that error is due only to random phenomena and considering Eq. (11), the error is the inverse of the signal-to-noise ratio. Backpropagation algorithms are known to be robust and work even with SNR close to 0 dB. This means that we can use Backpropagation for qualitative imaging algorithm for all cylinders tested in Table 1 because the largest error reaches -7.4 dB, which corresponds to a SNR of 7.4 dB.

Our conclusions with respect to this study are as follows:

1. Multistatic scenarios are better suited for quantitative reconstruction because multistatic configuration captures the electric field all around the object, which brings diversity in the collected information.
2. 2D assumption can be used for quantitative imaging in multistatic scenarios.
3. 2D assumption can be used for qualitative imaging using backpropagation algorithm.

4.2 Choice of FODs for the Reconstructions

FODs are all kind of debris that can be found on the airport runway. It is of course difficult to design a system considering all possible objects that might fall on the ground, and regulation authorities of civil aviation work on standard objects for FOD systems [8]. This means that particular objects are defined that FOD systems must detect under any weather conditions. The choice of these objects is not straightforward since FOD systems are very diverse both in implementation and nature of the sensor. With respect to the implementation, we find systems placed on each side of the runway [1] and those located at each end [9]. This makes the detection problem quite different because of the distance issue. Moreover, sensors of FOD systems can be optically based [10] or FM-CW millimeter-Wave radars [1, 9]. Once again, the detection problem is quite different. For example, optical sensors have difficulties to detect the contrast with the ground when objects are gray. Similarly, the efficiency of optically based sensors is greatly decreased under bad weather conditions. In turn, mm-Wave radars are highly sensitive to incidence, polarization, material, clutter and the signal to noise ratio drastically decreases with phase noise that is one of the main problems in designing radars of FM-CW type. To compare FOD systems we must take care of

- The size of the objects that indicates distance limitations, signal to noise ratio and clutter influences.
- The color for optically based sensors.
- The shape and material for radar-based sensors.

Taking into account these constraints, regulation authorities have settled a small tube in metal as test object. We have seen in Sect. 4.1 that 2D assumption is valid for such objects.

The 8 mm-diameter metallic cylinder will be our first test FOD. The second one is a brake pad (Fig. 6). We chose this object because it is a real FOD, courtesy of DGAC

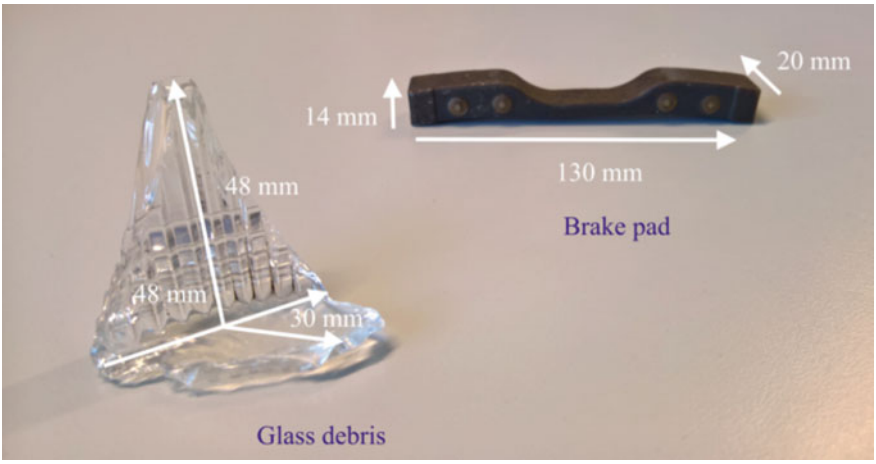


Fig. 6. Typical debris (courtesy of DGAC)

(French Civil Aviation). Dimensions of the brake pad are 12 mm width (Pb), 130 mm length (Lb) and 20 mm height (Hb).

Later on, we will also investigate the qualitative reconstruction of the glass debris also shown in Fig. 6.

4.3 2D Quantitative Reconstructions

The workflow of the reconstruction is as follows:

- We simulated 2D objects (direct problem described in Sect. 2).
- We corrupt the simulations with 10% Gaussian noise for avoiding an inverse crime.
- We perform the inversion (quantitative inverse algorithm described in Sect. 3).

4.3.1 Reconstruction of the Tube

Figure 7 shows the quantitative reconstruction of the tube. For metallic object, the real part of the permittivity is equal to 1 and the object is retrieved through its conductivity profile. First, we observe a good agreement between the expected shape of the FOD (dashed white curve) and the reconstruction. Second, the repartition of high permittivity values at the border of the tube clearly indicates that the object is metallic. Finally, the value of the conductivity is 80 S/m is much lower (several orders of magnitude) than the one of metallic targets but we have seen in previous studies [11] that the scattered electric field saturates with respect to the value of σ. Saturation is reached from 60 S/m. This points out the limitation of the inversion algorithm described in Sect. 3.2.

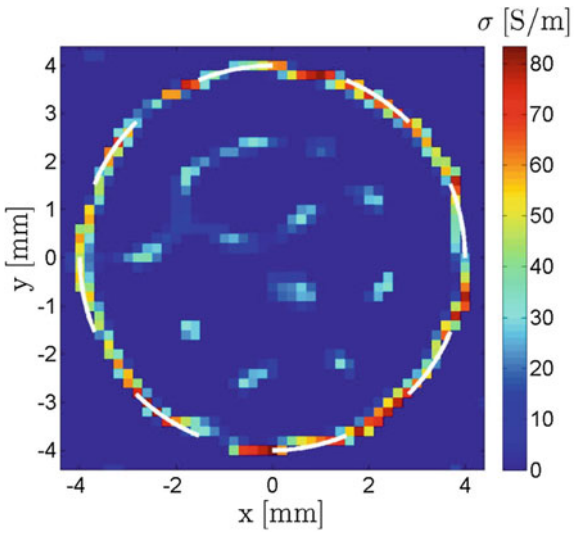


Fig. 7. Quantitative reconstruction (conductivity profile) of the tube at 94 GHz—TM-case, 1000 iterations

4.3.2 Reconstruction of Brake Pad

Prior to reconstruction, we investigate the most favorable configuration for quantitative imaging. For this, we start with 3D simulations. We first sketch the brake pad and perform simulations with incidence along *y*-axis (blue axis in Fig. 8) and vertical polarization (TM case). We chose several distances between transmitter and receiver, both distances are in the far-field as for the real application. Results are presented in Fig. 9. First, we see that the distance provokes a linear drop (in dB) of the amplitude as expected from the radar equation. This means that transmitter and receivers can be

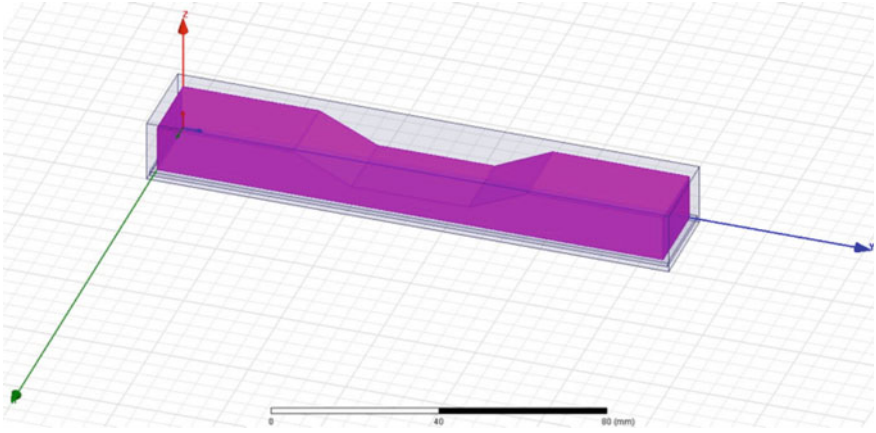


Fig. 8. Quantitative reconstruction (conductivity profile) of the brake pad at 94 GHz—1000 iterations

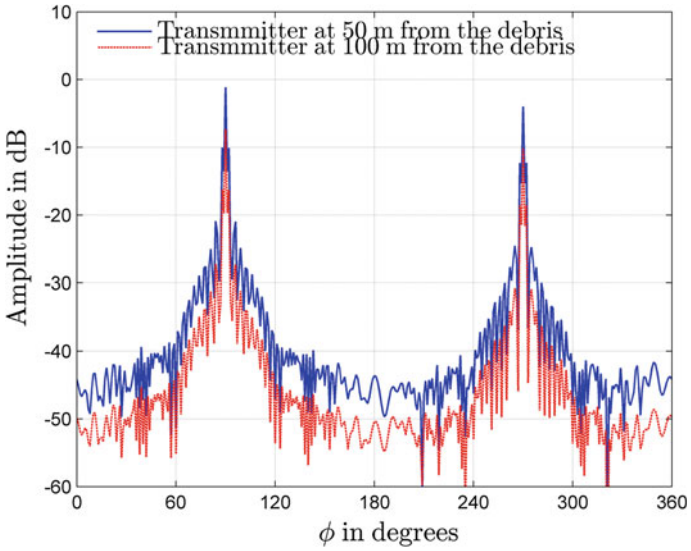


Fig. 9. Amplitude response for incidence along (conductivity profile) the brake pad at 94 GHz —TM case

placed at any location provided that they are in the far field of the object. Second, the brake response has two high peaks for $\phi = 90°$ and 270°, which are directions along the axis of the incidence. In the reconstruction we chose this cross section $\phi = 90°$ as reconstruction plane. It corresponds to the section Hb*Pb in Fig. 6.

Figure 10 shows the quantitative reconstruction of the conductivity profile. Note that axes have been changed compared to Fig. 9, the *xOy* plane of the reconstructed conductivity profile being the previous *xOz* plane in Fig. 9. Same conclusions apply as in Sect. 4.3.2: contour and metallic structure of the FOD are well retrieved whereas the value of σ is too small.

In this section, we have seen that it is possible to retrieve with good accuracy the shape of metallic FODs with quantitative imaging under 2D-assumption and multistatic configuration. It is also possible to predict the metallic nature of the FOD looking at the repartition of the conductivity profile even though the value of the conductivity is far away from the real one. Nevertheless, we have not discussed the reconstruction time. It strongly depends on the computer used, memory and code implementation. Neither of them was optimized for this study and the inversion took two hours on a workstation. Even if it would be possible to be much faster, for example using High parallel Computing (HPC), quantitative methods still are time consuming compared to qualitative ones. In the next section we will present results for qualitative imaging and discuss performance compared to quantitative imaging in the case of FODs.

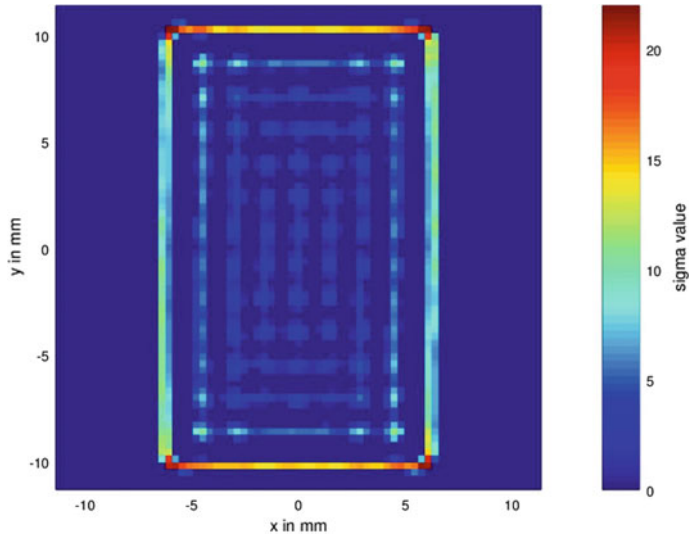


Fig. 10. Quantitative reconstruction (conductivity profile) of the brake pad at 94 GHz—TM case, 1000 iterations

5 Qualitative Imaging

Apart from the reconstruction time, qualitative imaging can be performed easily in 2D and 3D according to Eq. (6). We investigate the reconstruction of three objects: both FODs used in Sect. 4 and for the glass debris. All results are obtained from measurements performed with the 3D scanner of our laboratory [12].

5.1 Reconstruction of the Tube

For the tube, we simply make a 2D scan (in the horizontal plane) around the object placed vertically for several incidences taken from 0 to 350°. The receivers cannot cover 360° due to the physical presence of the transmitter. The scan range of receivers is [−145, 145°]. Figure 11 shows the image of the 8 mm-diameter tube with 2D Backpropagation. The left image corresponds to the raw result whereas we apply a threshold at half-power (3 dB or 0.707 in amplitude) for the right one. Results using threshold are as good as for the quantitative imaging for the shape reconstruction and conductivity around the object border as for metallic target.

5.2 Reconstruction of the Brake Pad

Results are obtained using 3D scans according to the description in Fig. 12. TM and TE cases are obtained by changing the polarization of the transmitter respectively, from vertical to horizontal. We conduct monostatic and multistatic measurements. Note that for multistatic measurements, the incidence is unique as shown by the red arrow in Fig. 12.

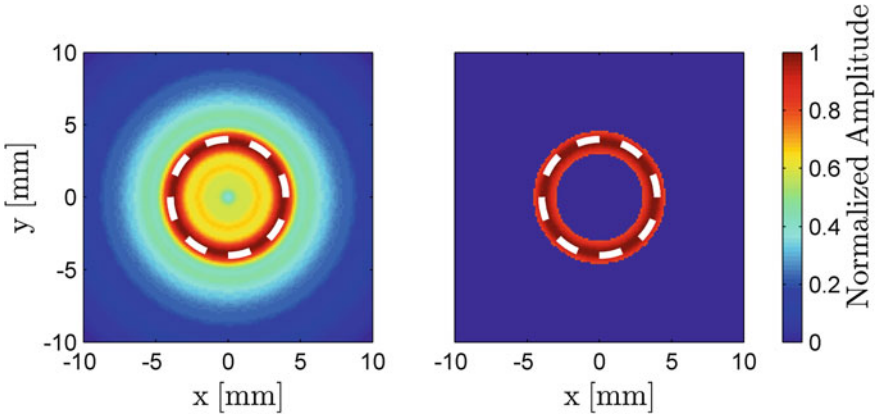


Fig. 11. Amplitude of the backpropagation at 94 GHz—TM case

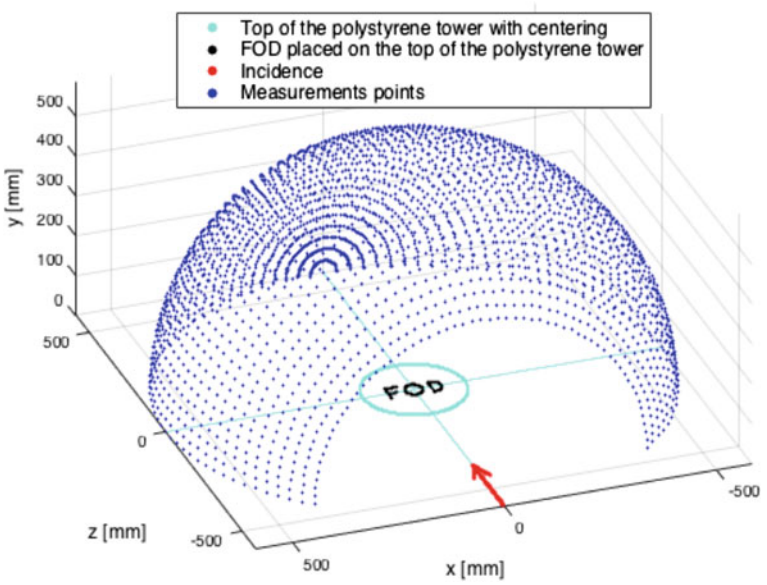


Fig. 12. 3D scan

Figure 13 shows the image obtained with 3D Backpropagation on monostatic (Fig. 13a) and multistatic (Fig. 13b) measurements. In the case of brake pad, both images provide the accurate FOD location but not its shape. A possible improvement would be to work with multi-incidences but we have to consider about the total measurement time.

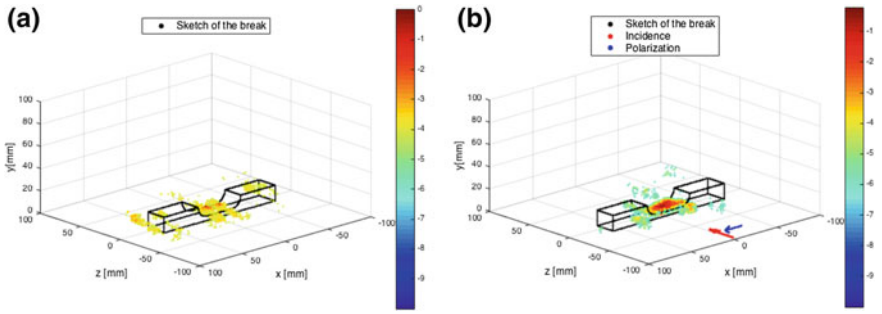


Fig. 13. Amplitude of the backpropagation (90 GHz)—TE, brake pad. **a** Monostatic measurements. **b** Multistatic measurement

5.3 Reconstruction of the Glass Debris

We measure the glass debris using same 3D scan conditions as for the brake pad. Results are presented in Fig. 14. Monostatic results are not presented because they provide no reconstruction. This matches our previous measurements results obtained with a monostatic real radar system. The glass debris was almost impossible to detect. On the contrary, using multistatic configuration, we can see that the reconstruction is very good.

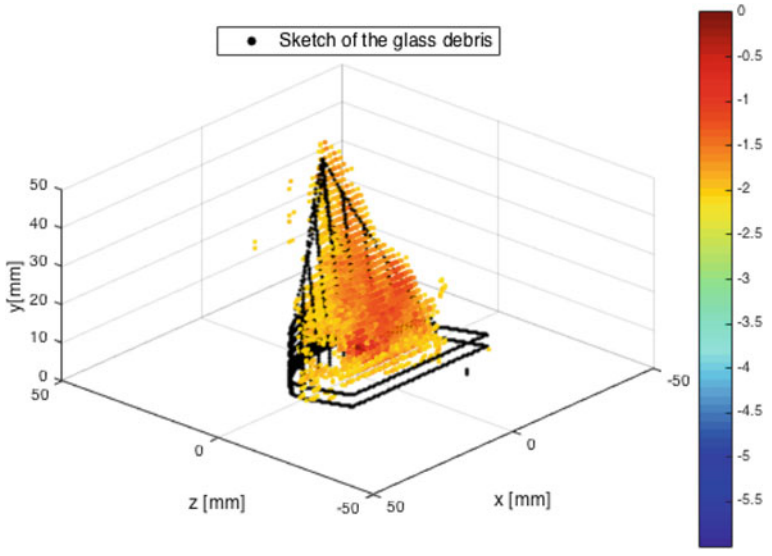


Fig. 14. Amplitude of the backpropagation (90 GHz)—Addition of multistatic TE and TM cases, glass debris

For a better understanding, we represent the measured scattered field of the TE case for monostatic (Fig. 15) and multistatic (Fig. 16) measurements. It is clear that the

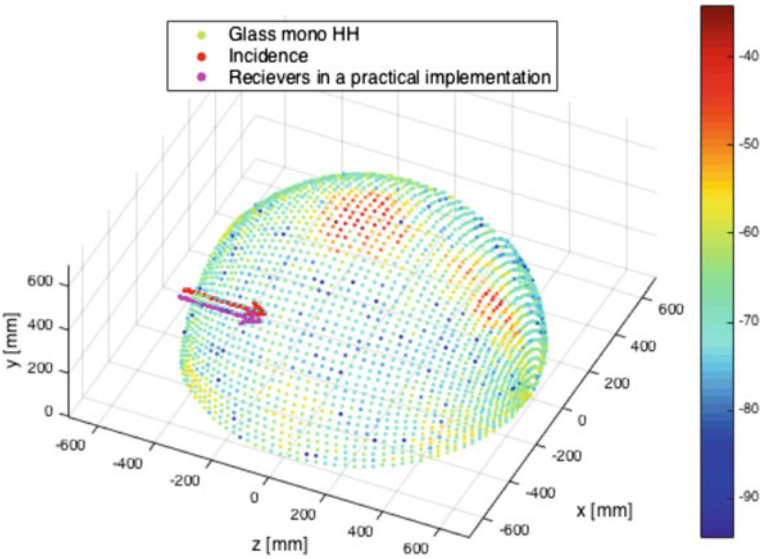


Fig. 15. Monostatic measurement. Amplitude of the measured scattered field at 90 GHz—TE case, glass debris

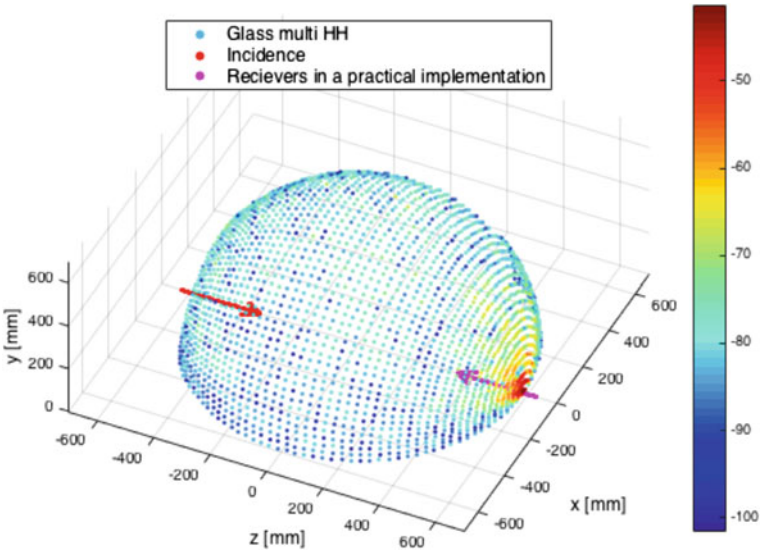


Fig. 16. Multistatic measurement. Amplitude of the measured scattered field at 90 GHz—TE case, glass debris

main radiation direction of the scattered field in the monostatic case is on the top of the debris where no receivers are placed in practical implementation. On the contrary, for the multistatic case, the maximum is in the direction opposite to the transmitter

(here we show only the TE but it is the same for the TM). This field is the most important one for the reconstructions because it has propagated through the debris and, therefore has the highest content of useful information.

6 Conclusion

In this paper, we investigate W-band qualitative and quantitative reconstructions of foreign objects debris (FODs) for runway surveillance application, based on Back-propagation and an inverse method, respectively. Typical dielectric and metallic objects found on runways are studied from synthetic data and experimental measurements to show the performance of such algorithms for imaging FODs. We emphasize that multistatic configuration enables to reconstruct shapes of FODs with accuracy, either with quantitative imaging under 2D assumption as well as with qualitative imaging from 3D measurements. Progress in mm-Wave circuits design, integration and long range transmission, will push next generation of FOD systems toward multistatic operation and will offer new perspectives for imaging.

Acknowledgements. Authors would like to thank DGAC (French Civil Aviation) for providing FODs.

References

1. Mazouni K, Zeitler A, Lanteri J, Pichot Ch, Dauvignac J-Y, Migliaccio C, Yonemoto N, Khomura A, Futatsumori S (2012) 76.5 GHz millimeter-wave radar for foreign objects debris detection on airport runways. Int J Microw Wirel Technol 4:317–326
2. Frischen A, Hasch J, Waldschmidt C (2015) Contour recognition with a cooperative distributed radar sensor network. In: 2015 IEEE-MTT ICMIM, Heidelberg-Germany, 27–29 April 2015
3. Weller JR (2014) FOD detection system. In: Wildlife and foreign object debris (FOD) workshop, Cairo, Egypt, 2014
4. Nsengiyumva F, Pichot Ch, Aliferis I, Lanteri J, Migliaccio C (2015) Millimeter-wave imaging of foreign objects debris detection (FOD) based on two-dimensional approach. In: Proceedings of the 2015 IEEE CAMA, Chiang Mai, Thailand, paper 1088
5. Balanis CA (1989) Advanced engineering electromagnetics. Wiley, New York
6. Le Brusq E, Aliferis I, Dauvignac J-Y, Pichot Ch (2004) Iterative nonlinear reconstruction of isotropic and anisotropic 2D objects using TE and TM polarizations. In: Proceedings of the URSI EMTS2004, Pisa, Italy, pp 703–705
7. https://www.ansys.com/products/electronics/ansys-hfss
8. O'Donnel MJ (2010) Airport foreign object debris (FOD) detection equipment, advisory circular, Federal aviation administration (FAA). https://www.faa.gov/documentlibrary/media/advisory_circular/150_5210_24.pdf
9. Beasly SP, Binns G, Hodges R, Badley RJ (2004) Tarsier, a millimetre wave radar for airport runway debris detection. In: Proceedings of the EURAD'04, Amsterdam
10. Xsight (2009) Xsight—advanced radar and optic sensors for FOD detection and homeland security. http://www.xsightsys.com

11. Zeitler A, Migliaccio C, Moynot A, Aliferis I, Brochier L, Dauvignac J-Y, Pichot Ch (2013) Amplitude and phase measurements of scattered fields for quantitative imaging in the W-band. IEEE Trans Antennas Propag 61(7):3927–3931
12. Jawad H, Lanteri J, Migliaccio C, Pichot Ch, Platt I, Tan A, Eccleston T, Woodhead I (2017) Microwave modeling and experiments for non destructive control improved quality of fruit. In: 2017 IEEE CAMA, Tsukuba, Japan, 4–6 December 2017, pp 124–127

300-GHz Bistatic Radar System by Radio Over Fiber Network Technology

Atsushi Kanno[1(✉)], Norihiko Sekine[1], Akifumi Kasamatsu[1], Naokatsu Yamamoto[1], and Tetsuya Kawanishi[1,2]

[1] National Institute of Information and Communications Technology, Tokyo, Japan
kanno@nict.go.jp
[2] Faculty of Science and Engineering, Waseda University, Tokyo, Japan
kawanishi@waseda.jp

Abstract. We demonstrated a 300-GHz bistatic radar system based on a radio signal over a fiber network to enhance the detection probability under a transmitter and two-receiver configuration. Fiber network management using optical frequency-modulated continuous waves is also discussed with fiber length measurement technique.

Keywords: Bistatic radar · Terahertz band · FM-CW · Optical FM-CW

1 Introduction

Nondestructive, high-precision imaging technology enhances safety and security in daily life. For example, some logistic centers have X-ray-based scanners to investigate concealed explosives in baggage to prevent terrorism attacks. In this scenario, a high-frequency radar system is an attractive means for detecting small objects and examining their surface structures with small invasiveness compared to X-rays. Actually, human body scanners that use millimeter-wave radio signals for security checkpoints have been rapidly developed and installed in airports [1]. However, these shorter wavelength radio signals have relatively large transmission losses because of their large free-space propagation and atmospheric attenuation coefficient, which degrade the obtained signal-to-noise power ratio (SNR). A multistatic radar system has great advantages for the effective collection of radio signals using several receivers to enhance the actual SNR, which will improve the target detection rate [2]. For instance, the detection rate of a foreign object debris detection system for airport runway surface surveillance should be >95%, as regulated in minimum aviation system performance standards [3]. This system is also useful for enhancing the precision of a radiolocation system [4]. However, this high-precision radar system requires a high-precision radar signal source, which increases the cost of a high-frequency radar system with high precision.

Radio over fiber (RoF) technology is a promising solution to deliver the radio signal generated at a central office via an optical fiber network. This helps realize a dramatic cost reduction by enabling the utilization of a single ultra-high-precision radar source set in a central office [5, 6]. Moreover, the low-loss feature of an optical fiber

Electronic Navigation Research Institute, *Air Traffic Management and Systems III*, Lecture Notes in Electrical Engineering 555,
https://doi.org/10.1007/978-981-13-7086-1_18

cable with <1 dB/km enables a long-distance remote radar head configuration. For instance, a foreign object debris detection system for airport runway surface surveillance could be configured with a combination of an optical fiber network and remote radar heads [7, 8]. A fiber length greater than 5 km could be achieved between the central office and remote radar heads.

In principle, the range resolution of a radar system is directly related to the bandwidth of the radar signal. Particularly in a frequency-modulated continuous-wave (FM-CW) radar system, a simple detector/receiver configuration can easily reduce the cost of the system. However, signals with a broad bandwidth must be generated to enhance the range resolution. Therefore, the use of higher frequency bands such as terahertz-band radio would be a straightforward way to achieve precise ranging [9].

In this study, a 300-GHz FM-CW-based bistatic radar system was evaluated using RoF technologies [10]. An FM-CW feeder network configured by the RoF network distributed simultaneous signals to each radar head, including the transmitter and receivers, with considerable losses. The optical fiber length management for future high-functional radar systems is also discussed.

2 Concept of Multistatic Radar System Based on Radio over Fiber Network

Figure 1 shows conceptual schematics of multistatic radar systems. In principle, an optically driven transmitter (Tx) and receiver (Rx) should be configured by launching the radar signal via an optical fiber network. The Tx is comprised of a photomixer

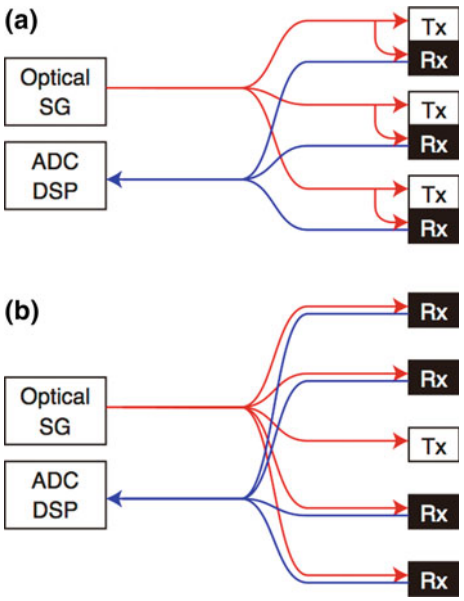


Fig. 1. Concept of terahertz-band multistatic radar system under **a** TRx and **b** Tx-Rx separation configurations. Red and blue arrows denote LO and RF signal lines and IF lines, respectively

(PM) and radio front-end, including an amplifier and multipliers. On the contrary, particularly in the FM-CW configuration, the Rx should also have a local oscillator (LO) signal input for an intermediate frequency (IF) downconversion with the radio front-end. In a sense, an optical LO port has the same signal waveform as the signal launching into the Tx. When using a conventional radar transceiver (TRx), the configuration of a multistatic radar system includes both the Tx and Rx connected to an optical signal generator (SG) for the delivery of the RoF signal. Downconverted IF signals are transmitted over the optical fiber cable to an analog-to-digital converter (ADC) and a digital signal processor (DSP) set in a central office. In this configuration, each TRx can be operated independently. Thus, the DSP provides functionality based on the multistatic configuration, such as the enhancement of the ranging and increase in the probability.

On the contrary, a configuration that separates the Tx and Rxs is a promising solution to enhance the detection probability and reduce the complexity of the optical fiber network configuration. The Tx should have only one input port for the optical LO/RF signal from the optical SG. The distributed receiver configuration has a passive double star network topology. Of course, with this configuration, the DSP can improve the ranging precision. Moreover, beamforming detection is also realized on the receiver side by tuning the LO transmission line distance to each Rx. A coherent summation of the IF signals in the ADC/DSP domain is also available when the optical fibers have close lengths within the small phase differences in the radar section. Therefore, a multistatic radar system with a Tx-Rx separation configuration has advantages for enhancing the function by optimizing the optical network domain. This study evaluated a configuration with one Tx and two Rxs in the terahertz band.

Optical path length management is also a key to realize the coherent summation of the signals received in Rxs. A measurement technique for the path differences will be discussed in Sect. 4.

3 Terahertz Bistatic Radar Demonstration

3.1 Experimental Setup

To conduct proof-of-concept experiments on a bistatic radar system in the 300 GHz band, an experimental setup was configured as shown in Fig. 2. In this system, a Tx and two Rxs are set under a quasi-coaxial configuration to evaluate the improvement in the detection probability. The laser diode (LD) of an optical signal generator provides a CW lightwave signal at a wavelength of 1548.5 nm. A Mach–Zehnder interferometer-type intensity modulator (MZM) operated by an arbitrary waveform generator (AWG) modulates the CW light to produce an optical frequency comb signal. The AWG generates a saw-type FM-CW electrical signal at a center frequency of 12.5 GHz with a chirp bandwidth of 0.8 GHz and pulse duration of 10 μs. An erbium-doped fiber amplifier (EDFA) boosts the optical power level output from the MZM for launching to an optical filter bank (OFB). The OFB picks up two optical components from the optical frequency comb signal with a frequency separation of 300 GHz. Then, the EDFA set after the OFB optimizes the optical power for distribution. First, the

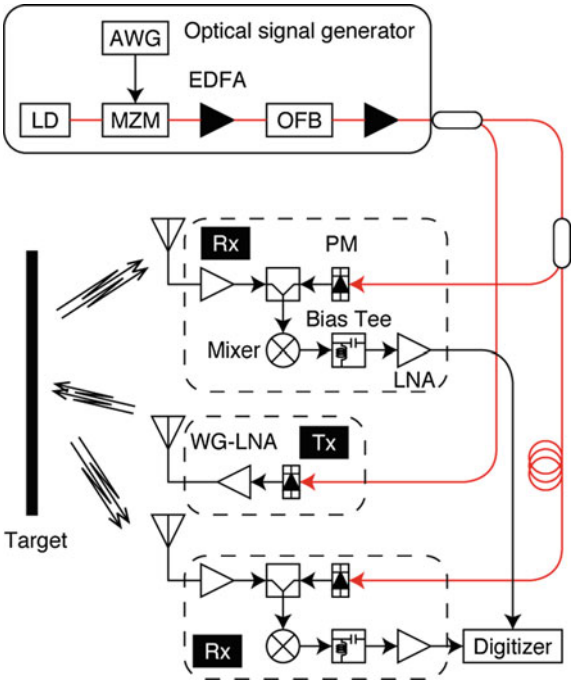


Fig. 2. Experimental setup

optical coupler splits the signal into two components: one for the Tx and the other for the Rxs. In the Tx, a photomixer (PM) receives the 300-GHz-separated two-tone optical signal and converts it into the 300-GHz signal. A waveguide-type low-noise amplifier (WG-LNA) optimizes the power level of the signal at 300 GHz to −10 dBm. In contrast, each Rx is configured with a PM, directional coupler, WG-LNA, 300-GHz fundamental mixer, bias-tee, and low-noise amplifier (LNA) for the IF component. A reflected 300-GHz signal captured by an antenna is input to the coupler and combined with an LO component converted by the PM. The mixer performs a frequency downconversion to the IF component. It should be noted that the mixer is operated with a bias voltage, and the bias-tee is inserted at the IF port of the mixer. The IF signal is amplified by the LNA, and finally, a digitizer acquires the time-evolution data of the IF signal. To easily identify the IF signal, the two Rxs have different fiber lengths for the LO signals. The insertion of the fiber can shift the IF beat frequency of the received FM-CW signal. In the proof-of-concept experiments, the fiber length difference was set at 5 m. Each antenna had an antenna gain of approximately 23 dBi, with a pyramidal horn structure. A metal plate was used as a target in the test.

3.2 Results

Based on the results of the optical domain evaluation, Fig. 3 shows the optical spectra of the generated optical frequency comb signals. Using this system, we obtained ±14th

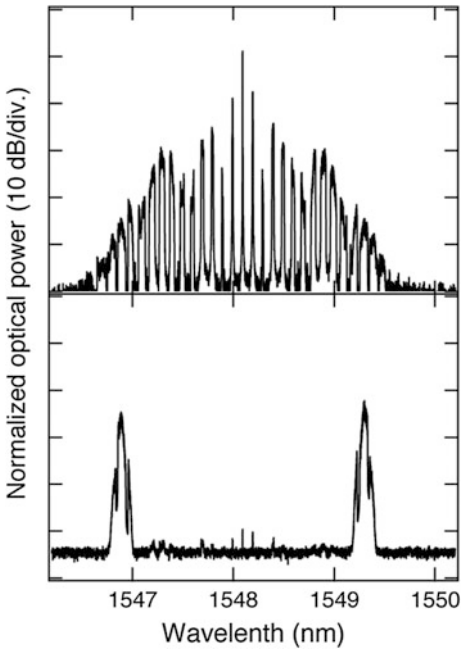


Fig. 3. Obtained optical spectra of (top) FM-CW-signal-launched optical frequency comb and (bottom) two-tone signals passed after OFB

order components using over-driven MZM modulation. However, when the order of the harmonic component was less than the 12th, a separation was observed between adjacent channels. This separation gradually disappeared as a result of overlapping with neighboring channels. This was because the bandwidth of the 0.8 GHz FM-CW signal provides an 11.2 GHz bandwidth at the 14th-order harmonics, which came close to a comb separation of 12.5 GHz. For the 300 GHz signal generation, ± 12th-order components were suitable for the generation of two-tone components with a separation of 12.5 GHz × 12 × 2 = 300 GHz: 24th-order multiplication in the frequency domain. The passband of the OFB was approximately 15 GHz, and the 300 ± 7.5-GHz signal could be converted by the PM. It should be noted that leaked components from adjacent channels were also converted by the PM. However, the suppression ratio of the unwanted components was larger than 10 dB, and therefore, they could not cause significant degradation of the radar configuration.

Figure 4 shows the obtained IF spectrum and ranging results with various target distances from the Tx under the configuration with one Tx and one Rx. In the received signal, an IF peak structure has a signal-to-noise ratio greater than 10 dB. There is no significant spurious component in the obtained signals. In addition, the IF peak frequencies with the various target distances have linear relationships. Thus, the FM-CW ranging was successfully demonstrated in 300 GHz band using a fundamental-mixer-based Rx. It should be noted that the power of the received signals was not calibrated because of the high frequency of 300 GHz. The conversion efficiency of the receiver

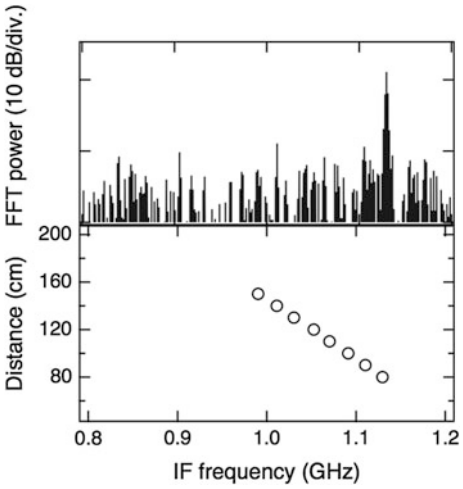


Fig. 4. (top) IF spectrum obtained by fast Fourier transform of digitized Rx signal under target distance of 80 cm and (bottom) ranging results with various target distances

mixer and gain of the IF amplifier were approximately 4000 V/W and 40 dB, respectively. On the contrary, the power observed at an IF frequency of 1.15 GHz in Fig. 4 (top), which is obtained by a fast Fourier transform (FFT), is −90 dBm. Thus, the power input to the heterodyne mixer of the corresponding peak could be estimated at approximately −62 dBm. This high conversion loss degrades the resultant SNR.

For a bistatic radar evaluation, we used two Rxs for the same target detection. Figure 5 shows the observed intermediate frequency spectra in the Rxs. In this configuration with two Rxs and a metal plate as a target, a small misalignment of the reflection angle could prevent the reflection of the signal to the Rxs. For instance, the IF signal could be detected at one Rx at a target distance of 100 cm, but not at the other (see the red signals in Fig. 5). In principle, increasing the Rxs would directly improve the precision of the ranging results because the FFT-based IF frequency estimation limits the range resolution by a bin of the FFT size. The resulting ambiguity could be reduced using statistical techniques. In this case, a detectable Rx signal helps to estimate the undetected signals at the other Rxs. Therefore, it also helps to enhance the precision of the ranging system. Thus, a bistatic configuration for the 300 GHz band could be used for probability-optimized radar systems.

The obtained SNR of 10 dB in the experiments was limited by the low input power of the reflected signals. It should be noted that the power input to the mixer should be less than −40 dBm. Thus, the dynamic range of the system would be estimated to be 30 dB. Increasing the bit resolution of the scope from 8 bits to 16 bits and optimizing the antenna and amplifiers for the terahertz transceiver would help to realize a dynamic range comparable to an X-ray scanner of 40–60 dB.

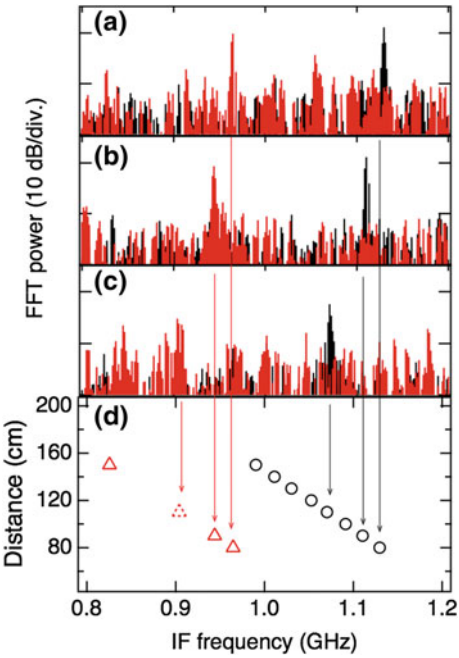


Fig. 5. **a–c** IF spectra obtained by FFT of digitized Rx signal with several target distances and **d** ranging results with various target distances. Red and black lines and circles show signals obtained from different Rxs

4 Path Length Management Technique

Management of the optical fiber path lengths is a key for optimizing the coherence of the radar signals transmitted in the RoF network. This is because higher functionality for a radar system such as a synthetic aperture radar system requires high coherency or phase information about the signal irradiated from the antennas. In a sense, a precise measurement technique for the fiber lengths is required for the RoF-based signal distribution network. However, it would not be easy to install additional equipment for such measurements in an existing RoF network because of the SNR management of the fiber network. Moreover, inputting an additional signal into the optical fiber may cause unwanted degradation from nonlinear effects. Thus, the transmitted FM-CW-modulated RoF signal should be used for both the radar and optical fiber length measurement. Figure 6 shows a concept block diagram of the simultaneous use of the FM-CW RoF signal for the fiber length measurement. The FM-CW RoF signal is passed through an optical circulator before transmission in the fibers. At a remote radar head, the RoF signal is split into two components using an optical tap coupler: one for the radar signal generation and the other for the fiber length measurement. For the length measurement, the RoF signal is input to an optical switch (SW) and then into a Faraday mirror (FM). The reflected signal from the FM is transmitted in the same fiber, and finally, the reflected RoF signal arrives at a central office. After mixing with the original RoF

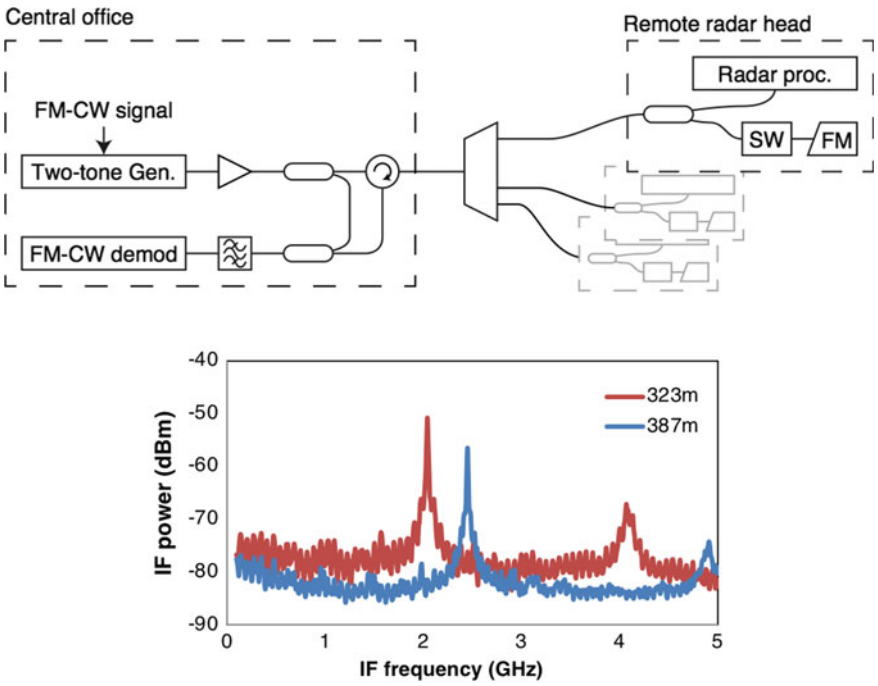


Fig. 6. (top) Conceptual block diagram of fiber length measurement system using optical FM-CW radar technique and (bottom) example IF signal spectra for different fiber lengths with 19.2 GHz bandwidth FM-CW signals

signal using a photodiode, the IF signal is obtained and analyzed in the same manner as the FM-CW radar detection. For multiple radar heads, the SWs are managed by the centralized controller under a time-division-multiplexed system to avoid interference.

Figure 6 also shows the obtained IF spectra with different fiber lengths [11]. The clear separation of the IF peaks is observed to identify the fiber lengths. It should be noted that the bandwidth of the FM-CW signal is 19.2 GHz for a proof-of-concept demonstration. This broader bandwidth provides a high-range resolution of approximately 0.5 cm at a fiber length. Even in a real radar application using the RoF system, the bandwidth of the optical FM-CW signal is approximately 1.3 GHz, which corresponds to an expected range resolution of 7.6 cm [7]. Its temporal resolution of 390 ps can maintain the synchronization of 10 μs class or shorter repetition rate FM-CW signals. It is believed that the installation of a variable optical delay line such as a fiber stretcher in each fiber cable will optimize the fiber length differences in the RoF-based distribution network.

5 Conclusion

A terahertz bistatic radar system based on an RoF network was discussed, configured, and demonstrated experimentally. An optical frequency comb source with a simple power splitting network easily enhanced the target detection probability of the radar system. Optical fiber length measurement in the RoF system was also evaluated with the simultaneous use of the optical FM-CW signal for future high-functional multistatic radar systems. The RoF-enabled multistatic radar system will enhance the detection probability as well as the detection of a complex-surface-structured target.

Acknowledgements. This study was conducted as a part of the "Research and development for expansion of radio wave resources," supported by the Ministry of Internal Affairs and Communications (MIC), Japan; the MIC/SCOPE #165003010; and JSPS Grants-in-Aid for Scientific Research Grant Number JP16K06406.

References

1. Ahmed SS (2014) The state of the art in personnel screening with mm-wave technology for security checkpoint. In: Defense, security and space forum in European microwave week 2014, Rome, Italy, September 2014
2. Hanle E (1986) Survey of bistatic and multistatic radar. IEEE Proc F Commun Radar Signal Process 133(7):587–595
3. EuroCAE ED-235 (2016) MASPS for foreign object debris detection system
4. Yin C, Xu S, Wang D (1996) Location accuracy of multistatic radars (TR^n) based on ranging information. In: Proceedings of CIE international conference radar, October 1996, pp 34–37
5. Yonemoto N, Kohmura A, Futatsumori S, Uebo T, Saillard A (2009) Broad band RF module of millimeter wave radar network for airport FOD detection system. In: 2009 international radar conference on surveillance for a safer world, Bordeaux, France
6. Kanno A, Kawanishi T (2014) Broadband frequency-modulated continuous-wave signal generation by optical modulation technique. J Lightw Technol 32(20):3566–3572
7. Kawanishi T, Kanno A, Shibagaki N, Yonemoto N, Angkaew T, Janpugdee P (2015) Field trial of radio-over-fiber based high-resolution radar. In: Proceedings of Thailand-Japan microwave, Bangkok, Thailand, Paper FR4-03, August 2015
8. Shibagaki N (2017) Experimental study of photonic based radar for FOD detection systems using 90 GHz-band. Air Traffic Manag Syst 420:239–248
9. Gente R, Jansen C, Geise R, Peters P, Gente M, Krumbholz N, Moeller C, Busch S, Koch M (2012) Scaled bistatic radar cross section measurements of aircraft with a fiber-coupled THz time-domain spectrometer. IEEE Trans Terahz Sci Technol 2(4):424–431
10. Kanno A, Sekine N, Kasamatsu A, Yamamoto N, Kawanishi T (2017) Optical-fiber-connected 300-GHz FM-CW radar system. In: Proceedings of SPIE. Passive and active millimeter-wave imaging XX, vol 10189, pp 101890F-1–6
11. Kanno A, Yamamoto N, Ohara K, Sotobayashi H, Kawanishi T (2018) Fiber length measurement for linear-cell distributed radar systems based on radio over fiber technique. In: IoT enabling Sensing/AI/Network and photonics conference in optics and photonics international congress, Yokohama, Kanagawa, IoT7-4, April 2018

HF-START: Application in Aid of Radio Communications/Navigation

Kornyanat Hozumi[1(✉)], Mamoru Ishii[1], Susumu Saito[2], Takashi Maruyama[1], Hiroyuki Nakata[3], and Takuya Tsugawa[1]

[1] Space Environment Laboratory, NICT, 4-2-1 Nukui-kita, Koganei, Tokyo 184-8795, Japan
{kukkai, mishii, tmaru, tsugawa}@nict.go.jp
[2] National Institute of Maritime, Port and Aviation Technology, ENRI, 7-42-23 Jindaiji-Higashi machi, Chofu, Tokyo 182-0012, Japan
susaito@enri.go.jp
[3] Graduate School of Engineering, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba-shi, Chiba 263-8522, Japan
nakata@faculty.chiba-u.jp

Abstract. HF-START is HF Simulator Targeting for All-users' Regional Telecommunications. This paper addresses the problems of worldwide existing radio propagation model and challenge on developing the HF-START. Its possible extension in the future is discussed. Meteorological information for aircraft in flight is broadcasted from worldwide radio stations in the frequency range of 3–15 MHz. Such frequency range is strongly affected from day-to-day bottom structure variation of the Earth's ionosphere, where it is influenced by both space weather and upper atmosphere activities. Space weather is thus significant to aeronautical users, who deal with the critical radio application. Space weather data is, however, difficult to be understood. To translate research level data to user level data, radio propagation simulator named HF-START is being developed. Space weather disturbances-triggered failure of communications/navigation is of high priority to forecast.

Keywords: Radio propagation simulator · HF-START · Space weather · Aeronautical users

1 Introduction

The aviation systems, currently, have been designed as self-contained systems for easily assessing the safety. However, disturbances associated with space weather are reported to limit the availability of those systems, especially at critical phases due to limited space weather and radio propagation information [1–3]. This is a large impact on worldwide economic situation and safety issue, and has been reviewed by the International Civil Aviation Organization (ICAO) meteorology expert group at a global level [4].

Even though satellite communication is getting popular, GEO satellites are not visible from polar region and the system itself is expensive. For aeronautical

Electronic Navigation Research Institute, *Air Traffic Management and Systems III*, Lecture Notes in Electrical Engineering 555,
https://doi.org/10.1007/978-981-13-7086-1_19

communications, High-Frequency (HF) radio communication is a classic and important means, especially for airplanes on oceanic en route and in polar routes. Moreover, meteorological information for aircraft in flight or VOLMET (French origin VOL (flight) and METEO (weather)) including Terminal Area Forecasts (TAF), Meteorological Terminal Aviation Routine Weather Report, or Meteorological Aerodrome Report (METAR) and SIGnificant METeorological Information (SIGMET), is broadcasted from worldwide radio stations in the frequency range of 3–15 MHz. Such frequency range is strongly affected from day-to-day bottom structure variation of the Earth's ionosphere, where is influenced by both space weather and upper atmosphere activities. Space weather is thus significant to aeronautical users, who deal with the critical radio application.

Space weather is the time-varying electromagnetic conditions beginning from the Sun to the vicinity of the Earth. Figure 1 shows the schematic picture of space weather. The Sun is a natural fusion plant that emits not only visible light and heat, which are essential for the life living, but also harmful radio emission and hot plasma named solar wind. The Earth has two Sun barriers, which are the magnetic field in the magnetosphere and the Earth's atmosphere. The protection of the magnetic field is, however, not perfect. When the magnetic field in the solar wind is southward, the solar wind plasma can penetrate the magnetosphere and affect the in-orbit spacecraft. In addition, it triggers ionospheric disturbances and induced currents particularly in the polar region that affect the power grid system [5, 6].

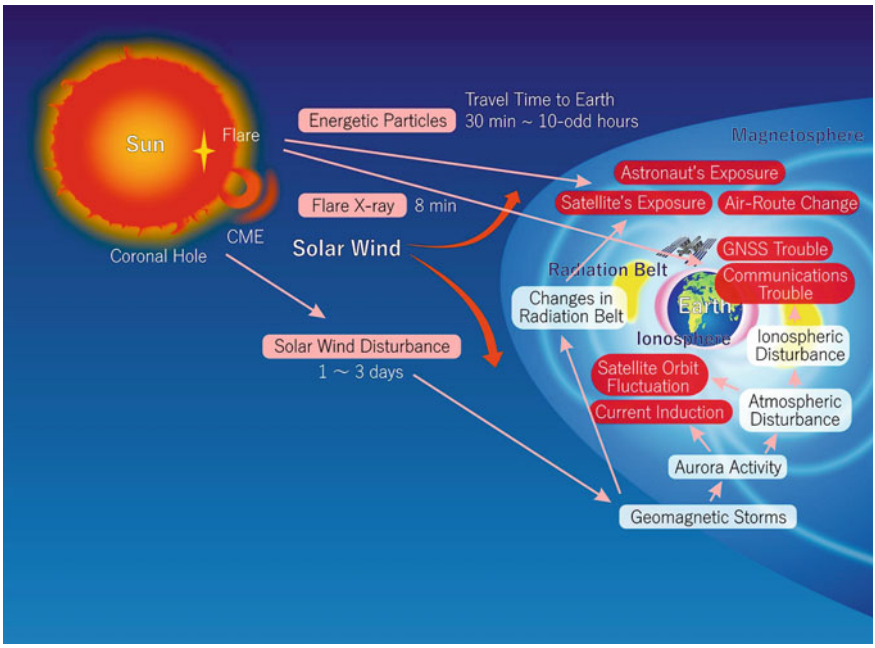


Fig. 1. Schematic picture of space weather

There are three categories of space weather impacts on aviation including HF telecommunications, Global Navigation Satellite System (GNSS) applications, and human radiation. The timeline of the social impact of space weather is shown in Table 1. At the first stage, about eight minutes after the solar eruption named "flare", the radio emissions, e.g., X-ray and Ultraviolet (UV) arrive at the Earth and enhance the highly absorbing layer, a so-called D-layer (around 60–80 km of altitude), of the Earth's ionosphere. Such absorption causes the disabilities of HF telecommunications. At the second stage, high-energy particles (proton and electron) arrive from 30 min to 2 days after the solar flare and make exposure for satellites, astronauts, and aircrews in high altitude and/or latitude. The Polar Cap Absorption (PCA) also occurs with the high-energy particles, which introduces the disabilities of HF telecommunications. At the third stage, high density and velocity ionized gas named Coronal Mass Ejection (CME) arrives to the Earth within 2 or 3 days after the solar flare. CME interacts with the Earth's magnetosphere when the magnetic field in CME is southward. The energy inputs into the magnetosphere consequently. It triggers the Earth's ionospheric disturbances those make the disability of HF telecommunications and increase error of GNSS. On the view of the aviation, the HF telecommunications and GNSS are affected at all stages, and human exposure in the high latitude is critical at the second stage.

Table 1. Three stages of space weather phenomena from the solar flare

	Time from solar flare	Source of the effect	Space weather phenomena	Social impact
1	~8 min later	Radio wave e.g., X-ray and ultraviolet	Dellinger effect (increase of electron density in D-region), solar radio burst	Disability of HF telecommunications/broadcast, increase errors/disability of GNSS, disability of airport surveillance radar, disability of mobile phone
2	~30 min to 2 days	High-energy particles	Exposure of satellite and human in high altitude, PCA (polar cap absorption)	Bit exchange on board in satellites, human exposure on high altitude/latitude aircrew, disability of HF telecommunications/broadcast
3	~2–3 days	CME (Coronal mass ejection)	Geomagnetic storm, ionospheric storm, expansion of atmosphere	Disability of HF telecommunications/broadcast, increase errors/disability of GNSS, impact on power grid, satellite anomaly, impact on trajectory of low-altitude satellite and space debris, extension of aurora area

The preparation of documents related to the space weather in ICAO [4] has been discussed since 2009 temporally in the working group "IAVWOPSG" (International

Airways Volcano Watch Operations Group). The structure of the working group was changed in 2014 and the discussion has been continued in a permanent subgroup in "WG-MISD" (Working Group-Meteorological Information and Service Development) [7]. There are three documents to be prepared in ICAO related to space weather: (1) Standard and Recommended Practices (SARPs) [4], (2) Concept of Operation (ConOps), and (3) space weather information manual. SARPs and ConOps have been approved in the Meteorology Panel of ICAO in July 2016. The discussion of manual continues and the approval of the manual is planned to be in 2018.

As a parallel process of the preparation of documents, the selection of ICAO Space Weather Center [7] has been discussed since about 2009. The center criteria were discussed and approved on the ICAO meteorology panel in 2016 [8]. The criteria are divided into the following four areas: (1) Institutional; (2) Operational; (3) Technical; and (4) Communications/Dissemination.

The technical criteria pertaining to the ability of a space weather information provider to provide the information required for the space weather information service. This includes the ability to meet all of the information provision requirements defined in the drafted SARPs and harmonize information with the space weather information providers for adjacent geographic areas, as necessary. The communications/dissemination criteria are intended to ensure that any potential global space weather information provider is able to distribute the global advisory message product to aviation decision makers through both such as OPerational METeorological (OPMET) data centers, and newer means of dissemination that is primarily Internet-based platform. Based on the criteria, ICAO sent a state letter to the member states to ask the interest of operating ICAO space weather center. The reply was closed on September 8, 2017. The selection process includes audit visits by the World Meteorological Organization (WMO) that will be completed by July 2018 and the operational provision of space weather information that will be started by November 2018.

Since variations of the ionosphere could, as noted above, change conditions of communications and navigation for aeronautical applications, information on the ionospheric conditions composes of an important part of the space weather information for aviation. Because the ionosphere has different characteristics at different regions, local ionospheric data should be properly inputted into the space weather information. In low latitudes, for example, ionospheric irregularity called plasma bubble frequently occurs. A well-developed plasma bubble structure poses problems on communications and navigation systems, i.e., Global Positioning System (GPS) delay, positioning error, failure of using SATellite COMmunications (SATCOM), etc. If it is not the worst-case scenario that the communications turns blackout, the Radio Frequency (RF) receiver with two or more frequencies can compensate such ionospheric error. However, most of the satellite communications/navigation systems equip only single frequency because of the certification and cost issues. Only the L1 signals of GPS and GLObal NAvigation Satellite System (GLONASS) have been standardized by ICAO. Though SARPs for Multi-Frequency and Multi-Constellation (MC-MF) GNSS is being discussed at ICAO, standardization and promulgation of MC-MF GNSS avionics would take many years. Therefore, the lower structure of plasma bubble is important as a precursor of a structure evolution of plasma bubble. Most areas in low latitudes near the Japanese meridian are Oceans where lack of ionospheric observation. Even though HF

Trans-Equatorial Propagation (HF-TEP) observation, which is able to provide azimuth and elevation angles of the refracted wave, is an important means to monitor the lower structure of the plasma bubble over the Ocean that is penetrating Japanese region [9], lack of information of altitude where the HF wave refracts causes an uncertainty of its usage. To predict an influence of variations of the ionosphere on aeronautical communication and navigation systems and provide necessary space weather information on the radio propagation, a radio propagation simulator that can account for the temporal and spatial variations of the ionosphere is one of the solutions.

One of the most popular radio propagation simulators around the world is Voice of America Coverage Analysis Program (VOACAP). It was supported by the US government. It can predict the RF propagation in a global scale. However, VOACAP has never been evaluated over Japan and Southeast Asia, for example. Also, there is no option to input the local ionospheric information. Then, the question raised from the research community is that how reliable is the VOACAP, especially in local scale.

Space weather users, who are from various fields, are increasing. No matter how good space weather data we have, it is meaningless if users cannot understand and effectively use it. To overcome this issue, research level data must be translated to user level data based on the users' needs in order to bridge over the valley of death between academic science and user needs. For this purpose, a new research framework in Japan named "Project for Solar-Terrestrial Environment Prediction (PSTEP)" was built based on a Grant-in-Aid for Scientific Research on Innovative Areas in 2015 [10]. The radio propagation simulator is one of the four products from PSTEP that have been being constructed for space weather users.

This paper introduces the PSTEP product, the local radio propagation simulator named HF-START, as a potential application in aid of radio communications/navigation. It is being developed mainly for Japan and Southeast Asia because our institute, National Institute of Information and Communications Technology (NICT), is operating ionospheric observational networks only in Japan and Southeast Asia. Coverage regions of the HF-START can be extended in the future if the ionospheric input from other regions becomes available and the HF-START can be properly evaluated in those regions. HF-START is a user-friendly HF propagation simulator that dedicates to the user's needs. Space weather disturbances-triggered failure of communications/navigation is of high priority to forecast. Currently, we simulate the HF-START and perform verification by assuming specific HF applications and evaluation by using the indices that users need for each application. Even if we keep developing the simulator for HF currently, the focused frequency range is planned to cover Very High Frequency (VHF) and microwave ranges as well.

2 Methodology

HF-START is implemented by the ray-tracing technique. The ray-tracing equations are derived from Maxwell's equations assuming that the free space wavelength goes to zero [11]. The observation point of the ray trajectories $(\vec{r})$ is started at $\bar{r} = \{x, y, z\}$. An initial direction vector $(\vec{n})$ of the ray is $\vec{n} = \{n_x, n_y, n_z\}$. The $|\vec{n}|$ is a refractive index n.

The propagation medium is defined by a Hamiltonian as a function of a refractive index as follows:

$$H(\vec{r},\vec{n},t) = \frac{\sqrt{n_x^2 + n_y^2 + n_z^2}}{n(\vec{r},\vec{n},t)} - 1 = 0 \tag{1}$$

where $n(x, y, z, n_x, n_y, n_z)$ is derived from the Appleton–Hartree formulation [12]. The ray-tracing equations are

$$\frac{dr}{d\tau} = \frac{c}{\omega}\nabla_n H(\vec{r},\vec{n},t) \tag{2}$$

and

$$\frac{dn}{d\tau} = -\frac{c}{\omega}\nabla_n H(\vec{r},\vec{n},t) \tag{3}$$

where τ is the group delay along the ray path, t is the time dependence, and ω is an angular frequency of the wave. Only, ordinary (O-mode) solutions to the ray and the reflective index are chosen.

For realistic simulation of local radio propagation, the background ionosphere that is a propagation medium is reconstructed from the real observation data. Figure 2 describes the model concept that the ionosonde data is used for the current simulator and the GNSS tomography, as well as the Ground-to-topside model of Atmosphere and Ionosphere for Aeronomy (GAIA) model [13], will be used in the future version. This allows retrieving a precise description of the local ionospheric profile, which is important for the interpretation of the HF propagation distance. Currently, the vertical ionospheric profile is reconstructed by using a Chapman profile for the F2 region and Gaussian functions for the F1 and the E regions [14]. Electron density profile in the F2 region is given by

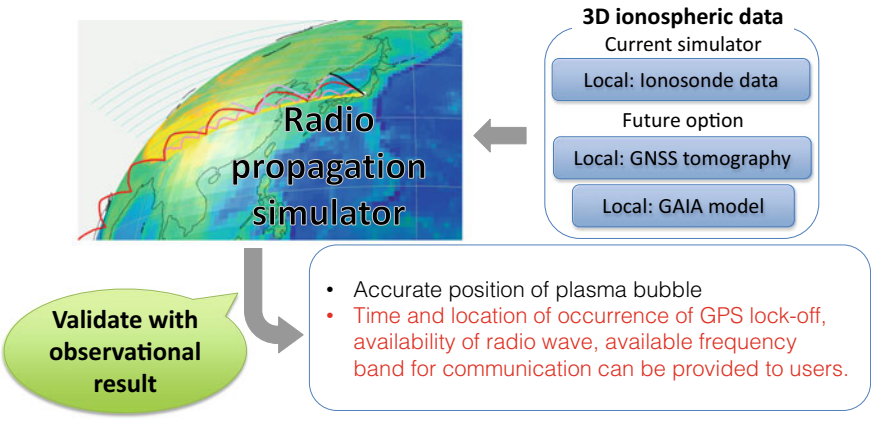


Fig. 2. Schematic figure to explain the model concept of HF-START

$$n_{e\_F2}(h) = \left(\frac{f_{F2}}{0.00898}\right)^2 \exp\left[\frac{1}{2}(1 - z_{F2} - e^{-z_{F2}})\right], \tag{4}$$

$$z_{F2} = \frac{h - h_{F2}}{H_{F2}} \tag{5}$$

where n_{e_F2} is electron density in the F2 region in cm^{-3}, h is any given height in the vertical profile in km, f_{F2} is plasma frequency at the F2 peak altitude in MHz, h_{F2} is peak height of the F2 region in km, and H_{F2} is scale height of the F2 region. Electron density profile in the F1 and E region is given by

$$n_{e\_L}(h) = \left(\frac{f_L}{0.00898}\right)^2 \exp\left(-z_L^2\right), \tag{6}$$

$$z_L = \frac{h - h_L}{H_L} \tag{7}$$

where n_{e_L} is electron density in the F1 (E) region in cm^{-3}, f_L is plasma frequency at the F1 (E) peak altitude in MHz, h_L is peak height of the F1 (E) region in km, and H_L is the scale height. The vertical electron density profile is estimated by

$$n_e(h) = n_{e\_F_2}(h) + n_{e\_L}(h). \tag{8}$$

In the Sect. 3.1 **Preliminary results**, the background ionosphere is assumed to be a perfect reflector and static. It is determined by 1401 × 3 × 721 grids with the grid size of 0.5 km, 0.5º, and 0.5º for altitude, latitude, and longitude, respectively. Equations 4–8 are employed to assimilate the 3D ionosphere.

In the Sect. 3.2 **Propagations during ionospheric perturbations**, the background ionosphere is assumed to be a perfect reflector but not static. It is determined by 63 × 17 × 20 grids with the grid size of 10 km, 5º, and 5º for altitude, latitude, and longitude, respectively. Equations 4–8 are employed to assimilate the 3D ionosphere. The ionospheric perturbation that frequently occurs in the Japanese region, a so-called Traveling Ionospheric Disturbance (TID) is consequently added to the 3D ionosphere by

$$n_{e\_TID} = w \cdot n_e(h) \cdot \left(1 + \sin\left(\frac{l}{2}y\right)\right), \tag{9}$$

where n_{e_TID} is 3D electron density with TID structure, w is weight, l is peak-to-peak longitudinal coverage of the TID in degree, and y is any given longitude in degree. It should be noted that this is a very simple representation of TIDs for evaluation of the HF-START. This model does not account for the real structure of TIDs that is aligned with the magnetic field line with about 45° inclination angle.

The target of the HF-START is to evaluate the ionospheric effect on shortwave communications. By using HF-START in aid of other ground-based observations,

accurate position of plasma bubbles in low latitudes will be clarified in the future version.

3 Results and Discussion

3.1 Preliminary Results

Altitude and plasma frequency information that can be obtained from the ionogram is used as an input of the HF-START. The peak heights of the F2, F1, and E layers were set at 255.58 km, 190 km, and 110 km, respectively. The critical frequencies of the F2, F1, and E layers were set at 8 MHz, 5 MHz, and 3.6 MHz, respectively. By employing Eqs. 4–8, the vertical profile is reconstructed as shown in Fig. 3.

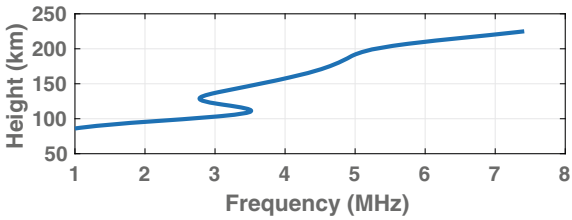


Fig. 3. Reconstructed ionospheric profile

The 3 and 7 MHz rays were launched with elevation angle of 80º. 10-hop propagation was set. The results of both frequencies are shown in Fig. 4a, b. The red dash line illustrates the Earth's curvature. It should be noted that the "Height" in the y-axis is height in the local coordinate at the origin of the radio wave transmission. The reflection height is found to depend on the plasma frequency. With the provided ionospheric condition, the reflection height of the 3 MHz is seen at about 100 km altitude, while the reflection height of the 7 MHz is seen at 220 km altitude. At 10th hop, the propagation distance of the 7 MHz goes beyond that of the 3 MHz because of its higher reflection height. Because the absorption is ignored in the current simulator, the reflection height is not changed even though the hop number progresses.

3.2 Propagations During Ionospheric Perturbations

Figure 5 shows the 3D ionospheric structure with TID reconstructed by using Eqs. 4–9. The peak heights of the F2, F1, and E layers were set at 350 km, 200 km, and 100 km, respectively. The scale heights of the F2, F1, and E layers were set at 70, 40, and 20, respectively. The critical frequencies of the F2, F1, and E layers were set at 15 MHz, 4 MHz, and 3 MHz, respectively. The peak-to-peak longitudinal coverage of the TID was set to be 20° in the F2 layer. In addition, random structure, which is assumed to represent any possible ionospheric instability below the F2 layer, was added to altitudes between 70 and 200 km. The random structure has random value for parameters h, l, and y in Eq. 9; and h is randomly set in the range of 70–200 km.

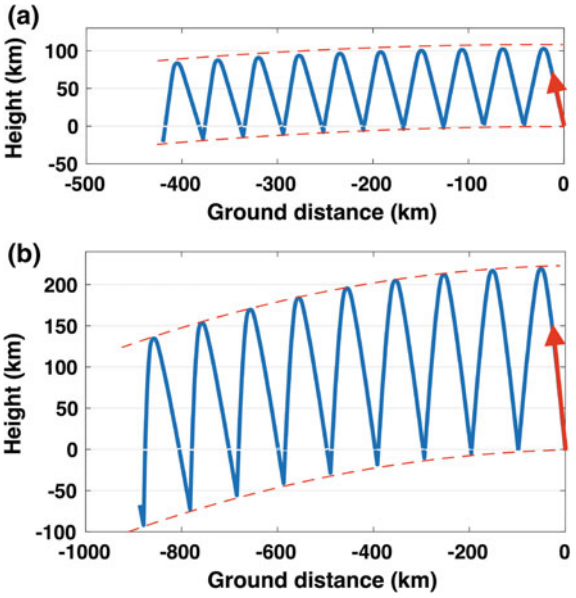


Fig. 4. Radio propagation results for **a** 3 MHz and **b** 7 MHz. The "Height" in y-axis is that in the local coordinate at the origin of the radio wave transmission

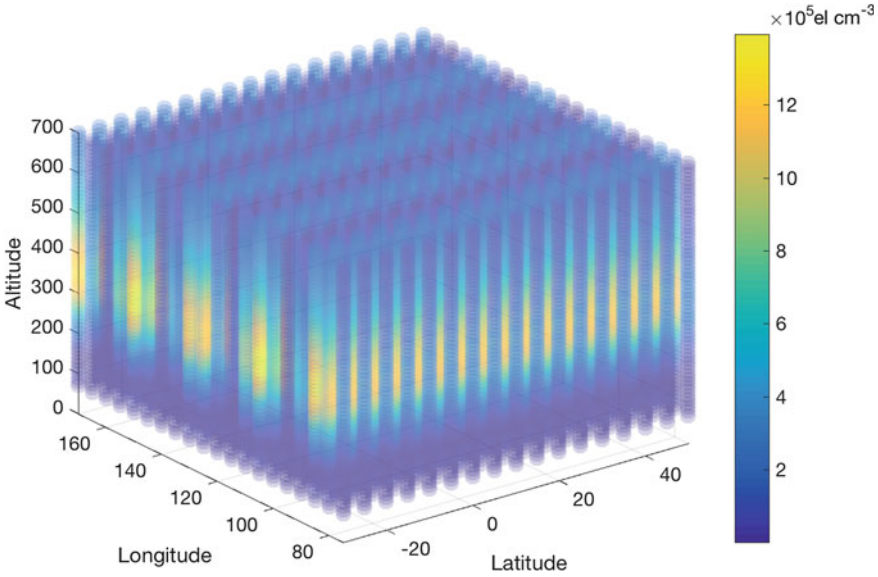


Fig. 5. Assimilated 3D ionosphere with TID structure

Figure 6 illustrates the side view of the ray path. Multi-hop mode of the 15 MHz frequency was launched at the location shown as the red star in Fig. 6. Taking the side view into account, the ray penetrates the ionosphere and reflects back to lower altitude but does not reach the ground. Instead, the ray reflects from the aforementioned random structure back to the higher altitude. Then, it reflects back from the F2 layer to the ground. If the ground-to-ground propagation is considered as one hop, the ray from the Tx location to the current location as seen in the dash red square in Fig. 6 is considered as one-hop propagation. Because ionospheric layers act as a waveguide, this one hop with duct mode propagation takes longer propagation distance and time is compared to the ray that does not experience such an ionospheric waveguide condition. With the longer propagation distance and time, the ionospheric waveguide could lead to larger skip zone, loss of communications, and higher signal attenuation in dB per hop. Because the absorption is ignored in the current simulator, the attenuation issue is not clarified.

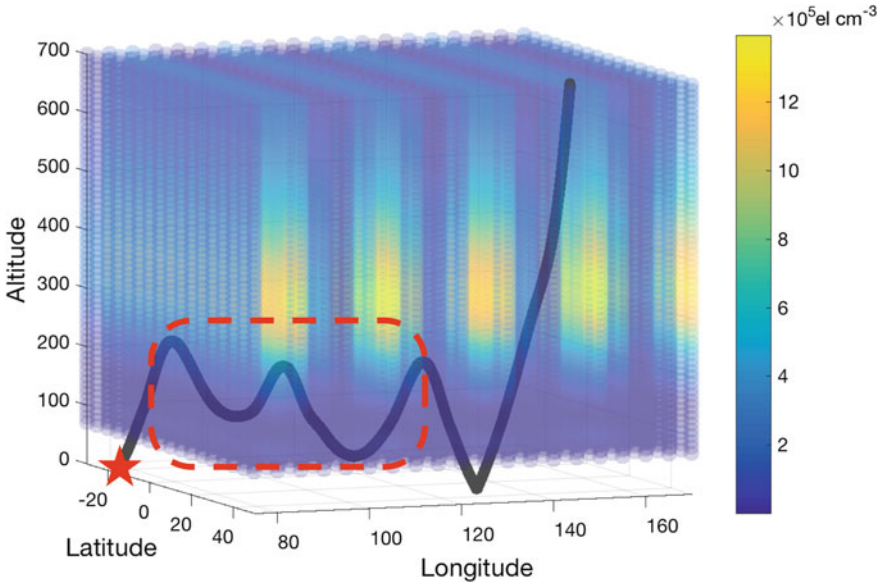


Fig. 6. The side view of the ray path of the 15 MHz frequency propagation during the TID

Japan is known to be a favorable location for sporadic E (E_s) to occur [15]. E_s and Medium-scale TID (MSTID) in Japan were found to have positive correlation [16, 17]. Saito et al. [16] reported coexistence of E_s and MSTID observed by different ground-based observational tools. Otsuka et al. [17] suggested that simultaneous existence of E_s and MSTID is a result of coupling process between E and F regions. If the E_s is treated as the aforementioned random structure in Fig. 5, there might be a risk of loss of communications due to the coexistence of E_s and MSTID. When such an issue happens without the space weather knowledge, it might generally be considered that the problem occurs due to an instrumental failure. It could lead to unnecessary replacement

of the RF receiver onboard the aircraft that is high but waste cost, for example. However, this kind of effect is not widely known and it may be just theoretical risk that needs evaluation. Therefore, it is reasonable to carry on the statistical study of the simultaneous occurrence of TID and E_s in Japan in order to reveal the probability of the duct mode propagation. The RF maintenance cost for the aircraft could reduce safely as a consequential outcome of the study.

3.3 Evaluation Plan

Because HF-START is the first simulator that employs the real ionospheric data observed in Japan as input, observation-based evaluation is mandatory. Evaluation campaign of HF-START has been launched in early 2018. This campaign employs Software-Defined Radio (SDR) based digital receiver in addition to existing domestic ionospheric observations. The campaign involves one Tx station at Nagara by Radio NIKKEI and four Rx stations at Wakkanai, Kokubunji, Yamagawa, and Okinawa as presented in Fig. 7. For each HF link, the experiment for evaluation will be performed every 15 min following the domestic ionosonde observation schedule in Japan. This allows observing the effect of the daily variations of the ionosphere on the measurements. Measurements will be performed at three different frequencies transmitted from radio NIKKEI channel 1 in the HF range of 3.925, 6.055, and 9.595 MHz. Propagation distance can be estimated by the time-lag information that is retrieved from waveform comparison between the two Rx stations those are receiving the same signal.

Fig. 7. Location of Tx (dark blue circle), planned Rx stations for the campaign (red circle), and NICT ionosonde stations (red circle). The light blue strips show the possible radio path of each radio link

3.4 Future Plan

In the future, the HF-START will be developed to be more realistic for real operation and service. The absorption will be included in the propagation simulator. Then, the

solar flare effect on the aeronautical communications and navigation will be clarified. Various conditions of ionospheric background such as during plasma bubble event will be tested. An uncertainty issue of the use of HF-TEP for monitoring of the bottom structure of plasma bubble will be solved. Then the HF-TEP can be used for the ICAO space weather information service. As shown in Fig. 8, the input space for 3D electron density will include not only the ionosonde information but also the near real-time GNSS tomography [18] and optionally, the ionospheric information predicted from an Artificial Intelligence (AI). Because the AI learns the prior information obtained from Japan [19, 20] and Southeast Asia [21, 22], ionospheric information from the two regions including over the ocean is expected to be predicted well and become a good source of input space for the HF-START. The web application will be developed regarding the user's needs and open to the public in the future.

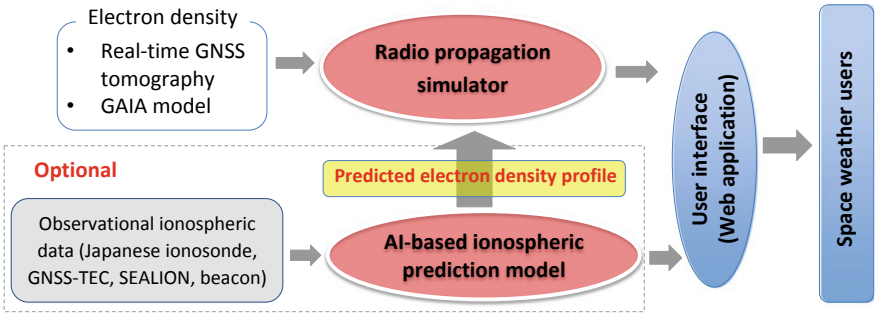


Fig. 8. Schematic figure of the future plan. The AI will be used for prediction of electron density profile. The user-friendly web interface will be developed

4 Summary

HF-START has been developed to incorporate arbitrary 3D ionosphere models to reflect real ionospheric conditions. 3D ionosphere was assimilated from ionosonde data by using Chapman and Gaussian functions. TID and random structures were successfully added to the assimilation in order to reveal the characteristics of the HF propagation with the perturbed ionosphere. The radio propagation parameters such as propagation distance and time delay have been simulated from the HF-START by the 3D ray-tracing technique. The elapsed time of the simulation is only few seconds. In the future, the input space of the simulator will include data imported from near real-time 3D GNSS tomography, GAIA model of NICT, and AI. The web application will be dedicated to the user's needs and open to the public in the future. The HF-START is an application in aid of radio communications/navigation for the space weather users.

Acknowledgements. A part of this work was supported by JSPS KAKENHI Grant Number JP15H05813, "The Project for Solar-Terrestrial Environment Prediction (PSTEP)".

References

1. Blagoveshchensky DV, Kalishin AS, Sergeyeva MA (2008) Space weather effects on radio propagation: study of the CEDAR, GEM and ISTP storm events. Ann Geophys 26:1479–1490
2. Cannon PS et al (2013) Extreme space weather: impacts on engineered systems and infrastructure. Rep., ISBN 1-903496-95-0, Royal Academy of Engineering, London
3. Cannon PS et al (2013) Extreme space weather: impacts on engineered systems and infrastructure – summary report. Rep., ISBN 1-903496-96-9, Royal Academy of Engineering, London
4. International National Civil Aviation Organization (2012) Report of the committee to the conference on agenda item 6. Twelfth Air Navigation Conference
5. Pirjola R, Viljanen A, Pulkkinen A, Amm O (2000) Space weather risk in power systems and pipelines. Phys Chem Earth Part C: Sol Terr Planet Sci 25(4):333–337. https://doi.org/10.1016/S1464-1917(00)00027-1
6. Fulneček J, Mach V, Vramba J (2015) Space weather effects on power grids. In: 2015 16th international scientific conference on electric power engineering (EPE), Kouty nad Desnou, pp 499–502. https://doi.org/10.1109/epe.2015.7161124
7. International Civil Aviation Organization (2016) METeorology Panel (METP) second meeting
8. International National Civil Aviation Organization (2016) Report of the twentieth meeting of the communications, navigation and surveillance of APANPIRG
9. Maruyama T, Kawamura M (2006) Equatorial ionospheric disturbance observed through a transequatorial HF propagation experiment. Ann Geophys 24:1401–1409
10. Ishii M (2017) Japanese space weather research activities. Space Weather 15:26–35. https://doi.org/10.1002/2016SW001531
11. Yeh KC, Liu CH (1972) Theory of ionospheric waves. Academic, New York
12. Jones RM, Stephenson JJ (1975) A versatile three-dimensional ray tracing computer program for radio waves in the ionosphere. U.S. Department of Commerce, OT Report, pp 75–76
13. Jin H, Miyoshi Y, Pancheva D, Mukhtarov P, Fujiwara H, Shinagawa H (2012) Response of migrating tides to the stratospheric sudden warming in 2009 and their effects on the ionosphere studied by a whole atmosphere-ionosphere model GAIA with COSMIC and TIMED/SABER observations. J Geophys Res 117:A10323. https://doi.org/10.1029/2012JA017650
14. Maruyama T, Yusupov K, Akchurin A (2016) Interpretation of deformed ionograms induced by vertical ground motion of seismic Rayleigh waves and infrasound in the thermosphere. Ann Geophys 34:271–278. https://doi.org/10.5194/angeo-34-271-2016
15. Taguchi S, Shibata H (1961) World map of f_oE_s. J Radio Res Lab 8:355–389
16. Saito S, Yamamoto M, Hashiguchi H, Maegata A, Saito A (2007) Observational evidence of coupling between quasi-periodic echoes and medium scale traveling ionospheric disturbances. Ann Geophys 25:2185–2194
17. Otsuka Y, Onoma F, Shiokawa K, Ogawa T, Yamamoto M, Fukao S (2007) Simultaneous observations of nighttime medium-scale traveling ionospheric disturbances and E region field-aligned irregularities at midlatitude. J Geophys Res 112:A06317. https://doi.org/10.1029/2005JA011548
18. Saito S, Susuki S, Yamamoto M, Saito A (2017) Real-time ionosphere monitoring by three-dimensional tomography over Japan. Inst Navig 64:495–504. https://doi.org/10.1002/navi.213

19. Maruyama T (2007) Regional reference total electron content model over Japan based on neural network mapping techniques. Ann Geophys 25:2609–2614
20. Maruyama T (2010) Solar proxies pertaining to empirical ionospheric total electron content model. J Geophys Res 115. https://doi.org/10.1029/2009ja014890
21. Watthanasangmechai K, Supnithi P, Lerkvaranyu S, Tsugawa T, Nagatsuma T, Maruyama T (2012) TEC prediction with neural network for equatorial latitude station in Thailand. Earth Planets Space 64:473–483. https://doi.org/10.5047/eps.2011.05.025
22. Wichaipanich N, Hozumi K, Supnithi P, Tsugawa T (2017) A comparison of neural network-based predictions of foF2 with the IRI-2012 model at conjugate points in Southeast Asia. Adv Space Res 59:2934–2950. https://doi.org/10.1016/j.asr.2017.03.023