Kanya Kusano *Editor*

Solar-Terrestrial Environmental Prediction

Springer

Solar-Terrestrial Environmental Prediction

Kanya Kusano
Editor

Solar-Terrestrial Environmental Prediction

Springer

Editor
Kanya Kusano
Institute for Space-Earth Environmental Research
Nagoya University
Nagoya, Aichi, Japan

ISBN 978-981-19-7764-0 ISBN 978-981-19-7765-7 (eBook)
https://doi.org/10.1007/978-981-19-7765-7

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

Preface

With the development of civilization and technology, the environment of human society is currently expanding even into space. Satellites play important roles in communication, safe aircraft operations, and weather and disaster monitoring. Satellite-based positioning and satellite broadcasting are presently indispensable infrastructures in our daily lives. Concurrently, various space probes have reached many planets and all parts of the solar system. Voyager 1 and Voyager 2, launched in 1977, have already escaped from the heliosphere and are navigating in interstellar space. Hayabusa and Hayabusa 2 have even enabled bringing samples from extra-terrestrial bodies to the Earth. Astronauts stay on the International Space Station continually, and it is expected that ordinary people will soon be able to go to space. There are plans for new manned missions to the moon and Mars. It can be stated that space is now the realm of human activity.

However, solar activity has significant impacts on the space around the Earth (geospace) and interplanetary space. The numerous energetic particles emitted by solar flares pose serious threats to the health of astronauts. In addition, solar flares and coronal mass ejections, which are explosive phenomena occurring on the Sun, significantly disrupt the electromagnetic environment not only in the space but also on the ground, and they may pose potential risks to electric power, communication, and aviation. In addition, many data suggest that long-term changes in sunspot activity, which can influence the space environment, also affect the weather and climate of the Earth. Therefore, the prediction of "space weather," i.e., the variations in solar activity and the space environment, is crucial for modern civilization to continue to develop and expand into space.

This textbook was edited for young researchers and graduate students based on the Project for Solar-Terrestrial Environment Prediction (PSTEP) in Japan, which was supported by the Ministry of Education, Culture, Sports, Science and Technology of Japan. The PSTEP consisted of four planned research groups (A01: Prediction System Group, A02: Solar Storm Group, A03: Terrestrial Electromagnetism

Group, and A04: Solar Cycle Activity Group) and many research units. For more information about the PSTEP and its research results, please refer to its website.

This textbook is divided into four parts corresponding to the planned researches of the PSTEP (Part I: Linkage between Space Weather and Society, Part II: Magnetosphere of Earth and Space Weather Prediction, Part III: Solar Storms and Space Weather Prediction, and Part IV: Prediction of Solar Cycle Activity and Climate Impacts). Because each chapter covers the basics and applications of each area, you can gain a broad understanding of the subject throughout the book. In addition, you can select and read the topics that you are interested in.

The planning of this textbook was realized on the basis of the PSTEP Summer School at Rikubetsu, 2017, which was conducted in Rikubetsu, Hokkaido, Japan. The contents of the summer school were published as a Japanese open textbook by Nagoya University Repository in 2021. This textbook has been newly edited by translating it into English and adding the latest research results.

Finally, I would like to express my sincere gratitude to all authors who collaborated on this project and wrote the different chapters despite their busy schedules. I would also like to thank Ms. Saori Haga for accurately carrying out the enormous amount of editing work and Ms. Motoko Kunieda, Ms. Akie Enomoto, and other members of the Secretarial Office of the Integrated Studies Division of the Institute for Space and Earth Environment Studies at Nagoya University. In addition, I acknowledge Mr. Akiyuki Tokuno and his coworkers from Springer for their continuous support and all of those who cooperated in the planning of the PSTEP and this textbook.

Nagoya, Aichi, Japan Kanya Kusano

Contents

Contributors

Ayumi Asai Astronomical Observatory, Graduate School of Science, Kyoto University, Kyoto, Japan

Yusuke Ebihara Research Institute for Sustainable Humanosphere, Kyoto University, Kyoto, Japan

Hitoshi Fujiwara Seikei University, Tokyo, Japan

Tada-nori Goto University of Hyogo, Himeji, Hyogo, Japan

Yoichiro Hanaoka National Astronomical Observatory of Japan, Tokyo, Japan

Hisashi Hayakawa Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Aichi, Japan
Institute for Advanced Research, Nagoya University, Nagoya, Aichi, Japan
RAL Space, Rutherford Appleton Laboratory, Harwell Campus, Science and Technology Facilities Council, Didcot, UK

Keisuke Hosokawa Department of Computer and Network Engineering, The University of Electro-Communications, Tokyo, Japan

Hideyuki Hotta Department of Physics, Graduate School of Science, Chiba University, Chiba, Japan

Kiyoshi Ichimoto Astronomical Observatory, Graduate School of Science, Kyoto University, Kyoto, Japan

Shinsuke Imada Department of Earth and Planetary Science, Graduate School of Science, The University of Tokyo, Tokyo, Japan

Mamoru Ishii National Institute of Information and Communications Technology, Tokyo, Japan

Kazumasa Iwai Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Aichi, Japan

Hidekatsu Jin National Institute of Information and Communications Technology, Tokyo, Japan

Yuto Katoh Graduate School of Science, Tohoku University, Sendai, Miyagi, Japan

Takashi Kikuchi Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Aichi, Japan

Yuki Kubo National Institute of Information and Communications Technology, Tokyo, Japan

Kanya Kusano Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Aichi, Japan

Takashi Minoshima Japan Agency for Marine-Earth Science and Technology, Yokohama, Kanagawa, Japan

Takefumi Mitani Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Sagamihara, Kanagawa, Japan

Hiroko Miyahara Humanities and Sciences/Museum Careers, Musashino Art University, Tokyo, Japan

Yasunobu Miyoshi Department of Earth and Planetary Sciences, Kyushu University, Fukuoka, Japan

Yoshizumi Miyoshi Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Aichi, Japan

Tsutomu Nagatsuma National Institute of Information and Communications Technology, Tokyo, Japan

Satoko Nakamura Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Aichi, Japan

Yuta Notsu Laboratory for Atmospheric and Space Physics, University of Colorado Boulder, Boulder, CO, USA
National Solar Observatory, Boulder, CO, USA
Department of Earth and Planetary Sciences, Tokyo Institute of Technology, Tokyo, Japan

Yuichi Otsuka Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Aichi, Japan

Shinji Saito National Institute of Information and Communications Technology, Tokyo, Japan

Susumu Saito Electronic Navigation Research Institute, National Institute of Maritime, Port and Aviation Technology, Tokyo, Japan

Takashi Sakurai Solar Science Observatory, National Astronomical Observatory of Japan, Tokyo, Japan

Tatsuhiko Sato Japan Atomic Energy Agency, Ibaraki, Japan

Toshifumi Shimizu Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Sagamihara, Kanagawa, Japan

Hiroyuki Shinagawa National Institute of Information and Communications Technology, Tokyo, Japan

Kazuo Shiokawa Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Aichi, Japan

Daikou Shiota National Institute of Information and Communications Technology, Tokyo, Japan

Takeshi Takashima Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Sagamihara, Kanagawa, Japan

Shin Toriumi Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Sagamihara, Kanagawa, Japan

Takuo Tsuda Department of Computer and Network Engineering, The University of Electro-Communications, Tokyo, Japan

Takuya Tsugawa National Institute of Information and Communications Technology, Tokyo, Japan

Satoru Ueno Astronomical Observatory, Graduate School of Science, Kyoto University, Gifu, Japan

Kyoko Watanabe National Defense Academy of Japan, Yokosuka, Kanagawa, Japan

Shinichi Watari National Institute of Information and Communications Technology, Tokyo, Japan

Seiji Yashiro Catholic University of America, Washington, DC, USA

Shigeo Yoden Institute for Liberal Arts and Sciences, Kyoto University, Kyoto, Japan

Kohei Yoshida Meteorological Research Institute, Japan Meteorological Agency, Ibaraki, Japan

Hisayoshi Yurimoto Department of Earth and Planetary Sciences, Hokkaido University, Sapporo, Hokkaido, Japan

Abbreviations

AC	Alternating current
ACE	Advanced composition explorer
AE	Auroral electrojet index
AEJ	Auroral electrojet
AIA	Atmospheric Imaging Assembly onboard SDO
AL	Auroral electrojet index-lower
AR5	The fifth assessment report
AU	Auroral electrojet index-upper
AUC	Area under curve
CACTus	Computer-aided CME tracking
CANS	Coordinated Astronomical Numerical Software
CDM	Conjunction data message
CEJ	Counter electrojet
CESM	Community Earth System Model of the National Center for Atmospheric Research
CGM	Corrected geomagnetic (coordinate)
CHIANTI	An atomic database for spectroscopic diagnostics of astrophysical plasmas
CIR	Corotating interaction region
CME	Coronal mass ejection
CMIP	Coupled Model Intercomparison Project
COPUOS	Committee on the Peaceful Uses of Outer Space
CPU	Central processing unit
CRAND	Cosmic ray albedo neutron decay
DC	Direct current
DDD	Displacement damage dose
DECK	Diagnostic, evaluation, and characterization of Klima
DLC	Diamond-like carbon
DP2	Type-2 geomagnetic disturbance of polar origin
Dst	Disturbance storm-time
EEJ	Equatorial electrojet

EIA	Equatorial ionization anomaly
EMA	Equatorial mass density anomaly
EMIC	Electromagnetic ion cyclotron
ENA	Energetic neutral atom
ENIAC	Electronic Numerical Integrator and Computer
ENSO	El Niño-Southern Oscillation
EPREM	Energetic particle radiation environment module
Es	Sporadic E
ESD	Electrostatic discharge
ESM	Earth System Model
ESTAR	Stopping power and range tables for electrons
EUV	Extreme ultraviolet
EVE	EUV Variability Experiment
F10.7	The solar radio flux at 10.7 cm (2800 MHz)
FAC	Field-aligned current
FAR	False alarm ratio
FFF	Force-free field
FISM	Flare irradiance spectral model
FN	False negative
foEs	Sporadic E critical frequency
FP	False positive
GAIA	Ground-to-topside model of atmosphere and ionosphere for aeronomy
GBAS	Ground-based augmentation system
GCR	Galactic cosmic ray
GDI	Gradient-drift instability
GEO	Geosynchronous equatorial orbit, also referred to as geostationary Earth orbit
GEONET	GNSS Earth Observation Network System
GIC	Geomagnetically induced current
GIE	Geomagnetically induced electric field
GLE	Ground level enhancement
GLONASS	Russian satellite navigation system
GNSS	Global navigation satellite system
GOES	Geostationary Operational Environmental Satellite
GONG	Global Oscillation Network Group
GPS	Global Positioning System
GSO	Geosynchronous orbit
HF	High frequency
HILDCAA	High-intensity, long-duration, continuous AE activity
HMI	Helioseismic and Magnetic Imager
HVDC	High-voltage DC
IBEX	Interstellar Boundary Explorer
ICAO	International Civil Aviation Organization

ICME	Interplanetary coronal mass ejection
ICRP	International Commission on Radiological Protection
IGY	International Geophysical Year
ILAT	Invariant latitude
IMF	Interplanetary magnetic field
IPCC	Intergovernmental Panel on Climate Change
IPS	Interplanetary scintillation
ISEE	Institute for Space–Earth Environmental Research
ISES	International Space Environment Service
ISS	International Space Station
JMA	Japan Meteorological Agency
Kp	Planetary geomagnetic index
LASCO	Large Angle and Spectrometric Coronagraph
LDEF	Long Duration Exposure Facility
LE	Leading edge
LEO	Low-Earth orbit
LF	Low frequency
LIMAS	Laser ionization mass nanoscope
LSTID	Large-scale traveling ionospheric disturbance
MC	Magnetic cloud
MDI	Michelson Doppler Imager
MDM	Midnight density maximum
MEO	Medium-Earth orbit
MHD	Magnetohydrodynamics
MI	Main impulse
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding
MLAT	Magnetic latitude
MLT	Magnetic local time
MMM	Multi-model mean
MPS	Magnetopause shadowing
MRI	Meteorological Research Institute of Japan
MRM	Multiple reanalysis mean
MSH	Millionth(s) of the solar hemisphere
MSTID	Medium-scale traveling ionospheric disturbance
MT	Magnetotellurics
MUF	Maximum usable frequency
NAO	North Atlantic Oscillation
NASA	National Aeronautics and Space Administration
NERC	North American Electric Reliability Council
NICT	National Institute of Information and Communications Technology
NLFFF	Non-linear force-free field
NOAA	National Oceanic and Atmospheric Administration

Nsat	Number of satellites
NWP	Numerical weather prediction
OSO	Orbiting Solar Observatory
PARMA	PHITS-based analytical radiation model in the atmosphere
Pc5	Geomagnetic pulsations with characteristic period 160–600 s
PCA	Polar cap absorption
PFU	Particle flux unit (/cm^2/s/sr)
PHITS	Particle and heavy ion transport code system
PI	Preliminary impulse
Pi2	Geomagnetic variations with characteristic period of 40–150 s and an irregular (damping) waveform
PIL	Polarity inversion line
PMS	Planar magnetic structure
POD	Probability of detection
POFD	Probability of false detection
PPP	Precise point positioning
PRC	Partial ring current
PSTAR	Stopping power and range tables for protons
PSTEP	Project for Solar-Terrestrial Environment Prediction
PVT	Position, velocity, and time
R1	Region 1
R1FAC	Region 1 field-aligned current
R2	Region 2
R2FAC	Region 2 field-aligned current
RC	Ring current
RCP	Representative concentration pathway
Re	Earth radius
ROC	Receiver operating characteristic, relative operating characteristic
RTA	Relativistic turning acceleration
SAA	South Atlantic anomaly
SAPS	Subauroral polarization stream
SBAS	Satellite-based augmentation system
SC	Solar cycle
SC	Sudden commencement
SCW	Substorm current wedge
SDDI	Solar Dynamics Doppler Imager
SDO	Solar Dynamics Observatory
SED	Storm enhanced density
SEE	Single-event effect
SEE	Solar EUV Experiment
SEM	Solar EUV Monitor
SEP	Solar energetic particle
SEU	Single-event upset
SFE	Solar flare effect

SFT	Surface flux transport
SI	Sudden impulse
SIMS	Secondary ion mass spectrometry
SIPS	Storm-induced plasma stream
SML	SuperMAG electrojet index-lower
SMM	Solar Maximum Mission
SMU	SuperMAG electrojet index-upper
SNMS	Secondary neutral mass spectrometry
SOHO	Solar and Heliospheric Observatory
SOLPENCO	Solar particle engineering code
SSA	Space situational awareness
SSC	Storm sudden commencement
SSI	Spectral solar irradiance
SSN	Sunspot number
STEREO	Solar Terrestrial Relations Observatory
SUSANOO	Space-weather-forecast-usable system anchored by numerical operations and observations
SWICS	Solar Wind Ion Composition Spectrometer
SXR	Soft X-ray
TAD	Traveling atmospheric disturbance
TEC	Total electron content
TECU	Total electron content unit
TID	Total ionizing dose
TID	Traveling ionospheric disturbance
TIE-GCM	Thermosphere-ionosphere-electrodynamics general circulation model
TIMED	Thermosphere Ionosphere Mesosphere Energetics and Dynamics satellite
TN	True negative
TOI	Tongue of ionization
TP	True positive
TSI	Total solar irradiance
UHF	Ultra high-frequency
UK	The United Kingdom
URA	Ultra-relativistic turning acceleration
UT	Universal time
UTC	Coordinated universal time
UV	Ultraviolet
VAPs	Van Allen Probes
VHF	Very high frequency
VLF	Very low frequency
VOR	VHF omni-directional radio range
WASAVIES	Warning system for aviation exposure to solar energetic particles
WDC	World Data Center

WPIA	Wave-particle interaction analyzer
WSA-ENLIL	Wang-Sheeley-Arge ENLIL solar wind prediction model
XBP	X-ray bright point
XRT	X-ray telescope onboard the Hinode satellite

Part I
Relation Between Space Weather and Society

Chapter 1
Social Impacts of Space Weather

Mamoru Ishii

The term "space weather" refers mainly to phenomena in near-Earth space and the atmosphere caused by the solar activity; in particular, it is frequently related to various events that affect human lives.[1] The Sun provides the heat and light necessary for life; however, it also emits electromagnetic waves that are dangerous (such as X-ray and ultraviolet ray), high-energy particle, and a high-speed plasma flow called the solar wind.

Electromagnetic wave causes photochemical reactions in the atmosphere of the Earth, ionizing the upper atmosphere and forming the ionosphere. Consequently, most of the electromagnetic waves with short wavelengths, which have adverse effects on life, do not reach the ground. In addition, the solar wind interacts with the magnetic field of the Earth and the ionosphere to produce a complex scenario. When the Sun-derived magnetic field in the solar wind is directed north, the magnetic field of the Earth acts as a protective barrier, and the influence of the solar wind near the Earth is limited. However, when the solar-derived magnetic field is southward, it magnetically reconnects with the magnetosphere of the Earth, and the energy of the solar wind reaches close to the Earth. In particular, it is known that solar flares and their associated energetic particles, coronal mass ejections (CMEs), the high-speed solar wind ejected from coronal holes, and the solar wind associated with filamentary ejections significantly affect the Earth and human life.

[1] According to the National Oceanic and Atmospheric Administration of the United States of America, "Space Weather refers to variations in the space environment between the sun and Earth (and throughout the solar system) that can affect technologies in space and on Earth" (https://www.regions.noaa.gov/central/index.php/noaas-space-weather-prediction-center-space-weather-toolkit/).

M. Ishii (✉)
National Institute of Information and Communications Technology, Tokyo, Japan
e-mail: mishii@nict.go.jp

K. Kusano (ed.), *Solar-Terrestrial Environmental Prediction*,
https://doi.org/10.1007/978-981-19-7765-7_1

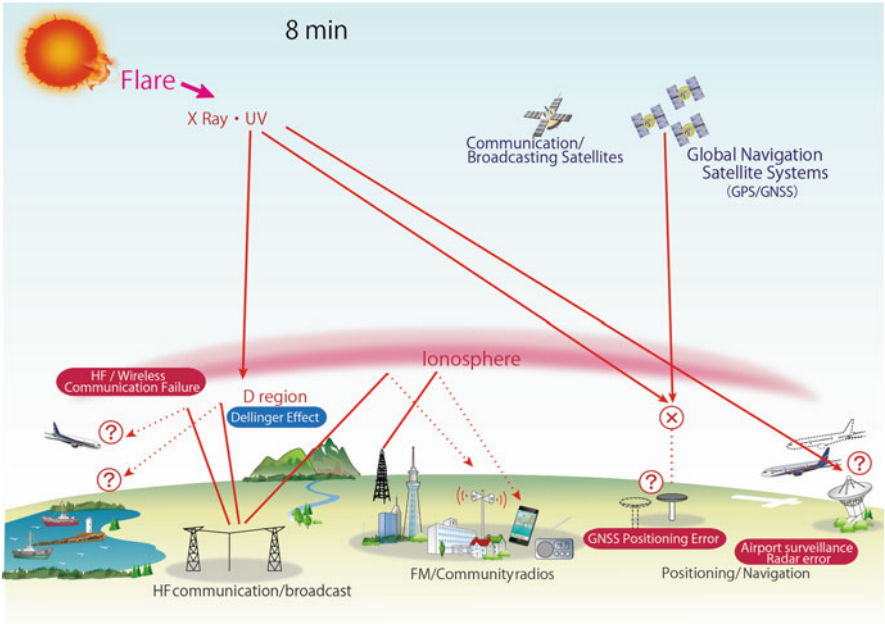

Fig. 1.1 Social impacts of electromagnetic waves

Because the propagation speeds of electromagnetic waves, energetic particles, and CMEs associated with solar flares are different, their effects on the Earth and society present time lags. Electromagnetic waves reach the Earth at the speed of light; therefore, their effects appear approximately 8 min after a flare occurs. High-energy particles reach the Earth in approximately 30 min–2 days, and CMEs frequently affect the Earth in 2–3 days, depending on their speeds.

X-rays and ultraviolet rays increase the electron density in the D region at relatively low altitudes in the ionosphere (Fig. 1.1). Because of the high density of the neutral atmosphere in this region, medium and short waves are absorbed, and their propagation is disturbed. This significantly hinders communication and broadcasting using these frequencies (called the Dellinger phenomenon or the radio blackout). It has also been reported that the use of satellite positioning signals and air traffic control radars becomes difficult because of the extremely wide bandwidth of the electromagnetic waves from the Sun. A characteristic of these phenomena is that they occur in an area where the Sun is visible, i.e., the daytime side.

The effects of high-energy particles are shown in Fig. 1.2. Human exposure to high-energy particles can occur during manned activities in space, and they are also considered to cause malfunctions via signal errors in the electronic circuits of flying objects such as satellites. Even in the atmosphere, it is known that at international aircraft altitudes, the risk of exposure is more than on the ground. High-energy particles also cause an increase in the electron density in the ionospheric D region,

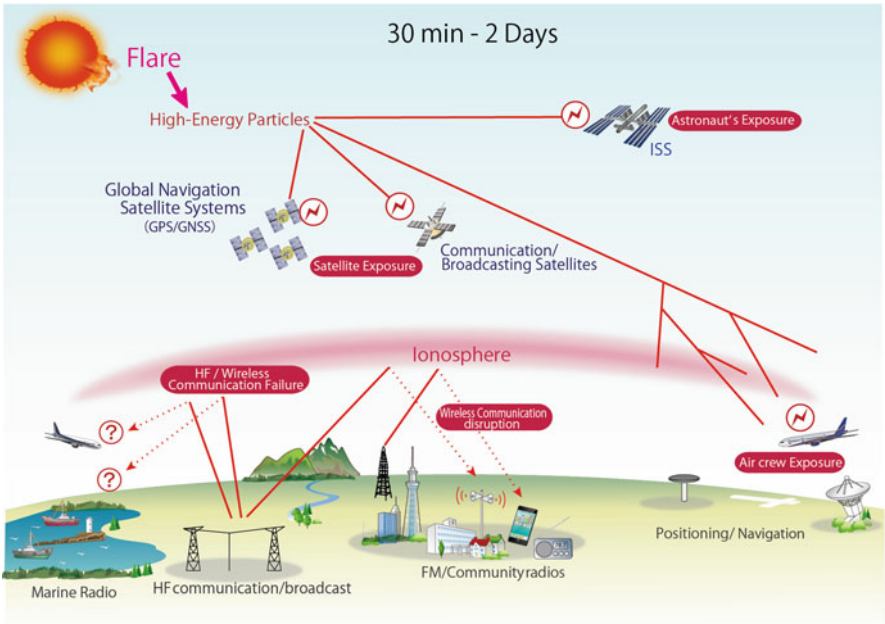

Fig. 1.2 Social impacts of high-energy particles

particularly in the polar regions, and restrict the use of communication and broadcasting, mainly in the shortwave band (called polar cap absorption).

The effects of CMEs, as shown in Fig. 1.3, include the charging phenomena of satellites, which may degrade satellite functions or cause the suspension of satellite operations. CMEs also generate large-scale auroras under currents between the magnetosphere and the ionosphere along the polar magnetic field lines as well as geomagnetically induced currents, which affect power grids and feeder systems. An aurora heats and expands the neutral atmosphere, affecting the orbits of low-altitude satellites, and also generates ionospheric storms, which cause ionospheric disturbances, which reduce the availability of satellite-to-ground and ionosphere-based communication and broadcasting. In particular, variations in the ionospheric delay can degrade satellite positioning accuracy, and the loss of lock due to scintillations can limit the use of a satellite positioning system.

Because the ionosphere is close to the surface of the Earth, it is affected by typhoons, fronts, earthquakes, and other meteorological phenomena other than the solar activity. Ionospheric disturbances caused by these meteorological phenomena are unrelated to the occurrence of solar flares. Typical examples are plasma bubbles and sporadic E (Es) layers, as shown in Fig. 1.4.

A plasma bubble is a low-electron density region locally in high density region that occurs symmetrically in the north–south direction after sunset near the magnetic equator. A satellite positioning signal passing through a plasma bubble shows a large difference from an assumed electron density model, which degrades the positioning

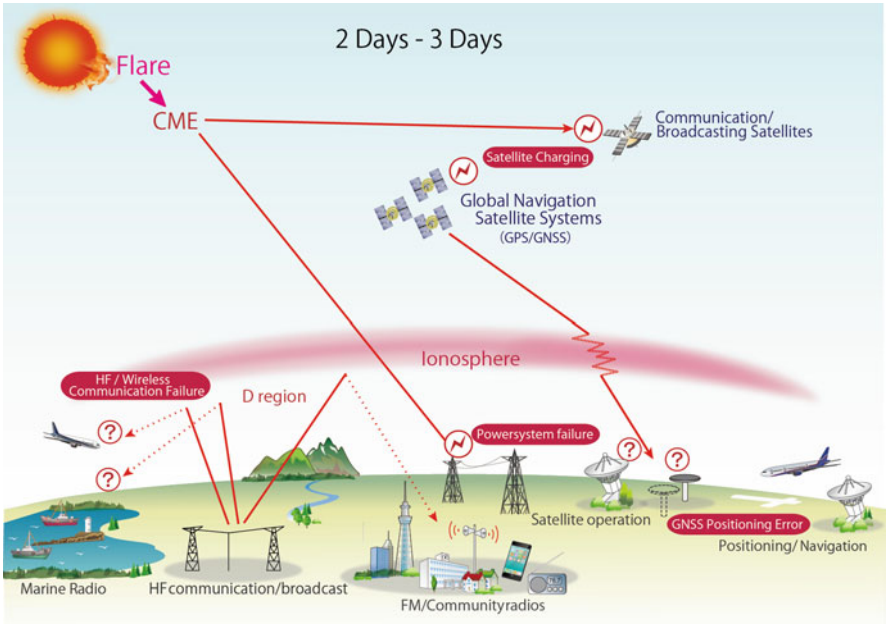

Fig. 1.3 Social impacts of CMEs

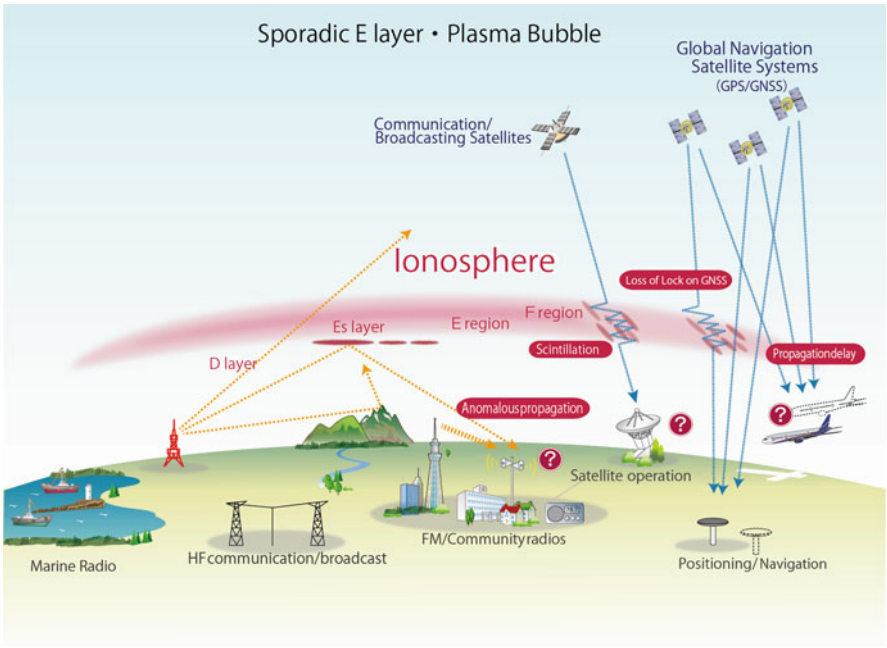

Fig. 1.4 Social impacts of space weather phenomena not caused by solar flares

accuracy of the satellite. In addition, the spatial variations in the electron density are very large near the boundary of a plasma bubble, causing scintillation of radio waves and limiting the use of satellite positioning because receiving a stable signal on the ground is impossible.

An Es layer is a phenomenon in which the ionospheric electron density increases locally at an altitude of approximately 100 km. It is known to increase high-frequency and very-high frequency reflections and increase the propagation distance more than expected.

In the following chapters, these phenomena and their implications for society are discussed in more detail.

Chapter 2
Impact of Space Weather on Various Fields

Yûki Kubo, Susumu Saito, Takuya Tsugawa, Yusuke Ebihara, Tsutomu Nagatsuma, Tatsuhiko Sato, and Hidekatsu Jin

2.1 Telecommunication and Satellite Navigation

This section is partially reprinted and revised from "Space weather benchmarks on Japanese society" authored by Ishii et al. (2021), which can be referred for a detailed discussion.

The upper atmosphere at altitudes of approximately 60–1000 km includes partly ionized gases, i.e., a plasma, produced by the absorption of the X-rays and extreme ultraviolet (EUV) radiation from the Sun, and is called the ionosphere. The ionosphere constantly varies under the influence of the Sun and the lower atmosphere and has properties that cause reflection and absorption of low-frequency (LF) and medium-frequency (MF) radio waves. Moreover, ultrahigh-frequency (UHF) radio waves, which are used for communication and positioning by satellites, are delayed and fluctuate depending on the density and structure of the ionospheric plasma (Fig. 2.1). Therefore, ionospheric variations and fluctuations can disturb essential social infrastructure systems that use radio wave propagation for our daily lives, such as the Global Navigation Satellite System (GNSS) including the Global Positioning System (GPS), communication and broadcasting satellites, and high-frequency (HF) radio communication and broadcasting. These systems have long been used

Y. Kubo (✉) · T. Tsugawa · T. Nagatsuma · H. Jin
National Institute of Information and Communications Technology, Tokyo, Japan
e-mail: kubo@nict.go.jp

S. Saito
Electronic Navigation Research Institute, National Institute of Maritime, Port and Aviation Technology, Tokyo, Japan

Y. Ebihara
Research Institute for Sustainable Humanosphere, Kyoto University, Kyoto, Japan

T. Sato
Japan Atomic Energy Agency, Ibaraki, Japan

K. Kusano (ed.), *Solar-Terrestrial Environmental Prediction*,
https://doi.org/10.1007/978-981-19-7765-7_2

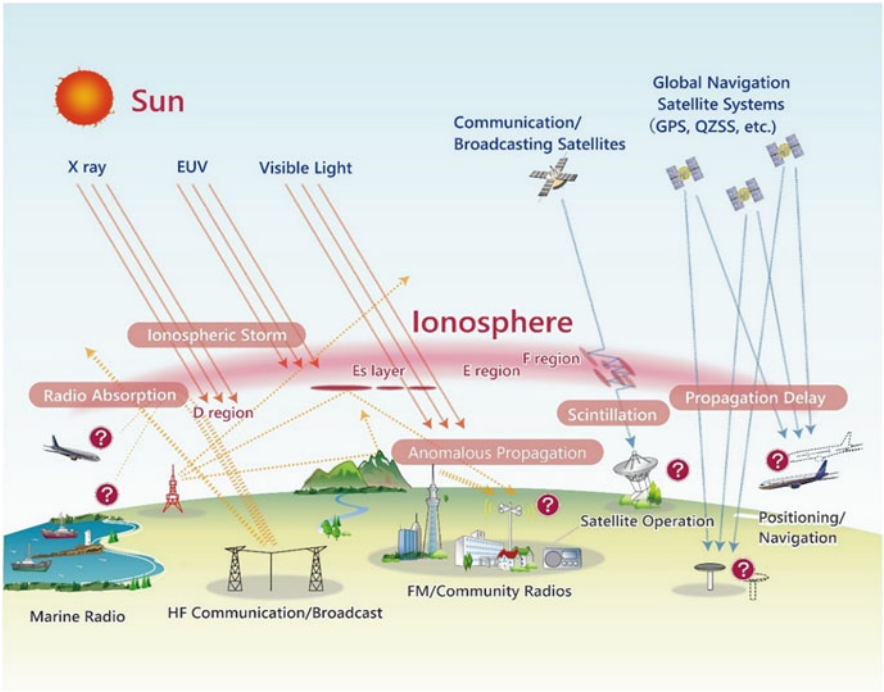

Fig. 2.1 Ionospheric effects on communication, broadcasting, and positioning

for communication beyond the line-of-sight range in aeronautical, maritime, and disaster prevention fields. In this chapter, first, the structure of the ionosphere is briefly described, and subsequently ionospheric effects on communication, broadcasting, and positioning and their impact scales are discussed.

2.1.1 Structure of Ionosphere

As shown in Fig. 2.2, the ionosphere is divided into several regions named based on the altitude and the composition: D, E, and F regions from the lowest to the highest (historically, they have been called D, E, and F layers, respectively). The electron density (or the plasma density) is highest in the F region at an altitude of approximately 300–400 km. Because the main ionizing sources in the ionosphere are basically electromagnetic waves (X-rays, EUV rays, etc.) from the Sun, the plasma density depends on the solar zenith angle and increases from sunrise, reaches a maximum during the daytime, and subsequently decreases during the night. The plasma density tends to be high at low latitudes where the solar zenith angle is small. However, as shown in Fig. 2.3, the equatorial ionization anomaly (EIA) is formed by the high plasma density in the latitudinal zone across the magnetic equator

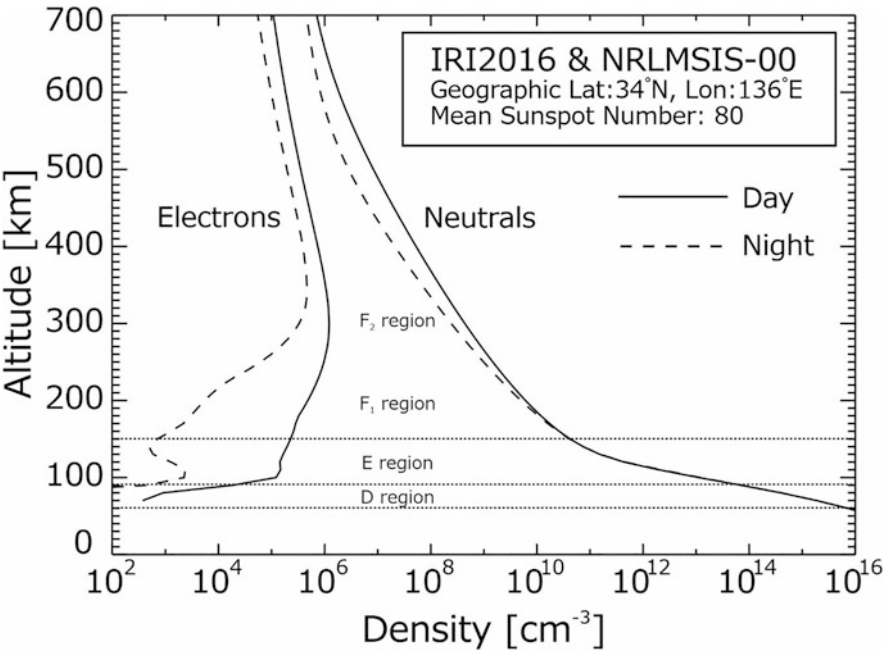

Fig. 2.2 Electron density profile in ionosphere

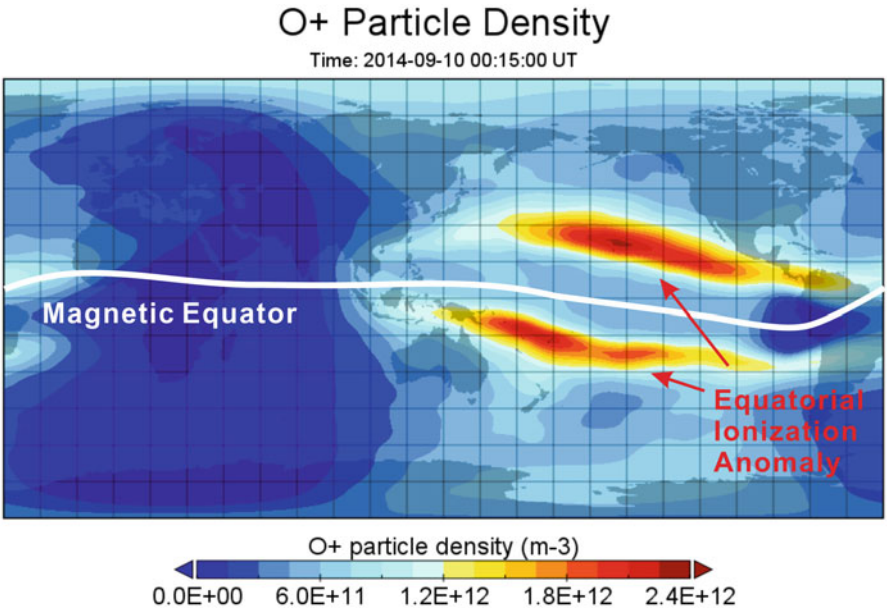

Fig. 2.3 Ionospheric F-region plasma density (O^+ particle density) distribution

(approximately ±15° in magnetic latitude). In the D and E regions, seasonal variations occur in which the plasma density becomes relatively high in summer when the solar zenith angle at mid-summer is basically small. In the F region, semiannual variations occur in which the diurnal plasma density reaches maximum in spring and autumn and seasonal anomalies are observed at mid- and high latitudes, where the daytime plasma density in winter is higher than that in summer. The actual ionosphere typically fluctuates under the influence of not only the solar radiation but also the solar wind, magnetospheric activity, and neutral atmospheric winds; see Chap. 4 for details.

2.1.2 *Communication and Broadcasting*

The propagation of radio waves used for communication and broadcasting from a transmission point to a reception point is classified depending on the frequency: waveguide, ionospheric, and line-of-sight propagation. The effects of the space weather phenomena in the ionosphere (ionospheric phenomena) are also different owing to the differences in the propagation of radio waves of various frequencies. Figure 2.4 shows the propagation of radio waves at different frequencies and their main applications, and Table 2.1 lists the ionospheric phenomena and their effects on communication and broadcasting. For details of each ionospheric phenomenon, please refer to Chap. 7. This chapter describes the effects of these phenomena and their impact scales.

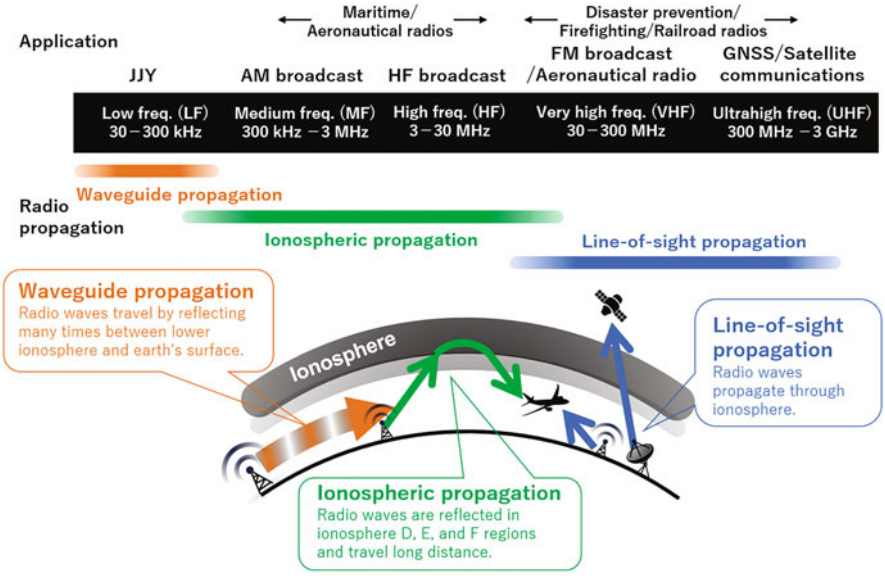

Fig. 2.4 Radio wave propagation and applications based on frequency

Table 2.1 Ionospheric disturbance phenomena and their impacts on communication and broadcasting

Application	Effect	Ionospheric disturbance phenomena
Communication/ broadcasting	Effects on HF radio waves	Dellinger phenomenon Negative ionospheric storms Plasma bubbles Polar cap absorption
	Effects on VHF radio waves	Plasma bubbles Sporadic E layers
	Effects on UHF radio waves	Plasma bubbles

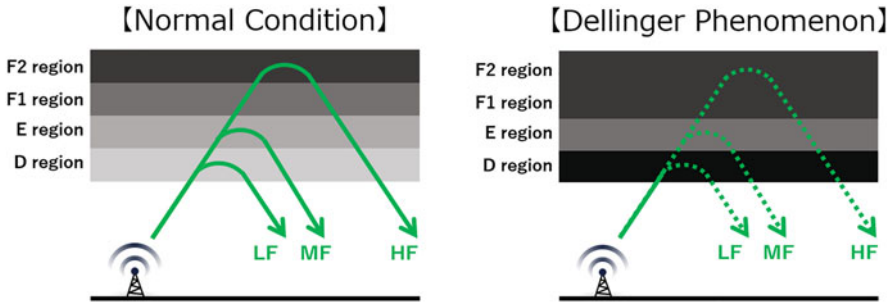

Fig. 2.5 Schematic of radio wave propagation during Dellinger phenomenon

2.1.2.1 Impacts on High-Frequency Radio Waves

Currently, with the development of communication satellites and submarine cables, the role of HF communication has become limited. However, it is still used for long-distance communication such as aeronautical, maritime, disaster prevention, and international radio broadcasting. Space weather phenomena in the ionosphere affecting HF communication include the Dellinger phenomenon, ionospheric negative storms, sporadic E layers, and polar cap absorption (PCA).

The Dellinger phenomenon is a phenomenon in which an abrupt increase in the X-ray intensity during a solar flare causes a rapid increase in the electron density in the ionospheric D region via photoionization by X-rays, resulting in the absorption of HF radio waves. The Dellinger phenomenon is named after its discoverer and also known as the shortwave fade-out. The D region at the bottom of the ionosphere absorbs radio waves by their collisions with the surrounding neutral atmospheric particles. The amount of absorption is large when the electron density is high, and the effect is large for low-frequency radio waves. Generally, HF radio waves are reflected in the F region of the ionosphere and propagate over long distances. However, when the Dellinger phenomenon occurs and the total amount of absorption along the radio propagation path in the D region is larger than the transmission strength of the radio waves, a communication blackout occurs (Fig. 2.5). The

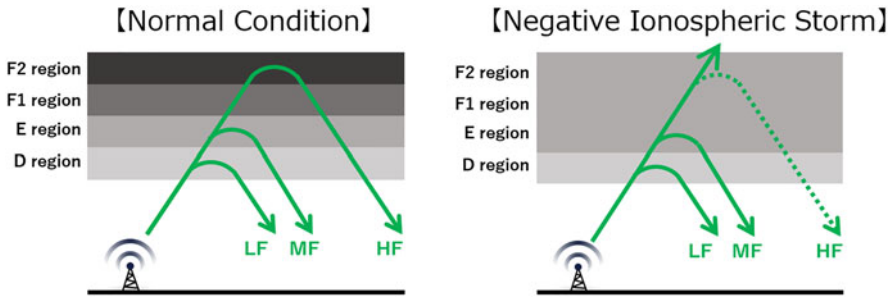

Fig. 2.6 Schematic of radio wave propagation during negative ionospheric storm

duration of the Dellinger phenomenon depends on the duration of the electromagnetic radiation associated with the corresponding solar flare and ranges from several tens of minutes to several hours.

When a magnetic storm occurs, a disturbance may be caused in the diurnal variation in the ionospheric plasma density, which is called an ionospheric storm. Ionospheric storms can be classified into positive ionospheric storms, in which the plasma density significantly increases, and negative ionospheric storms, in which it significantly decreases. Positive ionospheric storms occur when ionospheric ions and electrons are shifted to high altitudes. This shift is caused by the increase in the equatorward thermospheric wind associated with the heating of the polar thermosphere by magnetic storms and due to the eastward electric field in the polar regions from the magnetosphere and penetrating to low latitudes. Negative ionospheric storms are caused by global atmospheric composition changes due to heating of the polar thermosphere and a reduction in the ionospheric plasma density caused by chemical reactions. When a positive ionospheric storm occurs, HF radio waves, which generally pass through the ionosphere, are reflected by the ionosphere, and the frequency range that can be used for out-of-sight HF communication and broadcasting is expanded. Therefore, there is no effect in terms of failure. However, when a negative ionospheric storm occurs, HF radio waves, which are typically reflected in the F region, pass through the ionosphere without reflection, making the use of out-of-sight HF communication and broadcasting difficult (Fig. 2.6). This effect is similar to the Dellinger phenomenon. However, the difference is that the Dellinger phenomenon is caused by a solar flare and persists from several tens of minutes to several hours, whereas a negative ionospheric storm is caused by a magnetic storm and occurs from several hours to several days.

Plasma bubbles are “bubbles” of local low-electron density regions after sunset in the ionosphere at low latitudes near the magnetic equator (Fig. 2.7). They move eastward while expanding in the north–south direction along the magnetic field lines. When the solar activity is high or magnetic storms occur, they grow large and occasionally develop up to midlatitudes such as Japan. A large spatial variation in the electron density occurs inside and around a plasma bubble, which affects the radio propagation in HF–UHF bands. For HF radio waves, it is considered that

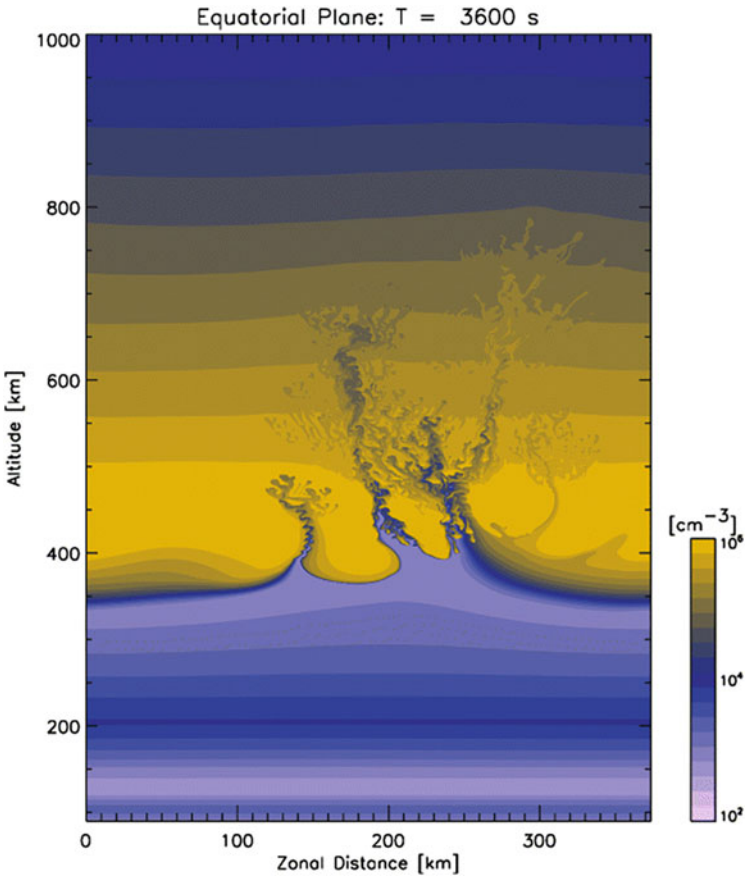

Fig. 2.7 Results of computer simulation of plasma bubbles (Yokoyama 2017)

changes are caused in their propagation direction by the bottom-side structure of a plasma bubble and that anomalous propagations occur along a plasma bubble structure; however, these effects have little impact on HF communication and broadcasting.

PCA is a phenomenon in which radio wave absorption is increased by the rapid ionization of the ionospheric D region in the polar cap regions caused by solar energetic particles (SEPs) ejected by solar flares. This phenomenon prevents HF communication and broadcasting (Fig. 2.8). It occurs 30 min to several hours after a solar flare and may remain for several days. It is similar to the Dellinger phenomenon in that HF radio waves are absorbed owing to anomalous ionization in the D region associated with a solar flare. However, the Dellinger phenomenon occurs only in the dayside, whereas PCA occurs in both day and nightsides in the polar cap regions. Therefore, aeronautical radio communication with an aircraft traveling along a polar route is affected.

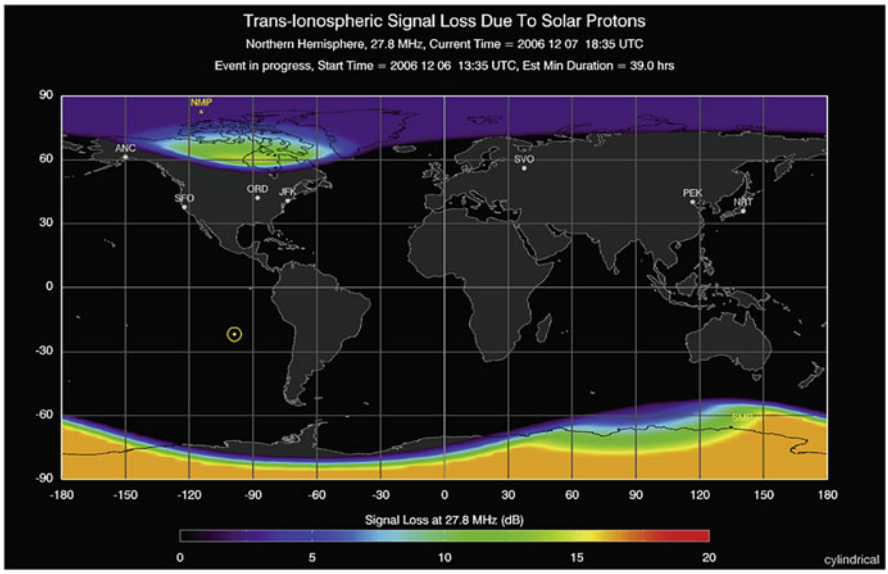

Fig. 2.8 Polar cap absorption in 27.8 MHz band due to SEP event on December 6, 2006 (Sauer and Wilkinson 2008)

In the following, we discuss briefly the occurrence frequencies and impact scales of the above ionospheric phenomena affecting HF communication and broadcasting. The Dellinger phenomenon is caused by the electromagnetic radiation such as X-rays associated with solar flares, and PCA mainly results from high-energy particles emitted by the Sun. Ionospheric storms associated with magnetic storms are highly correlated with solar coronal mass ejections (CMEs) and high-speed solar wind, which are the main factors of magnetic storms. Therefore, these phenomena are highly dependent on the solar activity in the 11-year cycle. In contrast, a plasma bubble is considered to be formed by the Rayleigh–Taylor instability in the lower part of the ionospheric F region, as described in Chap. 7. The growth rate of the Rayleigh–Taylor instability is high when the eastward electric field is large; the plasma is at a high altitude, and the altitude gradient of the plasma density is large. Therefore, the occurrence frequency of a plasma bubble increases with the solar activity. It may develop to mid- to high latitudes if a strong eastward electric field is applied to the ionosphere in the main phase of a magnetic storm. The occurrence frequency of plasma bubbles is high when the declination of the magnetic field lines of the Earth coincides with that of the sunset terminator, leading to marked seasonal and longitudinal variations in it. In East Asia, including Japan, the occurrence frequency is high in spring and autumn (Burke et al. 2004). Figure 2.9 shows the year-to-year variations in the sunspot number and the occurrence rate of plasma bubbles (number of occurrence days/number of target days) in March and September in Okinawa based on total electron content (TEC) observations by the GNSS Earth

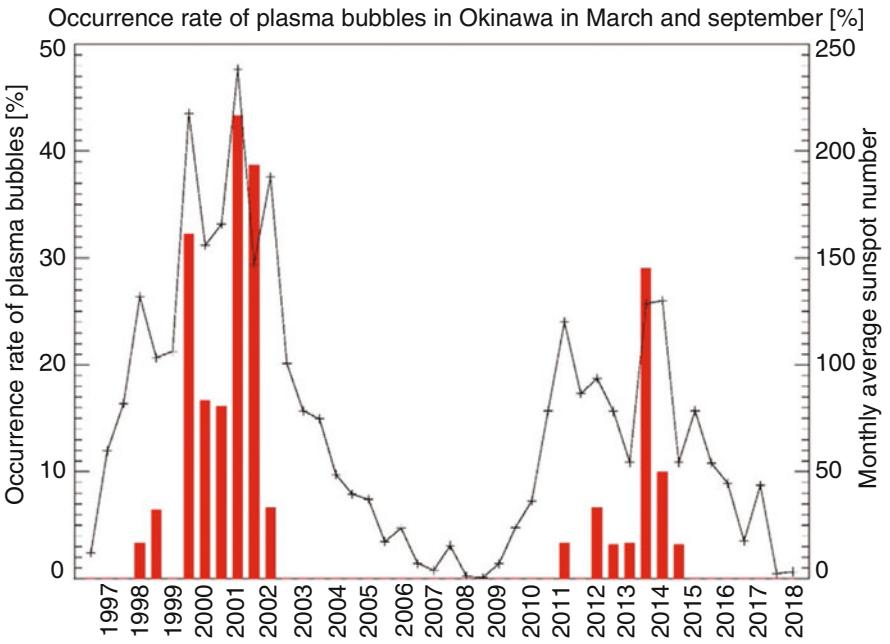

Fig. 2.9 Year-to-year variation in sunspot number and occurrence rate of plasma bubbles in Okinawa based on GEONET TEC observations (number of days of occurrence/number of days of observations [%])

Table 2.2 Occurrence frequencies and impact scales of ionospheric phenomena affecting HF communication and broadcasting

Ionospheric phenomenon	Scale parameter	Once per year	Once per 10 years	Once per 100 years	Once per 1000 years
Dellinger phenomenon	Absorption (long distance) [dB][a]	71	83	93	100
	Duration [h][b]	0.63–1.6	1.8–3.6	4.0–6.8	7.4–11.9
Plasma bubble	Duration [h]	4–5	6		

[a] Based on long-range HF communication line (for effective solar zenith angle of 0° at 6.6 MHz)
[b] Based on ionosonde observations (for 1–30 MHz "B" and dfmin ≥2.5 MHz). Value of dfmin was defined as difference between value of fmin and 27-day running median at same LT

Observation Network (GEONET). It can be seen that the occurrence rate of the plasma bubbles that develop in midlatitudes such as Japan is strongly correlated with the solar activity level.

From Tables 2.2 and 2.3, the relationship between the occurrence frequency and impact scale of each ionospheric phenomenon affecting HF communication and broadcasting can be seen, where the results are derived based on statistical analysis and simulation of long-term observation data. The details of the derivation are described in Ishii et al. (2021), and the references therein, and are omitted here.

Table 2.3 Occurrence frequency and impact scale of polar cap absorption

Ionospheric phenomenon	Scale parameter	Daily–several per year	Once per year	Once per 10 years	Once per 100 years	Once per 1000 years
Polar cap absorption	Proton flux ≥10 MeV [pfu][a]	10^{1-2}	10^{3}	10^{4}	4×10^{5}	10^{6}
	Kp index[b]	4 (7)	7 (8)	8 (9)	9	9
	Magnetic latitude (daytime) at 6.6 MHz, 10 dB[c]	65° (60°)	57° (55°)	53° (52°)	51°	50°
	Magnetic latitude (nighttime) at 6.6 MHz, 10 dB[c]	65° (60°)	60° (58°)	56° (55°)	53°	53°
	Minimum duration [h][d]	22	46	71	83	89

[a] Estimated setting. According to Kurt et al. (2004) and Jiggens et al. (2018). 1 pfu = $1/cm^2/s/sr$

[b] Estimated setting. Not necessarily corresponds to magnetic storms. Numbers in parentheses represent National Oceanic and Atmospheric Administration scale

[c] Latitude at which PCA extends. According to Sauer and Wilkinson (2008) and Jiggens et al. (2018)

[d] Duration of SEP event where flux of high-energy particles (≥10 MeV) is larger than 10 pfu. According to Sauer and Wilkinson (2008)

For example, in the case of the Dellinger phenomenon listed in Table 2.2, an HF communication failure due to the absorption of approximately 83 dB occurs once in 10 years and remains for 1.8–3.6 h in the case of radio waves of 6.6 MHz frequency used by commercial aircrafts. Depending on the scale of the PCA, the area where the HF communication and broadcasting are affected expands. For example, in Table 2.3, "Magnetic latitude (daytime) at 6.6 MHz, 10 dB" lists the latitude at which the 6.6 MHz frequency band is attenuated by 10 dB. At the once per year scale, it extends to 57° magnetic latitude on the daytime side of the Earth, and the effect continues for 46 h. Comparatively, at the once per 10 years scale, the influence extends to 53° magnetic latitude at lower latitudes and remains for 71 h.

2.1.2.2 Impacts on Very High-Frequency Communication

Radio waves in the very high-frequency (VHF) band are typically not reflected in the ionosphere; therefore, the method of radio wave propagation used for communication and broadcasting is line-of-sight propagation. It is mainly employed for frequency modulation radio broadcast, aeronautical navigation, disaster prevention, and firefighting radio broadcasting. Sporadic E layer (Es layers) and plasma bubbles are the ionospheric phenomena that affect VHF communication and broadcasting.

An Es layer is a dense layer of electrons that occurs locally and abruptly at an altitude of approximately 100 km and tends to be formed frequently in summer in Japan (Taguchi and Shibata 1961). An extremely developed Es layer has a larger

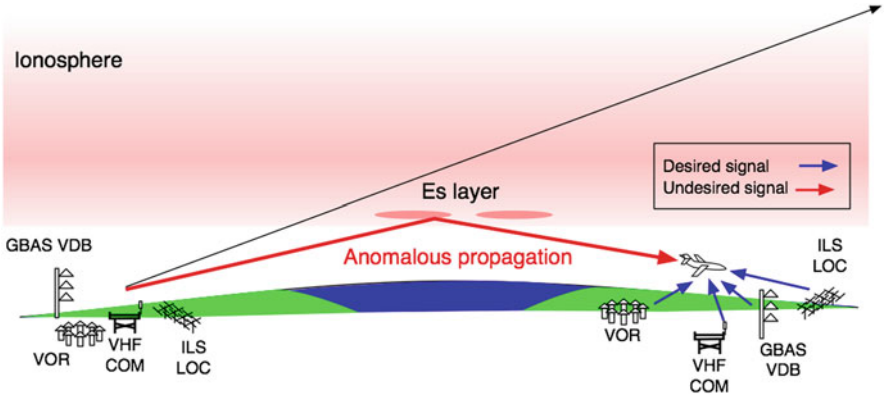

Fig. 2.10 Schematic of possible interference to VHF aeronautical radio waves (Sakai et al. 2019)

critical frequency than the F region. When the direction of radio wave propagation into the ionosphere becomes oblique and the angle of incidence increases, the frequency of radio waves that can be reflected increases. Therefore, even VHF radio waves, which are typically not reflected in the ionosphere, are reflected by an Es layer when the angle of incidence becomes extremely large. When VHF radio waves are reflected by an Es layer and propagate abnormally, they can interfere with distant communication and broadcasting.

Figure 2.10 shows a schematic of the effects of an Es layer on radio waves used in aeronautical navigation. An Es layer allows VHF radio waves (108–137 MHz for civil aviation) transmitted from ground stations outside the line of sight to reach an aircraft, causing interference. The monitoring observations of VHF aeronautical navigation message propagation in Kure City, Hiroshima Prefecture, showed the receipt of VHF radio waves from many out-of-sight transmission points (Sakai et al. 2019).

Large spatial variations occur in the electron density inside and around plasma bubbles, as mentioned earlier. VHF radio waves are scattered by the irregular structure of the plasma in a plasma bubble or propagate along the structure of a plasma bubble, with a part of them possibly propagating anomalously out of the line of sight. However, the strength of the anomalous propagation in the VHF band caused by a plasma bubble is smaller than that of the anomalous propagation in the line of sight or the anomalous propagation reflected by an Es layer. Thus, it is considered that an anomalous propagation in the VHF band does not cause any severe effect on VHF communication and broadcasting.

From Table 2.4, the relationship between the occurrence frequency and impact scale of an Es layer affecting VHF communication and broadcasting can be seen, which is derived based on long-term observation data of ionosondes. This relationship is derived by the following steps (see Ishii et al. (2021) for details): (1) the cumulative occurrence frequency distribution of the daily maximum Es layer critical frequencies (foEs) in Tokyo is derived using 62 years of ionosonde observation data

Table 2.4 Occurrence frequency and impact scale of sporadic E layer

Ionospheric phenomenon	Scale parameter	Once per year	Once per 10 years	Once per 100 years
Es layer	MUF for propagation of great-circle distance of 1000 km [MHz]	100	110	>130[a]
	MUF for propagation of great-circle distance of 2000 km [MHz]	130	>180[a]	>180[a]
	Duration for foEs >13.3 MHz[b] [h]	~3	~5	
	Duration for foEs >18 MHz[c] [h]	1–2	~3	

[a] Inequality is due to the fact that critical frequency that can be observed with ionosondes is up to 30 MHz
[b] Criterion for horizontal propagation of 80 MHz radio waves more than 2000 km
[c] Criterion for horizontal propagation of 80 MHz radio waves more than 1000 km

(manual scaling data) from 1957 to 2018; (2) the cumulative occurrence frequency distributions of the maximum usable frequencies (MUFs) for the propagation of great-circle distances of 1000 km and 2000 km are derived. From Table 2.4, it can be seen that the MUF at which an anomalous propagation over a horizontal distance of 1000 km (2000 km) is possible approximately once per year under the effect of an Es layer is 100 MHz (130 MHz). For the VHF band (108–118 MHz) used in an aeronautical radio (VHF omnidirectional radio range: VOR), this implies that an Es layer that causes anomalous propagation over horizontal distances of 2000 km and 1000 km occurs more than once per year and approximately once per 10 years, respectively. The occurrence frequency of Es layers in Table 2.4 reflects the propagation frequency at one set of transmitting and receiving stations. However, when considering the influence on the actual VOR, it is necessary to consider the spatial structure of all surrounding transmitting stations and the Es layer for one receiving station. It has been reported that the occurrence frequency of the anomalous propagation due to the reflection of the VOR in an Es layer is approximately 25% in the summer season when Es layers are frequently formed in Japan (Sakai et al. 2019). As noted from Table 2.4, the durations of strong Es layers are estimated to be 5 h and 3 h once per 10 years for foEs >13.3 MHz and foEs >18 MHz, respectively. For foEs >13.3 MHz, 80 MHz radio waves are reflected by an Es layer and propagate more than 2000 km horizontally, and for the latter foEs, the same radio waves propagate more than 1000 km horizontally.

An Es layer has little relation to the solar activity, and it is highly seasonally dependent, e.g., its occurrence frequency increases in summer in Japan. In addition, it is very difficult to observe and predict the exact location and time of occurrence of an Es layer because it occurs very locally and abruptly. To avoid the influence of an Es layer on VHF communication and broadcasting, it is practical to understand the current Es layer observation scenario and the interference of the corresponding VHF band and to take measures such as changing the frequency, if necessary.

2.1.2.3 Impact on Ultrahigh-Frequency Communication

UHF radio waves used in satellite communications typically pass through the ionosphere; however, short-period fluctuations may be caused in the signal reception strength by the irregular structure of the plasma associated with a plasma bubble and other phenomena. This phenomenon is called "ionospheric scintillation" or "scintillation," and it degrades the communication quality between the ground and a satellite, disturbing the imaging data transmitted from the satellite. This phenomenon has been known for a long time, and it is known to affect satellite communication such as Inmarsat in Japan (Karasawa et al. 1985). In March 2002, during Operation Anaconda, an operation of the US military to clean up al-Qaeda in Afghanistan, a failure in UHF satellite communication led to human casualties on the US military side, and it has been pointed out that this failure may have been caused by plasma bubbles (Kelly et al. 2014). The scintillations caused by the ionosphere decrease as the frequency increases and become almost negligible in a satellite communication system above 10 GHz compared to other effects such as in the troposphere. However, in L-band (1.5 GHz) and S-band (2 GHz) satellite communication systems, the effects of the scintillations caused by a plasma bubble cannot be ignored. A long radio wave propagation path passing through the structure of a plasma bubble (extending north–south along the magnetic field lines) implies a large impact. Therefore, in Japan, plasma bubble impact is expected to be large for satellite communication at low elevation angles to the south, such as geostationary Earth orbit (GEO) satellites.

2.1.3 Positioning

2.1.3.1 Satellite Positioning and Error Factors

Positioning implies determination of the satellite position, velocity, and time, which can be achieved by various approaches. However, in this section, we focus on positioning using satellite signals, which has been extensively used in recent years; representative examples are the GPS of the USA and various positioning systems based on the GNSS. The details of the GNSS, such as its concept, positioning principle, signal configuration, and receiver configuration, are described well in Misra and Enge (2001).

In the GNSS, the position of a receiving antenna and the time of a receiver are determined by measuring the propagation times of signals broadcasted from multiple GNSS satellites, which yield the distances to the satellites. It is necessary to receive signals from four or more satellites simultaneously to determine three-dimensional positions and times. In general, a large number of satellites from which signals are received and uniformity of their distribution in the sky imply a small positioning error and good positioning results.

Table 2.5 Typical values of errors in GNSS pseudoranges (GPS L1 frequency) (Misra and Enge 2001)

Error source	Error value
Ionospheric delay	Several meters to several tens of meters in vertical, changing by satellite elevation angle by factor up to 3
Tropospheric delay	2.3–2.6 m in vertical at sea level, changing by satellite elevation angle by factor up to 10
Satellite clock (after corrected by broadcast ephemeris)	Typically, 3 m
Random error in code pseudorange (receiver thermal noise and multipath)	0.5–1.0 m
Random error in phase pseudorange (receiver thermal noise and multipath)	0.5–1.0 cm

GNSS satellites mainly broadcast their signals in the UHF band in the frequency range from 1.0 to 1.6 GHz. GPS satellites broadcast their signals modulated by pseudorandom codes at three frequencies: L1 (1.57542 GHz), L2 (1.22760 GHz), and L5 (117645 GHz). GNSS receivers use these signals to measure the distance to a satellite, which contains various errors. A pseudorange measured by code modulation (code pseudorange, ρ) and that using accumulated carrier phases (phase pseudorange, $\lambda\phi$) can be expressed as follows:

$$\rho = r + \delta I + T + c(\delta t_u - \delta t_s) + \epsilon_\rho$$
$$\lambda\phi = r - \delta I + T + c(\delta t_u - \delta t_s) + \lambda N + \epsilon_\phi$$

where λ is the wavelength of the GNSS signal, c is the speed of light in vacuum, r is the geometric range, δI is the propagation delay error due to the difference in the refractivity in the ionosphere relative to that in the vacuum (ionospheric delay), and T is the propagation delay error due to the difference in the refractive index in the troposphere relative to that in the vacuum (tropospheric delay). δt_u and δt_s are receiver and satellite clock errors, respectively, ϵ_ρ and ϵ_ϕ are random errors in the code and phase pseudoranges (multipath errors due to reflections from the ground and surrounding objects, thermal noise, etc.), respectively, and N is the integer ambiguity of the phase. Here, the receiver clock error, δt_u, is determined in positioning together with the receiver (antenna) position included in the geometric distance r.

Table 2.5 lists the typical values of the range errors in the GPS L1 signal. Tropospheric and ionospheric delays are errors associated with the propagation path, between which the former can be corrected relatively well by models, whereas the latter is highly variable and can cause significant errors in satellite positioning. The error due to an ionospheric delay is inversely proportional to the square of the radiofrequency (f [Hz]) and proportional to the number of electrons per unit area between the satellite and the receiver (TEC [m^{-2}]) and is expressed as

$$\delta I = \frac{40.3}{f^2} TEC\ [m]$$

At the GPS L1 frequency, it is 0.16 m per 1 TECU (10^{16} m^{-2}).

As mentioned earlier, the final objective of GNSS-based positioning is to obtain the position and time (often called "PVT (position, velocity, and time)") of a receiver (an antenna). These properties can be obtained by the least-squares method using four or more pseudoranges of satellites as the best solution. The position and time (four-element longitudinal vector) and the ranging data (N_{sat} (number of satellites) − element longitudinal vector) are connected by a $4 \times N_{sat}$ matrix (called the S matrix), which is determined by the satellite geometry. Therefore, the position and time error (positioning error) is obtained by converting the pseudorange errors for the used satellites using the S matrix (Misra and Enge 2001). Specifically, the positioning error depends not only on the pseudorange error but also on the satellite geometries. The elements of the S matrix depend on the geometries of the satellites and the positions of the individual satellites. When the pseudorange error of a certain satellite is $\delta\rho$, the error contribution from the satellite is a multiple of $\delta\rho$ and the elements of the S matrix corresponding to the satellite. Here, the error in the satellite position due to the error in the satellite ephemeris becomes operative. Because the values of the S matrix elements range from less than 1 to approximately 3 in a normal satellite geometry, the contribution to the positioning error may be smaller than or amplified with respect to the ranging error, depending on the satellite. Because this effect generally depends on the number of available satellites, it is very important to use sufficient number of available satellites for GNSS positioning.

2.1.3.2 GNSS Positioning Methods and Ionospheric Effects

GNSS positioning can be classified into two types: single-frequency positioning, which uses GNSS signals at only one frequency, and dual-frequency positioning, which uses GNSS signals at two frequencies. In dual-frequency positioning, ionospheric delay errors can be eliminated by utilizing the dependence of an ionospheric delay on the frequency. Although receivers and antennas capable of dual-frequency positioning are more complex and expensive than those of single-frequency positioning, the latter technique is used in many GNSS positioning applications. For air navigation, where safety is the first priority, only L1 signals of the GPS of the USA and the GLONASS of Russia, for which international standards (ICAO 2018a, b) have been defined, are available as of 2020. Standardization is being carried out to make signals in the L1 and L5 bands of the GPS, GLONASS, Galileo of Europe, and Beidou of China available for air navigation. Because replacing/upgrading an airborne equipment is time-consuming, an extensive use of dual-frequency positioning for air navigation will require a long time. Thus, single-frequency positioning is commonly used despite its susceptibility to ionospheric delays.

GNSS positioning can further be classified into mainly three types: single, relative (differential), and high-precision single positioning (precise point positioning, PPP). They have their own advantages and disadvantages and are chosen based on the required performance such as accuracy and reliability. Single positioning is performed using a single receiver, whereas relative positioning requires multiple receivers simultaneously to measure their relative positions and achieve high-precision positioning. Relative positioning is further categorized into two types: code-based positioning, in which the pseudoranges are measured by code delays, and carrier phase-based positioning, which uses the carrier phases of the satellite signals. Carrier phase-based positioning is also called kinematic positioning. The above methods are diverse with various techniques for enhancing positioning performance. Even single positioning can be implemented with carrier phase-based and code-based positioning techniques. Carrier phase-based single positioning, such as PPP, can achieve centimeter-level accuracy using various corrections.

In single-frequency single positioning, an ionospheric delay is frequently corrected using empirical models. In the case of the GPS, the correction model parameters are broadcast from a satellite, and a receiver uses them to correct the ionospheric delay. However, if the ionospheric electron density distribution is significantly different from normal owing to ionospheric disturbances, errors will occur in the correction. Although both positive and negative ionospheric storms can cause errors, positive storms generally have a larger impact than negative storms, because the ionospheric electron density cannot be less than zero.

In single-frequency relative positioning, when the ionospheric delays for the reference and user receivers are different, the spatial difference in the ionospheric delays (called the ionospheric gradient) becomes an error factor. Because a difference in the ionospheric delays of receivers is a problem, ionospheric phenomena that change rapidly in space can be major issues. Typical examples include a plasma bubble in which the ionospheric electron density decreases rapidly in space and a storm-enhanced density (SED) or a storm-induced plasma stream in which the density increases rapidly with a magnetic storm. In kinematic positioning based on carrier phase measurements, errors of the order of one wavelength (0.19 m for the GPS L1 signal) or less can be problematic. Therefore, ionospheric disturbances even with relatively small amplitudes, such as medium-scale traveling ionospheric disturbances (MSTIDs), which have an amplitude of approximately 1 TECU (equivalent to 0.16 m for the GPS L1 signal), can be issues in terms of the increased positioning error and convergence time. However, it should be noted that even if the ionospheric spatial variability is small, if the overall ionospheric delay increases very significantly, the upper limit of the possible correction values may be exceeded, resulting in a correction error.

In PPP, precise corrections are made for error factors to achieve centimeter-level accuracy. As can be seen from Table 2.5, ionospheric delays account for a large proportion of the error. Therefore, ionospheric disturbances that make ionospheric correction difficult can be problematic for PPP as they increase the positioning error and convergence time.

Because an ionospheric delay depends on the frequency, it can be mostly eliminated by combining signals of different frequencies, except for the higher-order terms that depend on the magnetic field. In dual-frequency positioning, the ionospheric delay term is eliminated, i.e., an ionosphere-free linear combination is used. It is expressed as

$$\rho_{IF} = \frac{f_1^2 \rho_1 - f_2^2 \rho_2}{f_1^2 - f_2^2}$$

where f_1 and f_2 are the first and second frequencies, respectively, and ρ_1 and ρ_2 are the pseudorange measurements corresponding to the first and second frequencies. Although an ionosphere-free linear combination can eliminate ionospheric delays, it has the disadvantage that other errors are amplified, as can be understood from the fact that the denominator of the above equation is the difference of the square of the frequency. Because of this noise increase, the convergence time of the solution may increase even in the nominal case of dual-frequency kinematic positioning and PPP compared to that of single-frequency positioning. Therefore, an ionospheric correction is the key to improve the performance of high-precision positioning even in dual-frequency cases, and increased accuracy of ionospheric information is desired.

An ionospheric delay is not the only effect of the ionosphere on GNSS-based positioning. When irregular structures (with spatial scales of 300–400 m in the case of GNSS signals) are present in the ionosphere, interference patterns are generated on the ground under the Fresnel diffraction, producing fluctuations in the received signal strength. This is called ionospheric amplitude scintillation, and it is mainly associated with plasma bubbles in magnetic low latitudes. In contrast, in magnetic high latitudes, irregular fluctuations of the carrier phase are due to an ionospheric irregular structure crossing a satellite signal propagation path at high speed. This is called ionospheric phase scintillation. Ionospheric scintillation makes it difficult for a receiver to track a signal and reduces the accuracy of ranging, and if it is sufficiently strong, tracking of a satellite signal is interrupted. This effect is an issue regardless of positioning being in single frequency or dual frequencies. Ionospheric scintillation degrades positioning accuracy by reducing the number of satellites available for use.

Thus far, the discussion has mainly focused on the accuracy of GNSS-based positioning. However, it is important to note that this is not the only important factor and that the important indices differ depending on the usage. For example, in fixed-point surveying, millimeter-level positioning accuracy can be achieved by kinematic positioning with continuous measurements for a long time, where infrequent signal tracking interruption is not a problem because it can be corrected using long-time data. However, the same method cannot be applied for dynamic cases with high speed because it is difficult to correct the impact of signal tracking interruption. In contrast, in air navigation, although the required accuracy is meter level, very high reliability is required for real-time operation. It is necessary to ensure that the positioning solution is within a certain range and to issue an alarm if reliability of

the solution cannot be guaranteed. It is also important that the position solution is obtained without interruption. In such a case, anomalies that may occur very infrequently must be considered, if necessary, in comparison with the required reliability level. Thus, although "accuracy" tends to be focused on in GNSS positioning, it is important to discuss it using appropriate indicators.

Table 2.6 summarizes ionospheric phenomena and the GNSS applications affected by them. As mentioned above, the occurrence of an ionospheric effect strongly depends on the level of accuracy, reliability, and ease of use required for each GNSS positioning application. It should be noted that Table 2.6 provides general information, and some systems may be affected by ionospheric phenomena without "x." Conversely, some systems may not be affected by the items marked with "x."

2.1.3.3 GNSS Applications in Air Navigation

Air navigation has been shifting from ground-based conventional radio navigation systems to GNSS-based navigation systems, which can provide more flexible flight paths with higher traffic capacity, accuracy, and safety. In air navigation, international technical standards have been established for GNSS-based navigation, which include only accuracy, as described above, as well as safety-related requirements. Safety is the most important factor in air navigation, and safety levels ranging from $1 - 10^{-7}$ to $1 - 10^{-9}$ (99.999999–99.999999%) are required. However, in terms of accuracy, even for Category I approach, which provides precise approach guidance down to 60 m above the runway, the 95th percentile value of vertical errors of up to 4 m is allowed (ICAO 2018a, b).

Because the required level of safety is so high, GNSS satellite signals cannot be used without augmentation systems that correct errors in the GNSS and ensure safety. There are several types of augmentation systems such as the ground-based augmentation system (GBAS) and satellite-based augmentation system (SBAS), which can provide guidance for approach and landing (Fig. 2.11). An augmentation system improves the accuracy by correcting errors in the GNSS as well as monitors the GNSS signals and removes the error factors that cannot be corrected, to prevent the occurrence of hazardous misleading information, which may be a threat to the safety of GNSS users (Fig. 2.12). Because both the currently available GBASs and SBASs (as of July 2020) are based on the single-frequency code differential GNSS, the correction error associated with the spatial variation in the ionosphere is the most significant factor of the positioning error.

However, the ionospheric spatial variability during quiet time can be corrected by a GBAS and an SBAS with sufficient accuracy. In fact, in a GBAS flight experiment conducted at Ishigaki Island, the 95th percentile value of the vertical position errors was 1.17 m when the ionosphere was in quiet conditions (Saito et al. 2015). To achieve a safety level of $1 - 10^{-7}$–$1 - 10^{-9}$, a very conservative error limit, which is called the protection level, is used in air navigation. This corresponds to an error range of 5.4–6.1σ, when the error is assumed to follow the normal distribution with a

Table 2.6 Ionospheric disturbance phenomena and impacted GNSS applications. Note that mark "x" indicates that system may be impacted in general but may be different for some specific cases

Ionospheric effect		Ionospheric total electron content					Ionospheric irregularity	
		Ionospheric delay		Ionospheric delay gradient			Amplitude scintillation	Phase scintillation
Ionospheric phenomenon		Positive storm	Negative storm	SED/ SIPS	Plasma bubble	MSTID	Plasma bubble	Aurora
Single freq.	Single	x	x	x	x		x	x
	Relative			x	x		x	x
Single freq. (precise)	Single (PPP)			x	x	x	x	x
	Relative (phase)			x	x	x	x	x
Dual freq.	Single						x	x
	Relative					x	x	x
Dual freq. (precise)	Single (PPP)			x	x	x	x	x
	Relative (phase)			x	x	x	x	x

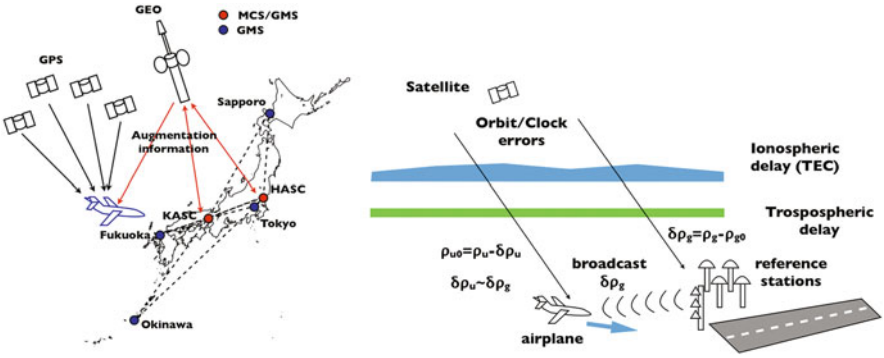

Fig. 2.11 Schematics of SBAS (left) and GBAS (right)

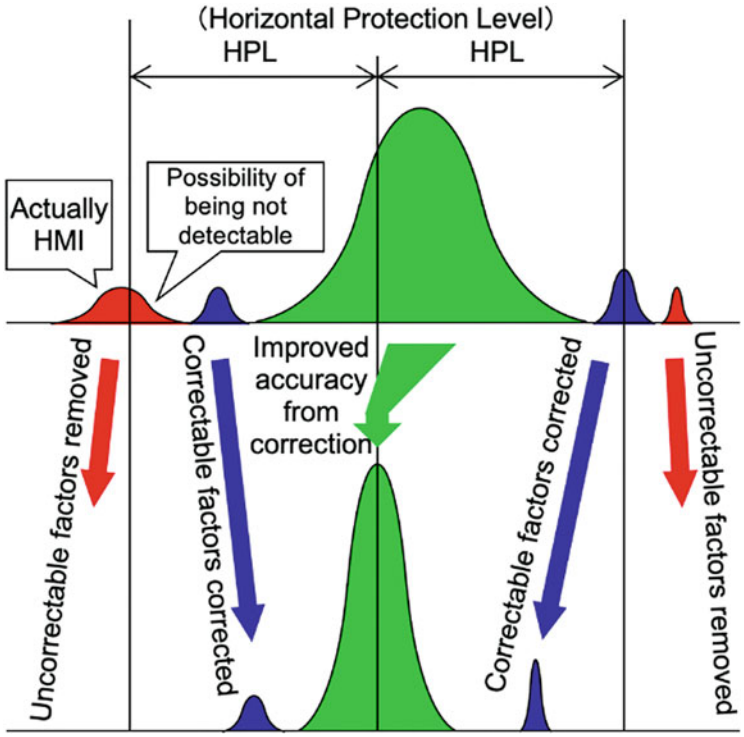

Fig. 2.12 Conceptual illustration of functions of augmentation system (Sakai et al. 2009)

standard deviation σ. In the calculation of the protection level, various conservative assumptions are made; the most notable of which is that about the existence of ionospheric gradients. In a GBAS and an SBAS, it is assumed that ionospheric gradients that can be threats to aircrafts always exist. This is because the ionosphere monitored by an augmentation system using the GNSS is not necessarily the same as

the ionosphere affecting the satellites used by an aircraft; thus, the augmentation system cannot determine the ionospheric state of an aircraft. Specifically, even if the probability is low, we cannot exclude the existence of ionospheric gradients exceeding the required level of safety, and therefore, ionospheric gradients are conservatively assumed to always exist. Although it will not be discussed in detail here, this assumption is the main reason for increasing the protection levels and occasionally exceeding the acceptable limit of potential errors corresponding to the operation (10–35 m for the Category I operation).

The above extremely conservative assumptions are made to make augmentation systems quantitatively evaluate risks independently of external systems. This is one of the major areas where space weather information can be utilized. The problem here is that augmentation systems can determine the state of the ionosphere with quantitative reliability only at "points" monitored by GNSS satellites. However, the presence of ionospheric gradients can be observed by various methods that have been developed for the study of ionospheric physics. For example, the spatial structure of a plasma bubble can be obtained using a VHF radar. A basic study to incorporate a VHF radar as an ionospheric monitoring tool for a GBAS showed that ionosphere-induced position errors in the GBAS could be reduced (Supriadi and Saito 2019).

In safety-critical GNSS applications, where various conservative assumptions are frequently made, space weather information can be utilized in many approaches. However, it should be noted that space weather information is obtained by an external system for the augmentation system; therefore, quantitative reliability for the space weather information is required. Specifically, the probability of missed detection of an ionospheric gradient that actually exists must be evaluated. However, because current augmentation systems assume that an ionospheric gradient always exists (i.e., the prior probability of occurrence is 1.0), the reliability of space weather information can be conservative. For example, a reliability of 90% (missed detection of one per ten opportunities) would be useful for augmentation systems by reducing the prior probability by one order.

Thus, it is very important to understand the technical background of targeted GNSS applications and incorporate the concept of reliability into space weather information.

2.1.3.4 Occurrence Frequencies of Ionospheric Phenomena that Degrade Positioning Accuracy

The ionospheric phenomena that degrade positioning accuracy depend on the positioning method. When an ionospheric delay (equivalent to the TEC) increases in a positive ionospheric storm, correction errors may be produced by empirical models, which may exceed the upper limit of the range of correction for relative positioning. Table 2.7 lists the estimated maximum TECs for different occurrence frequencies (resolution of 5 TECU) based on TECs observed in Japan in the past. The values are in the range of 70–90 TECU (equivalent to 11–17 m delay for a signal

Table 2.7 Ionospheric vertical TEC values corresponding to occurrence frequency over Japan (Ishii et al. 2021)

Location	Once per year	Once per 10 years	Once per 100 years
Hokkaido[a]	≥70 TECU	≥70 TECU	≥150 TECU
Tokyo	≥90 TECU[b]	≥130 TECU[c]	≥190 TECU
Kagoshima[d]	≥110 TECU	≥155 TECU	≥230 TECU

[a]Calculated as 0.8 times value of Tokyo based on statistical analysis
[b]Based on TEC observation data over 22 years
[c]Based on foF2 observation data over 62 years
[d]Calculated as 1.2 times value of Tokyo based on statistical analysis

at the GPS L1 frequency, the same as below) for once per year positive ionospheric storm events. They are 90–130 TECU (14–21 m) and 150–230 TECU (24–37 m) for once per 10-year and once per 100-year positive ionospheric storm events, respectively. Because the ionospheric electron density is maximum in the EIA crests at approximately ±15° magnetic latitude, the maximum TEC is larger in the southern part of Japan, which is located on the northern slope of an EIA crest.

MSTIDs, which may degrade the relative positioning performance using a carrier phase, are known to occur frequently in nighttime in summer and in daytime in winter in Japan. However, the characteristics of the occurrence frequency of MSTIDs that may degrade the positioning performance have not been studied well and remain to be explored in future research.

Plasma bubbles accompanying ionospheric gradients cause errors in relative positioning and result in ionospheric scintillation affecting GNSS positioning typically by reducing the number of available satellites. It is known that such plasma bubbles frequently occur in the period from sunset to after midnight in the spring and autumn equinoxes in the longitude around Japan. However, those degrading the positioning performance are observed relatively less frequently (Saito et al. 2017); therefore, a statistically relevant occurrence frequency for them has not been obtained yet. The occurrence frequency of ionospheric scintillation interfering with GNSS signal tracking (Saito and Yoshihara 2018), number of satellites simultaneously affected by scintillation and their probability (Beniguel and Adam 2007), and scintillation correlation with the ionospheric gradient, which affect the GNSS performance, still need to be statistically analyzed in detail.

2.2 Electric Power

2.2.1 Introduction

Geomagnetically induced currents (GIC) are quasi-DC caused by disturbances originating from the space environment, which can affect power grids, pipelines, and communication cables. For example, prolonged power outages were caused by GIC in Canada in March 1989 and Sweden in October 2003 (Bolduc 2002;

Pulkkinen et al. 2005). Once a power failure occurs, the social impact is enormous and the cost of restoration is high. Therefore, the establishment of guidelines for the evaluation of fault tolerance is desired. This section provides an overview of the impacts on power grids, causes, global trends, and current status of the numerical prediction of GIC.

2.2.2 Impacts of GIC on High-Voltage Transmission Network

A transformer is a component required to deliver power to a consumer and is used to change the AC voltage. When the excitation current exceeds a certain threshold, the transformer operates in the nonlinear region, and magnetic saturation occurs, distorting the output waveform. When a quasi-DC (GIC) with a sufficiently lower frequency than the frequency of a power transmission system (50 Hz or 60 Hz) increases, the half cycle of a sinusoidal wave saturates, whereas the remnant cycle remains unsaturated. This is called half-cycle saturation. When half-cycle saturation occurs, the reactive power increases, and harmonics with a frequency that is an integral multiple of the fundamental wave are generated. The generation of harmonics can cause the relays used to protect the power transmission system to malfunction. However, using digital protection relays, which are becoming increasingly well-known, can prevent the malfunctions due to above harmonics (Van der Zel 2007). Another problem is magnetic flux leakage to the outside environment, which causes eddy currents resulting in hot-spot heating. Hot-spot heating is known to deteriorate the insulating paper used in a transformer (Sakai et al. 2001b), and to generate bubbles in its insulating oil, which reduces the insulation strength and leads to dielectric breakdown (Sakai et al. 2001a). IEEE standard C57.91-1995 defines the upper limit of temperature associated with hot-spot heating. From the viewpoint of risk assessment, it is desirable to understand the relationship between GIC and the temperature of a transformer in advance. For more information on the effects of GIC on transformers, readers may refer to Van der Zel (2007).

2.2.3 Prevention of Hazards Due to GIC

2.2.3.1 United States of America

In March 2019, the President's Office released the National Space Weather Strategy and Action Plan (https://www.whitehouse.gov/wp-content/uploads/2019/03/National-Space-Weather-Strategy-and-Action-Plan-2019.pdf). Power outages due to GIC were identified as a space weather event of concern. Thus, rapid and accurate forecasting and planning for response and recovery from space weather events are required.

The North American Electric Reliability Council (NERC) has proposed various benchmark tests to ensure the reliability of transformers against GIC (https://www.nerc.com/pa/Stand/Pages/TPL0071RI.aspx). The reference values (scenarios) based on the events on March 13–14, 1989, were used to evaluate the geomagnetically induced electric field (GIE) and determine the GIC flowing through the North American transmission network. Hot-spot heating of each transformer was also predicted from the time series data of the estimated GIC. The same report (TPL-007-1) stated that a thermal impact assessment must be performed for transformers of 200 kV and above, if the GIC derived from the above scenario flows more than 75 A per phase. The report also recommended the protection and control of power grids against magnetic storms. (https://www.nerc.com/pa/RAPA/ra/Reliability%20Assessments%20DL/2012GMD.pdf).

The Space Weather Prediction Center of the National Oceanic and Atmospheric Administration (NOAA), an agency responsible for space weather operations of the US federal government, provides real-time maps of the GIEs in the United States of America (USA) (https://www.swpc.noaa.gov/products/geoelectric-field-1-minute). Such maps are useful for understanding the current status of GIC, in addition to the scale of magnetic storms (G-scale) and the Kp index.

The Community Coordinated Modeling Center of the National Aeronautics and Space Administration (NASA) has been conducting the Solar Shield project since 2007 to predict GIC using models and simulations together with CME data and real-time solar wind data (Pulkkinen et al. 2009). Currently, the project also provides GIE information and other data (https://ccmc.gsfc.nasa.gov/iswa/).

2.2.3.2 Canada

GIC are routinely monitored at transformers (Marti and Yin 2015). A GIC simulator has been developed to derive the GIC flowing through a substation in real time based on the geomagnetic field, auroral electrojet, and subsurface electrical conductivity (Boteler et al. 2014; Boteler 2014). In response to the NERC report, the Geological Survey of Canada published a report for the assessment of GIC flowing in Alberta (https://doi.org/10.4095/296909).

2.2.3.3 United Kingdom

The power grid suffered voltage drops, alarm activation of reverse-phase current compensators, transformer failures, and high-level harmonic currents due to transformer saturation during the magnetic storms that occurred on July 14, 1982, March 13–14, 1989, October 19–20, 1989, and November 8, 1991 (Erinmez et al. 2002). The impacts on the consumers were limited. In response to the above events, the transmission and distribution company, National Grid plc, undertook a risk assessment and risk management study on magnetic storms of various magnitudes. The British Geological Survey provides maps of GIE and an index called B_{GIC} in real

time (https://geomag.bgs.ac.uk/data_service/space_weather/gic_services.html). The Royal Academy of Engineering published a report on the impacts of extreme space weather on engineering systems in 2013, which described the effects of GIC on the power grid (Cannon 2013).

2.2.3.4 Finland

The Finnish Meteorological Institute provides real-time forecasts of geomagnetic variations, GIE, and GIC using geomagnetic data over the Scandinavian Peninsula, including Finland (http://swe.ssa.esa.int/web/guest/fmi-federated).

2.2.3.5 European Union

The European Risk from Geomagnetically Induced Currents project was conducted in 2011–2014 by a consortium of Finland, Hungary, Sweden, the UK, Russia, and the USA (Viljanen 2011). The project was aimed at predicting the temporal variation in the geomagnetic field with a lead time of 30 min and presenting a risk map for a wide area in Europe.

2.2.3.6 Australia

The GIE distributions over Oceania, including Australia, are provided using the GIC index (Marshall et al. 2011) in real time (https://www.sws.bom.gov.au/Geophysical/1/2/5). Since mid-2020, GIC measurements have been made at substations along the southeast coast of Australia in cooperation with power companies (Marshall et al. 2013).

2.2.3.7 Japan

By considering the geomagnetic latitude and ground conductivity in Japan, Tsuboi and Horiuchi (1988) suggested that GIC do not have a severe impact on the Japanese power grid. After the major blackouts caused by GIC in Canada in 1989, an impact assessment was conducted by Japanese electric companies. The Tokyo Electric Power Company conducted experiments by imposing artificial DC on a transformer (Takasu et al. 1993, 1994). GIC have also been measured (Ishikura et al. 2010). Response functions (impedance) that relate the geomagnetic variations and GIC measured at Shin-Hatano, Shin-Fuji, and Shin-Yamanashi substations have been obtained (Harada et al. 2009). Chubu Electric Power Co., Inc. also measured GIC. Reasonably high GIC values were recorded, such as approximately 7 A and 42 A at the Toei and Sunen substations during the magnetic storms on October 28, 1991 (Yumoto 1991) and November 6, 2001 (Tabata and Ogawa 2003), respectively.

Metatech, Inc. surveyed the GIC values at the substations of Chubu Electric Power Company and showed a good correlation between the GIC flowing in its power grid and the maximum amplitude of the Dst index (Kappenman 2010). The Dst index is a measure of a magnetic storm. The above result suggests that the ring currents are considered as major current sources of the GIC flowing in the Japanese power grid.

The Ministry of Economy, Trade and Industry published a report entitled "FY 2014 Research and Service Contract Report on Standards Related to Technical Standards for Electric Facilities" in March 2015. According to the report, the GIC flowing in the Japanese power grid increased to 129.09 A on October 31, 2003. However, no significant anomalies were detected in the power grid.

Under the Project for Solar–Terrestrial Environment Prediction (PSTEP), which was conducted from 2015 to 2020, GIC measurements began at four substations in 2017–2018 with the cooperation of electric power companies. Data were acquired at a sampling rate of 10 Hz to capture fast fluctuations in GIC (Watari et al. 2021). GIC variations associated with sudden commencements (SCs), magnetic storms, bay magnetic disturbances, and solar flare effects were shown. Ebihara et al. (2021) obtained coefficients related to the geoelectric field observed at the Kakioka Geomagnetic Observatory and the GIC at four substations. They showed that the calculated GIC were in good agreement with the observed ones with correlation coefficients of 0.81–0.91. Using the parameters suggested by Love and Swidinsky (2014), Ebihara et al. (2021) used the geomagnetic field variations observed in Bombay (currently Mumbai) in September 1859 (known as the Carrington Event) and thereby estimated the geoelectric field and the GIC that could flow in the Japanese power grid in case that the Carrington storm occurs again. The results are shown in Fig. 2.13. When the Carrington storm occurs again, the GIC will increase to 89 ± 30 A at Shin-Fukushima, 20 ± 7 A at Shin-Tsukuba, and 19 ± 27 A at Shin-Fuji. These values are lower limits because the number of substations investigated was limited.

2.2.4 Causes of GIC

The ultimate origin of a GIC is in the Sun. Here, we briefly review the processes causing GIC.

2.2.4.1 Sun

The solar disturbances that can cause large-amplitude GIC include flares, prominences, and coronal holes. A flare is an explosive phenomenon caused by a magnetic flux emerging from a growing sunspot group. Prominences are elongated structures that float in the corona, the outer atmosphere of the Sun, and may abruptly erupt into the interplanetary space, and are frequently accompanied by CMEs. Coronal holes are seen as "dark" regions at extreme UV wavelengths. High-speed solar wind blows

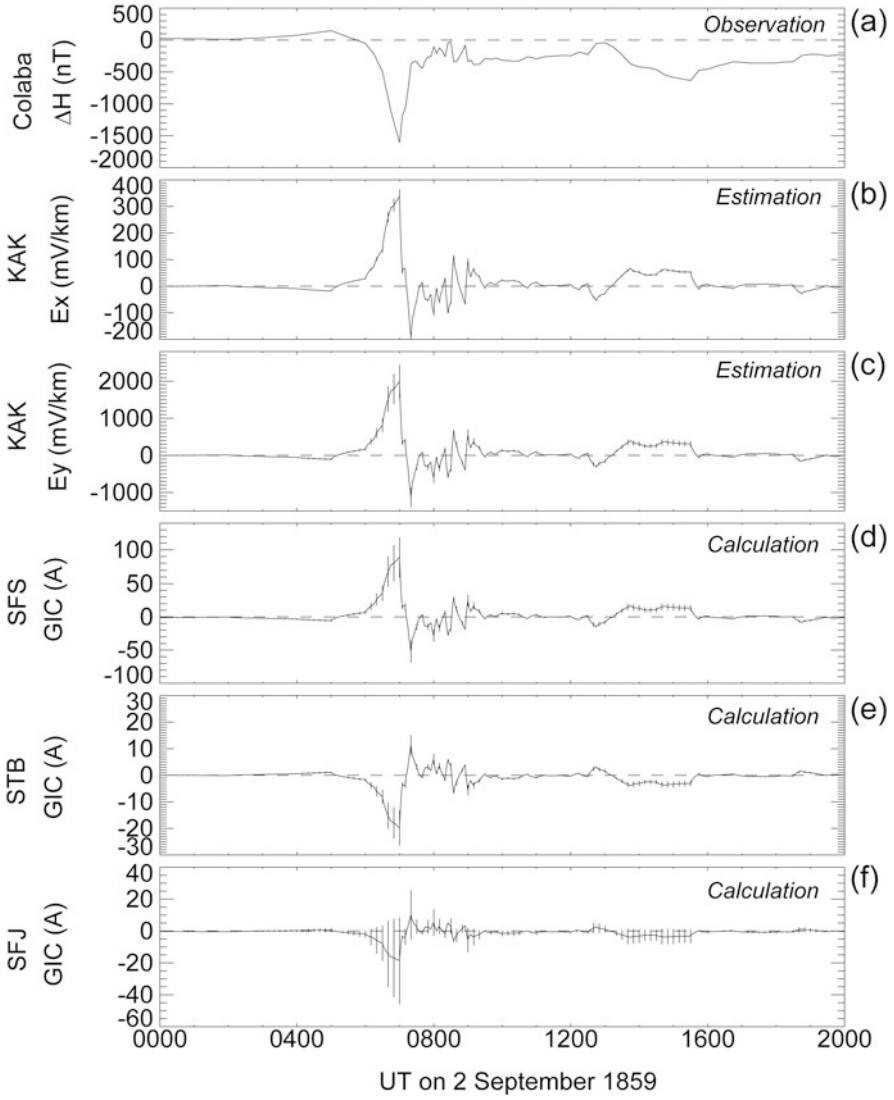

Fig. 2.13 (**a**) Geomagnetic variations recorded at Colaba, India, during Carrington Event on September 2, 1859, (**b**) north–south and (**c**) east–west components of geoelectric field expected to appear at Kakioka Geomagnetic Observatory during Carrington Event, (**d–f**) GIC expected to flow at three Japanese substations (Ebihara et al. 2021)

in coronal holes. Although large coronal holes frequently appear in the polar regions of the Sun, if they also appear near the equator, they are more probable to affect the Earth. The impacts of coronal holes on geomagnetic activities are smaller than those of CMEs. For more information on the structure of the Sun, please refer to Chap. 8.

2.2.4.2 Solar Wind

The Sun always emits ionized gases (a plasma) into the interplanetary space. The speed of the solar wind is variable. For example, the speeds of CMEs vary from 100 to 2000 km/s. A shock is generated in front of the main body of a CME. When this shock arrives at the Earth, the magnetosphere is rapidly compressed, resulting in geomagnetic variations known as SSCs or SIs. Large-amplitude GIC are known to flow during SSCs/SIs (Kappenman 2003). When the southward component of the magnetic field is embedded in a sheath (between the shock and the main body of a CME) and/or in the main body of a CME, it can intensify equatorial ring currents. This causes geomagnetic field distributions for a few days, known as magnetic storms. A corotating interaction region (CIR) is formed by the interaction of the high-speed solar wind blowing out from a coronal hole and its preceding low-speed solar wind. In the vicinity of the orbit of the Earth, forward and backward shock waves are frequently formed. The compression increases the density and the magnetic field, and the latter becomes turbulent. If the turbulent magnetic field contains a southward component, it can cause weak magnetic storms. Owing to the rotation of the Sun, the magnetic storms associated with CIRs tend to occur with a period of ~27 days. For more information on solar wind disturbances, please refer to Chap. 9.

2.2.4.3 Magnetospheric and Ionospheric Currents

The space dominated by the magnetic field of the Earth is called the magnetosphere. The magnetic field lines of the Earth are stretched to the opposite side of the Sun (the night side) by the solar wind and shaped like the tail of a comet. When a CME arrives at the Earth's magnetosphere, a shock wave in front of the CME can compress the magnetosphere abruptly. The compression of the magnetopause gives rise to the enhancement of the magnetopause current and the ionospheric current, resulting in rapid geomagnetic disturbances. The rapid geomagnetic disturbance is called a sudden storm commencement (SSC), or a sudden impulse (SI). Here, we simply refer to as a sudden commencement (SC) for the sake of simplicity. When the southward magnetic field component is embedded in CMEs or CIRs, the ring current and the ionospheric current develop almost simultaneously, resulting in geomagnetic disturbances. The ring current consists of the westward current (in the outer part) and the eastward current (in the inner part). Since the total amount of the westward current dominates that of the eastward current, the ring current gives rise to an increase in the horizontal component of the geomagnetic field. Twin current vortices are also developed in the ionosphere, which also give rise to change the geomagnetic field with its polarity and magnitude depending on the magnetic local time. These geomagnetic disturbances lasting for several days are called magnetic storms. At low latitudes, the horizontal component of the geomagnetic field decreases for a few to several hours and then recovers to the prestorm level for a few days. At high latitudes, large-amplitude, rapid geomagnetic disturbances are prominent. They

successively occur and each of the them lasts for several hours. They are primarily caused by auroral electrojet and accompanied by explosive auroral brightening, which are called substorms. During the substorm, the geomagnetic fields are significantly disturbed, in particular, at high latitudes. A noticeable geomagnetic disturbance also appears at low latitudes during the substorms. The horizontal component of the geomagnetic field gradually increases and decreases for about 1 h, which are caused by the field-aligned currents. This kind of the gradual geomagnetic disturbances is called a bay (or a magnetic bay) and is commonly observed at low and midlatitudes during the substorms.

2.2.4.4 Geomagnetically Induced Electric Field

The temporal variations of the geomagnetic field induce an electric field in the ground based on Faraday's law, which is called GIE. The spatial distribution of the GIE is, in general, nonuniform because of nonuniform source current, and nonuniform subsurface resistivity distribution. Because the subsurface resistivity varies significantly with the depth, the depth (skin depth) of the electromagnetic field penetrating the interior of the Earth depends on the frequency of the imposed magnetic disturbances. Thus, the amplitude of a GIE also depends on the frequency. In Japan, the Japan Meteorological Agency (JMA) had conducted observations of the geoelectric field at Kakioka, Memanbetsu, and Kanoya until 2021. For more information on the geoelectric field and the resistivity structure, please refer to Chap. 6.

2.2.4.5 Geomagnetically Induced Currents

When an electric field is induced in the ground, electric current flows in long conductors that are grounded such as power lines, pipelines, and communication cables, which is called GIC. In extra-high-voltage transmission lines, the amplitude of the GIE is generally high because their resistance and grounding resistance are, in general, small. GIC can be measured by measuring the current flowing in the conductor connecting the neutral point of a transformer to the ground or by measuring the magnetic field fluctuation around the transmission line.

2.2.5 Forecast of GIC

2.2.5.1 Sun

Typically, it takes a day to a few days for a CME from the Sun to reach the Earth. If we detect the CMEs heading toward the Earth and obtain their images by coronagraphs, we will be able to estimate the propagation speed and to predict GIC with a

lead time of a few days. Solar flare, which is one of the causes of CMEs, has been predicted by various organizations in various countries. While most of the predictions are based on empirical knowledge, the group of Prof. Kusano at Nagoya University has developed an algorithm for physical prediction (Kusano et al. 2020). For details, please refer to Chap. 10.

2.2.5.2 Solar Wind

By analyzing coronagraphs, one can approximately estimate the speed of a CME and its arrival time. A more accurate approach is to incorporate the information of the CMEs that have departed from the Sun into solar wind simulations, e.g., the WSA-ENLIL model (Odstrčil and Pizzo 1999) and SUSANOO-CME (Shiota and Kataoka 2016). The solar wind is monitored by the Advanced Composition Explorer and Deep Space Climate Observatory spacecrafts at Lagrange point L1, which is approximately 1.5 million km away from the Earth in the direction of the Sun. It takes several tens of minutes to an hour for the solar wind to reach the Earth from L1, whereas radio waves reach the Earth in approximately 5 s. Thus, we can obtain the magnetic field, speed, and density of the solar wind several tens of minutes to an hour in advance. Using the solar wind data obtained at L1, it is possible to evaluate its effects on the magnetosphere of the Earth.

2.2.5.3 Magnetospheric and Ionospheric Currents (Geomagnetic Variations)

Empirical and physics-based models are useful for predicting the geomagnetic disturbances caused by the magnetospheric and ionospheric currents. For the contributions from the magnetopause and the ring currents, an empirical model suggested by Burton et al. (1975) enables prediction of the Dst index (geomagnetic variability averaged over mid- and low latitudes in the longitude direction). The model uses the solar wind velocity, solar wind density, and interplanetary magnetic field as input parameters (Burton et al. 1975). Rastätter et al. (2013) tested 30 empirical models and physics-based models for the Dst index and concluded that the former provide better results than the latter. An auroral electrojet can cause large-amplitude geomagnetic variations and, hence, large-amplitude GIC in the polar regions (Pirjola and Viljanen 1989). As summarized by Liemohn et al. (2018), an auroral electrojet has been predicted using a linear impulse model (Clauer et al. 1981), neural networks (Gleisner and Lundstedt 2001; Amariutei and Ganushkina 2012), and empirical formulas (Luo et al. 2013). Recent magnetohydrodynamic simulations can reproduce an observed auroral electrojet reasonably well on the basis of the temporal variations in the solar wind parameters (Ebihara et al. 2019).

2.2.5.4 Geomagnetically Induced Electric Field

There are several methods to determine an electric field induced on the ground. To obtain a geoelectric field distribution, the finite element method (Rodi 1976) and the finite difference method (Mackie et al. 1994) have been used. Goto (2015) compared the results of one-, two-, and three-dimensional calculations and pointed out that three-dimensional calculations are important for evaluating GIC. Fujita et al. (2018) used the finite difference method to determine the global distribution of the geoelectric fields over a wide area of the Japanese archipelago. The above calculations were conducted in the frequency domain; however, in another study, the three-dimensional time evolution of the electromagnetic field was also directly solved by finite difference time domain simulations (Nakamura et al. 2018). A local geoelectric field can be obtained by the convolution of the geomagnetic variations (Cagniard 1953). The convolution method was shown to reproduce well the geomagnetic variations measured at the Kakioka Geomagnetic Observatory (Love and Swidinsky 2015).

2.2.5.5 Geomagnetically Induced Currents

There are two methods for deriving GIC numerically. The first one is to use transfer functions that relate a geoelectric (geomagnetic) field and a GIC (Ingham 1993; Pulkkinen et al. 2007). This method is employed for power grids, e.g., in Finland (Pulkkinen et al. 2007), New Zealand (Ingham et al. 2017), and Japan (Harada et al. 2009; Ebihara et al. 2021). It is applicable for substations at which the transfer functions are obtained in advance. The second method, which is appropriate for many substations, is to consider a power grid as a circuit and subsequently calculate the current flowing through the entire power grid based on the conservation of current (Viljanen and Pirjola 1994). In the PSTEP, the GIC flowing in the Japanese power grid with a voltage class of 187 kV or higher were modeled. An example is shown in Fig. 2.14. A slightly eastward (magnetic eastward) and uniform geoelectric field is applied to calculate the GIC flowing through the power grid. The radius of the circle represents the magnitude of GIC, and the color indicates its polarity (blue and red for GIC flowing in and out of the grid, respectively). In this case, the GIC tend to flow into the grid near its western end, whereas they flow out of the grid near the eastern end.

2.2.6 Summary

The generation mechanisms, impacts on power grids, and current status of forecast models of GIC are overviewed. The occurrence frequency of grid failures due to GIC is low, particularly in low and midlatitude regions such as Japan (Pulkkinen et al.

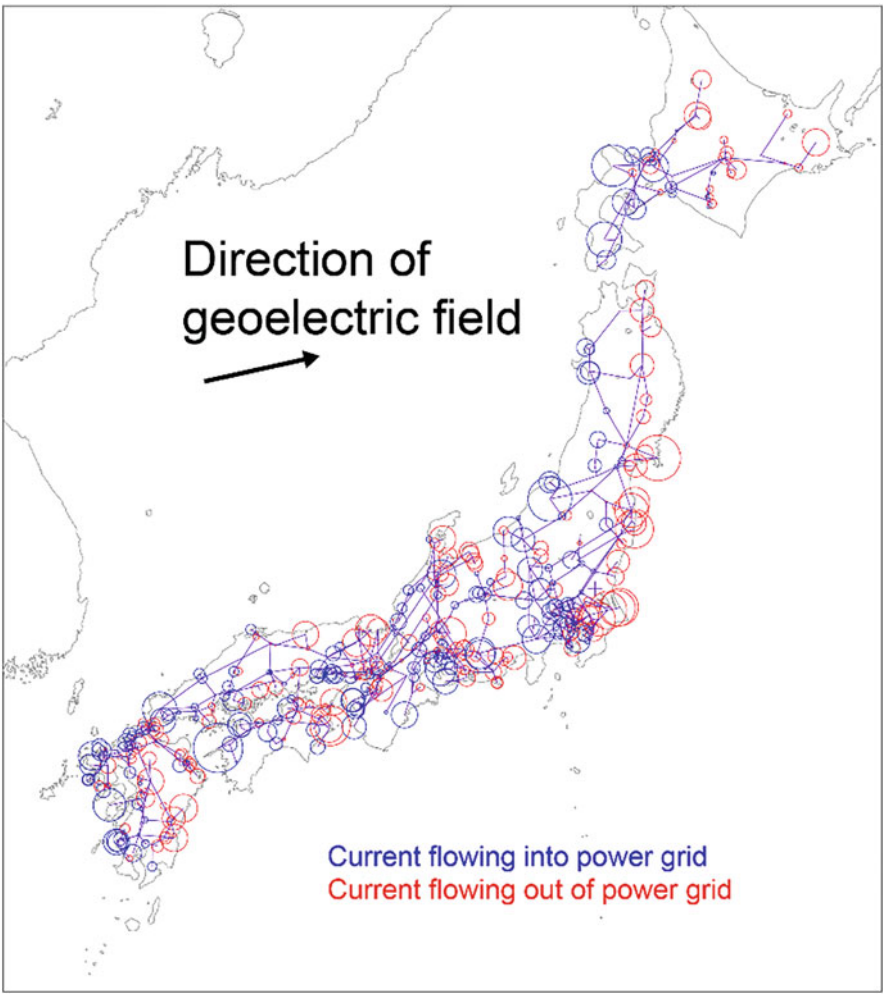

Fig. 2.14 Example of GIC flowing in Japanese power grid with voltage equal to or above 187 kV. Radius of circle indicates magnitude of GIC. Blue (red) color indicates GIC flowing into (out of) the grid

2012). Although large-amplitude GIC have been observed, such as 55.8 A in China (Liu et al. 2010) and 129 A in Japan (METI 2015), thus far, there have been no reports of blackouts caused by GIC, at least, in Japan. However, as described in Chap. 11, many high-risk super-magnetic storms have occurred in the past. For example, magnetic storms of similar or greater magnitude than the one in March 1989, which caused a prolonged power outage in Canada, have occurred at least five times in the past 160 years. The occurrence probability of such a super-magnetic storm in the near future is not zero. It is desirable to consider GIC as a low-frequency mega-disaster that is infrequent but can cause enormous damage if it occurs.

To accurately predict a GIC, it is essential to correctly understand the comprehensive processes from the Sun to the Earth as a chain reaction. The PSTEP from 2015 to 2020 has established a research infrastructure for predicting GIC by close collaboration among researchers in the individual research fields including the Sun, solar wind, magnetosphere, ionosphere, and solid Earth. This framework is expected to be maintained for improving the GIC prediction accuracy.

2.3 Satellites

The space environment differs significantly from the environment on the surface of the Earth where humans live. Space is a region of tenuous ionized gases (a plasma), where the mean free path of a particle is long and there is no effect of the absorption of the solar UV radiation by the atmosphere. Therefore, spacecrafts such as satellites are affected by high-energy particles such as galactic cosmic rays, SEPs, and radiation belt particles trapped in the magnetosphere of the Earth and by the solar UV radiation. Interactions between satellites and their surrounding plasma particles also occur in the space environment.

Because the characteristics of the plasma, such as temperature and density, differ depending on the region of the space environment, its effects on satellites are also different. The geospace, which is the space environment around the Earth, fluctuates significantly with time under solar activities such as solar flares and CMEs and disturbances in the magnetosphere of the Earth such as substorms and geomagnetic storms caused by the interaction between the solar wind and the magnetosphere. In addition, a local effect is caused by the nonuniformity of the geomagnetic field intensity called the south Atlantic anomaly (SAA) in low-Earth orbits (LEOs). Therefore, in the design phase of space systems such as satellites, it is necessary to adopt countermeasures considering the impacts of the space environment. To avoid the risks from disturbances in the space environment to spacecraft and manned space activities, we need to recognize the effects caused by the space environment and their risks using space weather forecasts and analyze the primary cause immediately when an anomaly occurs.

In this section, the impacts on satellites due to the space environment and their specific examples are explained. The effects of the upper atmosphere on the orbital motion of satellites in LEOs are described in Sect. 2.5.

2.3.1 Surface Charging

The charging of satellites by their interaction with solar irradiance and surrounding plasma is called surface charging. It is caused by the accumulation of electric charge from the currents that flow into and out of the satellite over time. These currents are due to the surrounding space plasma (electrons and ions) flowing into the satellite

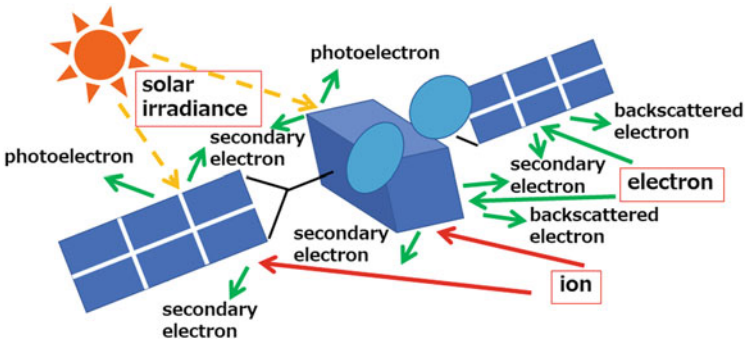

Fig. 2.15 Charged particles contributing to charging of satellite surface by interaction of satellite with sunlight and space plasma

surface, secondary and backscattered electrons caused by the interaction between the space plasma flowing in and the surface materials, and photoelectrons emitted from the satellite surface by solar irradiance (Fig. 2.15). The space plasma with energy ranging from several keV to several tens of keV mainly contributes to surface charging.

Charging against the surrounding space plasma is called absolute charging, and it is a floating potential formed by the equilibrium state of the currents flowing in and out between the satellite and the surrounding space plasma, as described above, and is frequently simply called the charging potential.

Because the ease of photoelectron and secondary electron emissions differs with the material, if the materials and structures of satellite surfaces are different, their respective potentials also differ. The potential difference between the satellite surface material and the satellite structure is called the differential charging potential, and the charging is called differential charging. When the differential charging potential reaches the discharge potential, a discharge occurs, causing satellite anomalies such as damage to the solar panels on the satellite surface and satellite malfunctions.

The satellite potential varies depending on the state of solar irradiance (sunlight, eclipse) and the surrounding plasma environment. In a GEO, a phenomenon called a substorm can rapidly increase the fluxes of electrons and ions of several kiloelectron volts to several tens of kiloelectron volts. The associated development of differential charging increases the risk of electric discharges that can cause satellite malfunctions.

The satellite anomalies caused by surface charging that have been reported thus far occurred during periods of high geomagnetic activity in spring and autumn (Choi et al. 2011) and from midnight to morning sector (McPherson and Schober 1975). This suggests that surface charging and discharge-induced anomalies are caused by the injection of hot electrons associated with substorms into GEOs.

Example of Substorm-Related Failures: On April 11, 1997, a failure occurred in the transponder and solar array of geostationary satellite TEMPO 2. It is believed to

have been caused by an electrical discharge in the solar array paddles associated with a substorm particle injection (Ferguson et al. 2015). This discharge accident triggered considerable research on spacecraft charging, and ground-based charging and discharging tests were conducted before a launch (Toyoda 2018).

On April 5, 2010, geostationary satellite Galaxy 15, near midnight, failed to respond to commands from the ground station. It is considered that high-temperature electrons associated with a large-scale substorm event were injected into Galaxy 15, which was in eclipse at the time of the failure, causing surface charging and discharging (Loto'aniu et al. 2015).

2.3.2 *Internal Charging*

A charged particle incident on a material loses its energy by ionizing and exciting the surrounding atoms. The depth at which a charged particle with a certain energy penetrates a material is called the range. If the thickness of the material is smaller than the range, the charged particle penetrates through the material. Figure 2.16 shows the range of electrons and protons penetrating aluminum. The stopping power and range tables for electrons (ESTAR) and stopping power and range tables for protons (PSTAR) of the National Institute of Standards and Technology in the USA were used to calculate the range (Berger et al. 1998). Because the typical thickness of

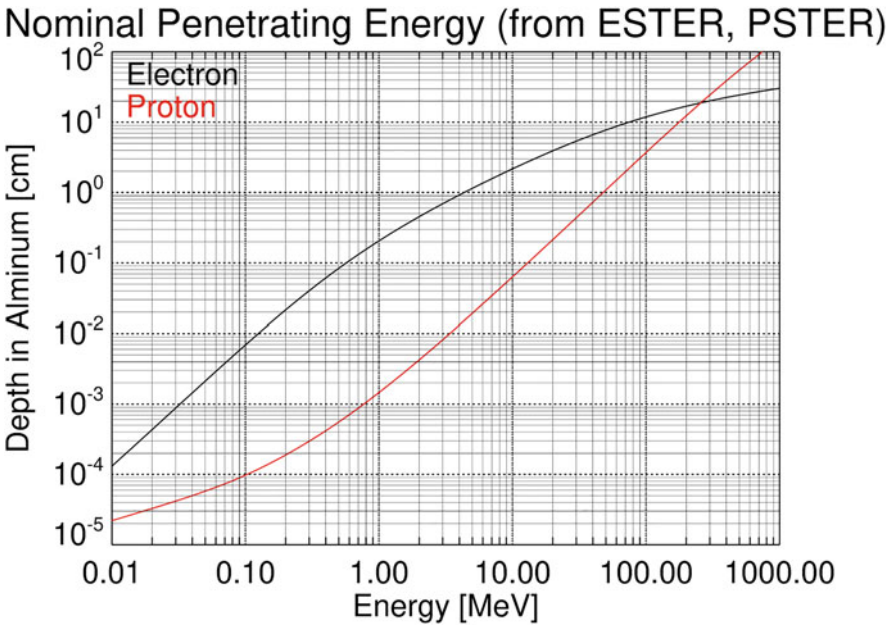

Fig. 2.16 Electron and proton distances in aluminum calculated by ESTAR and PSTAR

a satellite structure is more than 0.762 mm (30 mils) aluminum equivalent, electrons and protons of more than 0.5 MeV and 10 MeV, respectively, can penetrate a satellite structure and enter it. This phenomenon is called internal charging. The energy range of the particles that contribute to internal charging depends on the material and thickness of the charged area and the material and thickness of the satellite structure on the path of the particles that enter the area.

In the geospace, electrons with energies of several megaelectron volts are trapped in the outer radiation belt, which is located at an altitude of approximately two to seven times the radius of the Earth (6378.2 km). Moreover, they penetrate the structures of satellites in medium-Earth orbits (MEOs) and GEOs that pass through the outer radiation belt, causing internal charging. The contribution of protons to internal charging is almost negligible compared to that of the electron flux in the outer radiation belt, although protons of more than 10 MeV can increase during a solar proton event.

Internal charging develops when the electron flux in the outer radiation belt increases, because the charging factor in a satellite structure is almost only the high-energy electrons that enter it. Significant fluctuations of the electron flux in the outer radiation belt are due to the solar wind–magnetosphere interactions. For details of the behavior and mechanism of the variations, please refer to Chap. 5. Similar to surface charging, when the differential potential reaches the discharge potential by the development of charging, the electrons are discharged, causing problems in the internal integrated circuits. Specifically, the risk of satellite failure increases when the flux of the outer radiation belt electrons remains high and the number of particles (fluence) incident on the satellite structure increases. It is considered that the complexity of the elements associated with the high performance of components used in satellites increases the risk due to internal charging.

Satellite anomalies that can be attributed to internal charging are known to occur more frequently in the dayside and periods of low solar activity than in the nightside and periods of high solar activity, respectively (Wrenn et al. 2002; Lohmeyer and Cahoy 2013). The former is consistent with the facts that the dayside electron flux of outer radiation belt in GEO is higher than the nightside one, since the dayside of the magnetosphere of the Earth is compressed by the solar wind. The latter is consistent with the fact that a continuous increase in the outer radiation belt electrons tends to occur during the declining phase of the solar activity.

Examples of Anomalies Caused by Electron Enhancement in Outer Radiation Belt: Some examples of satellite anomalies that occur in response to electron enhancement in the outer radiation belt and are caused by internal charging are introduced below.

During January and February 1994, several communication and broadcasting satellites in GEOs experienced anomalies. On January 20, 1994, the Intelsat K communication satellite and the Anik E-1 and 2 communication satellites of Canada experienced a momentum wheel failure, which prevented communication. In addition, on the evening of February 22, 1994, the satellite broadcasting service of Japanese satellite BS-3a was interrupted for 60 min and subsequently resumed by switching to backup satellite BS-3b. During this period, high-energy electron

observations by Geostationary Operational Environmental Satellite (GOES) 7, a meteorological satellite, showed that in the GEOs, electron flux above 2 MeV was continuously high since January 13 and February 8, 1994, respectively, in association with recurrent geomagnetic disturbances caused by internal charging (Leach and Alexander 1995).

In addition, there was a total loss of geostationary satellite Telstar 4 on September 19, 2003. At that time, the high-energy electron flux in its GEO was at a high level for several days owing to a geomagnetic disturbance caused by the high-speed solar wind that arrived on September 16, 2003 (Choi et al. 2011).

2.3.3 Single-Event Effects

When high-energy particles such as protons and heavy ions (cosmic radiation) are incident on semiconductor devices in satellites, electron–hole pairs are generated along their trajectories of the particles. The generated electron–hole pairs cause transient currents and voltage fluctuations in the semiconductor devices, causing transient malfunctions (soft errors) or permanent damage of the devices and circuits (hard errors). Because their effects differ depending on the type of semiconductor device or circuit, the ionization effect associated with the injection of a single charged particle is called a single-event effect (SEE). Various SEEs are due to the differences in the device types and generation mechanisms, and new SEEs have been reported owing to the miniaturization of devices and the development of new ones. Major SEEs are as follows:

Single-Event Upset (SEU): It is a soft error phenomenon in which malfunction of a circuit in a large-scale integration such as a memory or flip-flop is due to an electrical charge generated by the radiation incident on it, resulting in data inversion.

Single-Event Transient: It is a soft error phenomenon resulting from a digital pulse propagating in a logic circuit owing to the output voltage fluctuation of the logic element caused by the generated charge when a high-energy particle is injected into the logic element.

Single-Event Latchup: It is a soft error phenomenon in which a thyristor structure, which is mainly contained in a complementary metal oxide semiconductor, is turned on by the noise current due to the incidence of high-energy particles, and there is a large current flow. If this large current flows for a long time without resetting the circuit, this error may become a hard error.

Single-Event Burnout: It is a hard error phenomenon that occurs when a large current flows in a power metal oxide semiconductor field effect transistor owing to an *off* device being turned *on* by radiation. This phenomenon also occurs in bipolar junction transistors and insulated gate bipolar transistors.

Single-Event Gate Rupture: It is a hard error phenomenon in which a gate dielectric film with a high electric field is destroyed by an incident high-energy particle.

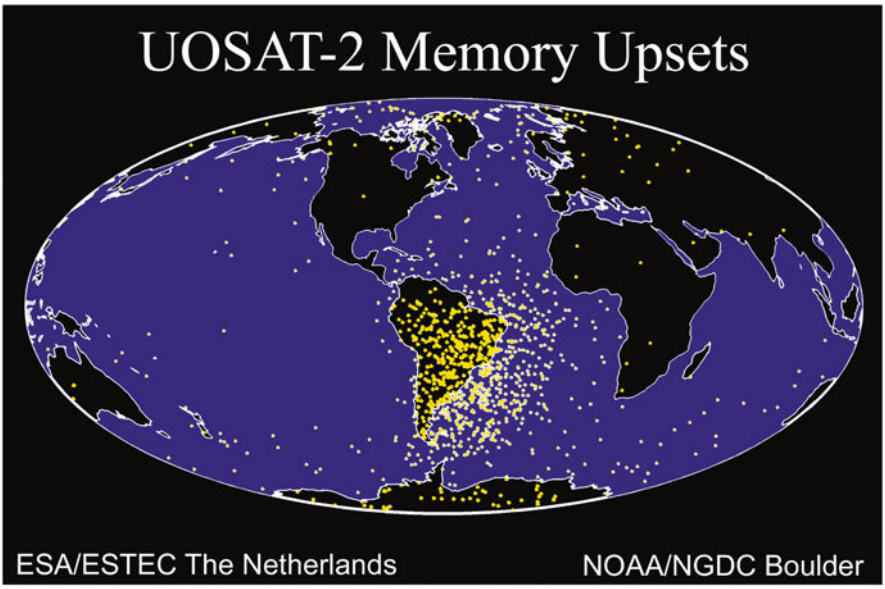

Fig. 2.17 SEU occurrence distribution for UoSAT-2 satellite (courtesy of NOAA/National Centers for Environmental Information)

Because soft errors cause errors and malfunctions of data in memory, countermeasures such as introducing error correction functions and methods such as triple majority voting have been considered. In contrast, hard errors cause functional loss of a satellite; therefore, measures such as thicker shielding for important elements and circuits are necessary.

The main particles contributing to SEEs differ depending on the satellite orbit: galactic cosmic rays and SEPs for MEO and GEO satellites and inner radiation belt protons for LEO satellites.

Example of SEE Anomaly: During a solar proton event, an SEE caused an anomaly in the attitude control of the BSAT-2a broadcasting satellite in a GEO at 4:17 p.m. Japan Standard Time (JST) on September 25, 2001, which interrupted the broadcast for approximately 1 h. At that time, a solar proton event with a maximum of 12,900 PFU of protons of 10 MeV or higher occurred in conjunction with the X2.6/2B flare in September 24. The SEU generated by this event is considered to be the cause of the above SEE anomaly. In addition, an anomaly occurred at 21:26 JST on January 17, 2005, when the attitude and telemetry signals of broadcasting satellite JCSAT-1B in a GEO, which was under orbit control, were lost. At this time, a solar proton event with more than 10 MeV protons reaching 5130 PFU occurred, and some command error might have been a cause of the anomaly.

As an example of an SEE in a LEO, Fig. 2.17 shows the distribution of the SEU occurrences in British satellite UoSAT-2. Owing to the nonuniformity of the geomagnetic field intensity, SEUs are frequently observed in the SAA region where the

inner radiation belt protons are distributed up to the altitude of a LEO and in the root region of the inner zone.

2.3.4 *Total Dose Effect/Cumulative Dose Effect*

Permanent damage to semiconductor components and materials of satellites caused by the cumulative effects of incoming energetic particles is called the total dose effect/cumulative dose effect. The total dose effect is mainly caused by the protons and electrons in the radiation belts and SEPs (protons). It is classified into the total ionizing dose (TID) effect and the displacement damage dose (DDD) effect.

The degradation of physical properties due to the accumulation of electron–hole pairs generated in semiconductor devices by the injection of high-energy charged particles is called the TID effect. The charge accumulated in the SiO_2 insulating film of a semiconductor device forms a fixed charge or interface states, which shift the threshold voltage, increase the supply current (leakage current), change the operation timing, and degrade various characteristics of the semiconductor device. The TID effect is proportional to the absorbed dose.

The DDD effect is a phenomenon in which lattice defects are created in semiconductor crystals by the injection of high-energy particles, resulting in changes in the physical properties. It can be observed in optical devices, charge-coupled devices, and solar cells. For example, degradation of a solar cell onboard a satellite is one of the DDD effects.

Example of Total Dose Effect: As a specific example of the TID effect, the results of ^{60}Co γ-rays irradiation tests conducted on a certain consumer central processing unit (CPU) are shown in Fig. 2.18. The horizontal axis represents the total irradiation amount (equivalent to the irradiation time), and the vertical axis denotes the current

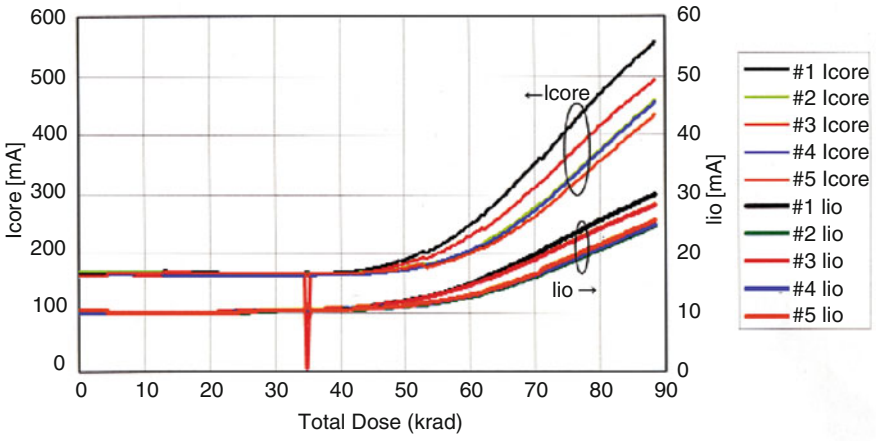

Fig. 2.18 Example of radiation irradiation test on CPU (Akioka 2009)

flowing through the CPU. The amount of current increases as the total irradiation amount increases (Akioka 2009). The same effect is expected to be caused in the space environment by the same radiation dose.

Specific examples of the DDD effect are the solar proton events on August 12, September 29, and October 19, 1989 (maximum values of 9200 PFU, 4500 PFU, and 40,000 PFU, respectively) that caused rapid reductions in the output voltage of the solar cells of Himawari-3, a meteorological satellite in a GEO (Research and Study Group on the Way of Space Environment Monitoring Using Radio Waves 1993).

2.3.5 *Others (Material Degradation and Property Changes in Space Environment)*

Thus far, we have introduced phenomena that affect the semiconductor devices and electronic circuits of satellites. Here, we discuss the material degradation caused by the atomic oxygen and optical property changes of glass materials under ultraviolet (UV) and other radiations.

Degradation of Materials by Atomic Oxygen: In the atmosphere of the Earth, oxygen accounts for approximately 21% by volume, and it is the second major component after nitrogen. In the troposphere and stratosphere, oxygen exists as molecules, whereas at higher altitudes, it dissociates into atoms (atomic oxygen) under the vacuum UV radiation of the Sun. Figure 2.19 shows the number density altitude profiles of the major atmospheric components (O, N_2, O_2, He, H, and N) calculated by the MSIS-E-90 atmospheric model (Hedin 1991). Atomic oxygen is dominant at altitudes between 200 and 500 km in the solar minimum period and between 200 and 700 km in the solar maximum period. If the orbital velocity of the satellite is V and the number density of atomic oxygen at the altitude of the satellite is

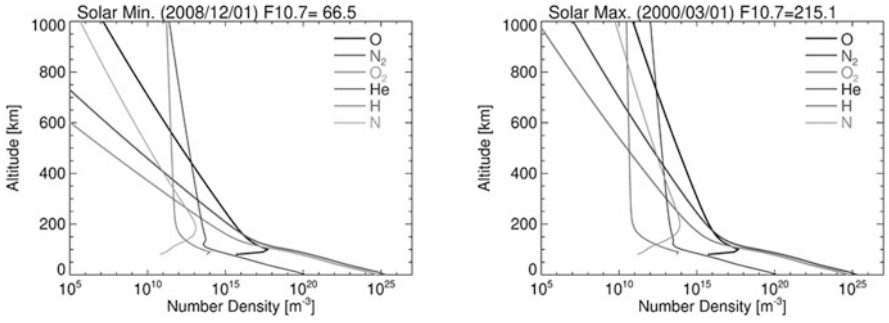

Fig. 2.19 Altitude distributions of number densities of major atmospheric components (O, N_2, O_2, He, H, and N) using MSIS-E-90 atmospheric model (Left, solar activity minimum; right, solar activity maximum)

n, the atomic oxygen collides with the satellite surface material at a flux (irradiation dose rate) of nV.

Atomic oxygen oxidizes the material it collides with owing to its high chemical activity. If the oxidized material is gaseous or has peeling properties, the material surface is eroded. If the oxidized material is solid or has stable properties, an oxide layer is formed on it. Both factors degrade the thermo-optical (solar absorption and infrared emissivity) and mechanical properties of the surface material. Therefore, when operating a satellite in an LEO, it is necessary to take measures such as improving the stability against atomic oxygen by coating the surface material with inorganic materials (Kimoto et al. 2009).

Case Study of Material Degradation by Atomic Oxygen: The problem of degradation of surface materials by atomic oxygen emerged during the early missions of the Space Shuttle Program of NASA (~1981). An investigation of the surface materials used in the payload bay of the space shuttle orbiter, which was exposed to space between space shuttle mission plans STS-1 and STS-3, showed degradation of the paints and polymers, particularly Kapton, used in the thermal blanket of the camera. An analysis of the causes showed that the degradation could be quantitatively explained by considering it as material degradation due to a flux of atomic oxygen at 200–300 km, the flight altitude of the space shuttle (Leger 1983).

In April 1984, NASA launched the Long Duration Exposure Facility (LDEF) in an LEO to investigate the effects of cosmic radiation, meteoric material, cosmic dust, and atomic oxygen on the mechanical, electrical, and optical properties of various materials and space systems exposed to space for long periods. The LDEF was originally scheduled to be recovered in March 1985 but was delayed by the Challenger accident in January 1986 and was recovered in January 1990. This experiment revealed the details of the effects of atomic oxygen on the materials used in spacecrafts (Levine 1991). Fig. 2.20 shows an example of the degradation of a polyimide film covering the edge of a solar array on the International Space Station (Banks et al. 2002).

Changes in Optical Properties of Glass Materials due to UV and Other Radiations: It is known that irradiation of glass materials with UV light, X-rays, γ-rays, and high-energy particles causes changes in optical properties such as transmission characteristics and the refractive index (Yokota 1957). Because the main component of an optical glass is SiO_2, the incident of high-energy electromagnetic waves or particles and the total dose effect generate electron–hole pairs in the glass, which changes its optical properties. This is called browning or solarization of the glass material.

Various additives are added to a glass material to adjust its optical properties. Glass materials containing lead oxide (lead glass) have a high transparency and a high refractive index; however, their optical properties are easily changed by radiation. In contrast, cerium oxide can be added to make a glass radiation-resistant (Sawai and Tashiro 1957).

Thus, it is necessary to use a glass material for the optical equipment used in the space environment that can maintain changes in the optical properties within acceptable ranges for the cumulative dose of space radiation during the planned mission

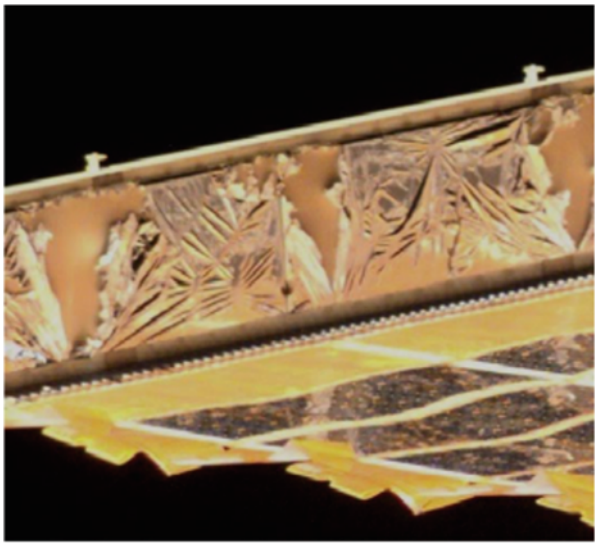

Fig. 2.20 Degradation of polyimide film covering edge of solar array on International Space Station (Banks et al. 2002)

period. Moreover, the properties must be validated in advance by conducting irradiation tests on the ground (Doyle and Czichy 1994; Nishino et al. 1998).

Examples of Changes in Optical Properties of Glass Materials due to UV and Other Radiations: Browning and solarization of glass in space have been investigated by the LDEF mission of NASA, as mentioned above. Figure 2.21 shows the changes of various glass materials after their irradiation tests. Although the five glass samples in the figure are irradiated with the same amount of radiation, their colorations significantly differ depending on the type of glass (Akioka 2009).

2.4 Exposure to Cosmic Radiation

2.4.1 Introduction

Cosmic radiation[1] is everywhere in the space, and the Earth is exposed to it. Therefore, spacecrafts, satellites, and astronauts in space are always exposed to it. Furthermore, cosmic radiation reaches the atmosphere of the Earth and increases the radiation dose at flight altitudes by producing nuclear reactions. Therefore, cosmic radiation exposure is not only a problem for astronauts but also an important issue from the viewpoint of radiation protection of aircrew, called aircrew exposure.

This chapter focuses particularly on the exposure of aircrew to cosmic radiation. First, the regulation for aircrew exposure is briefly explained. Subsequently, the

[1]The term "cosmic radiation" is frequently used when discussing exposure due to cosmic rays, although the term "cosmic rays" is commonly used in scientific papers.

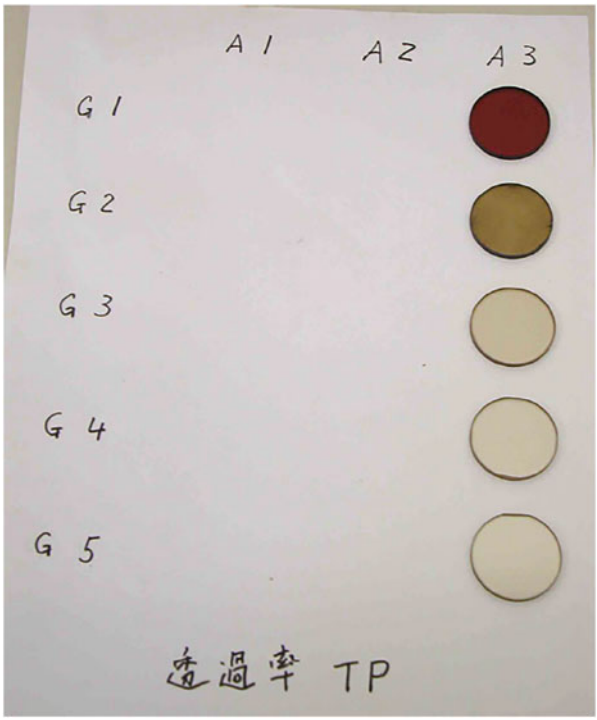

Fig. 2.21 Changes in transmittance after irradiation tests of different types of glass materials (Akioka 2009)

mechanism of aircrew exposure is described and the models developed to evaluate their radiation dose are outlined. Finally, the aircrew dose assessment models developed in Japan and their sample results are presented.

2.4.2 Regulation for Aircrew Exposure

In 1990, the International Commission on Radiological Protection (ICRP) issued a recommendation that "the operation of jet aircraft should be included as one of the occupational exposures to natural radiation" (ICRP 1991). Based on this recommendation, in 1996, the European Radiation Dosimetry Group published a guidance of 6 mSv per year as the constraint value of the radiation dose of aircrews (European Commission 1996). In Japan, a guideline on aircrew exposure was issued in April 2006, and the Ministry of Education, Culture, Sports, Science and Technology; Ministry of Health, Labor, and Welfare; and Ministry of Land, Infrastructure, Transport, and Tourism jointly notified the scheduled aviation associations to adopt measures in accordance with the guideline (Yasuda 2009). Following this guideline, domestic airlines are required to maintain the radiation doses of their

aircrews below 5 mSv per year. More detailed information of the regulation for aircrew exposure is given in ICRP (2016).

In 2007, the cosmic radiation doses of flight crews and cabin crews working for domestic airlines were assessed (Yasuda 2009; Yasuda et al. 2011). Based the results, the average annual doses were 1.68 mSv and 2.15 mSv for flight crews and cabin crews, respectively, and their corresponding maximum annual doses were 3.79 mSv and 4.24 mSv, which were lower than the constraint of the aircrew dose. However, 2007 was a period approaching the solar activity minimum, and no large-scale solar particle event (SPE) that could cause an increase in the radiation dose at flight altitudes occurred that year. If such an event had occurred during this period, the radiation dose of the aircrews could have easily approached or even exceeded the constraint value of 5 mSv.

In 2011, the International Civil Aviation Organization (ICAO), a specialized agency of the United Nations, began to consider the use of space weather information as essential information for aircraft operations. After several years of discussion, Amendment 78 of Annex 3 to the Convention on International Civil Aviation, "Meteorological Services for International Air Navigation," was introduced in 2018, and the use of space weather information was included (ICAO 2018a, b). The ICAO has recognized the importance of space weather information for aircraft operation owing to the following three points. The first is to avoid shortwave communication problems between aircrafts and ground control, the second is to prevent the increase in satellite positioning errors related to electronic navigation, and the third is to reduce the exposure of aircrew to cosmic radiation. In response to this, on November 7, 2019, an agency appointed by the ICAO following an inspection by the World Meteorological Organization (referred to as the ICAO Space Weather Center) began to distribute the space weather information specifically for commercial aircraft operations. When a large SPE occurs, the ICAO Space Weather Center is responsible for providing an advisory called *moderate* (30 μSv/h or higher at 25,000 ft–46,000 ft) or *severe* (80 μSv/h or higher at 25,000 ft–60,000 ft) as a part of the space weather information (as of December 2021). Therefore, a dose assessment model that can assess the dose rate due to cosmic radiation worldwide in real time is necessary, and the development of such a model is in progress in several countries globally.

2.4.3 Mechanism of Aircrew Exposure

The cosmic radiation from outer space can be classified into galactic cosmic rays (GCRs), which come continuously from outside the solar system, and SEPs, which are emitted abruptly from the Sun when solar flares and CMEs occur. The main components of both GCRs and SEPs are hydrogen nuclei (i.e., protons), and these are called primary cosmic rays.

GCRs have a very broad energy spectrum with a power-law distribution ranging from the megaelectron volt region to approximately 100 EeV (10^{20} eV). This energy

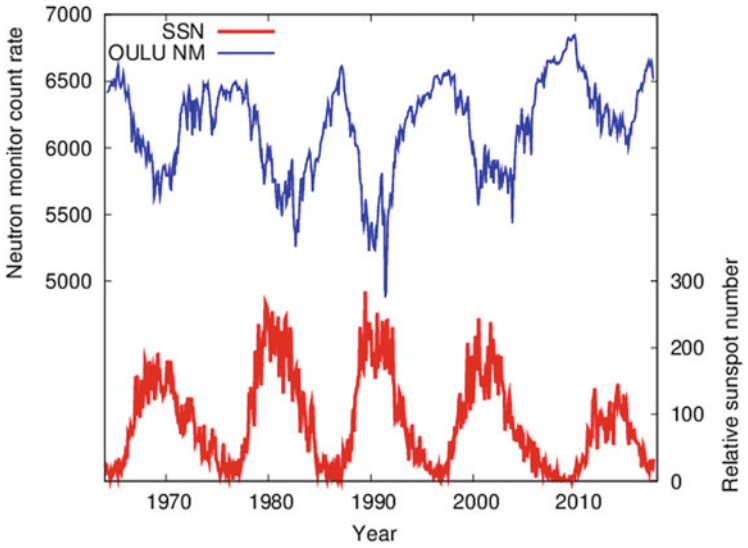

Fig. 2.22 Count rates of Oulu neutron monitor versus relative sunspot number

spectrum has an approximate power-law index of 2.7 up to approximately $10^{15.5}$ eV. GCRs with these energies are considered to be accelerated by supernova remnant shock waves in the galaxy and propagate into the solar system. GCRs with higher energies have a power-law index of approximately 3.0 up to approximately 10^{19} eV, and their acceleration sources are considered to be extragalactic. However, because the energies of cosmic rays contributing to aircrew exposure range from approximately 100 MeV to a few hundred gigaelectron volts, we focus our discussion on the GCRs in this energy range.

GCR fluxes in this energy range around the Earth change in accordance with the solar activity cycle with a period of approximately 11 years. They decrease during the solar maximum and increase during the solar minimum. Specifically, a GCR flux is inversely correlated with the solar activity (Fig. 2.22). However, a GCR flux does not change much on the short timescale of a few days. Therefore, the radiation dose from a GCR also does not change on a short timescale, and is a steady exposure that cannot be avoided.

In contrast, an SEP flux increases abruptly with large solar flares and CMEs, followed by a decrease on the timescale of a few hours. It can reach approximately 10,000 times higher than a GCR flux in extreme cases. It is mainly accelerated by solar flares, coronal shock waves, and interplanetary shock waves (see Chap. 9 for details of the acceleration mechanism). Particularly, SEPs with energies of more than 100 MeV, which can be the source of aircrew exposure, are mainly accelerated by solar flares and coronal shock waves (i.e., near the Sun) and propagate in the interplanetary space to the vicinity of the Earth. The energy spectra of these SEPs observed near the Earth present approximately a power-law distribution. However, the power-law exponent is not typically the same as that for a GCR, significantly

differing for different events. Soft spectra (large exponent) have relatively fewer high-energy components than hard spectra (small exponent). This difference is important for the dose assessment of aircrew exposure to cosmic radiation. In addition, a SEP spectrum changes significantly with time. Immediately after the onset of an SPE, only high-energy particles arrive at the Earth from the Sun; therefore, the energy spectrum does not show a noticeable power-law distribution during this time; instead it presents little flux in the low-energy side. With time, low-energy particles also arrive at the Earth, and the energy spectrum gradually approaches a power-law distribution. Furthermore, SEPs reach the Earth with pitch angle scattering due to the disturbance of the interplanetary magnetic field as it propagates from the Sun to the Earth. Therefore, immediately after the start of an SPE, the particles that are hardly scattered during the propagation arrive at the Earth. Moreover, the majority of the particles are oriented away from the Sun along the magnetic field lines, showing a very large anisotropy of the pitch angle distribution. With time, the pitch angle distribution of the particles gradually becomes isotropic. Thus, because SEP fluxes, energy spectra, and angular distributions vary significantly from event to event and on short timescales, the estimation of the dose rate due to SEPs is difficult compared with that due to GCRs, which can be considered constant on short timescales.

Primary cosmic rays such as GCRs and SEPs propagate in the interplanetary space and eventually reach the Earth, where they propagate in the magnetosphere and reach the upper atmosphere. Because primary cosmic rays are charged particles, their propagation orbit is affected by the magnetic fields when they propagate through the magnetosphere. Therefore, their fluxes differ depending on the latitude and longitude, when they reach the upper atmosphere. For example, their fluxes are generally larger at higher latitudes such as the North and South Poles, whereas they are smaller at lower latitudes near the equator. Thus, even if the flux distribution of primary cosmic rays is generally uniform outside the magnetosphere, latitudinal and longitudinal dependence of the flux distribution occurs in the upper atmosphere.

When primary cosmic rays that pass through the magnetosphere and reach the upper atmosphere enter the atmosphere, they cause a cascade of nuclear reactions with various nuclei composing the atmosphere, called an air shower. It generates numerous secondary cosmic rays such as neutrons, muons, and photons (Fig. 2.23). These secondary cosmic rays predominantly increase the radiation dose rate at flight altitudes. An increase in the neutrons produced by an air shower due to the arrival of SEPs can be observed by ground-based neutron monitors, which is called ground-level enhancement (GLE) and frequently used as an indicator of the increased radiation dose at flight altitudes. Some systems for automatically detecting GLE from neutron monitor data and for issuing alerts have been developed (Kuwabara et al. 2006; Souvatzoglou et al. 2014).

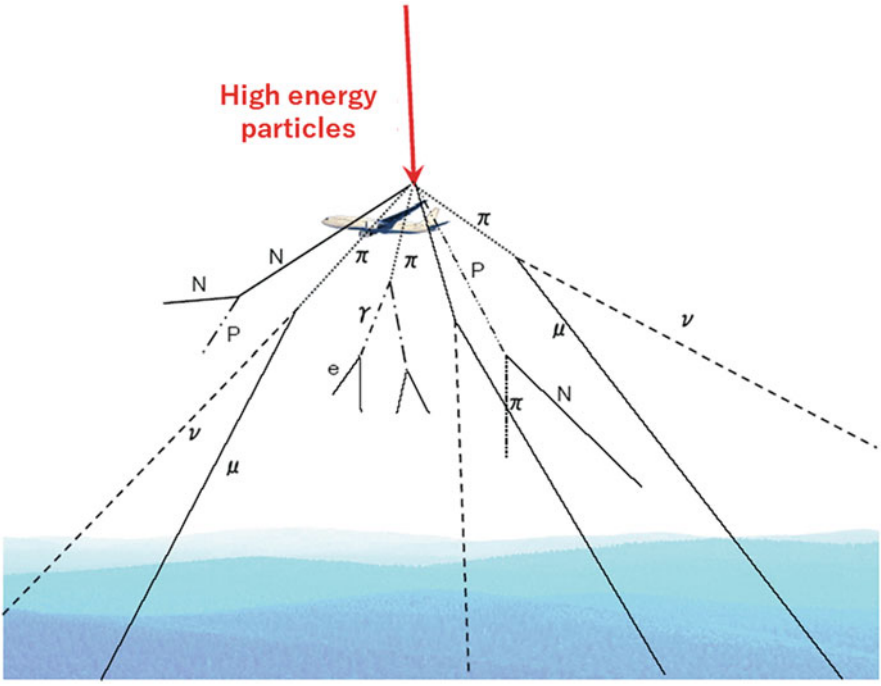

Fig. 2.23 Schematic of air shower reaction in atmosphere (courtesy of Dr. Hiroshi Yasuda, Hiroshima University)

2.4.4 Aircrew Dose Assessment Models

Aircrew dose is rarely measured because it is impractical to mount a radiation monitor on each airplane. Therefore, the radiation dose of an aircrew should be evaluated by numerical calculations, and models for aircrew dose estimation from GCRs and SEPs have been developed in many countries. For example, in the USA, NAIRAS (Mertens et al. 2013) is being developed by NASA, and a CARI-7-based model (Copeland 2018) is being developed by the US Federal Aviation Administration. In addition, in Europe, AVIDOS (Latocha et al. 2016) and SiGLE (Lantos 2006) are being built by Seibersdorf Laboratories, Austria, and the Paris Observatory, France, respectively.

In Japan, a program called JISCARD has been established at the National Institute of Radiological Sciences to calculate the radiation doses from GCRs for arbitrary flight routes and is used for the dose estimation of aircrews by Japanese airline companies (Yasuda et al. 2011). JISCARD is based on the Particle and Heavy Ion Transport code System (PHITS)-based analytical radiation model in the atmosphere (PARMA), which can calculate the fluxes of primary and secondary cosmic rays in the atmosphere and estimate the radiation dose rate at any point in the

atmosphere (Sato 2015). The PARMA is based on the results of air shower simulations conducted using the PHITS (Sato et al. 2018) and used not only for dose assessment but also for the estimation of semiconductor soft error rates and cosmogenic nuclide yields. In addition, to issue an alert for aircrew exposure due to SEPs, a system called the warning system for aviation exposure to SEPs (WASAVIES) has been developed by collaboration of several organizations in Japan such as the Japan Atomic Energy Agency, National Institute of Information and Communications Technology, and National Institute of Polar Research. The WASAVIES is a system that calculates not only the GCR dose rate using the PARMA but also abrupt increases in the SEP dose rate in real time immediately after a GLE occurrence.

Except for the WASAVIES, the aforementioned numerical models estimate aircrew doses by extrapolating observation data from either ground-based neutron monitors or the GOES. In contrast, the WASAVIES estimates the radiation dose rate at any location on the Earth from the surface to an altitude of 100 km by interpolating between the observed data of ground-based neutron monitors and the GOES. In the WASAVIES system, three numerical simulations are used for the interpolation. Brief descriptions of all three simulations are provided below.

1. SEP fluxes in the near-Earth region are calculated by solving the focused transport equation, which describes the propagation of an SEP emitted from the Sun to the Earth through the interplanetary space (Kubo et al. 2015).
2. SEP fluxes at the top of the atmosphere of the Earth are calculated by tracing the motions of the SEPs in the magnetosphere of the Earth (Kataoka et al. 2014).
3. SEP fluxes including their secondary particles in the atmosphere are calculated by reproducing the air showers induced by SEPs using PHITS.

There are four free parameters in these simulations, and their numerical values are determined in real time by observing ground-based neutron monitors and the GOES. The radiation doses are estimated from the secondary cosmic ray fluxes calculated by the third simulation by following the ICRP 2007 recommendations (ICRP 2007).

2.4.5 Examples of Aircrew Dose Rates Estimated by WASAVIES

Figure 2.24 shows global dose rate maps at the conventional flight altitude (altitude of 12 km = 39,370 ft) estimated by the WASAVIES during the GLE associated with the X7.1 flare on January 20, 2005 (GLE69). It presents the map at 6:55 universal time coordinated (UTC) when the highest dose rate is observed. It is noticeable that the radiation dose rates in the polar regions are very high. The maximum radiation dose rate shown in the map is approximately 460 μSv/h over Antarctica, which is approximately 120 times higher than the corresponding GCR dose rate, which is approximately 4 μSv/h.

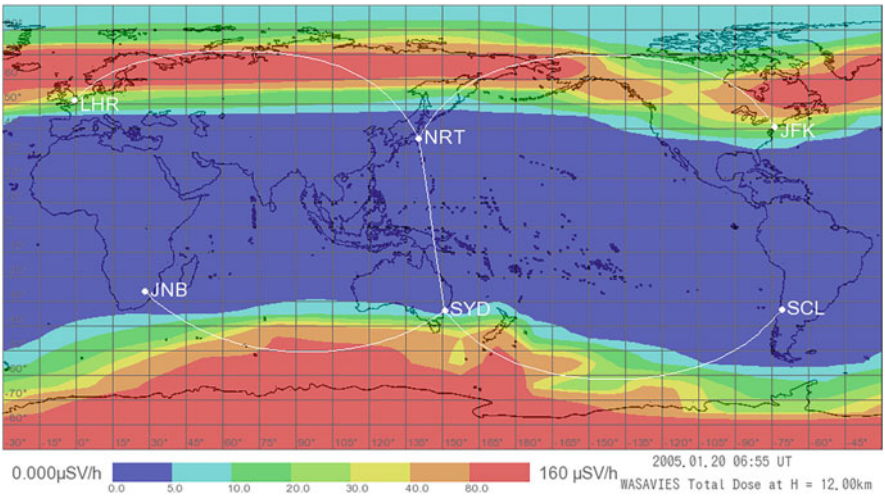

Fig. 2.24 Radiation dose rate map at altitude of 12 km at peak of GLE69 estimated using WASAVIES

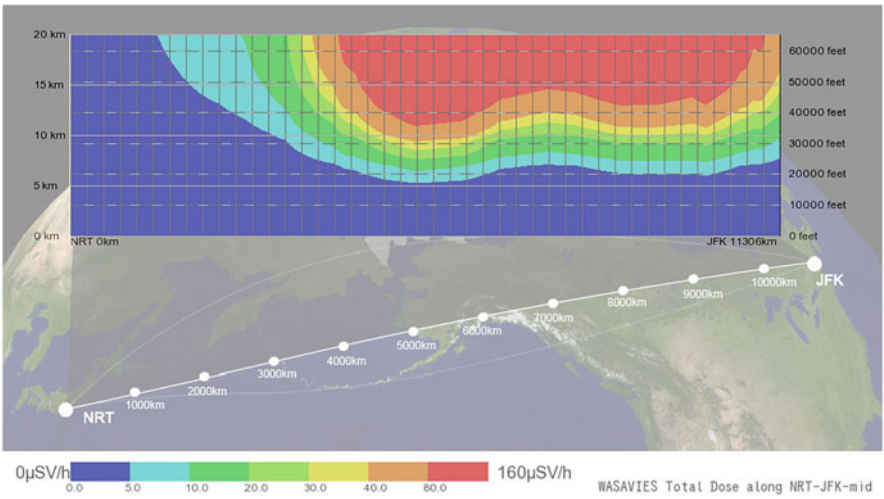

Fig. 2.25 Radiation dose rate along flight route between Tokyo (NRT) and New York (JFK) at peak of GLE69 estimated using WASAVIES

The five white curves in the radiation dose rate map represent typical flight routes, and the WASAVIES can also calculate the radiation dose rates along these routes. Figure 2.25 shows, as an example, the dose rates along a flight route between Tokyo (NRT) and New York (JFK) at 6:55 UTC on January 20, 2005. It can be seen that if the aircraft had been flying at 40,000 ft. (= 12.2 km) above Alaska at that time, the crew on the flight would have received radiation dose rate of over 80 μSv/h.

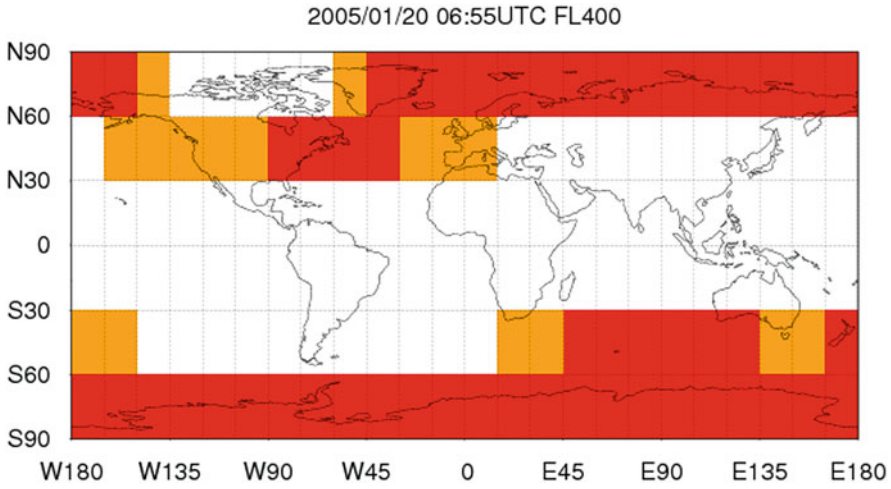

Fig. 2.26 ICAO radiation advisory map developed using WASAVIES at 40,000 ft. altitude at peak of GLE69. ICAO radiation advisory is evaluated on a 30° latitude × 15° longitude grid

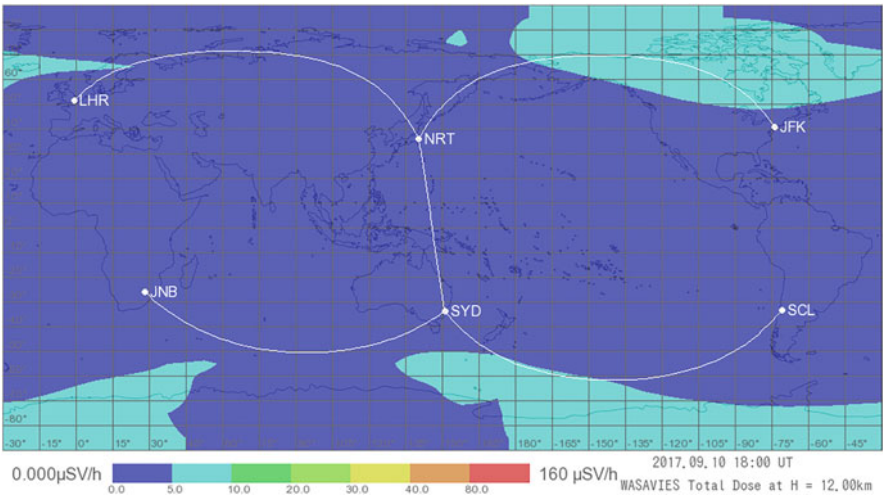

Fig. 2.27 Radiation dose rate map at altitude of 12 km at peak of GLE72 developed using WASAVIES

Figure 2.26 shows the ICAO radiation advisory map at 6:55 UTC and 40,000 ft. altitude estimated by the WASAVIES, although the WASAVIES had not yet been developed. The areas shown in orange and red are *moderate* and *severe*, respectively, and the SEREVE area extends to both polar regions.

Figure 2.27 shows the dose rate map at an altitude of 12 km at the peak of a GLE associated with the X8.2 flare (GLE72), which occurred on September 10, 2017 (Kataoka et al. 2018). Although there is a slight increase in the radiation dose rates at

both polar regions, this increase is much smaller than that in case of GLE69. However, the scales of the flares in both events (X7-8) are similar, and the locations of the flares are approximately 60–90° longitude in the western hemisphere of the solar surface. This is because the energy spectra of the SEPs generated during GLE69 are very hard in comparison to those during GLE72, which is considered to be closely related to the acceleration and transport mechanism of the SEPs; however, an answer has been clearly obtained yet.

It should be mentioned that the increase in the dose rate for people living on the ground is very small even when a large increase is observed in the SEP dose rate at flight altitudes. For example, the SEP dose rate on the ground during the peak of GLE69 is estimated to be approximately the same or at most four times higher than the corresponding GCR dose rate. Considering that public is always exposed to GCRs, its total radiation dose on ground due to SEP exposure is trivial.

2.4.6 Summary

Cosmic radiation dosimetry is a complex research topic involving various fields including solar physics, space weather, nuclear physics, and radiological protection. The WASAVIES is a successful example of such interdisciplinary research. In principle, the WASAVIES can predict the radiation dose rate from immediately after the detection of a GLE to several hours ahead. However, the prediction accuracy of the radiation dose rate is not yet sufficiently high to support actual operation, and improving it is an issue. To achieve this, further studies are necessary for elucidating the acceleration and transport processes of SEPs.

2.5 Atmospheric Drag

2.5.1 Introduction

The outermost part of the atmosphere of the Earth is called the thermosphere (which varies from 90 km in altitude to 400–800 km depending on the solar activity). It is a region in which various fluctuations are driven by waves transmitted from the lower atmospheric layer and electromagnetic waves and energetic particles from the Sun and the geomagnetosphere (phenomena occurring in the thermosphere are explained in Chap. 14). Since Sputnik 1 was launched into this region in 1957, satellites and space stations have been sent into space to support the maintenance and development of our social activities such as monitoring of weather and environmental changes, disaster monitoring, communication, positioning, and scientific observations. The number of spacecrafts sent into space has been increasing. Particularly in recent years, the number of satellites is expected to increase further owing to satellite constellation projects planned with the lowering of the cost of equipment and launch.

Simultaneously, the number of space debris, such as unwanted satellites, parts of rocket bodies, and fragments, is increasing (NASA, 2020a). As of February 2020, 20,000 spacecrafts and debris with a size of 10 cm or larger existed in the orbit of the Earth. Most of the satellites and debris in LEOs pass through the thermosphere, where they are subject to the atmospheric drag, which distorts their orbits and attitudes. Although there are several external forces that affect their orbits, the atmospheric drag is the main external force in the thermosphere. Models of the thermosphere are used to estimate the atmospheric drag in orbit prediction (empirical or physical models; see below). These models are explained subsequently in this section. However, they cannot accurately predict various variations in the thermosphere, and the atmospheric drag is the main uncertainty in orbit prediction. Therefore, the influence of the atmospheric drag cannot be avoided in near-Earth space utilization.

In this section, we first present the orbits of satellites and debris (Sect. 2.5.2) and the effects of external forces on these orbits (Sect. 2.5.3) as the basic information. Subsequently, we explain the factors on which the atmospheric drag depends and how it affects orbits (Sect. 2.5.4). Following this, we describe the relationship between the thermospheric variations and the atmospheric drag (Sect. 2.5.5) and show how the atmospheric drag specifically affects space utilization (Sect. 2.5.6). Finally, we introduce a model for predicting thermospheric variability (Sect. 2.5.7).

2.5.2 *Basis 1: Orbits of Objects Orbiting the Earth*

Before describing how satellites and space debris are affected by the background thermospheric atmosphere, we briefly explain how the orbits of objects orbiting the Earth are determined and the orbital elements that define the positions of such objects. For details, please refer to textbooks on orbital mechanics (King-Hele 1987, etc.).

An object orbiting the Earth is primarily under the influence of the universal gravitational force and is in Keplerian motion. Its equation of motion is as follows:

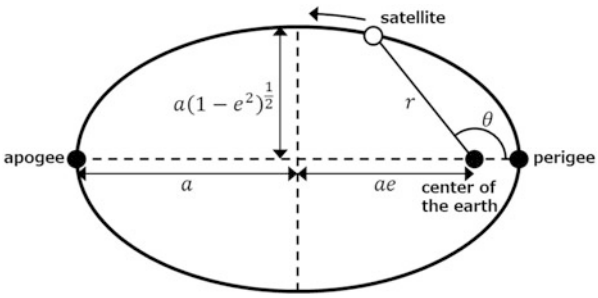

Fig. 2.28 Orbit of object (satellite) orbiting Earth

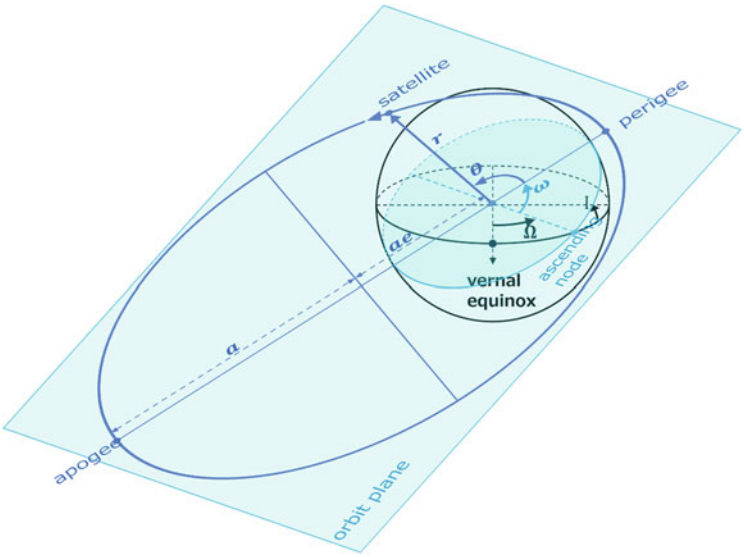

Fig. 2.29 Six orbital elements that define orbit of object (satellite) orbiting Earth

$$\ddot{r} - r\dot{\theta}^2 = -\frac{\mu}{r^2} \tag{2.1}$$

$$\frac{1}{r}\frac{d}{dt}\left(r^2\dot{\theta}\right) = 0 \tag{2.2}$$

Above, as shown in Fig. 2.28, the coordinate system is a two-dimensional polar coordinate system centered on the Earth in the orbital plane of the object (r and θ), t is time, and μ is a constant ($\mu = GM = 3.986004 \times 10^5 \mathrm{km}^3\mathrm{s}^{-2}$, where G is the universal gravitational constant and M is the mass of the Earth). Equations (2.1) and (2.2) yield the r and θ directions, respectively, and their solutions are the following equations representing the orbits of objects (the derivation process is omitted):

$$\ddot{r}r = \frac{p}{1 + e\cos\theta} \tag{2.3}$$

$$r^2\dot{\theta} = (\mu p)^{\frac{1}{2}} \tag{2.4}$$

Above, $p = a(1 - e^2)$, where a and e are arbitrary constants that appear in the process of derivation and correspond to the long radius and eccentricity of the elliptical orbit (see Fig. 2.28). In addition, the velocity of the object, v, can be obtained from the following equation:

$$v^2 = \dot{r}^2 + \left(r\dot{\theta}\right)^2 = \frac{\mu}{p}\left(1 + e^2 + 2e\cos\theta\right) \tag{2.5}$$

There are six basic parameters that determine the shape and position of the orbit of an object orbiting the Earth, which are called orbital elements (Fig. 2.29). These are semimajor axis a, eccentricity e, true anomaly θ (which indicates the position of the orbit with respect to the perigee), right ascension of the ascending node Ω, inclination angle of the orbit I, and argument of perigee ω (which is defined as the angle between the ascending intersection direction and the perigee).

2.5.3 Basis 2: Change in Trajectory Due to External Force

In the absence of forces except the universal gravitational force, the orbital elements listed in the previous section except θ are not time-varying. However, in reality, there are higher-order gravitational components originating from the nonuniformity of the mass distribution of the Earth, gravitational forces from other celestial bodies such as the Sun and the Moon, solar radiation pressure, and atmospheric drag when passing through the thermosphere. When an orbit is subjected to these external forces, the orbital elements except θ also change. The equation of motion and the change in the angular momentum of an object under external forces are as follows:

$$\ddot{\boldsymbol{r}} + \frac{\mu}{r^3}\boldsymbol{r} = \boldsymbol{f} \tag{2.6}$$

$$\frac{d}{dt}\left\{(\mu p)^{\frac{1}{2}}\boldsymbol{n}\right\} = \boldsymbol{r} \times \boldsymbol{f} \tag{2.7}$$

Let $\boldsymbol{n}$ be the unit vector in the direction normal to the orbital plane and $\boldsymbol{f}$ denote the external force vector acting per unit mass. When Eqs. (2.6) and (2.7) are expanded and transformed to each directional component, a system of equations showing the time variation in the orbital elements is derived as follows:

$$\dot{a} = \frac{2a^2 v}{\mu} f_T \tag{2.8}$$

$$\dot{e} = \frac{1}{v}\left\{2f_T(e + \cos\theta) - f_N\frac{r}{a}\sin\theta\right\} \tag{2.9}$$

$$\dot{\Omega}\sin i = (\mu p)^{-\frac{1}{2}} r f_n \sin(\omega + \theta) \tag{2.10}$$

$$\frac{di}{dt} = (\mu p)^{-\frac{1}{2}} r f_n \cos(\omega + \theta) \tag{2.11}$$

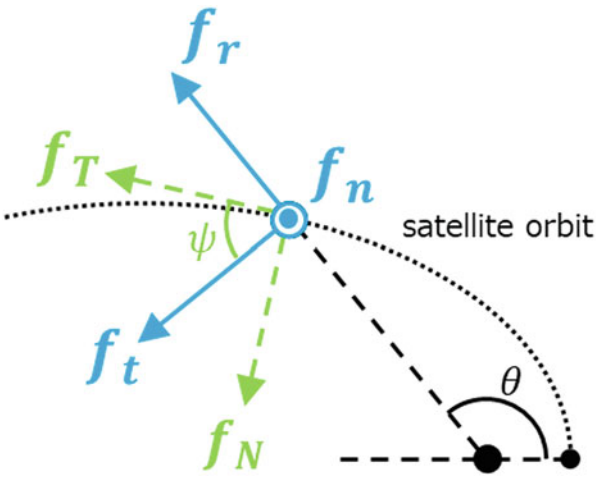

Fig. 2.30 Components of external forces acting on object orbiting Earth in each direction

$$\dot{\omega} + \dot{\Omega}\cos i = \frac{1}{e}\left(\frac{p}{\mu}\right)^{\frac{1}{2}}\left\{ -f_r \cos\theta + f_\theta\left(1 + \frac{r}{p}\sin\theta\right)\right\} \tag{2.12}$$

where f_r and f_θ are r and θ components of the external forces on the orbital plane, respectively, and f_n is the normal direction component of the external force (Fig. 2.30). In Eqs. (2.8) and (2.9), the directional component along the trajectory of the object (velocity direction) f_T and its perpendicular inward direction component f_N are used, instead of f_r and f_θ. This is highly convenient because the atmospheric drag acts mainly in the direction along the trajectory, as described below.

2.5.4 *Factors That Determine Magnitude of Atmospheric Drag*

Objects in LEOs (altitudes below 2000 km) are subject to the atmospheric drag because they collide with atoms and molecules in the upper atmosphere of the Earth. The closeness of an object to the Earth implies high density of the atmosphere and high frequency of collisions, which make the atmospheric drag an important external force. In addition, the uncertainty of the atmospheric drag leads to difficulty in dealing with it in space utilization. In this section, we describe the factors that determine the magnitude of the atmospheric drag and its effect on the change in the trajectory of an object.

The force per unit mass due to the atmospheric drag can be expressed as follows:

$$\boldsymbol{f}_{\mathbf{drag}} = -\frac{1}{2}B\rho|V_r|\boldsymbol{V_r} \tag{2.13}$$

where ρ is the mass density of the background atmosphere at that point, $\boldsymbol{V_r}$ is the relative velocity vector between the object and the atmosphere, and B ($=C_dS/m$) is the inverse ballistic coefficient. C_d is the dimensionless atmospheric drag coefficient, S is the cross-sectional area of the plane perpendicular to the trajectory of the object, and m is the mass of the object. B is frequently referred simply as the ballistic coefficient. The details of these parameters are discussed by King-Hele (1987) and others, and only a brief description is provided here. The drag coefficient, C_d, in the expression of B depends on how the momentum is exchanged between the object and the surrounding atmosphere. At altitudes above 150 km, the exchange is between the object and individual atmospheric particles. The mean free path of the atmosphere is on the order of meters to kilometers, which is longer than the size of the object, and the atmosphere can be regarded as a collection of particles, instead of as a fluid. In this scenario, C_d depends on the composition of the atoms and molecules attached to the object and the surrounding atmosphere, their average velocities, and the shape and rotation of the object. If the object is spherical or cylindrical and the main components of the atmosphere are oxygen atoms (from approximately 200 km altitude to 550–750 km altitude), then C_d~2.2, whereas at higher altitudes when the main atmospheric components are hydrogen atoms, $C_d > 4$ (Afonso et al. 1985). For the cross section of an object, S, the average approximation is considered, because most objects move with a changing orientation of the plane perpendicular to the orbit. If the shape of the object is known to be spherical or cylindrical, the approximation can be estimated relatively simply. An example complex shape is the CHAMP satellite (Liu et al. 2005).

For actual applications such as orbit calculation, it is difficult to obtain accurate values of C_d and S; thus, estimated values of B are used in the orbit prediction using the orbit propagation model, simplified general perturbations model 4 (Hoots and Roehrich 1988). However, the estimates obtained include the errors in the fitting process, and accurate B and atmospheric density estimation need correction (Bowman 2002; Picone et al. 2005).

The relative velocity between an object and the background atmosphere in Eq. (2.13), V_r, can be approximated as follows:

$$|V_r|\boldsymbol{V_r} \cong \left(1 - \frac{rw}{v}\cos i\right)^2 |v|\boldsymbol{v} \tag{2.14}$$

$\boldsymbol{v}$ is the velocity vector of the object and w is the angular velocity of the rotation of the Earth. Here, it is assumed that the object is moving sufficiently rapidly compared to the atmosphere and that the atmosphere is corotating with the Earth. In many cases, the error in the approximation using Eq. (2.14) is estimated to be within 3% (Picone et al. 2005). However, at high latitudes, during magnetic storms, the speed of the thermospheric atmosphere exceeds the speed of the rotation of the Earth (~400 m/s) (see Chap. 14), Eq. (2.14) is inapplicable, and the force perpendicular

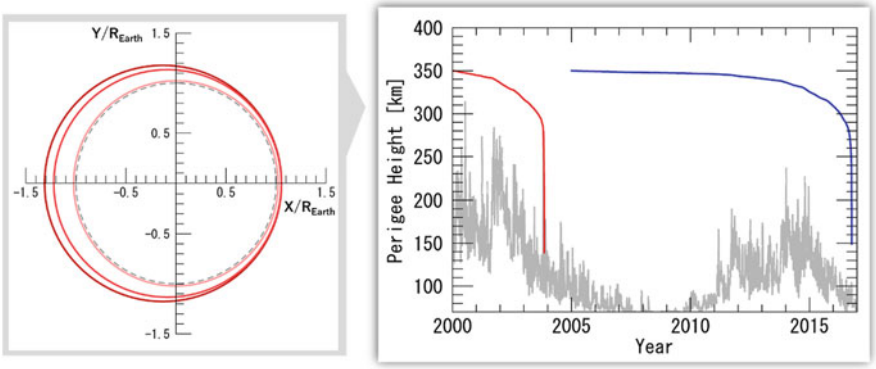

Fig. 2.31 Example of calculated orbit transition of satellite due to atmospheric drag. (Left) Change in orbit in orbital plane. Three solid red lines indicate, in order from outer line, initial orbit, intermediate orbit (1/2 of lifetime), and orbit immediately before reentry into atmosphere. Inner dotted line denotes position of surface of the Earth. Two solid lines (red and blue) show results of calculations in which ballistic coefficient and initial orbit of the satellite are the same but time of injection into orbit is shifted (high and low solar activity periods). Gray solid line shows variation in F10.7 index, which is an index of solar activity

to the orbital plane must be considered (King-Hele 1987, Chap. 8). Using Eq. (2.14), Eq. (2.13) becomes

$$\boldsymbol{f}_{\mathbf{drag}} \cong -\frac{1}{2}BF\rho|v|\boldsymbol{v} \tag{2.15}$$

Above, $F\ (=(1-r_W/_v\cos i)^2)$ is a parameter related to the relative velocity with the atmosphere. Eq. (2.15) shows that the atmospheric drag acts mainly in the direction of the velocity of the object (f_T component) and changes the orbit length radius and eccentricity through Eqs. (2.8) and (2.9).

Here, we present an example of orbit change due to the atmospheric drag. Figure 2.31 shows the calculated change in the orbit of an object by considering the atmospheric drag as the only force other than the universal gravitational force acting on the object. In this example, a hypothetical satellite (B is set as 0.05 m^2kg^{-1} and the orbital inclination is set as 60°) is studied from the initial position at a perigee altitude of 350 km and an apogee altitude of 2000 km until it reenters the atmosphere (reaching an altitude of less than 150 km). An empirical model ("empirical model" is described subsequently in Sect. 2.5.7), NRLMSISE-00 (Picone et al. 2002), was used for the density distribution of the thermosphere. Only variations in the atmosphere due to the season and the solar UV intensity were considered, and parameters corresponding to the effects of magnetic storms were turned off. Three orbits are shown in Fig. 2.31 (left): the initial orbit, orbit in the first half of the lifetime leading up to atmospheric reentry, and orbit immediately before atmospheric reentry. It can be seen that the orbital radius of the satellite shrinks under the atmospheric drag and the shape of the orbit changes from an ellipse to a circle (the eccentricity becomes

smaller). Figure 2.31 (right) also shows the time series of the perigee altitude. As the orbital radius shrinks (the altitude decreases) under the atmospheric drag, the atmospheric drag increases as the background atmospheric density increases, and the degree to which the orbital radius shrinks (the altitude decreases) increases. The two solid lines in the figure show that the results of each calculation shift only at the start of the calculation (the ballistic coefficient and initial orbit of the satellite are the same). The red and blue lines represent periods of relatively high solar activity (January 1, 2000) and low solar activity (January 1, 2005), respectively. A high solar extreme UV intensity implies a dense atmosphere in the thermosphere and a large effect of the atmospheric drag. Therefore, the satellite reenters the atmosphere 3.9 years after the start of the calculation in the former case and 11.8 years in the latter case, causing a large difference in the satellite lifetime.

Table 2.8 Typical temporal and spatial variations observed in thermosphere and their induced displacements of satellite orbit after 1 day (adapted from He et al. (2020); only ※ part is from Le et al.). In second row, amplitude represents range of atmospheric density variation at 400 km altitude. Third and fourth rows list satellite orbits calculated assuming circular orbit at altitude of 400 km and using $B = 0.0023\ m^2kg^{-1}$ ($Cd = 2.3$, $S = 0.5\ m^2$, $m = 500$ kg). Orbital displacement values are calculated 90 times by changing ascending intersection equinox and orbital inclination angles in ranges of 0–360° and 0–90°, respectively, and the average values and standard deviations are shown

Thermospheric variations	Amplitude of atmosphere density variation %	Displacement of orbit(m/day)	
		Solar maximum (F10.7 = 200)	Solar minimum (F10.7 = 70)
Annual variation	5–15	300 ± 20	25 ± 5
Semiannual variation	20	800 ± 40	70 ± 15
Daily variations (period: 24 h)	10–40	50 ± 30	10 ± 10
(period: 12 h)	10	100 ± 100	10 ± 10
(period: 8 h)	<3	1 ± 1	0.5 ± 0.5
27-day variation due to solar rotation	<100	–	–
Disturbance due to solar flare	<30※	–	–
Disturbance due to geomagnetic storm	100–200	–	–
Spatial variations (in latitude)	10–20	50 ± 40	15 ± 10
(in longitude)	<5	1 ± 1	0.5 ± 0.5
MDM	20–30	150 ± 30	15 ± 6
EMA	1–6	50 ± 30	5 ± 3

2.5.5 *Variation in Thermospheric Atmospheric Density and its Contribution to Atmospheric Drag*

As expressed by Eq. (2.15), the atmospheric density and atmospheric drag in the thermosphere are directly linked. Various temporal and spatial variations occur in the thermosphere, which change the atmospheric drag and the orbits of the satellites and the debris. The details of the physical processes that occur in the thermosphere are described in Chap. 14. In this section, we describe the different types of atmospheric density variations and their contributions to the atmospheric drag.

In Table 2.8, which is based on the articles by He et al. (2020) and Le et al. (2012), the first column lists the typical types of variabilities found in the thermosphere. These are the (quasi-)regular temporal variabilities (annual and daily variabilities), variabilities derived from the solar activity (changes in the active region of the solar surface and its movement due to rotation, increased radiation intensity due to solar flares, and magnetic storms due to CMEs and other sources), spatial variabilities, and other characteristic spatiotemporal variabilities. In the table, MDM (i.e., midnight density maximum) refers to the phenomenon of density increase at the equator at midnight (e.g., Ruan et al. 2014). Moreover, EMA (i.e., equatorial mass density anomaly) refers to the density distribution increase at mid- and low latitudes compared to that at the equator during the day (Liu et al. 2005). The table does not include variations longer than 1 year (e.g., variations on the scale of several decades due to global warming and variations on the scale of approximately 11 years due to the solar activity). The second column provides the approximate ranges of the atmospheric density variations listed in the first column (at an altitude of 400 km). Columns 3 and 4 present the orbital displacements of satellites in circular orbits (400 km altitude) after 1 day due to the variations in column 1. The difference in columns 3 and 4 is due to the solar activity level, indicating that a high solar activity level implies a dense atmosphere (see Chap. 14) and a large effect of the atmospheric drag. Regarding the difference due to the thermosphere variabilities, the values of the orbital displacements due to the annual variation in the thermosphere (1-year and half-year cycles), MDM, and 12-hour cycle are large except for the variations due to the solar surface activity and solar flares. In contrast, the contributions of the 24-hour cycle and latitudinal spatial variations to the orbital displacement values are small. This is because the orbital displacement cancels out as the satellite passes through the peaks and valleys of the atmospheric tides and latitudinal distribution in the thermosphere within the local time range and latitudinal range of the satellite in one orbit. For example, a polar orbit passes through two local time ranges that are 12 hours apart. One half of it is in phase with 12-hour-period atmospheric tides, whereas the other half is out of phase with 24-hour-period atmospheric tides; therefore, the changes cancel out. In the case of long-term thermospheric variations such as annual variations, the range of atmospheric variations per day is small, whereas the orbital displacement after 1 day is large because of the global increase or decrease.

Berger et al. (2020) used the empirical thermospheric model, NRLMSISE-00, to estimate the orbit displacements of existing satellites at altitudes of 200–650 km during magnetic storms. Their estimates showed that a magnetic storm of Kp = 6-equivalent could result in orbital displacements of the orders of a few kilometers and 12 km if it lasts for 3 h and 18 h, respectively.

Although not included in Table 2.8, changes in the thermosphere caused by the propagation of phenomena in the lower atmosphere (troposphere and stratosphere) to the upper atmosphere can also contribute to the atmospheric drag. A phenomenon called stratospheric sudden warming occurs from zero to several times per winter, mainly in the Northern Hemisphere (from November to March). In each occurrence, the atmospheric density in the thermosphere changes in amplitude and phase with daily variations for several days (particularly, the amplitude of the 12-hour cycle significantly increases (<tens of percent)). Moreover, it has been reported that the global mean atmospheric density decreases by 3–7% depending on the altitude (Liu et al. 2013; Yamazaki et al. 2015).

2.5.6 *Influence of Atmospheric Drag on Space Utilization*

The previous section explained that the variation in the atmospheric density in the thermosphere changes the orbits of satellites and space debris through the atmospheric drag. In this section, we briefly describe the effects of the atmospheric drag on space applications.

2.5.6.1 Lifetime of Satellite

As described in Sect. 2.5.4, the atmospheric drag reduces the orbital radius of a satellite. Figure 2.31 also shows that the atmospheric density around the satellite changes depending on the altitude and the solar activity. This, in turn, changes the effect of the atmospheric drag, i.e., the degree of reduction in the orbital radius, and affects the timing of the atmospheric reentry of the satellite. Table 2.9 shows the

Table 2.9 Periods until reentry into atmosphere for satellites in circular orbits below 500 km altitude. Orbital inclination is set as 60°, and solar activity is calculated by fixing F10.7 = 70 or 200. Bracketed numbers are longer than timescale of solar activity variation and are not highly relevant unless variation in F10.7 is considered

	Solar maximum(F10.7 = 200)		Solar minimum (F10.7 = 70)	
Initial height (km)	B = 0.05	B = 0.01	B = 0.05	B = 0.01
500	220 days	2.9 years	(8.8 years)	(44 years)
400	43 days	210 days	330 days	4.5 years
300	6.4 days	31 days	23 days	110 days
200	0.41 days	2.0 days	0.67 days	3.2 days

estimated time for a satellite in a circular orbit ($e = 0$) of 500 km altitude or less to reenter the atmosphere (reaching an altitude of 150 km or less). NRLMSISE-00 (Picone et al. 2002) was used to model the atmospheric density of the background thermosphere. As can be seen from the table, a low or a high solar activity implies a high atmospheric density, and a large ballistic coefficient B implies a strong atmospheric drag and a short lifetime of the satellite. This table does not include the thermospheric variations caused by solar surface phenomena as described in Sect. 2.5.5. However, if we assume that a geomagnetic storm with Kp = 8 occurs for 1 day, the lifetime of the satellite at initial altitudes of 200, 300, and 400 km will be shortened from 0.41 days to 0.29 days, from 6.4 days to 5.4 days, and from 43 days to 42 days, respectively, for the case of F10.7=200 and B=0.05.

2.5.6.2 Satellite Tracking

In a satellite operation organization, the managed satellites are tracked by ground antenna and orbit prediction calculation. When a satellite enters a visible range from a nonvisible range in which the antenna cannot track, antenna acquisition becomes impossible if the actual position of the satellite significantly deviates from the prediction by the orbit calculation. The potential of the atmospheric drag to cause such orbit prediction errors is discussed in the science recommendations for space weather hazards (Ishii et al. 2021). For simplicity, if the aperture diameter of a ground antenna is 10 m and the frequency is the X band (8 GHz), the beamwidth is 0.2625°, which corresponds to a range of 31 km at an altitude of 400 km. This value is sufficiently longer than the orbital displacement due to the atmospheric drag, as listed in Table 8; however, it is insufficient to interfere with satellite tracking within the range of thermospheric variations during quiet conditions. The above hazard map examines the case of a thermospheric disturbance assuming a magnetic storm, and the estimate concludes that the effect will be insufficient to interfere with satellite tracking above an altitude of 300 km. The US Air Force regularly observes satellites and space debris larger than 10 cm in size using the Space Surveillance Network (SSN) and continues to update a catalog that includes the orbital elements of each object. According to the testimony of an operator, the SSN temporarily lost sight of most of the LEO satellites during the large magnetic storms in October–November 2003, and it took several days for the operation to restore the catalog (Berger et al. 2020).

2.5.6.3 Prediction of Collisions with Satellites and Space Debris

As described in Sect. 2.5.1, the number of LEO satellites and debris in the near-Earth space environment continues to increase, and it is important to use space considering their collision risk. Therefore, space agencies in various countries have been monitoring satellites and debris and analyzing their orbits as a part of their Space Situational Awareness (SSA) activities. In the USA, the 18th Space Control

Squadron (18 SPCS) is responsible for SSA data sharing and sends conjunction data messages (CDMs) to satellite operating agencies. The Aeolus satellite and one of SpaceX's constellation satellites are at high risk of collision, and the European Space Agency, which received a CDM from 18 SPCS, is increasing the altitude of the Aeolus satellite to avoid the risk (https://www.esa.int/Safety_Security/ESA_spacecraft_dodges_large_constellation).

Estimates of the atmospheric drag affect the assessment of collision risk. The collision risk assessment method of NASA (Hejduk and Snow 2018) and the collision risk assessment model of the University of Michigan (Bussy-Virat et al. 2018) take the following steps. First, whether the satellite under evaluation will come close to any of the numerous objects in space is investigated. Subsequently, the collision risk for the cases determined to be "close" in the first step is analyzed. Specifically, the collision probability, P_c, is calculated as the probability when the distance between two objects is within a collision distance (several meters) based on the trajectories of the two objects and their uncertainty. Satellite operation agencies use the value of P_c in evaluating the collision risk. The uncertainty of the orbit prediction due to the uncertainty of the atmospheric drag has a significant impact on P_c. In Sect. 2.5.4, the factors that determine the atmospheric drag were specified as the ballistic coefficient, B, and the atmospheric density. These estimates are derived from tracking data and models up to the time when the collision risk is assessed, and if these variables are small in the forecast period (~10 days), the uncertainty is limited. However, in reality, as mentioned in Sect. 2.5.5, some thermospheric variations have a significant range of variability for a period of ~10 days (e.g., variations associated with the 27-day solar rotation cycle, sudden solar flares, magnetic storms, and lower atmosphere phenomena). The uncertainty increases because a thermospheric model used for the calculation cannot predict these variations with sufficient accuracy. The errors in the thermospheric model itself and the uncertainties in the predicted values of the input parameters, such as the solar radiation intensity index (F10.7) and the geomagnetic disturbance index (Ap), affect the assessment of the collision risk (Hejduk and Snow 2018; Bussy-Virat et al. 2018).

2.5.6.4 Analysis of Atmospheric Entry

We have already discussed the effects of the atmospheric drag on the lifetimes of satellites and the timing of their atmospheric reentry. However, the effects of atmospheric deceleration and heating are even greater after objects reenter the atmosphere (e.g., the US FAA tutorial document, "Returning from Space: Re-entry" n.d.). At altitudes below 150 km, the mean free path of the atmosphere becomes smaller than the size of an object, and the atmosphere is treated as a fluid, instead of as particles. In addition, because the speed of an object exceeds the speed of sound, a shock wave is formed in front of it. With such a change in the atmospheric condition around the object, the coefficient of the atmospheric drag, C_d, also changes (King-Hele 1987) and affects the trajectory and velocity of the object (e.g., Smith et al.

2005). The air inside the shock wave is very hot owing to adiabatic compression, and the object surface also becomes hot (>1000 K), which melts many materials including metals such as iron. Satellite operation organizations are analyzing how many constituents of the objects that reenter the atmosphere will reach the surface of the Earth without burning up and their speeds and extents of impacts, to evaluate their risk to humans living on the ground (e.g., Smith et al. 2005).

2.5.6.5 Satellite Attitude

The effects of the atmospheric drag on the orbit of an object are discussed above; however, its influence on the attitude of an object can also affect satellite operation. If the center of gravity is shifted from the center point where the atmospheric drag acts, torque $\boldsymbol{T}_{\mathbf{drag}} = \boldsymbol{r}_{\mathbf{rel}} \times \boldsymbol{f}_{\mathbf{drag}}$ operates to rotate the satellite. $\boldsymbol{r}_{\mathbf{rel}}$ is the relative distance vector between the center point where the atmospheric drag acts and the center of gravity and $\boldsymbol{f}_{\mathbf{drag}}$ is the atmospheric drag per unit mass. Generally, the attitude of a satellite is maintained by the attitude control function. The X-ray astronomy satellite, ASCA, is an example of a satellite that may have been affected by the torque due to the atmospheric drag. ASCA, which had been in continuous operation after the completion of its originally scheduled mission period, suffered the effects of the X6.1 class solar flare on July 14, 2000. At that time, a large-scale magnetic storm (Dst index <300 nT) occurred near the Earth owing to the arrival of a coronal gas ejection from the Sun. This caused the thermospheric atmosphere to be heated and expanded by energetic particles and electric currents. Although the orbit of the ASCA was within ±30° of latitude at an altitude of 440 km perigee, it is estimated that the density of the surrounding atmosphere rapidly increased several times owing to the atmospheric expansion caused by the magnetic storm. This may have led the torque due to the atmospheric drag to exceed the attitude control capability of the satellite, significantly disturbing its attitude, leading to the suspension of further observations (JAXA's website: http://www.isas.jaxa.jp/j/japan_s_history/06/02/06.shtml).

2.5.7 Conclusion: Prediction Models for Thermospheric Variability

In this section, we explained how the atmospheric drag in the thermosphere changes the orbits of LEO satellites and debris and how it specifically affects space applications. As introduced in Sect. 2.5.6, a model that describes the distribution of and the variations in the atmospheric drag is necessary to predict the orbits of satellites and debris, estimate the timing of atmospheric reentry, evaluate the collision risk, and evaluate the risk of atmospheric reentry. There are two types of such models of the thermospheric atmosphere. The first type is empirical models, which describe the

average behavior of the atmosphere, which is statistically estimated based on actual observation data with variables such as the location, season, solar activity, and geomagnetic disturbance index. The second type is physical models, which numerically solve the equations representing the physical laws of the atmosphere.

Some general characteristics and examples of empirical models of the thermosphere are listed below.

- **Reproduces regular and average variations with good accuracy.**
  Because an empirical model is based on observation data, it can reproduce regular and average variations in data-rich regions with relatively good accuracy. For example, in Table 8, variations associated with the solar activity cycles, annual variations, and (monthly average) daily variations are summarized. The accuracy is low for regions with few observation data (e.g., the lower thermosphere below a 200 km altitude).
- **Variations associated with geomagnetic disturbances reproduce the average pattern.**
  As the scale of a magnetic storm increases, the errors in the model increase owing to the lack of observation data and the differences in behavior of the thermosphere in different events.
- **Temporal and spatial resolutions are low.**
  The shortest period in the standard empirical model is 8 h. In the latitudinal direction, up to sixth-order components of the spherical harmonic function are included, which corresponds to a resolution of approximately 20°. The minimum wavelength of the variation in the longitude direction is 120°.
- **Computational cost is low.**
- **Example models:**
  NRLMSISE-00 (Picone et al. 2002), JB2008 (Bowman et al. 2008), and DTM2013 (Bruinsma 2015).

The general features and examples of physical models of the thermosphere are listed below.

- **Capable of reproducing not only regular but also irregular variations.**
  As with empirical models, regular variations in the thermosphere can be reproduced; however, the accuracy of the reproduction is generally inferior. A physical model that includes the troposphere can reproduce daily variations in the thermosphere and sudden changes due to a stratospheric temperature rise.
- **Fluctuations associated with geomagnetic disturbances can reproduce event-specific behaviors.**
- **High temporal and spatial resolutions.**
  As a standard, temporal variations can be reproduced on the order of minutes, and spatial resolution is less than 5° in the latitude and longitude directions. For example, detailed temporal and spatial developments can be reproduced for the variability associated with solar flare emissions and the effects of magnetic storms, as can be seen from Table 8. In addition, the MDM and the EMA can be reproduced only by physical models in terms of the resolution.
- **Computational cost is high.**

- **Examples of models:**
  TIE-GCM (Richmond et al. 1992) and CTIPe (Fuller-Rowell et al. 2002) for the thermosphere and ionosphere. WAM (Akmaev et al. 2008), WACCM-X (Liu et al. 2010), and GAIA (Miyoshi and Fujiwara 2008; Jin et al. 2011) include the troposphere and the thermosphere (and the ionosphere).

Empirical models are used in actual space applications, as described in Sect. 2.5.6, because they have better reproducibility for regular and average variations and lower computational cost than physical models. The accuracies of physical models have been improved recently by the introduction of data assimilation. Both types of models depend on the future solar radiation and geomagnetic disturbances in predicting thermospheric variability. For example, the estimation of the collision risk, as mentioned in Sect. 2.5.6, requires prediction up to 10 days in the future; however, presently, the prediction error of thermospheric models is large and contributes to the uncertainty of satellite and debris orbits. Therefore, it is desirable to improve the prediction accuracies of thermospheric models including the prediction of the solar radiation and geomagnetic disturbances.

References

Afonso, G., Barlier, F., Berger, C., Mignard, F., Walch, J.J.: Reassessment of the charge and neutral drag of LAGEOS and its geophysical implications. J. Geophys. Res. **90**(B11), 9381–9398 (1985). https://doi.org/10.1029/JB090iB11p09381

Akioka, M.: Space radiation effect on satellite. J. Natl. Inst. Infor. Commun. Technol. **56**(1-4), 49–55 (2009)

Akmaev, R.A., Fuller-Rowell, T.J., Wu, F., Forbes, J.M., Zhang, X., Anghel, A.F., Iredell, M.D., Moorthi, S., Juang, H.-M.: Tidal variability in the lower thermosphere: comparison of whole atmosphere model (WAM) simulations with observations from TIMED. Geophys. Res. Lett. **35**, L03810 (2008). https://doi.org/10.1029/2007GL032584

Amariutei, O.A., Ganushkina, N.Y.: On the prediction of the auroral westward electrojet index. Ann. Geophys. **30**(5), 841–847 (2012)

Banks, B., Lenczewski, M., Demko, R.: Durability issues for the protection of materials from atomic oxygen attack in low earth orbit, 53rd IAC (2002)

Beniguel, Y., Adam, J.-P.: Effects of scintillations in GNSS operation, Chapter 3.5. In: Lilensten, J. (ed.) Space Weather - Research towards Applications in Europe. Springer, Dordrecht (2007)

Berger, M. J., Coursey, J. S., Zucker, M. A., Chang, J.: Stopping-Power & Range Tables for Electrons, Protons, and Helium Ions, NIST Standard Reference Database 124, (1998) https://doi.org/10.18434/T4NC7P

Berger, T.E., Holzinger, M.J., Sutton, E.K., Thayer, J.P.: Flying through uncertainty. Space Weather. **18**, e2019SW002373 (2020). https://doi.org/10.1029/2019SW002373

Bolduc, L.: GIC observations and studies in the Hydro-Québec power system. J. Atmos. Sol. Terr. Phys. **64**(16), 1793–1802 (2002). https://doi.org/10.1016/S1364-6826(02)00128-1

Boteler, D.: Methodology for simulation of geomagnetically induced currents in power systems. J. Space Weather Space Clim. **4** (2014). https://doi.org/10.1051/swsc/2014018

Boteler, D.H., Pirjola, R., Blais, C., Foss, A.: Development of a GIC simulator. 2014 IEEE PES General Meeting | Conference & Exposition (2014). https://doi.org/10.1109/PESGM.2014.6939778

Bowman, B. R.: True Satellite Ballistic Coefficient Determination for HASDM, AIAA/AAS Astrodynamics Specialist Conference and Exhibit, (2002) https://doi.org/10.2514/6.2002-4887

Bowman, B. R., Tobiska, W. K., Marcos, F. A., Huang, C. Y., Lin, C. S., Burke, W. J.: A new empirical thermospheric density model JB2008 using new solar and geomagnetic indices. In AIAA/AAS Astrodynamics Specialist Conference and Exhibit, pp. 6438. (2008), https://doi.org/10.2514/6.2008-6438

Bruinsma, S.: The DTM-2013 thermosphere model. J. Space Weather Space Clim. **5**, A1 (2015). https://doi.org/10.1051/swsc/2017008

Burke, W.J., Huang, C.Y., Gentile, L.C., Bauer, L.: Seasonal-longitudinal variability of equatorial plasma bubbles. Ann. Geophys. **22**, 3089–3098 (2004)

Burton, R.K., McPherron, R.L., Russell, C.T.: An empirical relationship between interplanetary conditions and Dst. J. Geophys. Res. **80**(31), 4204–4214 (1975). https://doi.org/10.1029/JA080i031p04204

Bussy-Virat, C.D., Ridley, A.J., Getchius, J.W.: Effects of uncertainties in the atmospheric density on the probability of collision between space objects. Space Weather. **16**, 519–537 (2018). https://doi.org/10.1029/2017SW001705

Cagniard, L.: Basic theory of the magneto-telluric method of geophysical prospecting. Geophysics. **18**(3), 605–635 (1953). https://doi.org/10.1190/1.1437915

Cannon, P.S.: Extreme Space Weather: Impacts on Engineered Systems - A Summary. Royal Academy of Engineering, London (2013)

Choi, H.-S., Lee, J., Cho, K.-S., Kwak, Y.-S., Cho, I.-H., Park, Y.-D., Kim, Y.-H., Baker, D.N., Reeves, G.D., Lee, D.-K.: Analysis of GEO spacecraft anomalies: Space weather relationships. Space Weather. **9**, S06001 (2011). https://doi.org/10.1029/2010SW000597

Clauer, C.R., et al.: Solar wind control of auroral zone geomagnetic activity. Geophys. Res. Lett. **8** (8), 915–918 (1981)

Copeland, K.: Federal Aviation Administration Technical report, DOT/FAA/AM-18/6 (2018)

Doyle, D.B., Czichy, R.H.: Influences of simulated space radiation on optical glasses. SPIE. **2210**, 434–448 (1994)

Ebihara, Y., Tanaka, T., Kamiyoshikawa, N.: New diagnosis for energy flow from solar wind to ionosphere during substorm: global MHD simulation. J. Geophys. Res. Space Physics. **124**(1), 360–378 (2019). https://doi.org/10.1029/2018JA026177

Ebihara, Y., Watari, S., Kumar, S.: Prediction of geomagnetically induced currents (GICs) flowing in Japanese power grid for Carrington-class magnetic storms. Earth Planets Space. **73**(1) (2021). https://doi.org/10.1186/s40623-021-01493-2

Erinmez, I.A., Kappenman, J.G., Radasky, W.A.: Management of the geomagnetically induced current risks on the national grid company's electric power transmission system. J. Atmos. Sol. Terr. Phys. **64**, 743–756 (2002). https://doi.org/10.1016/S1364-6826(02)00036-6

European Commission, Radiation Protection, No.85, 1996-01, EC, Luxembourg (1996)

Ferguson, D.C., Worden, S.P., Hastings, D.E.: The space weather threat to situational awareness, communications, and positioning systems. IEEE Trans. Plasma Sci. **43**(9), 3086–3098 (2015). https://doi.org/10.1109/TPS.2015.2412775

Fujita, S., Fujii, I., Endo, A., Tominaga, H.: Numerical modeling of spatial profiles of geomagnetically induced electric field intensity in and around Japan. Tech. Rep. Kakioka Magn. Observ. **14**(2), 35–50 (2018)

Fuller-Rowell, T.J., Millward, G.H., Richmond, A.D., Codrescu, M.V.: Storm-time changes in the upper atmosphere at low latitudes. J. Atmos. Sol.-Terr. Phys. **64**, 1383–1391 (2002)

Gleisner, H., Lundstedt, H.: A neural network-based local model for prediction of geomagnetic disturbances. J. Geophys. Res. Space Physics. **106**(A5), 8425–8433 (2001)

Goto, T.-N.: Numerical studies of geomagnetically induced electric field on seafloor and near coastal zones incorporated with heterogeneous conductivity distributions. Earth Planets Space. **67**(1) (2015). https://doi.org/10.1186/s40623-015-0356-2

Harada, M., Izutsu, J., Uetake, T., Terayama, T., Tominaga, K., Nagao, T.: Monitoring of the variation of underground resistivity structure by measuring the neutral current of 500 kV electric power lines. Bull. Inst. Oceanic Res. Dev. Tokai Univ. **30**, 45–57 (2009)

He, C., Yang, Y., Carter, B., Zhang, K., Hu, A., Li, W., et al.: Impact of thermospheric mass density on the orbit prediction of LEO satellites. Space Weather. **18**, e2019SW002336 (2020). https://doi.org/10.1029/2019SW002336

Hedin, A.E.: Extension of the MSIS thermospheric model into the middle and lower atmosphere. J. Geophys. Res. **96**, 1159 (1991)

Hejduk, M.D., Snow, D.E.: The effect of neutral density estimation errors on satellite conjunction serious event rates. Space Weather. **16**, 849–869 (2018). https://doi.org/10.1029/2017SW001720

Hoots, F. R., Roehrich R. L.: Models for propagation of NORAD element sets, Aerospace Defense Command Spacetrack Rep. 3, Peterson AFB, CO (1988)

Ingham, M.: Analysis of variations in cathodic protection potential and corrosion risk on the natural gas pipeline at Dannevirke. Report for the Natural Gas Corporation, 94 (1993)

Ingham, M., Rodger, C.J., Divett, T., Dalzell, M., Petersen, T.: Assessment of GIC based on transfer function analysis. Space Weather. **15**(12), 1615–1627 (2017). https://doi.org/10.1002/2017sw001707

International Civil Aviation Organization: Annex 3 to the Convention on International Civil Aviation, Twentieth Edition, 3–5 (2018a)

International Civil Aviation Organization: International standards and recommended practices, Annex 10 to the Convention on the International Civil Aviation, 7th edition, November (2018b)

International Commission on Radiological Protection: ICRP Publication, 60. Ann. ICRP. **21**(1–3) (1991)

International Commission on Radiological Protection: ICRP Publication, 103. Ann. ICRP. **37**(2–4) (2007)

International Commission on Radiological Protection: ICRP Publication, 132. Ann. ICRP. **45**(1) (2016)

Ishii, M., Shiota, D., Tao, C., Ebihara, Y., Fujiwara, H., Ishii, T., Ichimoto, K., Kataoka, R., Koga, K., Kubo, Y., Kusano, K., Miyoshi, Y., Nagatsuma, T., Nakamizo, A., Nakamura, M., Nishioka, M., Saito, S., Sato, T., Tsugawa, T., Yoden, S.: Space weather benchmarks on Japanese society. Earth Planets Space. **73**, 108 (2021)

Ishikura, S., Takeuchi, M., Nakatani, H., Okuda, M., Tamaki, H.: Development of measurement equipment for effects of GIC (Geomagnetically-Induced Currents). Nissin Electr. Rev. **55**(2), 37–42 (2010)

Jiggens, P., Heynderickx, D., Sandberg, I., Truscott, P., Raukunen, O., Vainio, R.: Updated model of the solar energetic proton environment in space. J. Space Weather Space Clim. **8**, 22 (2018)

Jin, H., Miyoshi, Y., Fujiwara, H., Shinagawa, H., Terada, K., Terada, N., Ishii, M., Otsuka, Y., Saito, A.: Vertical connection from the tropospheric activities to the ionospheric longitudinal structure simulated by a new Earth's whole atmosphere–ionosphere coupled model. J. Geophys. Res. **116**, A01316 (2011). https://doi.org/10.1029/2010JA015925

Kappenman, J.G.: Storm sudden commencement events and the associated geomagnetically induced current risks to ground-based systems at low-latitude and midlatitude locations. Space Weather. **1**(3) (2003). https://doi.org/10.1029/2003sw000009

Kappenman, J.: Geomagnetic storms and their impacts on the US power grid. M(805) (2010)

Karasawa, Y., Yasukawa, K., Yamada, M.: Ionospheric scintillation measurements at 1.5 GHz in mid-latitude region. Radio Sci. **29**, 643–651 (1985)

Kataoka, R., et al.: Radiation dose forecast of WASAVIES during ground-level enhancement. Space Weather. **12**, 380–386 (2014)

Kataoka, R., et al.: Radiation dose nowcast for the ground level enhancement on 10–11 September 2017. Space Weather. **16**, 917–923 (2018)

Kelly, M.A., Comberiate, J.M., Miller, E.S., Paxton, L.J.: Progress toward forecasting of space weather effects on UHF SATCOM after Operation Anaconda. Space Weather. **12**, 601–611 (2014). https://doi.org/10.1002/2014SW001081

Kimoto, Y., Miyazaki, E., Ishizawa, J., Shimamura, H.: Atomic oxygen effects on space materials in low earth orbit and its ground evaluation (in Japanese). J. Vac. Soc. Jpn. **52**(9), 475–483 (2009). https://doi.org/10.3131/jvsj2.52.475

King-Hele, D.G.: Satellite orbits in an atmosphere. Blackie and Son Ltd. (1987)

Kubo, Y., et al.: Interplanetary particle transport simulation for warning system for aviation exposure to solar energetic particles. Earth. Planets. Space. **67**, 117–129 (2015)

Kurt, V., Mavromichalaki, H., Gerontidou, M.: Statistical analysis of solar proton events. Ann. Geophys. **22**, 2255–2271 (2004)

Kusano, K., Iju, T., Bamba, Y., Inoue, S.: A physics-based method that can predict imminent large solar flares. Science. **369**(6503), 587–591 (2020). https://doi.org/10.1126/science.aaz2511

Kuwabara, T., et al.: Development of a ground level enhancement alarm system based upon neutron monitors. Space Weather. **4**, S10001 (2006)

Lantos, P.: Radiation doses potentially received on-board aeroplanes during recent solar particle events. Radiat. Prot. Dosim. **118**, 363–374 (2006)

Latocha, M., et al.: Cosmic radiation assessment at ESA's space weather portal with AVIDOS. In: 16th European conference on radiation and its effects on components and systems (RADECS), pp. 1–4, Bremen (2016)

Le, H., Liu, L., Wan, W.: An analysis of thermospheric density response to solar flares during 2001–2006. J. Geophys. Res. **117**, A03307 (2012). https://doi.org/10.1029/2011JA017214

Leach, R. D., Alexander, M. B.: Failures and Anomalies Attributed to Spacecraft Charging, NASA Technical Report NASA-RP-1375, pp. 30, August, (1995)

Leger, L.J.: Oxygen atom reaction with shuttle materials at orbital altitudes. NASA Tech. Memo. **58246** (1983)

Levine, A. S.: LDEF 69 Months in Space. First Post-Retrieval Symposium, NASA-CP-3134 (1991).

Liemohn, M.W., et al.: Model evaluation guidelines for geomagnetic index predictions. Space Weather. **16**(12), 2079–2102 (2018)

Liu, H., Luhr, H., Henize, V., Köhler, W.: Global distribution of the thermospheric total mass density derived from CHAMP. J. Geophys. Res. **110**, A04301 (2005). https://doi.org/10.1029/2004JA010741

Liu, H.-L., Foster, B.T., Hagan, M.E., et al.: Thermosphere extension of the whole atmosphere community climate model. J. Geophys. Res. **115**, A12302 (2010). https://doi.org/10.1029/2010JA015586

Liu, H., Jin, H., Miyoshi, Y., Fujiwara, H., Shinagawa, H.: Upper atmosphere response to stratosphere sudden warming: Local time and height dependence simulated by GAIA model. Geophys. Res. Lett. **40**, 635–640 (2013). https://doi.org/10.1002/grl.50146

Lohmeyer, W.Q., Cahoy, K.: Space weather radiation effects on geostationary satellite solid-state power amplifiers. Space Weather. **11** (2013). https://doi.org/10.1002/swe.20071

Loto'aniu, T.M., Singer, H.J., Rodriguez, J.V., Green, J., Denig, W., Biesecker, D., Angelopoulos, V.: Space weather conditions during the Galaxy 15 spacecraft anomaly. Space Weather. **13**, 484–502 (2015). https://doi.org/10.1002/2015SW0001239

Love, J.J., Swidinsky, A.: Time causal operational estimation of electric fields induced in the Earth's lithosphere during magnetic storms. Geophys. Res. Lett. **41**(7), 2266–2274 (2014). https://doi.org/10.1002/2014GL059568

Love, J.J., Swidinsky, A.: Observatory geoelectric fields induced in a two-layer lithosphere during magnetic storms. Earth Planets Space. **67**(1), 58–58 (2015). https://doi.org/10.1186/s40623-015-0213-3

Luo, B., et al.: Prediction of the AU, AL, and AE indices using solar wind parameters. J. Geophys. Res. Space Physics. **118**(12), 7683–7694 (2013)

Mackie, R.L., Smith, J.T., Madden, T.R.: Three-dimensional electromagnetic modeling using finite difference equations: The magnetotelluric example. Radio Sci. **29**(4), 923–935 (1994). https://doi.org/10.1029/94rs00326

Marshall, R.A., Smith, E.A., Francis, M.J., Waters, C.L., Sciffer, M.D.: A preliminary risk assessment of the Australian region power network to space weather. Space Weather. **9**(10) (2011). https://doi.org/10.1029/2011sw000685

Marshall, R.A., Gorniak, H., Van Der Walt, T., Waters, C.L., Sciffer, M.D., Miller, M., Dalzell, M., Daly, T., Pouferis, G., Hesse, G., Wilkinson, P.: Observations of geomagnetically induced currents in the Australian power network. Space Weather. **11**(1), 6–16 (2013). https://doi.org/10.1029/2012sw000849

Marti, L., Yin, C.: Real-time management of geomagnetic disturbances: hydro one's eXtreme space weather control room tools. IEEE Electrific. Mag. **3**(4), 46–51 (2015). https://doi.org/10.1109/mele.2015.2480637

McPherson, D.A., Schober, W.R.: Spacecraft charging at high altitudes: the SCATHA satellite program. J. Spacecraft Rockets. **12**(10), 621–626 (1975). https://doi.org/10.2514/3.57027

Mertens, C.J., et al.: Space Weather. **11**, 603–635 (2013)

METI: fy 2014 report on specification researches of standard technologies of electrical power equipment, summarized by the institute of applied energy. Ministry of Economy, Trade and Industry of Japan (2015)

Misra, P., Enge, P.: Global Positioning System, Signals, Measurements, and Performance. Ganga-Jamuna Press, Lincoln, MA (2001)

Miyoshi, Y., Fujiwara, H.: Gravity waves in the thermosphere simulated by a general circulation model. J. Geophys. Res. **113**, D01101 (2008). https://doi.org/10.1029/2007JD008874

Nakamura, S., Ebihara, Y., Fujita, S., Goto, T., Yamada, N., Watari, S., Omura, Y.: Time domain simulation of geomagnetically induced current (GIC) flowing in 500-kV power grid in japan including a three-dimensional ground inhomogeneity. Space Weather. **16**(12), 1946–1959 (2018). https://doi.org/10.1029/2018sw002004

Nishino, Y., Suematsu, Y., Tsuneta, S., Ichimoto, K., Kobiki, T., Takeyama, N.: Radiation testing of optical glasses and crystals for Solar-B optical instruments (in Japanese). Rep. Natl. Astron. Observ. Jpn. **3**, 145–150 (1998)

Odstrčil, D., Pizzo, V.J.: Three-dimensional propagation of coronal mass ejections (CMEs) in a structured solar wind flow: 1. CME launched within the streamer belt. J. Geophys. Res. Space Phys. Ther. **104**(A1), 483–492 (1999). https://doi.org/10.1029/1998ja900019

Picone, J.M., Hedin, A.E., Drob, D.P., Aikin, A.C.: NRLMSISE-00 empirical model of the atmosphere: Statistical comparisons and scientific issues. J. Geophys. Res. **107**(A12), 1468 (2002). https://doi.org/10.1029/2002JA009430

Picone, J.M., Emmert, J.T., Lean, J.L.: Thermospheric densities derived from spacecraft orbits: Accurate processing of two-line element sets. J. Geophys. Res. **110**, A03301 (2005). https://doi.org/10.1029/2004JA010585

Pirjola, R., Viljanen, A.: On geomagnetically-induced currents in the finnish 400 kV power system by an auroral electrojet current. IEEE Trans. Power. Deliv. **4**(2), 1239–1245 (1989)

Pulkkinen, A., Lindahl, S., Viljanen, A., Pirjola, R.: Geomagnetic storm of 29-31 October 2003: Geomagnetically induced currents and their relation to problems in the Swedish high-voltage power transmission system. Space Weather. **3**(8) (2005). https://doi.org/10.1029/2004SW000123

Pulkkinen, A., Pirjola, R., Viljanen, A.: Determination of ground conductivity and system parameters for optimal modeling of geomagnetically induced current flow in technological systems. Earth Planets Space. **59**(9), 999–1006 (2007). https://doi.org/10.1186/BF03352040

Pulkkinen, A., Hesse, M., Habib, S., Van der Zel, L., Damsky, B., Policelli, F., Fugate, D., Jacobs, W., Creamer, E.: Solar shield: forecasting and mitigating space weather effects on high-voltage power transmission systems. Nat. Hazards. **53**(2), 333–345 (2009). https://doi.org/10.1007/s11069-009-9432-x

Pulkkinen, A., Bernabeu, E., Eichner, J., Beggan, C., Thomson, A.W.P.: Generation of 100-year geomagnetically induced current scenarios. Space Weather. **10**(4), 1–19 (2012). https://doi.org/10.1029/2011SW000750

Rastätter, L., Kuznetsova, M.M., Glocer, A., Welling, D., Meng, X., Raeder, J., Wiltberger, M., Jordanova, V.K., Yu, Y., Zaharia, S., Weigel, R.S., Sazykin, S., Boynton, R., Wei, H., Eccles, V., Horton, W., Mays, M.L., Gannon, J.: Geospace environment modeling 2008-2009 challenge: Dst index. Space Weather. **11**(4), 187–205 (2013). https://doi.org/10.1002/swe.20036

Research and Study Group on the Way of Space Environment Monitoring Using Radio Waves: The second phase report - Problems of heliospheric environment – (in Japanese), (July 1993).

Returning from Space: Re-Entry: Advanced Aerospace Medicine On-line: Section III-Space Operation, Federal Aviation Administration https://www.faa.gov/about/office_org/headquarters_offices/avs/offices/aam/cami/library/online_libraries/aerospace_medicine/tutorial/section3/spacecraft_design/ (n.d.)

Richmond, A.D., Ridley, E.C., Roble, R.G.: A thermosphere/ionosphere general circulation model with coupled electrodynamics. Geophys. Res. Lett. **19**, 601–604 (1992)

Rodi, W.L.: A technique for improving the accuracy of finite element solutions for magnetotelluric data. Geophys. J. Int. **44**(2), 483–506 (1976). https://doi.org/10.1111/j.1365-246X.1976.tb03669.x

Ruan, H., Lei, J., Dou, X., Wan, W., Liu, Y.C.-M.: Midnight density maximum in the thermosphere from the CHAMP observations. J. Geophys. Res. Space Physics. **119**, 3741–3746 (2014). https://doi.org/10.1002/2013JA019566

Saito, S., Yoshihara, T.: Impact Assessment of Ionospheric Scintillation Associated with Plasma Bubbles on GAST-D Ground Integrity Monitors, Proceeding of Institute of Navigation GNSS+ 2018, 2186–2184, (2018).

Saito, S., Yoshihara, T., Kezuka, A., Saitoh, S., Fukushima, S., Otsuka, Y.: GAST-D Flight Experiment Results with Disturbed and Quiet Ionospheric Conditions, Proceedings of Institute of Navigation GNSS+ 2015, 1494–1499, (2015).

Saito, S., Sunda, S., Lee, J., Pullen, S., Supriadi, S., Yoshihara, T., Terkildsen, M., Lecat, F., ICAO APANPIRG Ionospheric Studies Task Force: Ionospheric delay gradient model for GBAS in the Asia-Pacific region. GPS Solut. **21**, 1937–1947 (2017). https://doi.org/10.1007/s10291-017-0662-1

Sakai, M., Daikuhara, N., Maki, K., Taniguchi, Y., Yanabu, S.: A research of critical temperature on bubble formation under overloaded oil-immersed transformer. IEEJ Trans. Power Energy. **121**(4), 500–506 (2001a). https://doi.org/10.1541/ieejpes1990.121.4_500

Sakai, M., Kobayashi, N., Daikuhara, N.: Research on insulation paper thermal deterioration in oil-immersed transformer over-load operation. IEEJ Trans. Power Energy. **121**(4), 507–513 (2001b). https://doi.org/10.1541/ieejpes1990.121.4_507

Sakai, T., Matsunaga, K., Yoshihara, T., Saito, S.: Air navigation with global navigation satellite systems and the ionospheric effects. J. NICT. **55**, 231–242 (2009)

Sakai, J., Hosokawa, K., Tomizawa, I., Saito, S.: A statistical study of anomalous VHF propagation due to the sporadic-E layer in the air-navigation band. Radio Sci. **54**, 426–439 (2019)

Sato, T.: Analytical model for estimating terrestrial cosmic ray fluxes nearly anytime and anywhere in the world: extension of parma/expacs. PLoS One. **10**(12), e0144679 (2015)

Sato, T., et al.: Features of particle and heavy ion transport code system (PHITS) version 3.02. J. Nucl. Sc. Technol. **55**, 684–690 (2018)

Sauer, H., Wilkinson, D.: Global mapping of ionospheric HF/VHF radio wave absorption due to solar energetic protons. Space Weather. **6** (2008). https://doi.org/10.1029/2008SW000399

Sawai, I., Tashiro, J.: Glass as a radiation shielding material (in Japanese). J. Soc. Mater. Sci. Jpn. **6**(49), 638–644 (1957). https://doi.org/10.2472/jsms1952.6.638

Shiota, D., Kataoka, R.: Magnetohydrodynamic simulation of interplanetary propagation of multiple coronal mass ejections with internal magnetic flux rope (SUSANOO-CME). Space Weather. **14**(2), 56–75 (2016). https://doi.org/10.1002/2015SW001308

Smith, R., Bledsoe, K., Dobarco-otero, J., Rochelle, W., Johnson, N., Pergosky, A., Weiss, M.: Reentry survivability analysis of the Hubble space telescope (HST). 4th Eur. Conf. Space Debris. **4**(1) (2005)

Souvatzoglou, G., et al.: Space Weather. **12**, 633 (2014)

Supriadi, S., Saito, S.: Simulation study of mitigation of plasma bubble effects on GBAS using a VHF radar. Navigation. **66**, 845–855 (2019). https://doi.org/10.1002/navi.330

Tabata, Y., Ogawa, S.: The installation of the geomagnetically-induced currents measurement and actual observation, The Annual Meeting record I.E.E. Japan, 192 (2003).

Taguchi, S., Shibata, H.: World maps of foEs. J. Radio Res. Lab. **8**, 355–389 (1961)

Takasu, N., Miyawaki, F., Saitou, S., Fujiwara, Y.: DC excitation of transformer cores due to geomagnetically-induced currents. IEEJ Trans. Power Energy. **113**(4), 435–444 (1993). https://doi.org/10.1541/ieejpes1990.113.4_435

Takasu, N., Oshi, T., Miyawaki, F., Saito, S., Fujiwara, Y.: An experimental analysis of DC excitation of transformers by geomagnetically induced currents. IEEE Trans. Power Del. **9**(2), 1173–1182 (1994). https://doi.org/10.1109/61.296304

Toyoda, K.: International standardization of the worst plasma environment for spacecraft charging (in Japanese), Society of Japanese Aerospace Companies Aviation and Space, No. 774, 1–3, (2018)

Tsuboi, A., Horiuchi, S.: J. Elect. Soc. **108**(3), 233–236 (1988)

Van der Zel, L.: Review and Development of Improved Techniques for GIC Measurement and Mitigation. In: Technical Update, vol. 1013801. Electric Power Research Institute (2007)

Viljanen, A.: European project to improve models of geomagnetically induced currents. Space Weather. **9**(7) (2011). https://doi.org/10.1029/2011sw000680

Viljanen, A., Pirjola, R.: Geomagnetically induced currents in the Finnish high-voltage power system. Surv. Geophys. **15**(4), 383–408 (1994). https://doi.org/10.1007/BF00665999

Watari, S., Nakamura, S., Ebihara, Y.: Measurement of geomagnetically induced current (GIC) around Tokyo, Japan. Earth Planets Space. **73**(102) (2021). https://doi.org/10.1186/s40623-021-01422-3

Wrenn, G.L., Rodgers, D.J., Ryden, K.A.: A solar cycle of spacecraft anomalies due to internal charging. Ann. Geophys. **20**(7), 953–956 (2002). https://doi.org/10.5194/angeo-20-953-2002

Yamazaki, Y., Kosch, M.J., Emmert, J.T.: Evidence for stratospheric sudden warming effects on the upper thermosphere derived from satellite orbital decay data during 1967–2013. Geophys. Res. Lett. **42**, 6180–6188 (2015). https://doi.org/10.1002/2015GL065395

Yasuda, H.: Isotope News (in Japanese). **663**, 8–12 (2009)

Yasuda, H., et al.: Management of cosmic radiation exposure for aircraft crew in Japan. Radiat. Prot. Dosim. **146**, 123–125 (2011)

Yokota, Y.: Effect of radiation on glass (in Japanese). J. Cerma. Soc. Jpn. **65**(740), C244–C252 (1957). https://doi.org/10.2109/jcersj1950.65.740_C244

Yokoyama, T.: A review on the numerical simulation of equatorial plasma bubbles toward scintillation evaluation and forecasting. Prog Earth Planet Sci. **4**, 37 (2017). https://doi.org/10.1186/s40645-017-0153-6

Yumoto, K.: On the effect of GIC (in Japanese). Kaiyo. Mon. **23**, 264–269 (1991)

Chapter 3
Severe Space Weather Disasters

Mamoru Ishii

3.1 Scale and Occurrence Rate of Space Weather Disasters

This chaper is based on "Assessment of the impact of space weather phenomena on society for scientific suggestions" published under the Project for Solar-Terrestrial Environment Prediction [PSTEP, 2020, Ishii et al. 2021]. The most severe space weather phenomenon in the history is the Carrington event. In the Carrington event, auroral observations were recorded 17 h after the onset of a solar flare. Specifically, a high-speed large-scale coronal mass ejection (CME) reached the Earth in 17 h and generated a severe magnetic storm. The scale of a magnetic storm is generally expressed in terms of the Dst index, which is the intensity of the ring current generated during a magnetic storm, and the peak Dst index during the Carrington event was estimated to be −1760 nT or −850 nT (Tsurutani et al. 2003).

The largest event since the international geological year of 1957–1958, when observations were improved and expanded, was a CME triggered by an X4 class solar flare in March 1989, which produced a severe magnetic storm (Dst index = −589 nT) called the Quebec storm (Cid et al. 2014). Geomagnetically induced currents (GICs) affected the power grid of the Hydro-Québec Power Corporation in Canada, causing a power outage for approximately 10 h, and many cases of satellite failures occurred. Other examples include the Bastille event in July 2000 (Dst = −301 nT due to a CME associated with an X5 class solar flare) and a series of magnetic storms following the Halloween event in October–November 2003. The latter were Dst = −353 nT due to the X17 flare in October 28, Dst = −383 nT due to the X10 flare in October 29, no magnetic storm related with the X28 flare in

M. Ishii (✉)
National Institute of Information and Communications Technology, Tokyo, Japan
e-mail: mishii@nict.go.jp

K. Kusano (ed.), *Solar-Terrestrial Environmental Prediction*,
https://doi.org/10.1007/978-981-19-7765-7_3

November 4, and Dst = −472 nT due to the M-class flare in November 18. A study by the National Aeronautics and Space Administration reported numerous satellite failure cases, in which approximately 59% scientific satellites and spacecrafts were affected during this entire period and approximately 24% missions adopted safety measures such as shutting down equipment (Goka 2006).

A Carrington event-class CME occurred on the opposite side of the Sun from the Earth on July 23, 2012; however, it did not affect the Earth. If this event had occurred after 1 week, it could have hit the Earth, in which case the Dst index of the magnetic storm would have been approximately −1182 nT (Baker et al. 2013).

The measurement of carbon isotope ^{14}C in Yakusugi cedar is one of the methods to determine past space weather phenomena. Specifically, carbon isotope ^{14}C produced by cosmic rays and absorbed into Yakusugi cedar as $^{14}CO_2$ is measured in its annual rings, revealing events of increase in the intensity of cosmic rays. In the annual rings of a Yakusugi cedar corresponding to A.D. 774–775 and 993–994, the ^{14}C concentration increased rapidly. This suggests that there was an increase in the cosmic radiation in these years and that a large-scale solar storm occurred that caused an increase in solar energy particles (SEPs). Table 3.1 shows the frequency and size of space weather phenomena in solar activity and geomagnetic activity. The scale of that solar storm was estimated to be ten to several tens of times larger than the Carrington event (Miyake and Masuda 2014).

Space weather events with scales that can affect social infrastructures such as power grids and satellites have a probability of occurring once in 10–100 years. To examine the social impacts of space weather events, it is necessary to estimate the scales of the disturbances that may occur. However, the relationship between the scale and the frequency of occurrence is different for different phenomena such as solar flares, energetic particles, CMEs, and magnetic storms. Furthermore, as shown in the case of the Halloween event in October–November 2003, the largest events in different categories do not show a one-to-one correspondence, and the causal relationship between them is not self-evident.

The definitions of the frequency of occurrence and scale of each phenomenon in the ionosphere and the polar cap absorption are provided in Tables 3.2 and 3.3, respectively.

A substorm is evaluated by the magnitude of the westward-flowing auroral jet current.

The upper and lower envelopes of the superposed variations in the horizontal elements of the geomagnetic field observed at multiple auroral regions are called AU and AL indices, respectively, which represent the magnitude of a westward auroral jet current. Because the auroral zone moves to lower latitudes with increasing geomagnetic activity, the AL index is apparently underestimated when geomagnetic data observed at a particular magnetic latitude are used. The SuperMAG electrojet index overcomes this problem by determining the magnitude of an auroral jet current using the variations in the horizontal component of the geomagnetic field observed at several locations (Newell and Gjerloev 2011). This index is also called the SMU index for the upper envelope and the SML index for the lower envelope, and the latter, in particular, has become commonly used instead of the AL index. Here, we

Table 3.1 Frequency of space weather phenomena and summary of phenomenon size[a]

Phenomena	Scale parameter	Daily–several times per year	Once per year	Once per 10 years	Once per 100 years	Once per 1000 years
Solar flare	X-ray peak flux (1–8 A)	M1–X1	X7.6	X12	X44	X101
	Energy [erg][b]	$<2.0 \times 10^{30}$	1.0×10^{31}	1.0×10^{32}	1.0×10^{33}	1.0×10^{34}
CME	Velocity [km/s][c]	250–450	2000	3000	3800–4500	4700–6600
	Kinetic energy [erg][d]				4.4×10^{33}	9.8×10^{33}
SEPs	Fluence [10^{10} p/cm^2] (>10 MeV)[e]		~0.1 ※1	~1.0 ※1	~2.4–5.1	~3.8–14
	Ell [%· h]			~190	~6600	~223,000
Solar radio burst [sfu][f]	Very high frequency (VHF) (30–300 MHz)				$<2.8 \times 10^{9}$	
	Ultrahigh frequency (300–3000 MHz)				$<1.2 \times 10^{7}$	
	Global Positioning System (1176–1602 MHz)				$<1.2 \times 10^{7}$	
	F10.7 (2800 MHz)				$<1.3 \times 10^{7}$	
	Microwave (4000–20,000 MHz)				$<3.7 \times 10^{7}$	
Variations in relativistic electron flux in outer radiation belt	Electron fluence (>2 MeV) [10^{10} cm^{-2} sr^{-1}] in GEO	~0.1	~0.6	~3.0	~10[g]	~26[g]
Magnetic storm	Dst index [nT][h]		~−220 ※2	~−450 ± 30	~−650 ± 110	~−930 [−320, −1520] ※3
Substorm	AL index [nT][i]		~−3700	~−4100	~−4100	~−4200
	SML index [nT]	−460	−3000	−4000	−5000	−5900

[a] Note that value of once per 1000 years is an extrapolation of Weibull function
[b] According to [1] with N[year −1] = 10^{31} × E[erg] −1
[c] According to Gopalswamy (2018), frequency with which CMEs of higher than tabulated speeds occur
[d] According to Gopalswamy (2018)
[e] According to Gopalswamy (2018), 1 is value read visually from Fig. 1–27
[f] According to National Science and Technology Council, [2018]
[g] Estimated by extreme value statistical analysis and should be treated with caution
[h] According to Tsubouchi and Omura (2007)
[i] According to Nakamura et al. (2015)

Table 3.2 Frequency of ionospheric disturbance scale occurrence and phenomenon scale summary

Phenomena and parameters		Once per year	Once per 10 years	Once per 100 years	Once per 1000 years
Total electron content (TEC)	Tokyo [TECU][a]	≥90	≥130	≥190	
	Kagoshima [TECU][b]	≥110	≥160	≥230	
	Hokkaido [TECU][b]	≥70	≥105	≥150	
Dellinger phenomenon	Absorption at 6.6 MHz (long circuit) [dB][c]	71	83	93	~100
	Event duration [h][d]	0.6–1.6	1.8–3.6	4.0–6.8	7.4–11.9
Sporadic E layer	Frequency propagating 1000 km [MHz][e]	100	110	>130	
	Frequency propagating 2000 km [MHz][e]	130	>180	>180	
	Duration for foEs[f] > 18 MHz [h]	1–2	~3		
	Duration for foEs >13.3 MHz [h]	~3	~5		
Plasma bubble	Event duration [h]	4–5	6		

[a] The estimation of once per year is based on 22 years of TEC dataset. The estimation of once per 10 years and 100 years are based on 66 years of foF2 dataset
[b] Values of Kagoshima and Hokkaido are calculated as 1.2 and 0.8 times the size of Tokyo, respectively
[c] Based on long-range shortwave line (at effective solar zenith angle of 0° at 6.6 MHz)
[d] Based on ionosonde observations (1–30 MHz, for blackout "B" and the difference between the value of the minimum frequency and the mean minimum frequency for reference days (dfmin) ≥ 2.5 MHz)
[e] Value indicates lower limit of upper frequency limit
[f] foEs, maximum ordinary mode frequency for sporadic E layer

Table 3.3 Frequency and magnitude of polar cap absorption

Phenomena and parameter	Daily–several times per year	Once per year	Once per 10 years	Once per 100 years	Once per 1000 years
Proton flux above 10 MeV [pfu][a]	10^{1-2}	10^3	10^4	4×10^5	10^6
Kp index[b]	4 (7)	7 (8)	8 (9)	9	9
Magnetic latitude (day) at 6.6 MHz 10 dB[c]	65° (60)	57° (55)	53° (52)	51°	50°
Magnetic latitude (night) at 6.6 MHz 10 dB[c]	65° (60)	60° (58)	56° (55)	53°	53°
Minimum duration [h][d]	22	46	71	83	89

[a] Estimated from Kurt et al. (2004) and Jiggens et al. (2018). 1 pfu = $1/cm^2/s/sr$
[b] Estimated value. It does not necessarily correspond to magnetic storms. Numbers in parentheses represent National Oceanic and Atmospheric Administration scale
[c] Latitude at which polar cap absorption expands. According to Sauer and Wilkinson (2008) and Jiggens et al. (2018)
[d] Duration of SEP event in which particle flux above 10 MeV is >10 pfu. According to Sauer and Wilkinson (2008)

use the AL index, which has been extensively used, and the SML index, which has come to be commonly used instead of the AL index, as indices of the scale of a substorm phenomenon.

3.2 Actions of National and International Organizations for Space Weather Disasters

3.2.1 National Activities

Swiss Re, a Swiss insurance company, estimated the economic losses that could result from a Carrington-class event in the current information society. Based on their study, the estimated damage is approximately $300 billion, which is more than the economic loss caused by the Great East Japan Earthquake ($100–$250 billion) (Swiss Re 2014). In 2015, more than 50 experts from over 20 organizations in the United States (US) formulated the National Space Weather Strategy (National Science and Technology Council 2015) and Space Weather Action Plan (National Science and Technology Council 2019). Since 2016, in response to the Action Plan, the US Department of State, in collaboration with other countries, has been actively developing a framework for international cooperation in case of extreme events. "Space Weather Phase 1 Benchmarks" was published in 2018, providing benchmarks for induced ground currents, electromagnetic radiation, ionospheric disturbances, solar radio bursts, and atmospheric expansion (National Science and Technology Council 2018).

In Europe, the United Kingdom has been actively documenting the social impacts of space weather. In 2013, the Royal Academy of Engineering published "Extreme space weather: impacts on engineered systems and infrastructure" (Royal Academy of Engineering 2013), which reported on the social impacts of extreme events. This was followed by the publication of the National Risk Register by the Cabinet Office (2015a) and the Space Weather Preparedness Strategy (2015b).

In Asia, the Ministry of Science for the Future of Korea published the "Space Radio Interference Crisis Management Manual" in 2013.

3.2.2 Activities of International Organizations

As an international organization, the International Space Environment Service (ISES) has been active since 1962 under the International Council of Science Union. As of January 2020, 20 countries and the European Space Agency were members of the ISES, which is a federation of organizations that regularly disseminate space weather information.

In 2010, the World Meteorological Organization (WMO) established the Interprogramme Coordination Team on Space Weather as a temporary organization focusing on the distribution of space weather information in the WMO Information System. In 2015, the Inter-programme Team on Space Weather Information, Systems, and Services was established as a regular organization.

The International Civil Aviation Organization (ICAO) has been considering the incorporation of space weather information as weather information for use in aviation operations. Its aim is to avoid the risk of degradation of high-frequency (HF) communication and satellite positioning and increase in human exposure by space weather events, owing to the increase in the use of polar routes since the beginning of the twenty-first century.

In November 2011, the International Transportation Association issued a letter to the ICAO requesting consideration of space weather information for aircraft operations, following which discussions were initiated. Consequently, the Concept of Operation and the Third Annex, which specifies the meteorological information to be used for aeronautical operations, were revised. In parallel, the selection of organizations to provide the corresponding information has been underway. In June 2017, the ICAO released a state letter to ICAO Member States regarding their interest in providing the above information, and 22 states expressed their interest. Following written and in-person reviews, three organizations were approved as ICAO Space Weather Global Centers in November 2018 (United States, PECASUS (Finland, Austria, Belgium, Cyprus, Germany, Italy, Netherlands, Poland, and United Kingdom), ACFJ (Australia, Canada, France, and Japan). After a year of coordination meetings, the space weather information for aviation operations was launched on November 7, 2019. In 2020, the China and Russia Consortium joined the framework as the fourth global center.

In the United Nations, the Committee on the Peaceful Uses of Outer Space (COPUOS) has been studying the promotion of space weather monitoring and forecasting by countries. In particular, in 2018, space weather was selected as a focus area for the 50th anniversary of UNISPACE, the predecessor of COPUOS. A survey of the efforts of Member States in space weather monitoring and forecasting and the development of recommendations based on its results are underway.

3.3 Social Impacts of Space Weather Phenomena

Space weather phenomena have various social impacts depending on their scale, and occasionally they can cause damage. A schematic of the social damages caused by space weather phenomena is shown in Fig. 3.1, and a summary of their caused impacts, damages, and disturbances is provided in Table 3.4.

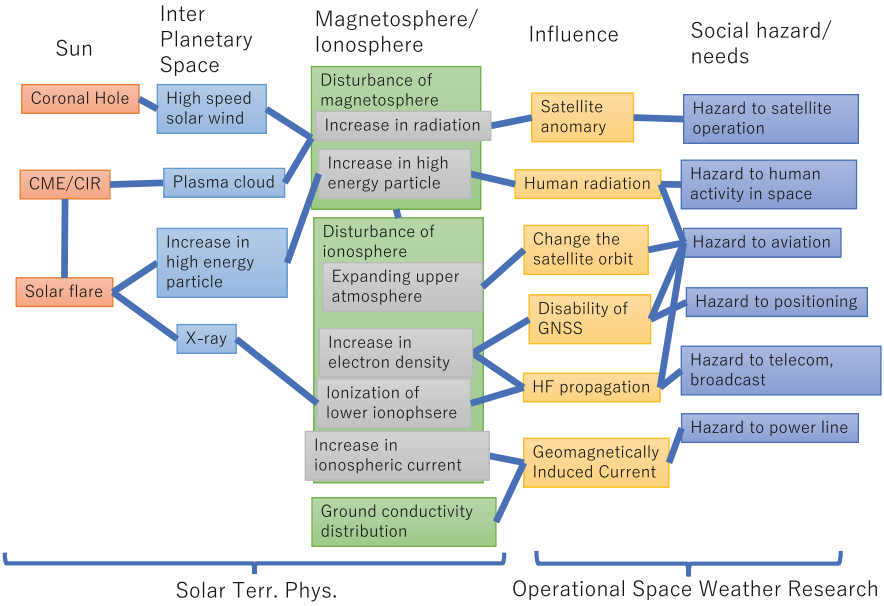

Fig. 3.1 Space weather phenomena and their impacts on society [Ishii et al. 2021]

In the hazard map [1], the potential social impacts of a massive space weather event were examined for different fields. Table 3.5, and the discussion of the frequency of occurrence and scale of a phenomenon in Sect. 1, summarizes the extents of the effects on the infrastructure of the Japanese society caused by different space weather phenomena, such as once per 10 years or once per 100 years.

Table 3.4 Sectors affected by space weather disasters and their impacts

Field	Effects and hazards	Causal space weather phenomena
Electric power	Power failure (overcurrent to power lines)	GIC
Satellite operation	Satellite charging (surface charge, internal charge) (malfunctions and other problems, with satellite loss in worst case)	Electrons in radiation belt and substorm high-temperature electrons
	Single-event upset (malfunction or other problems, worst-case satellite loss)	SEPs
	Solar panel degradation	
	Orbit and attitude disturbances (satellite attitude control (magnetic torque converter) anomaly and satellite atmospheric drag) (satellite loss in worst case)	Heating of upper atmosphere
	Impacts on satellite communication	Plasma bubble
Communication and broadcasting	Shortwave (HF) communication disorders	Dellinger phenomenon Ionospheric negative storm and polar cap absorption (PCA) Plasma bubble
	Disorders of very short-frequency (VHF) communication	Sporadic E layer and plasma bubble
Satellite positioning	Degradation of single-frequency positioning accuracy	Ionospheric positive storm, plasma bubble, fast propagation of total electron content (TEC) structure in auroral region, and solar radio burst
	Degradation of dual-frequency positioning (general) accuracy	Plasma bubble, fast propagation of TEC structure in auroral region, and solar radio burst
	Degradation of dual-frequency positioning (phase) accuracy	Medium-scale propagating ionospheric disturbance and solar radio burst
Aviation	Communication failure	Ionospheric negative storm, Dellinger phenomenon, PCA, plasma bubble, and sporadic E layer
	Degradation of positioning accuracy	Ionospheric positive storm, plasma bubble, and fast propagation of TEC structure in auroral region
	Human exposure	SEPs
	Communication errors in electronic devices	
	Radar disturbance	Solar radio burst (microwaves)
Manned space activities	Human exposure	SEPs
Earth life	Exposure on ground	SEPs
	Soft errors in electronic devices	
	Malfunction of signal equipment	GIC

Table 3.5 Space weather impact matrix

Field	Effects and hazards	Causal space weather phenomena	Occurrence probability and effects				
			Daily–several per year	Once per year	Once per 10 years	Once per 100 years	Once per 1000 years
Electricity	Blackout (overcurrent in power lines)	GIC					*
Satellite operation	Satellite surface charging	Substorm electrons				*	*
	Satellite internal charging	Energetic electrons				*	*
Radio communication and broadcasting ■	HF hazard	Ionospheric negative storm				*	*
		Dellinger phenomenon				*	*
		Polar cap absorption (PCA)				*	*
		Plasma bubble				*	*
	Very-high-frequency hazard	Sporadic E layer	△	△	△	△	△
		Plasma bubble				*	*
Positioning ■	Poor positioning accuracy	Ionospheric positive storm				*	*
		Plasma bubble				*	*
		Fast propagation of total electron content (TEC) structure in auroral region				*	*

(continued)

Table 3.5 (continued)

Aviation	Communication failure	Ionospheric negative storm				*	*
		Dellinger phenomenon				*	*
		PCA				*	*
		Plasma bubble			*	*	*
		Sporadic E layer			*	*	*
	Poor positioning accuracy	Ionospheric positive storm	*	*	*	*	*
		Plasma bubble	*	*	*	*	*
		Fast propagation of TEC structure in auroral region	*	*	*	*	*
	Aircrew exposure	SEPs					*
Human space activity	Astronaut exposure	SEPs					*
Daily life on surface of earth	Exposure on ground	SEPs					*

Color scale

Effects are negligible	Effects require adequate measures, e.g., backup preparation	Events prevent continuous operations and problems have long duration	Unevaluated effects

Notes: ■ influence depends on system of infrastructure; * may revise with future studies; △: "anomalies" instead of "damage"

References

Baker, D.N., Li, X., Pulkkinen, A., Ngwira, C.M., Mays, M.L., Galvin, A.B., Simunac, K.D.C.: A major solar eruptive event in July 2012: defining extreme space weather scenarios. Space Weather. **11**(10), 585–591 (2013)

Cabinet Office, UK National Risk Register of Civil Emergencies, 2015. [Online]. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/419549/20150331_2015-NRR-WA_Final.pdf (2015a).

Cabinet Office, UK Space Weather Preparedness Strategy, 2015. [Online]. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/449593/BIS-15-457-space-weather-preparedness-strategy.pdf (2015b).

Cid, C., Palacios, J., Saiz, E., Guerrero, A., Cerrato, Y.: On extreme geomagnetic storms. J. Space Weather Space Clim. **4**(A28), 10 (2014)

Goka, T.: Dictionary of space environment risk, Maruzen, p. 179, (2006).

Gopalswamy, N.: Extreme Solar Eruptions and their Space Weather Consequences. In: Extreme Events in Geospace: Origins, Predictability, and Consequences, vol. 2018, pp. 37–63. Elsevier Inc. (2018)

Ishii, M., Shiota, D., Tao, C., Ebihara, Y., Fujiwara, H., Ishii, T., Ichimoto, K., Kataoka, R., Koga, K., Kubo, Y., Kusano, K., Miyoshi, Y., Nagatsuma, T., Nakamizo, A., Nakamura, M., Nishioka, M., Saito, S., Sato, T., Tsugawa, T., Yoden, S.: Space weather benchmarks on Japanese society. Earth Planets Space. **73**(1), 108 (2021). https://doi.org/10.1186/s40623-021-01420-5

Jiggens, P., Heynderickx, D., Sandberg, I., Truscott, P., Raukunen, O., Vainio, R.: Updated model of the solar energetic proton environment in space. J. Space Weather Space Clim. **8**, 22 (2018)

Kurt, V., Mavromichalaki, H., Gerontidou, M.: Statistical analysis of solar proton events. Ann. Geophys. **22**, 2255–2271 (2004)

Miyake, F., Masuda, K.: Cosmic phenomena inscribed in the Yakusugi cedar: cosmic ray intensity anomalies in A.D. 774-775 and 993-994 (from recent studies). J. Physical Soc. Japan. **69**(2), 93–97 (2014)

Nakamura, M., Yoneda, A., Oda, M., Tsubouchi, K.: Statistical analysis of extreme auroral electrojet indices. Earth Planets Space. **67**(153) (2015)

National Science & Technology Council: US Government: National Space Weather Strategy, October 2015. https://www.whitehouse.gov/sites/default/files/microsites/ostp/final_nationalspaceweatherstrategy_20151028.pdf (2015).

National Science & Technology Council: US Government: Space Weather Phase 1 Benchmarks, June 2018. https://www.whitehouse.gov/wp-content/uploads/2018/06/Space-Weather-Phase-1-Benchmarks-Report.pdf (2018).

National Science & Technology Council: US Government: National Space Weather Strategy and Action Plan 2019. https://www.whitehouse.gov/wp-content/uploads/2019/03/National-Space-Weather-Strategy-and-Action-Plan-2019.pdf (2019).

Newell, P.T., Gjerloev, J.W.: Substorm and magnetosphere characteristic scales inferred from the SuperMAG auroral electrojet indices. J. Geophys. Res. **116**, A12232 (2011)

PSTEP: A01 group, Hazard map for science recommendation for space weather disaster, Grant-in-Aid for Scientific Research on Innovative Areas "Prediction of Solar-Terrestrial Environment" (2020).

Royal Academy of Engineering: Extreme space weather: impacts on engineered systems and infrastructure, 3 Carlton House Terrace. Royal Academy of Engineering, London (2013)

Sauer, H., Wilkinson, D.: Global mapping of ionospheric HF/VHF radio wave absorption due to solar energetic protons. Space Weather. **6**(12) (2008)

Swiss Re.: Space Weather Impacts a Risk to Society? http://www.swpc.noaa.gov/sites/default/files/images/u33/NOAA-MASTER.pdf (2014).

Tsubouchi, K., Omura, Y.: Long-term occurrence probabilities of intense geomagnetic storm events. Space Weather. **5**(S12003) (2007)

Tsurutani, B.T., Gonzalez, W.D., Lakhina, G.S., Alex, S.: The extreme magnetic storm of 1–-2 September 1859. J. Geophys. Res. **108**(A7) (2003)

Part II
Geospace and Space Weather Forecast

Chapter 4
Introduction of Space Weather Research on Magnetosphere and Ionosphere of the Earth

Kazuo Shiokawa

4.1 Structures of and Variations in Magnetosphere and Ionosphere of the Earth

As shown schematically in Fig 4.1, various space weather events occur in the Earth's magnetosphere and ionosphere, due to emissions and plasma flow from the sun and due to atmospheric waves from the lower atmosphere. Figure 4.2 schematically shows the north-south cross-section of the magnetosphere of the Earth from the side (equatorial direction) and the positions of major satellites in the magnetosphere. The magnetosphere extends to approximately ten times the radius of the Earth (Re) in the solar direction (left side of the figure) (10 Re = ~60,000 km) and has a long tail in the antisolar direction up to several hundred Re. On average, the plasma density in the solar wind is approximately 1–10 cm^{-3}, whereas it is 0.1 cm^{-3} or less in the magnetosphere close to the solar wind. Thus, the magnetosphere has a low-density bubble-like structure protected by the magnetic field of the Earth within a sea of dense plasma (solar wind) emitted by the Sun. As the magnetosphere reaches a few Re, the plasma density is increased by the outflows from the ionosphere.

Figure 4.2 shows that on the dayside (left side in the figure), a bow shock caused by the high-speed solar wind hitting the magnetosphere is present in front of the magnetosphere. On the downstream side of this bow shock, there is a magnetosheath, in which the magnetic field is highly disturbed. A magnetotail consists of a plasma sheet filled with plasma with energy of several kiloelectron volts (keV), sandwiched between a lobe with a low-density plasma and the strong magnetic field. There is a noticeable boundary between the plasma sheet and the lobe (plasma sheet boundary layer). The plasma sheet extends not only to the nightside (right side in Fig. 4.2) but also to the dayside. The energy of the plasma particles in

K. Shiokawa (✉)
Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Aichi, Japan
e-mail: shiokawa@nagoya-u.jp

K. Kusano (ed.), *Solar-Terrestrial Environmental Prediction*,
https://doi.org/10.1007/978-981-19-7765-7_4

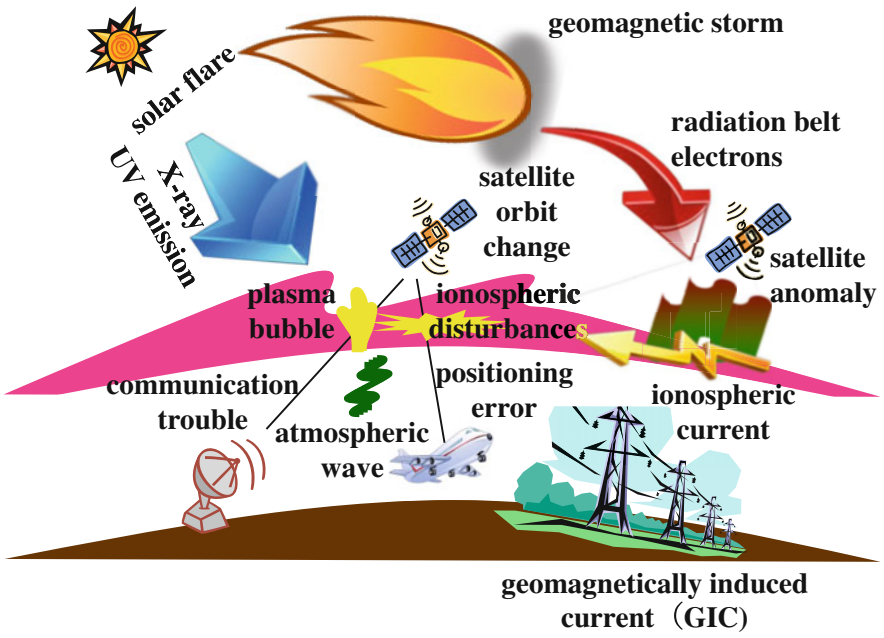

Fig. 4.1 Schematic of typical space weather phenomena occurring in magnetosphere and ionosphere of the Earth

the plasma sheet increases as they approach the Earth, forming a ring current region having energy of several tens of keV and radiation belts of several hundred keV to several tens of megaelectron volts (MeV). The distribution of the radiation belt electrons is divided into inner and outer belts by a slot region where the density is low at approximately 2–3 Re. In contrast, a cold and dense ($>10^3$ cm^{-3}) plasma with energy of a few electron volts (eV) flows out from the ionosphere of the Earth, spreading out to ~4 Re and forming a plasmasphere. The outer boundary of the plasmasphere (plasmapause) is frequently a well-defined density boundary. This is because the orbit of the plasma flowing out from the ionosphere is divided into an inner part, which orbits with the rotation of the Earth, and an outer part, which flows away toward the magnetopause by convection (Nishida 1966).

The above magnetospheric structures are mainly characterized by the magnetic field and plasma of the magnetosphere, which are invisible to the naked eye. However, they can be illustrated as in Fig. 4.2 because many satellites have observed the magnetosphere over a long period and statistically investigated the characteristics of its magnetic field and plasma. The solar wind in front of the magnetosphere has a Lagrangian point (L1) where satellites can remain because the gravity of the Sun and the Earth are balanced. The solar wind and the interplanetary magnetic field (IMF), which determines the disturbance levels of the magnetosphere and

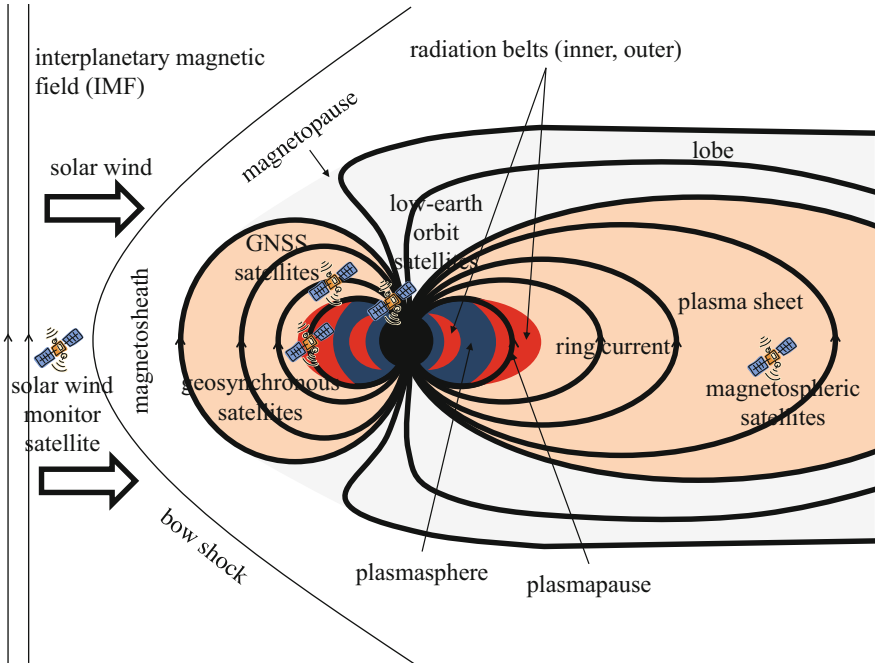

Fig. 4.2 Schematic of north–south cross section of magnetosphere of the Earth viewed from the side (equatorial direction) and locations of major satellites in it

ionosphere of the Earth, can be constantly monitored at L1 point, and they are the main information for space weather forecasting. Geostationary satellites and the Global Navigation Satellite System (GNSS) satellites, which are frequently used for space applications, are located at altitudes of 20,000–40,000 km and are strongly affected by the high-energy plasma in the outer radiation belt. Low-earth orbit (LEO) satellites, such as the International Space Station and Earth observation satellites that image the ground, are also strongly affected by the radiation belts when they pass through orbits near the North and South Poles.

The plasma in the magnetosphere of the Earth is nonstationary; it changes with time with the variations of the solar wind and the IMF. In particular, when the IMF is southward, as shown in Fig. 4.3, the IMF and the magnetic field of the Earth form a magnetic reconnection at the daytime magnetopause, which facilitates the penetration of the magnetic field energy and plasma particles into the magnetosphere (Dungey, 1962). Simultaneously, these reconnected magnetic field lines are swept into the magnetotail and accumulate there, which begins to be compressed from both north and south. This pressure causes the magnetic fields in the Northern and Southern Hemispheres to reconnect in the plasma sheet of the magnetotail. This magnetic reconnection in the magnetotail causes the plasma to be launched in the earthward and counter-earthward directions by the magnetic tension force. The

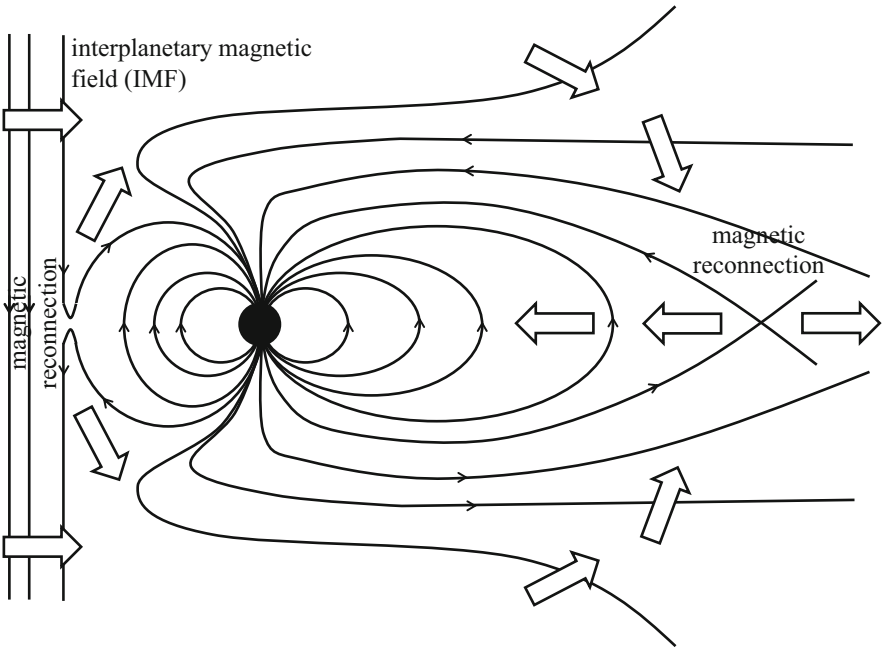

Fig. 4.3 Schematic of magnetic recombination between day- and nightsides and resulting magnetic field and plasma convection when IMF is southward

earthward plasma and the magnetic field are accelerated in space around the Earth. A part of them is lost into the atmosphere, and the remainder go around the dayside magnetopause. Such a series of plasma convections in the magnetosphere is considered to occur on a timescale of several to 10 h for one revolution.

This magnetic reconnection on the nightside does not necessarily occur continuously; it may occur abruptly after accumulation of the magnetic field flux in the magnetotail. Such an abrupt event is called a magnetospheric substorm, which occurs once or several times a day and lasts from one to several hours. It is also called an auroral substorm, because an aurora in the polar regions shines brightly owing to the injection of high-energy plasma particles into the atmosphere. In contrast, when an IMF turns significantly southward for longer than a few hours, a series of the plasma convections occurs continuously, resulting in a magnetic storm with the intrusion of large amounts of energetic plasma near the Earth. A magnetic storm occurs approximately once a month on average and lasts from one to several days. Magnetic storms occur when systematic plasma and IMF structures are in the solar wind, such as CMEs and co-rotating interaction regions associated with coronal holes, in which the strong southward IMF persists for long periods. Although substorms and magnetic storms have different timescales, both are typical magnetospheric disturbances and intrinsically important for understanding space weather around the Earth.

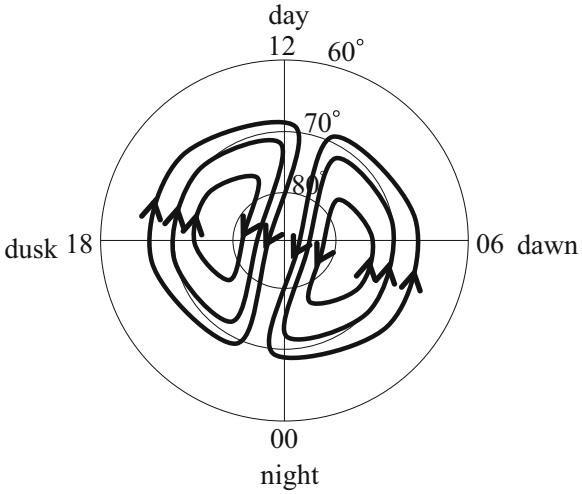

Fig. 4.4 Schematic of plasma convection in ionosphere at high latitudes (thick solid line). Figure shows view of the Earth from above the North Pole. Outermost circle represents 60° magnetic latitude

In a substorm, an intense auroral variability and ionospheric currents (auroral electrojet currents) are known to occur in the polar regions. The magnitude of the variability in the horizontal (northward) component of the geomagnetic field caused by the abovementioned ionospheric currents in the auroral zone is indexed by the AU (positive variability), AL (negative variability), and AE (= AU − AL) indices (Davis and Sugiura 1966; WDC-Kyoto et al. 2015a). However, when magnetic storms occur, a ring current, which flows westward in the longitude direction around the Earth, becomes stronger. This current is known to change the geomagnetic field in mid- and low latitudes in a southward direction (decreasing the background northward geomagnetic field). The strength of a magnetic storm is frequently expressed by how negatively the Dst index changes (Sugiura 1964; WDC-Kyoto et al. 2015b). The Dst value indexes the abovementioned mid- and low-latitude variations in the horizontal (northward) component of the geomagnetic field.

Figure 4.4 schematically shows the plasma flow in the magnetosphere illustrated in Fig. 4.3, projected along the magnetic field lines into the ionosphere. The figure looks down on the Earth from above the North Pole, and the magnetic North Pole is at the center. The plasma flow from day to night around the magnetic North Pole corresponds to the plasma that is magnetically reconnected with the IMF on the dayside magnetopause and is swept to the nightside. It can be seen that the plasma injected toward the Earth by the nightside magnetic reconnection flows around the morning and evening sides of the Earth at lower latitudes and subsequently returns to the dayside. Such plasma convections have been observed by high-frequency (HF) radars, which can monitor plasma motions in the ionosphere, and by low-altitude satellites flying in the ionosphere (e.g., Heppner and Maynard 1987; Weimer 1995; Ruohoniemi and Baker 1998).

Figure 4.5 shows the typical motion of plasma sheet particles injected from the magnetotail toward the Earth as a result of its magnetic reconnection. The figure is a

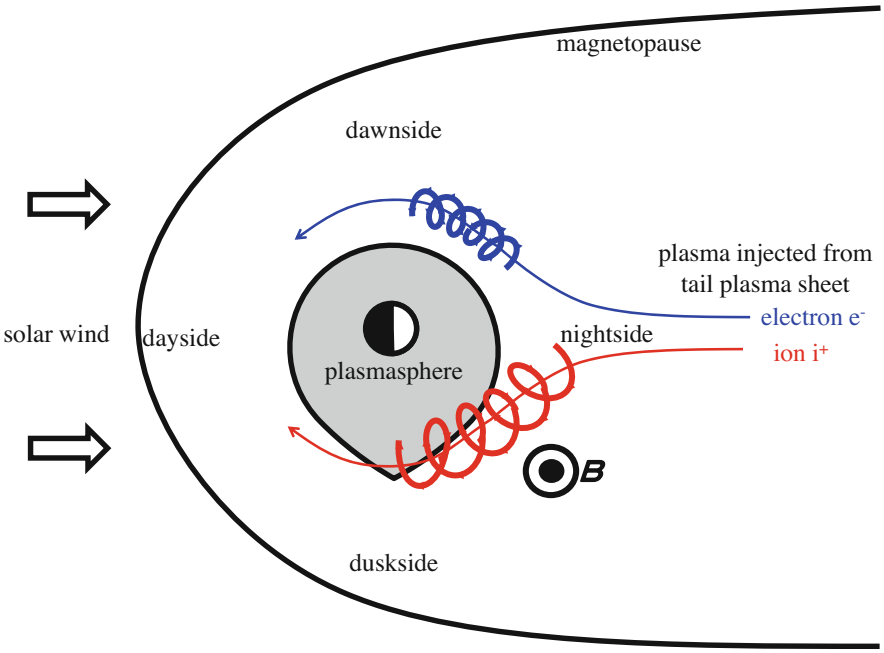

Fig. 4.5 Schematic of plasmasphere and motion of plasma sheet particles injected from magnetotail. The figure shows equatorial cross section of magnetosphere of the Earth viewed from above the North Pole. Schematics of cyclotron motions of ions (red) and electrons (blue) are also shown

schematic of the inner magnetosphere, which is within ten times the Re, in the equatorial cross section viewed from above the North Pole. The electrons and ions constituting the plasma start to drift eastward (morning side) and westward (evening side), respectively, when they approach the Earth. This is because the radius of the cyclotron motion decreases on approaching the Earth and increases on going far from it, as shown by the red and blue rotating lines in the figure. This is due to the earthward gradient of the magnetic field strength and the curvature of the geomagnetic field. Because the rotation directions of the cyclotron motions of the ions and the electrons are opposite, they are separated into westward and eastward directions, respectively. For the theoretical derivation of this magnetic field gradient drift and the curvature drift velocity, please refer to plasma physics textbooks such as by Nicholson (1986). The abovementioned longitudinally drifting high-energy plasma of several keV to several tens of keV spatially overlaps with low-energy (less than several eV) dense plasmas in the plasmasphere and interacts with them to produce various phenomena in the inner magnetosphere. In addition, some of these injected electrons are further accelerated from several hundred keV to several MeV in energy by interaction with electromagnetic waves, forming the radiation belts. The generation and propagation of these electromagnetic waves are influenced by the

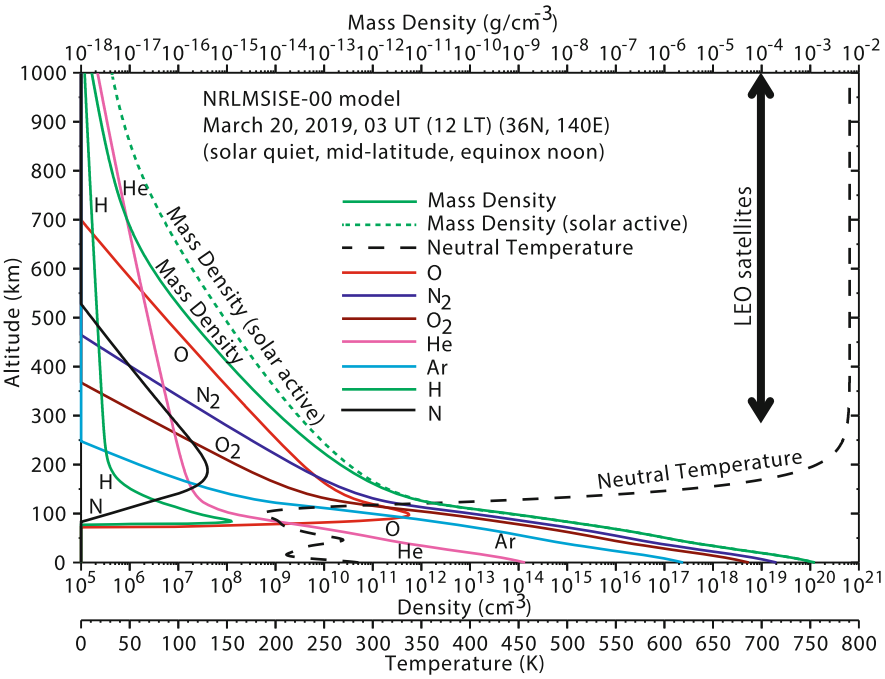

Fig. 4.6 Altitude profiles of neutral atmosphere density and its main composition and temperature at midlatitude (36°N, 140°E) and vernal noon (03 universal time (UT) = 12 LT) during solar activity minimum (March 20, 2019) obtained using from NRLMSISE-00 model. Green dotted line is mass density at same time and location during solar activity maximum (March 20, 2003), as calculated by the Community Coordinated Modeling Center at Goddard Space Flight Center of the National Aeronautics and Space Administration (https://ccmc.gsfc.nasa.gov/modelweb/)

low-energy plasma in the plasmasphere. Thus, the inner magnetosphere is an important region for space weather prediction. This is because the complex interaction of plasma particles with energies of more than six orders of magnitude from eV to MeV with the electromagnetic field and the electromagnetic waves generates the radiation belts, which affect safe and secure space utilization.

Thus far, we have presented various characteristics of the magnetosphere of the Earth, which significantly vary as it interacts with the solar wind. The magnetic field lines of the magnetosphere are concentrated toward the Earth and connected to the atmosphere, as shown in Fig. 4.2. The upper layer of the atmosphere of the Earth is ionized by the solar ultraviolet radiation and forms the ionosphere, which supplies plasma to the magnetosphere, similar to the case of the plasmasphere. Moreover, the ionosphere is electrically connected to the magnetosphere through magnetic field-aligned currents, and acts mainly as a load on the magnetosphere, in terms of an electric circuit.

To understand the structure of the ionosphere, we first show the altitudinal distribution of the atmospheric density and its composition in Fig. 4.6. This figure

shows the atmospheric density at midlatitudes (over Japan) at 12:00 (local time (LT)) on March 20, 2019 (the vernal equinox of the solar activity minimum) obtained using the NRLMSISE-00 model (Picone et al. 2002). Up to an altitude of approximately 100 km, the atmosphere of the Earth consists mainly of nitrogen (N_2) and oxygen (O_2) molecules. From 100 km upward, oxygen molecules are converted to oxygen atoms (O) by photodissociation in sunlight, and gravitational diffusion separation (molecular diffusion) becomes more effective. Thus, oxygen (O), helium (He), and hydrogen (H) atoms are the main components in order of their weights from low to high altitudes. The altitude range in which the atmospheric density becomes 1/e (scale height) is approximately 6–8 km below 120 km, whereas it increases rapidly 30–50 km above 120 km.

The temperature of the atmosphere (the black dashed curve in Fig. 4.6) decreases with distance from the ground (a heat source) up to approximately 10 km of altitude. This region is called the troposphere, in which convection is caused by this negative temperature gradient. Above the tropopause, ozone absorbs the solar ultraviolet radiation and becomes another heat source, temporarily raising the temperature to form the stratosphere and the mesosphere (regions of positive and negative temperature gradients, respectively). Above the mesopause at an altitude of approximately 90 km, the atmosphere is heated, because it is the altitude region where the atmosphere mainly absorbs the ultraviolet radiation from the Sun, and the temperature increases rapidly with the altitude to form the thermosphere. The temperature at the upper end of the thermosphere is approximately 800 K during the solar minimum, whereas it rises to more than 1000 K during the solar maximum.

Most of the LEO satellites and space stations, which are frequently used in space applications, fly in an altitude range of 300–1000 km, as shown by the arrows in Fig. 4.6. This is because it is difficult to maintain a satellite in space for a long period below ~300 km as it can undergo large collisions with the atmosphere (atmospheric drag). Comparing the solid green (solar minimum period; March 20, 2019) and dotted (solar maximum period; March 20, 2003) lines in Fig. 4.6, in the solar maximum period, the mass density of the thermosphere in which a satellite flies increases more than twice than in the solar minimum period. This suggests that the solar activity plays a major role in the lifetime of a satellite.

Figure 4.7 shows the altitudinal distributions of the density of the ionized atmosphere and its composition and electron and ion temperatures. It presents the electron density in the ionosphere at 12:00 noon (03 UT) at midlatitudes (over Japan) on March 20, 2019 (the vernal equinox of the solar activity minimum) using the IRI-2012 model (Bilitza et al. 2014). The temperature of the neutral atmosphere shown at the same time is based on the CIRA-86 model (Fleming et al. 1990). The difference between the terms “electron density” and “plasma density” is that in a plasma consisting of monovalent ions and electrons, the electron density (density of negative charges) and the ion density (density of positive charges) are typically equal. Therefore, the two terms are frequently used synonymously. Because the plasma density of the ionosphere has historically been measured as the electron density by the reflection of radio waves contributed by electrons, the plasma density is generally expressed as the electron density.

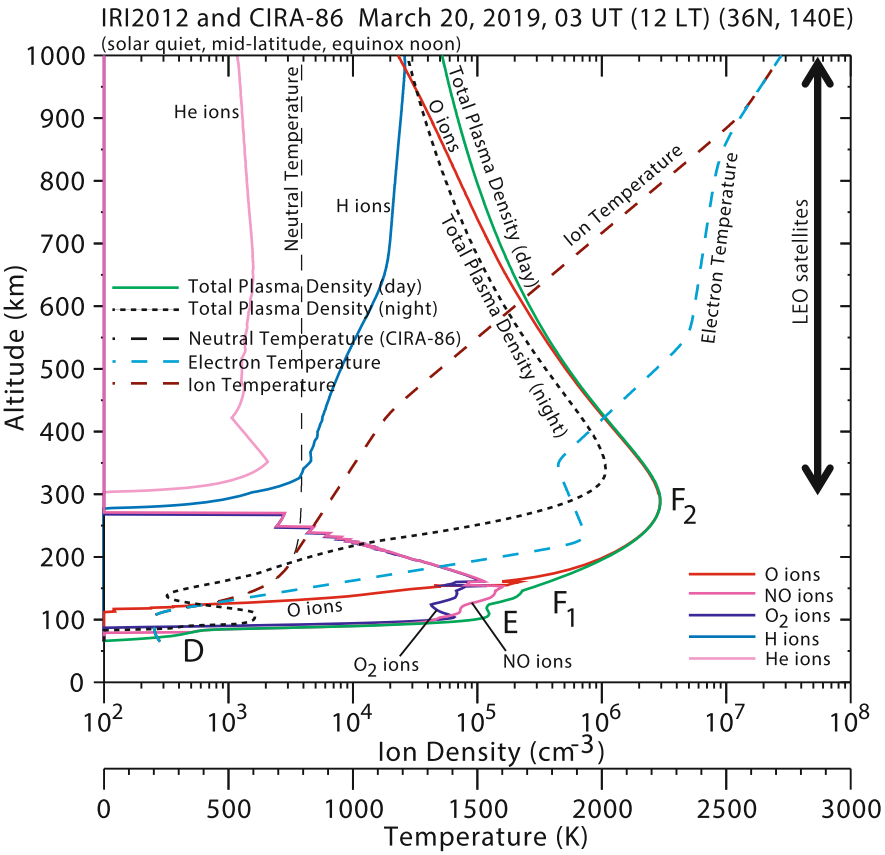

Fig. 4.7 Altitude profiles of ionospheric electron density, its representative composition, and ion and electron temperature from the IRI-2012 model at midlatitudes during solar minimum, spring equinox daytime, as calculated by the Community Coordinated Modeling Center at Goddard Space Flight Center of National Aeronautics and Space Administration (https://ccmc.gsfc.nasa.gov/modelweb/)

The daytime ionosphere has local peaks and bulges in the D, E, F_1, and F_2 layers from the bottom. These can be attributed to the wavelength distribution of the corresponding solar ultraviolet radiation, compositional distribution of the atmosphere, and recombination reactions between the neutral and ionized atmospheres. The E layer is dominated by NO^+ and O_2^+, whereas the F layer is dominated by O^+. H^+ becomes the dominant component at heights above 1000 km. The D layer (D region) is dominated by complex molecular ions called hydrated and cluster ions, which are formed from $NO^+$and O_2^+ by various chemical reactions.

The electron density at the peak height of the F layer, where the density of the ionosphere is the highest, is approximately 10^6 cm^{-3}. The density of the neutral atmosphere at the same altitude of ~300 km is approximately 10^9 cm^{-3}, i.e., only

1 out of 1000 atoms and molecules in the atmosphere is ionized at the highest level. Therefore, the dynamical variability in the ionospheric plasma is significantly influenced by the winds of the neutral atmosphere.

The ionospheric electron density at 00:00 (15 UT) on the same day and at the same location is also shown by a dotted line in Fig. 4.7. Because there is no ionization source at night, the recombination due to the collision between the ionospheric plasma and the neutral atmosphere gradually progresses from the lower part. Moreover, the peak height of the ionospheric F layer increases and the overall electron density drops to less than half. At night, the E layer almost disappears owing to the recombination; however, as described in Chap. 7, metal ion accumulation is caused by the vertical wind shear of the neutral atmosphere, and a layer called the sporadic E layer abruptly emerges.

LEO satellites and space stations fly above the peak altitude of the ionospheric F layer for the most part because of their altitudes above 300 km. Therefore, the communication with the ground is strongly affected by the plasma in the ionosphere.

The electron density in the ionosphere is generally high near the equator, where the solar ultraviolet radiation is the strongest, and decreases to less than half as it approaches both poles. The latitudes at which the density reaches its maximum are approximately 5–15° north and south of the magnetic equator and are called the equatorial ionization anomaly. The equatorial ionization anomaly occurs because the plasma generated at the magnetic equator is lifted by the eastward electric field during the daytime and redistributed along the magnetic field lines to high latitudes by the gravitational force of the Earth. Thus, the density of the ionosphere is at its maximum at the equatorial ionization anomaly, where the impact on positioning and communication also becomes maximum.

Similar to the magnetospheric plasma, the ionospheric plasma dynamically varies. The ionospheric electron density is broadly determined by the balance between the production by the solar ultraviolet radiation, quenching by recombination with the neutral atmosphere, and diffusion along the geomagnetic field lines. In particular, the ionospheric plasma in the F layer is subject to an $\mathbf{E} \times \mathbf{B}$ drift due to the east–west electric field and the northward magnetic field and flows with velocity $\boldsymbol{v} = (\boldsymbol{E} \times \boldsymbol{B})/B^2$ in the direction of the product of the electric field, $\boldsymbol{E}$, and the magnetic field, $\boldsymbol{B}$. The eastward electric field increases the altitude of the ionosphere, whereas the westward electric field decreases it. Both electric fields may be transmitted to the ionosphere by solar wind–magnetosphere interactions, magnetospheric storms, and substorms.

The plasma in the ionospheric F layer is also shifted up and down along the magnetic field lines by the meridional winds in the neutral atmosphere. Because the plasma in the ionosphere cannot move across a magnetic field line owing to its cyclotron motion, the poleward winds push the plasma down along the geomagnetic field lines, whereas the equatorward winds push it up at midlatitudes. The winds in the neutral atmosphere (thermosphere) are dominated by thermospheric tides, which have a 1-day cycle, and they blow from the warmed dayside atmosphere to the cooler nightside atmosphere. Therefore, the daytime ionosphere tends to be pushed down by the poleward winds, whereas the nighttime ionosphere tends to be lifted by the

equatorward winds. In addition, atmospheric waves propagating from the lower atmosphere (such as atmospheric gravity waves, harmonics of atmospheric tides, and planetary waves) and heating of the upper atmosphere at high latitudes due to magnetic storms also produce neutral wind variations in the thermosphere. When the ionosphere is pushed down by the electric field and neutral winds, the density of the ionospheric plasma decreases as it recombines with the higher-density neutral atmosphere at lower altitudes.

In addition to these external ionospheric disturbances, the ionospheric plasma itself causes fluid-like plasma instabilities, as described in Chap. 7. Most of them are Rayleigh–Taylor-type instabilities, such as plasma bubbles at equatorial latitudes and medium-scale traveling ionospheric disturbances at midlatitudes, both increased by an $\mathbf{E} \times \mathbf{B}$ drift due to the polarized electric field. A gradient drift instability, which is caused by the density gradient of the ionospheric plasma and the current flowing orthogonally to it, is also broadly a Rayleigh–Taylor-type plasma instability. These ionospheric plasma instabilities are important for understanding space weather, because they produce subkilometer-scale plasma structures in the ionosphere, which cause interferences in the radio waves used for satellite-to-ground communication and satellite positioning.

4.2 Measurement of Magnetosphere and Ionosphere of the Earth

The structures of and variations in the magnetosphere and ionosphere of the Earth described thus far have been clarified by various measurements using instruments on board satellites and remote sensing instruments on the ground. Thus, the knowledge of the magnetosphere and the ionosphere is strongly connected with the measurement techniques, and understanding of these measurement techniques is indispensable for the understanding of the magnetosphere and the ionosphere. In this section, we briefly introduce the representative measurement techniques for the terrestrial magnetosphere and ionosphere.

The plasma and the electromagnetic field in the magnetosphere of the Earth, except the electromagnetic waves propagating to the ground, are difficult to observe from the ground by remote sensing. Thus, they have been measured by direct (in situ) observations using satellites. Fig. 4.8 shows the instruments and their energy and frequency ranges used to measure the plasma and the electromagnetic field using the Exploration of Energization and Radiation in Geospace (also called Arase) satellite of Japan, which observes the magnetosphere around the Earth (Miyoshi et al. 2018). Typical examples of the data observed by these instruments are presented in Fig. 4.9. The Arase satellite is a satellite that observes the inner magnetosphere with an apogee altitude of ~32,000 km, a perigee altitude of ~440 km, and an orbital period of ~10 h. This satellite is a typical magnetospheric satellite that carries major plasma and electromagnetic field instruments besides

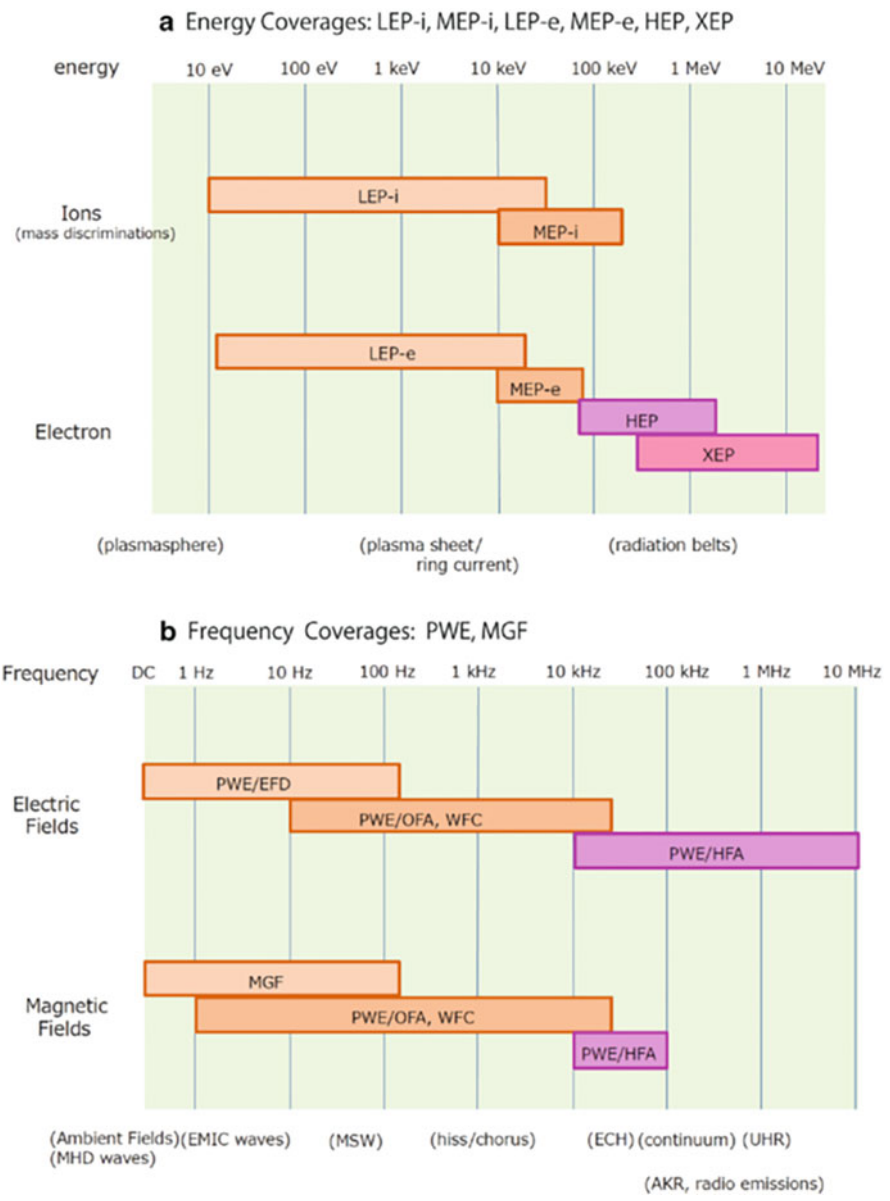

Fig. 4.8 (**a**) Measured energy ranges of particle instruments and (**b**) measured frequency ranges of electromagnetic field instruments on board Arase satellite (Miyoshi et al. 2018)

optical instruments. Other such satellites are the Time History of Events and Macroscale Interactions during Substorms and Radiation Belt Storm Probes satellites of the United States and the Geotail satellite of Japan. Figure 4.8a shows six instruments that measure ions (i) and electrons (e) and their measured energy ranges.

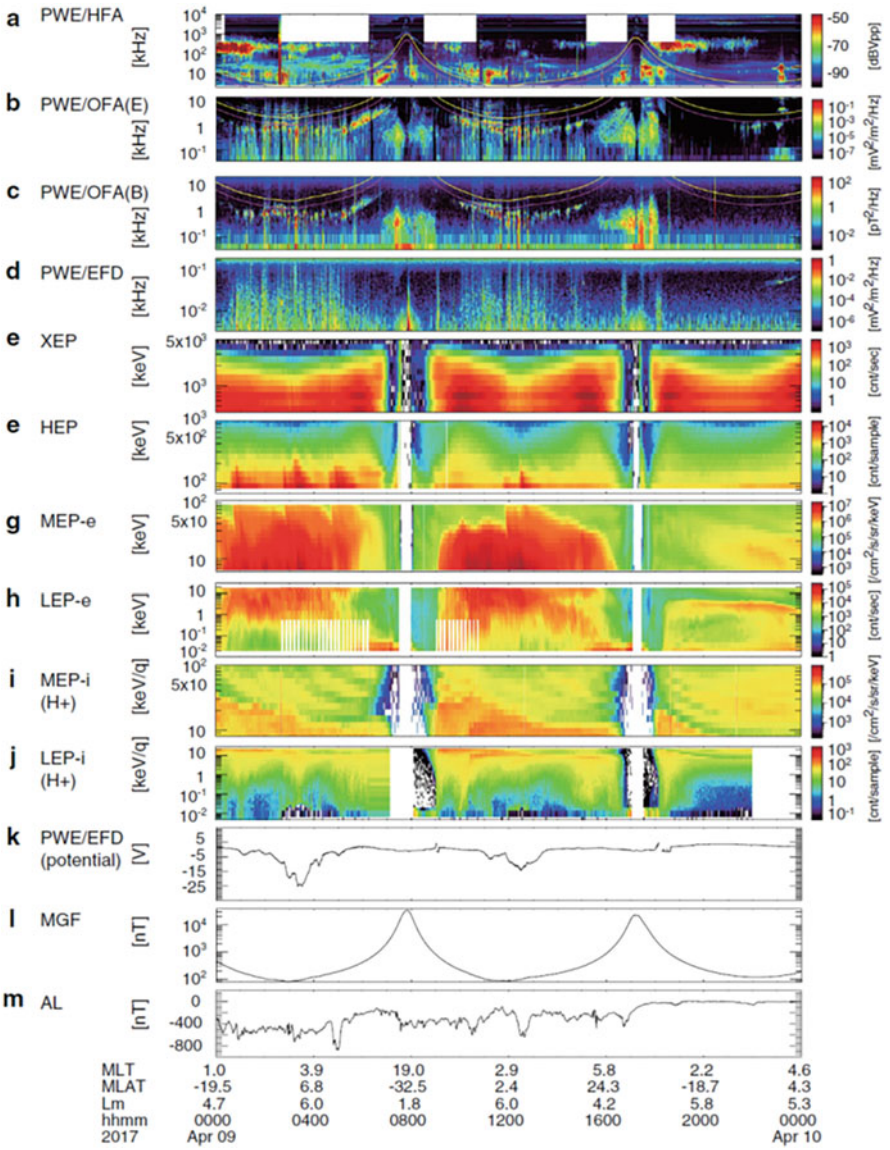

Fig. 4.9 (**a–d**) Frequency spectrum of electromagnetic waves, (**e–j**) energy–time spectrum of plasma particles, (**k**) satellite potential, (**l**) magnetic field strength, and (**m**) AL index observed by Arase satellite (Miyoshi et al. 2018)

Low-energy particle experiments, ion mass analyzer (LEP-i); low-energy particle experiments, electron analyzer (LEP-e); medium-energy particle experiments, ion mass analyzer (MEP-i); and medium-energy particle experiments, electron analyzer (MEP-e), at the low-energy side use the principle of an electrostatic analyzer, which

is a device that applies a constant electric field between two parallel plates and counts only the charged particles that pass between them. MEP-i and MEP-e also use a new technology called an avalanche photodiode (Kasahara et al. 2018; Yokota et al. 2017). Using an electrostatic analyzer, the number of particles at an energy is measured by changing the strength of the electric field between two parallel plates stepwise. However, to avoid discharge accidents, the strength of the electric field between the two plates is typically limited to be less than a few kilovolts. Thus, the energy of the particles that can be measured has an upper limit of tens to hundreds of keV. For particles with higher energies, a solid-state detector is used to directly count them and simultaneously measure their energy from the magnitude of the output pulse. Because the measurement principle and energy range of the different instruments vary, simultaneous measurement of the energy distribution function of plasma particles over a wide energy range from 20 eV to more than 10 MeV is realized by installing multiple instruments.

Panels (e)–(j) in Fig. 4.9 show the particle flux measured by these plasma particle detectors in color. The vertical axis is the energy, and the horizontal axis is the time, with the satellite position moving from the ionosphere near the Earth (0800 universal time (UT) and 1800 UT) to the magnetosphere with an orbital period of approximately 10 h. Note that the units of the different colors are slightly different. In MEP-i and MEP-e, whose instruments are calibrated, the color unit is [/cm^2/s/sr/keV], which indicates the number particles entered per unit area [cm^2], unit time [s], unit solid angle [sr], and unit energy [keV]. By calibrating the sensitivity of each particle detector and adjusting it to the same units, a uniform particle distribution function over a wide energy range can be obtained. However, the calibration of the absolute sensitivity of particle detectors may include systematic errors due to various factors, and their fields of view and energy resolutions are different. Thus, comparing and matching the distribution functions obtained by different instruments are difficult tasks, which could become research topics.

Figure 4.8b shows six instruments that measure the electric and magnetic fields and their ranges of measurement frequencies. The magnetic field is measured over a wide range of frequencies using fluxgate magnetometers (MGFs), which measure low frequencies from DC to 128 Hz, and search coil magnetometers (plasma wave experiment (PWE)/onboard frequency analyzer (OFA), PWE/high-frequency analyzer (HFA)), which measure high frequencies. These magnetometers are mounted on the top of a 5-meter-long mast to avoid the magnetic field noise generated by the satellite body. The electric field is measured by four (two pairs) 15-meter-long wire antennas stretched out in the satellite spin plane. The electric field variations in space obtained from these antennas are divided into three frequency bands by three types of receivers: PWE/EFD, PWE/OFA-WFC, and PWE/HFA. Although the electric field perpendicular to the spin surface cannot be measured, the x, y, and z components of the magnetic field can be obtained. By assuming that the electric field along a magnetic field line is zero, we can obtain the third component of the electric field from the equation, $\boldsymbol{E} \cdot \boldsymbol{B} = 0$.

Figure 4.9a, b, and d shows the electric field variations in the abovementioned three frequency bands, and Fig. 4.9c and i presents the wave and DC components of

the magnetic field measured by a search coil magnetometer and a fluxgate magnetometer, respectively. The satellite potential shown in Fig. 4.9k is not the potential difference between the two ends of a single-wire antenna but the potential at the root of the antenna, which represents the potential of the satellite relative to that of the surrounding plasma. This value varies with the density and temperature of the plasma around the satellite, and the plasma density can be estimated from this value.

As presented in Fig. 4.8a, the particle detectors cannot measure the distribution function of the plasma below 10–20 eV. This is because if the particle energy is extremely low, it is affected by the potential of the satellite body, and accurate measurement is impossible. Such low-energy plasmas, as represented by the plasmasphere, account for most of the plasma density in space and are important phenomena that have significant effects on the excitation and propagation of electromagnetic waves. The background plasma density can be accurately estimated from the frequency, f_{UHR}, of the upper hybrid resonance (UHR), which is characteristically strong at a single frequency of 10–1000 kHz. The f_{UHR} is a function of the plasma frequency, f_p, and the cyclotron frequency, f_c, as $f_{\mathrm{UHR}}^2 = f_p^2 + f_c^2$. The value of f_c is obtained from the magnetic field strength measured by a fluxgate magnetometer. Therefore, we can estimate the plasma frequency, f_p, and, thus, the ambient plasma density, from the f_{UHR} measured from a high-frequency wave spectrum at 10–1000 kHz.

As described above, satellites have performed numerous direct in situ observations of the magnetosphere of the Earth, providing a schematic of the dynamic variations in the magnetosphere, which is presented in Sect. 2.1.1. In addition, to visualize the global picture of the magnetosphere, resonant scattering of sunlight by plasmasphere ions and high-energy neutral particles from ring current ions moving away from the magnetic field lines have been received by satellites in combination with wide-area auroral images from high altitudes, such as those by the IMAGE satellite (Mitchell et al. 2000; Mende et al. 2000; Goldstein and Sandel 2005). In the future, there are plans to image the magnetosphere using electromagnetic waves that propagate straight, irrespective of the ambient magnetic field and the plasma density, such as X-rays (Sibeck et al. 2018).

In the ionosphere, which is closer to the Earth than the magnetosphere, direct measurements from LEO satellites using plasma particles and electromagnetic field instruments have been conducted, similarly to those in the magnetosphere. Remote sensing measurements from the ground have also been conducted for a long time. Figure 4.10 summarizes the measurement altitude ranges and measurement parameters of major ground-based remote sensing instruments. Broadly, from left to right, these are radars, which emit radio waves from the ground and measure the reflected waves from the ionosphere; radio receivers, which receive radio waves emitted by satellites at altitudes higher than the ionosphere on the ground; radio receivers which measure natural radio waves in the very-low-frequency (VLF) band, low-frequency (LF) band, and millimeter-wave band and magnetic field fluctuations of natural origin; spectrometers and interferometers, which receive atmospheric light emitted

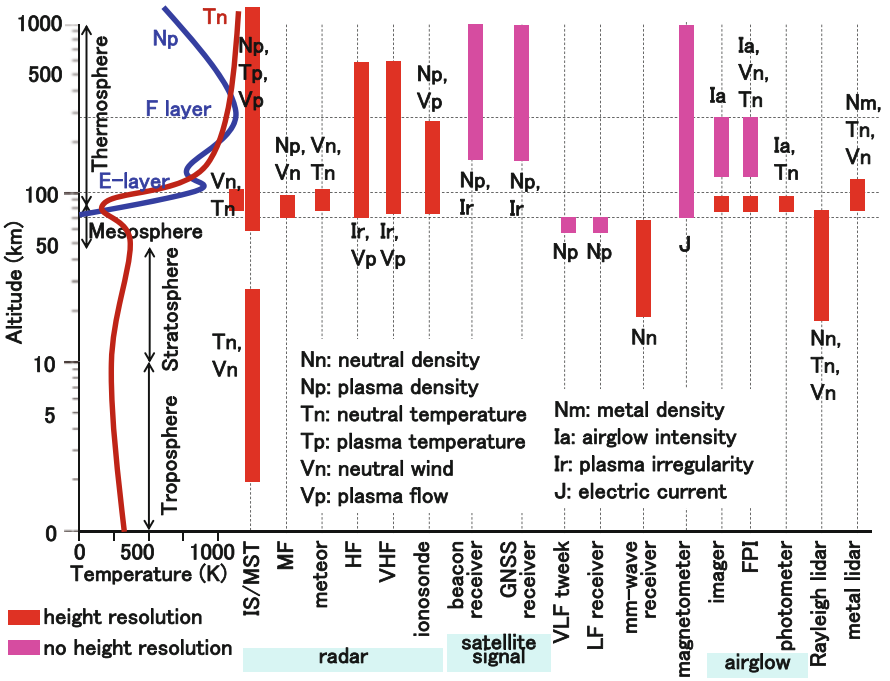

Fig. 4.10 Schematic of ground-based instruments and measurement altitudes in upper atmosphere and ionosphere (Oberheide et al. 2015)

in the ionosphere; and lidars, which emit laser beams and measure the reflected lights. Typical instruments are briefly introduced below.

1. Ionosonde: It is a radar that has long been used to observe the ionosphere. It emits radio waves in a frequency range from 2 to 20 MHz from the ground and measures the reflection from the ionosphere. By measuring the time difference between the emission and the reflection while changing the frequency, the altitude of each electron density in the lower ionosphere (e.g., h'E and h'F) and the peak electron density in the ionosphere (e.g., foE and foF) can be determined. This instrument has been operated in many countries since the International Earth Observation Year in 1957 and has significantly contributed to the development of the ionospheric standard model.
2. Incoherent scattering radar: It is a radar that emits powerful radio waves from the ground and captures the radio waves that are reflected very weakly by the random thermal motion of the electrons in the ionosphere. It can measure the electron density distribution and the velocity of the plasma motion in the ionosphere at all altitudes, from lower to upper levels. It requires a large antenna with a large aperture and a high output power of several megawatts or more, making it a large-scale facility of several billion yen in value. Therefore, there are only less than ten such devices in the world, such as the European Incoherent Scatter (EISCAT)

radar in Scandinavia, Jicamarca radar in Peru, Arecibo radar in Puerto Rico, and middle and upper atmosphere (MU) radar in Japan.

3. Coherent scattering radar: When there are repeating thick and thin electron density structures in the ionosphere and the spatial scale of the repetition is the same as the half-wavelength of a radar radio wave, the radio wave can be scattered efficiently. Because this scattering is much stronger than the incoherent scattering described above, the ionosphere can be observed with a relatively inexpensive radar with an output power of approximately several kilowatts. However, observation is impossible unless electron density structures are present in the ionosphere. Therefore, radar radio waves do not always return, and the measurement direction and time are limited. The Super Dual Auroral Radar Network (SuperDARN) radars, which are commonly used in the polar regions, and VHF radars, which are employed in the equatorial regions, are such coherent scatter radars with interferometric capability to have two-dimensional field of views.
4. GNSS receiver: Radio signals from the Global Navigation Satellite System (GNSS) satellites, which are commonly used for positioning, present a delay proportional to the ionospheric electron density when they pass through the ionosphere. The total electron content (TEC) of the ionosphere, integrated in the direction of altitude, can be obtained by measuring this delay. In recent years, to measure crustal movement, more than 1000 GNSS receivers have been installed over Japan by its Geospatial Information Authority. This multipoint network enables two-dimensional imaging observations of the TEC in the ionosphere. The TEC distribution in the ionosphere along the orbit of a satellite can also be measured in approximately 10 min by receiving radio waves from LEO satellites, which move with quite a high speed (~8 km/s) when viewed from the ground, instead of the GNSS satellites.
5. VLF tweeks and LF radio receivers: LF-band standard radio waves have been used to set the times of radio clocks. VLF-band tweek radio waves are emitted by lightning discharges. These LF/VLF waves spread across the Earth while propagating through the duct between the ground and the ionospheric D layer. By receiving these radio waves, altitude changes in the lower ionospheric D layer, which are difficult to measure with radars such as ionosondes, can be monitored from the ground.
6. Millimeter-wave receiver: Atmospheric minor constituents such as ozone and nitrogen oxides in the upper atmosphere emit radio waves in the millimeter-wave band. By receiving these radio waves, the amounts and fluctuations of these minor constituents can be measured from the ground.
7. Magnetometer: Magnetometers measure the geomagnetic variations that are mainly caused by the currents in the magnetosphere and the ionosphere. In particular, the current flowing in the ionosphere contributes the most to the magnetic field variations on the ground. By installing magnetometers at multiple points on the ground, a two-dimensional distribution of the ionospheric currents and its time evolution can be observed. Combining this distribution with magnetic field fluctuations observed by satellites, there are also efforts to obtain a

three-dimensional structure of the ionospheric, magnetospheric, and field-aligned electric currents. The measurement of geomagnetic variations is essential as the input information for the measurement and modeling of GICs, which are major phenomena in space weather. Geomagnetic variations also present characteristic oscillations called geomagnetic pulsations, which have periods of approximately 0.2 s–10 min. From these geomagnetic pulsations, it is possible to diagnose phenomena in the magnetosphere and ionosphere of the Earth, such as predicting the occurrence of substorms using Pi2 geomagnetic pulsations and measuring the plasmasphere density through field-line resonance frequencies.

8. Airglow imager and Fabry–Perot interferometer: Because nocturnal airglows and auroras originate at altitudes of 80–300 km with a constant emission layer at each wavelength, their imaging observation is possible from the ground. Fabry–Perot interferometers have been used to measure the neutral and plasma velocities in the upper atmosphere by the Doppler shift of the emission lines of auroras and airglows.
9. Lidar: An optical laser beam is launched into the sky to measure scattered laser lights, which can be used to measure the density, temperature, and wind speed of the scattering atmosphere. In particular, a metal lidar, which measures the resonant scattering of metal ions such as sodium, iron, and calcium at altitudes of 80–110 km, is an important instrument that can measure the atmosphere at altitudes of 100 km or higher, which is difficult by other instruments.

As described above, the ionosphere and the upper atmosphere have been observed from the ground by various instruments as well as by satellites. Although satellites can directly observe geophysical quantities in situ, because they move at high speeds, they cannot determine changes over time at a fixed point. In contrast, remote sensing from the ground can measure for a long time at a fixed point; however, the physical quantity that can be measured varies depending on the instrument, and there are several measurement limitations. To understand the changes in the ionosphere and the upper atmosphere, it is very important to combine the advantages of various instruments, instead of being limited to measurements of one type of instrument.

References

Bilitza, D., Altadill, D., Zhang, Y., Mertens, C., Truhlik, V., Richards, P., McKinnell, L.-A., Reinisch, B.: The International Reference Ionosphere 2012 - A model of international collaboration. J. Space Weather Space Clim. **4**(A07), 1–12 (2014). https://doi.org/10.1051/swsc/2014004

Davis, N., Sugiura, M.: Auroral electrojet activity index AE and its universal time variations. J. Geophys. Res. **71**(3), 785–801 (1966). https://doi.org/10.1029/JZ071i003p00785

Dungey, J.W.: The structure of the exosphere or adventure in velocity space. In Geophysics, p. 505. Gordon and Breach Science Publishers, London (1962)

Fleming, E.L., Chandra, S., Barnett, J.J., Corney, M.: Zonal mean temperature, pressure, zonal wind, and geopotential height as functions of latitude, COSPAR International Reference

Atmosphere: 1986, Part II: middle atmosphere models. Adv. Space Res. **10**(12), 11–59 (1990). https://doi.org/10.1016/0273-1177(90)90386-E

Goldstein, J., Sandel, B.R.: The global pattern of evolution of plasmaspheric drainage plumes. In: Burch, J.L., Schulz, M., Spence, H. (eds.) Inner Magnetosphere Interactions: New Perspectives from Imaging, p. 1. American Geophysical Union, Washington, DC (2005). https://doi.org/10.1029/2004BK000104

Heppner, J.P., Maynard, N.C.: Empirical high-latitude electric field models. J. Geophys. Res. **92**(A5), 4467–4489 (1987). https://doi.org/10.1029/JA092iA05p04467

Kasahara, S., Yokota, S., Mitani, T., Asamura, K., Hirahara, M., Shibano, Y., Takashima, T.: Medium-energy particle experiments - electron analyser (MEP-e) for the exploration of energization and radiation in Geospace (ERG) mission. Earth Planets Space. **70**(1), 1–6 (2018). https://doi.org/10.1186/s40623-018-0847-z

Mende, S.B., et al.: Far ultraviolet imaging from the IMAGE spacecraft. 2, wideband FUV imaging. Space Sci. Rev. **91**, 271–285 (2000)

Mitchell, D.G., et al.: High energy neutral atom (HENA) imager for the image mission. Space Sci. Rev. **91**, 67–112 (2000)

Miyoshi, Y., et al.: Geospace exploration project ERG. Earth Planets Space. **70**, 101 (2018). https://doi.org/10.1186/s40623-018-0862-0

Nicholson, D.R.: Introduction to plasma theory, Wiley (1986), 日本語訳:プラズマ物理の基礎、小笠原正忠・加藤鞆一共訳、丸善(1986)

Nishida, A.: Formation of plasmapause, or magnetospheric plasma knee, by the combined action of magnetospheric convection and plasma escape from the tail. J. Geophys. Res. **71**(23), 5669–5679 (1966). https://doi.org/10.1029/JZ071i023p05669

Oberheide, J., Shiokawa, K., Gurubaran, S., Ward, W.E., Fujiwara, H., Kosch, M.J., Makela, J.J., Takahashi, H.: The geospace response to variable inputs from the lower atmosphere: a review of the progress made by task group 4 of CAWSES-II. Progr. Earth Planet. Sci. **2**, 2 (2015). https://doi.org/10.1186/s40645-014-0031-4

Picone, J.M., Hedin, A.E., Drob, D.P., Aikin, A.C.: NRLMSISE-00 empirical model of the atmosphere: statistical comparisons and scientific issues. J. Geophys. Res. **107**(A12), 1468 (2002). https://doi.org/10.1029/2002JA009430

Ruohoniemi, J.M., Baker, K.B.: Large-scale imaging of high-latitude convection with super dual auroral radar network HF radar observations. J. Geophys. Res. **103**(9), 20797–20811 (1998). https://doi.org/10.1029/98JA01288

Sibeck, D., et al.: Imaging plasma density structures in the soft X-rays generated by solar wind charge exchange with neutrals. Space Sci. Rev. **214**(4), 1–24 (2018). https://doi.org/10.1007/s11214-018-0504-7

Sugiura, M.: Hourly values of equatorial Dst for the IGY. In Annals of the International Geophysical Year, vol. 35, pp. 7–45. Pergamon Press, Oxford (1964)

Weimer, D.R.: Models of high-latitude electric potentials derived with a least error fit of spherical harmonic coefficients. J. Geophys. Res. **100**, 19595–19607 (1995)

World Data Center for Geomagnetism, Kyoto, M. Nose, T. Iyemori, M. Sugiura, T. Kamei, Geomagnetic AE index, https://doi.org/10.17593/15031-54800 (2015a)

World Data Center for Geomagnetism, Kyoto, M. Nose, T. Iyemori, M. Sugiura, T. Kamei, Geomagnetic Dst index, https://doi.org/10.17593/14515-74000 (2015b)

Yokota, S., Kasahara, S., Mitani, T., Asamura, K., Hirahara, M., Takashima, T., Yamamoto, K., Shibano, Y.: Medium-energy particle experiments--ion mass analyzer (MEP-i) onboard ERG (Arase). Earth Planets Space. **69**, 172 (2017). https://doi.org/10.1186/s40623-017-0754-8

Chapter 5
Space Radiation

Yoshizumi Miyoshi, Yuto Katoh, Shinji Saito, Takefumi Mitani, and Takeshi Takashima

5.1 Time Variations in Radiation Belts

The radiation belts are high-energy charged particle populations that are distributed around the Earth. Their entire region ranges from a few hundred kilometers in altitude to the vicinity of the geostationary Earth orbit (GEO). Typical energies range from a few hundred kiloelectron volts (keV) to a few megaelectron volts (MeV) for electrons and from a few hundred keV to a few tens of MeV for ions (Li and Hudson 2019, Kanekal and Miyoshi 2021, for review).

Figure 5.1 shows the typical structure of the radiation belts observed by the Arase satellite (Miyoshi et al. 2018). The typical electron energy is more than several hundred keV. It has been believed that the electrons have a double belt structure, and the outer belt is centered at approximately 4–4.5 times the radius of the Earth. However, the recent Van Allen Probes observations showed that electrons above 1 MeV are present in the outer belt and only very low-flux electrons are found in the inner belt (Li et al. 2015; Fennell et al. 2015). Baker et al. (2014) also pointed out that under typical conditions, the outer belt has an "impenetrable barrier" at approximately $L = 3$ and the MeV electron flux is low at $L < 3$. The electrons in the inner zone are electrons produced by the cosmic ray albedo neutron decay (CRAND)

Y. Miyoshi (✉)
Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Aichi, Japan
e-mail: miyoshi@isee.nagoya-u.ac.jp

Y. Katoh
Graduate School of Science, Tohoku University, Sendai, Miyagi, Japan

S. Saito
National Institute of Information and Communications Technology, Tokyo, Japan

T. Mitani · T. Takashima
Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Sagamihara, Kanagawa, Japan

K. Kusano (ed.), *Solar-Terrestrial Environmental Prediction*,
https://doi.org/10.1007/978-981-19-7765-7_5

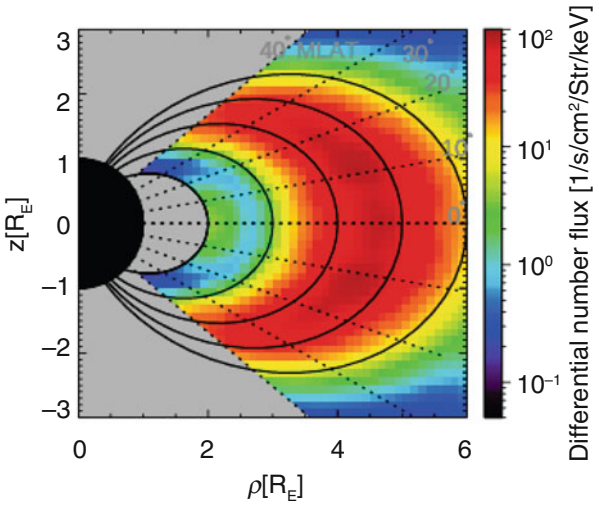

Fig. 5.1 Spatial distribution of radiation belt electrons (2 MeV) observed by the Arase satellite. Colors represent differential flux (/cm^2 s str keV)

(Li et al. 2017) and the radial diffusion from the outer belt into the lower *L*-shell during large magnetic storms (Claudepierre et al. 2019). The majority of the outer belt electrons are expected to originate from the plasma sheet in the magnetosphere, and their flux is highly variable with the solar wind variations, which is discussed subsequently.

The radiation belt ions above a few MeV are stably distributed, whereas ions up to a few hundred keV show large variations during magnetic storms. In addition, when solar energetic particles penetrate the radiation belts, they contribute to radiation belt ions.

5.1.1 Average Variations in Radiation Belt Electrons

Radiation belt electrons vary in their structures and fluxes on various time scales. In this section, we describe the basic properties of several time variations in the outer belt electrons.

5.1.1.1 Transient Variations of Approximately Minutes to 1 h

When the magnetopause of the magnetosphere is changed by interplanetary shocks or abrupt changes in the dynamic pressure, an MHD fast mode wave propagates from the dayside magnetopause to the nightside, which interacts with electrons drifting around the Earth and causes their radial transport. In particular, when the magnetopause is compressed toward the Earth, the fast mode wave propagates with the westward electric field, interacting with electrons drifting eastward, causing their

transport and acceleration toward the Earth (e.g., Foster et al. 2015). Electron injections of more than 10 MeV were observed near $L = 2$ during a large-amplitude storm sudden commencement (SSC) on March 24, 1991 (Li et al. 1993). Although no solar wind was observed at this SSC event, it is believed that the above electron injections were caused by interactions with fast mode waves from the dayside magnetopause with the arrival of a coronal mass ejection (CME).

5.1.1.2 Variations During Substorms

The flux of the outer belt electrons near GEO changes during a substorm. The typical change observed by geostationary orbit satellites is that in the growth phase of a substorm, the flux of MeV electrons initially decreases with the deformation of the background magnetic field and subsequently returns to the original flux level with the substorm onset. These changes are considered to be adiabatic with the change in the background magnetic field. In contrast, for electrons with energies from several tens of keV to 300 keV, a sharp flux enhancement called an injection occurs during a substorm onset. Although the mechanism of this injection has been discussed, recent observations by the Time History of Events and Macroscale Interactions during Substorms (THEMIS) and Magnetosphere Multistage (MMS) satellites suggest that it is caused by electrons associated with a dipolarization flux bundle. The electron flux frequently increases by more than two orders of magnitude with an injection and is considered to be a part of the seed population that forms the outer belt electrons.

5.1.1.3 Changes During Magnetic Storms

During magnetic storms associated with the occurrence of CMEs and corotating interaction regions (CIRs), the flux and structure of the electrons in the outer radiation belt undergo significant changes. Typically, in the main phase of a magnetic storm, the flux in most regions of the outer belt is significantly reduced, resulting in the disappearance of the outer belt. The mechanisms responsible for the disappearance are considered to be as follows: (i) electron energy decrease and electron moving to the outer L value due to the decrease in the background magnetic field occurring in the main phase of the storm (Kim and Chan 1997), (ii) escape from the dayside magnetopause, and (iii) scattering to the atmosphere by the interaction with plasma waves. (ii) and (iii) are discussed in Sect. 2.2.4. In the main phase of a magnetic storm, an increase in the flux frequently occurs in the inner region of the outer belt. Although the mechanism of this increase has not been clarified, it has been suggested that the acceleration is caused by fast mode waves and by the strong electric field in the outer radiation belt.

Subsequently, during the recovery phase of a magnetic storm, the flux of the outer belt increases from the inner L value to outward. Reeves et al. (2003) showed that in approximately 50% magnetic storms, the flux is increased compared to that before their start, whereas in 25% magnetic storms, it is comparatively decreased. In the

remaining 25% magnetic storms, no change occurs in the flux. In addition, Miyoshi and Kataoka (2005) showed that the flux outside the outer belt increases significantly during CIR-driven magnetic storms, whereas that inside it increases significantly during severe CME-driven magnetic storms.

The acceleration in the recovery phase is considered to be a radial diffusion process based on the drift resonance between fast mode waves and drifting electrons (external process, adiabatic acceleration) and also considered to be a cyclotron resonance between electrons and whistler mode waves (internal accelerations, nonadiabatic acceleration). Observations by the Van Allen Probes satellite in the 2010s showed that acceleration by a cyclotron resonance increases the radiation band electrons. Furthermore, Boyd et al. (2018) pointed out that more than 70% outer belt electron-increase events during magnetic storms are internal accelerations. These acceleration mechanisms are not exclusive, and different mechanisms can be triggered simultaneously. In addition, it is considered that the mechanisms may change depending on the phase of the magnetic storm.

5.1.1.4 Semiannual and Solar Cycle Variations

It is known that the activity of the magnetosphere of the Earth increases around the spring and fall equinoxes owing to the Russell–McPherron effect (Russell and McPherron 1973). Semiannual variations also occur in the electron flux in the extraterrestrial radiation belt, which increase in spring and autumn (Baker et al. 1999). Moreover, it is known that a recurrent flux increase occurs in association with CIR magnetic storms induced by recurrent coronal holes (Miyoshi et al. 2004).

Regarding the dependence on the solar activity, the flux in the outer zone, such as in GEO, decreases near the solar activity maximum and subsequently increases during the solar activity declining phase. Miyoshi et al. (2004) showed that the position of the center of the outer belt changes with the solar activity and that the outer belt moves closer to or further away from the Earth depending on the solar activity. Miyoshi et al. (2004) and Miyoshi and Kataoka (2011) proposed a model in which the outer belt moves closer or further away from the Earth depending on the solar activity, resulting in a solar cycle phase-dependent change in the L value. In contrast, in a low-altitude region close to the Earth, changes in the solar activity cycle are triggered by changes in the upper atmosphere and CRAND-inducing cosmic rays (Xiang et al. 2019).

5.2 Solar Wind–Radiation Belt Coupling

It is well known that variations in the radiation belt electrons, particularly in the outer belt, are frequently associated with changes in the solar wind. Therefore, studies have been conducted in terms of which parameters of the solar wind are related to the

above changes and how they depend on large-scale structures, CIRs, and CMEs, which cause magnetic storms.

5.2.1 Solar Wind Parameter Dependence

Paulikas and Blake (1979) pointed out that there is a strong correlation between the increase in the electrons in the radiation belts and the change in the solar wind velocity, and this trend has been confirmed by many studies since then. In addition, a linear predictive filter algorithm (Baker et al. 1990a, b) and a neural network have been developed using the solar wind velocity as the input, which demonstrates the forecast of the radiation belt flux variations. Figure 5.2 shows the flux variations of the radiation belt electrons observed by the Arase satellite and the solar wind speed. The figure shows that the arrival of fast winds originating from CMEs and coronal hole flows causes an increase in the radiation belts. However, it also shows that the flux increase does not always occur when fast solar winds arrive. Reeves et al. (2011) analyzed long-term data on solar wind speed and the radiation belt electron flux and showed that their relationship is not a simple correlation. In fact, for the same solar wind speed, the values of the radiation belt electron flux are distributed over several orders of magnitude. It has also been pointed out by Miyoshi and Kataoka (2005) and Li et al. (2011) that the radiation belt electron flux does not always increase even when the solar wind speed is high and that the solar wind speed is a sufficient but not a necessary condition for the radiation belt increase.

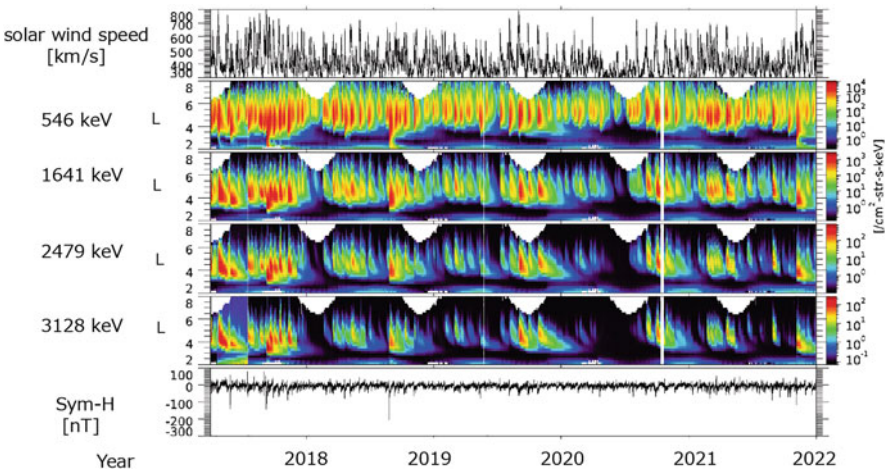

Fig. 5.2 L-time diagram of radiation belt electrons observed by Arase satellite for each energy. From top to bottom, solar wind velocity, 550 keV electrons, 1350 keV electrons, 2500 keV electrons, 3900 keV electrons, and the Sym-H index. Colors represent electron flux (/cm^2 s str keV)

In contrast, with respect to the dependence on the interplanetary magnetic field (IMF), Miyoshi and Kataoka (2008a) focused on the fact that the average north–south IMF in high-speed solar wind depends on the Russell–McPherron effect (Russell and McPherron 1973) and found that by considering this effect, it is possible to discriminate between the southward IMF dominant coronal hole and the northward IMF dominant coronal hole. The statistical results showed that a significant increase in the radiation belt electron flux occurs associated with the southward IMF dominant coronal hole streams. Miyoshi et al. (2013, 2016) applied same analysis about IMF dependence in coronal hole streams to whistler mode waves, the plasmasphere thermal electron density, and high-temperature plasma sheet electrons. They found that the key parameters of the acceleration by a cyclotron resonance, in particular, depend on the IMF Bz, and these key parameters enhance largely during the southward IMF dominant coronal hole streams.

When the average of Alfvénic fluctuation of IMF embedded in the fast stream is the southward, HILDCAA (high-intensity, long-duration, continuous AE activity; Tsurutani and Gonzalez 1988) is occured. Many of the key parameters involved in a cyclotron resonance are enhanced during substorms, whose relatively long durations are considered to be important for the increase in the radiation belt electrons. The importance of substorms persisting over several days for the radiation belt increase was also pointed out by the Van Allen Probes satellite observations (Jaynes et al. 2015).

5.2.2 Solar Wind Structure Dependence

In the CME-driven magnetic storm, the main phase is triggered by the southward IMF in the CME sheath or magnetic cloud, and the magnetic storm gradually recovers after the passage of the magnetic cloud. In the CIR-driven magnetic storm, the main phase is driven by the southward IMF in the stream interface between the slow stream and the fast coronal hole stream. Moreover, the Alfvénic fluctuations in the coronal hole stream cause small to medium geomagnetic activities that persist for several days or more, with successive substorms.

During magnetic storms driven by CME and CIR, the radiation belts show drastically different variations. Electrons in the outer radiation belt increase significantly during the recovery phase of the CIR-driven magnetic storm, i.e., the period when the fast coronal hole flow arrives. However, this increase is relatively small in the recovery phase of the CME-driven magnetic storm (Miyoshi and Kataoka 2005). The increase in the radiation belt electron flux in the outer belt, including GEO, does not depend on the magnetic storm size. Note that such an increase also depends on the L value (Miyoshi and Kataoka 2005; Turner et al. 2019): in CIR-driven magnetic storms, the electron flux in the outer belt ($L > 4$) increases significantly, whereas in large CME-driven magnetic storms, it increases on approaching the Earth, such as in the inner part of the outer belt and in the slot region ($L \sim 3$).

5.3 Acceleration Mechanisms

Acceleration mechanisms of the radiation belt electrons are categorized into adiabatic and nonadiabatic processes.

5.3.1 Adiabatic Acceleration Process

An "adiabatic acceleration" process is based on the conservation of the adiabatic invariants associated with the radial transport of electrons from the plasma sheet to the radiation belts. The first adiabatic invariant is defined as the ratio of the perpendicular component of the kinetic energy to the background magnetic field strength. Considering a dipole field for the intrinsic magnetic field of the Earth, the background magnetic field strength depends on the distance from the Earth. Therefore, electrons transported from the plasma sheet to the inner magnetosphere experience an increase in the background magnetic field strength during their radial transport. The conservation of the first adiabatic invariant in the transport process increases the kinetic energy perpendicular to the background magnetic field.

The main driver of the radial transport of plasma is the convection electric field in the magnetosphere. Solar wind–magnetosphere interactions induce the magnetic reconnection in dayside magnetopause, and the electric field in the solar wind induces the convection electric field in the magnetosphere. The convection electric field directs from dawn to dusk in the magnetic equatorial plane, and the direction of the $\mathbf{E} \times \mathbf{B}$ drift becomes earthward. As the speed of the solar wind increases, the magnitude of the convection electric field increases, thereby increasing the magnetospheric convection.

The magnitude of the convection electric field is determined by solar wind parameters such as the solar wind speed and the IMF, and the convection electric field is not essential for the radial transport for high-energy charged particles because magnetic gradient and curvature drifts of these particle are so strong. In contrast, MHD waves (fast mode waves) in the magnetosphere cause a drift resonance between charged particles. Consequently, high-energy charged particles are transported inward through the drift resonance. This process is called "radial diffusion" and is considered to cause radial transport of the radiation belt electrons associated with an adiabatic energy change.

As described in Sect. 2.2.1, a rapid inward transportation, called an injection, of electrons and ions in the energy range from a few keV to hundreds of keV is observed after the substorm onset, which leads to a rapid enhancement of high-energy particles from the plasma sheet to the inner magnetosphere. Although an injection plays an important role in the increase in the ring current ion and electron fluxes, its effect on the increase in the radiation belt electron flux has not been clarified quantitatively. However, it has been commonly understood that an injection

plays an important role in the excitation of plasma waves such as whistler mode chorus emissions, as described below.

5.3.2 *Nonadiabatic Acceleration Process*

Adiabatic invariants are defined as quantities that are conserved over a time scale corresponding to the typical period of the motion of charged particles. The first, second, and third adiabatic invariants are related to the gyration of charged particles, bounce motion between the mirror points of the Northern and Southern Hemispheres, and period of the drift motion around the Earth, respectively. A "nonadiabatic" process is a process in which the periodic motion of charged particles is disturbed by electromagnetic field fluctuations on a time scale shorter than the period for which an adiabatic invariant is defined, and it results in significant energy changes.

As a physical process of a nonadiabatic acceleration of the radiation belt electrons, the resonant interaction with whistler mode plasma waves has been considered (Summers et al. 1998; Miyoshi et al. 2003; Horne et al. 2005). Whistler mode waves, which appear in the wave electromagnetic field spectra in the frequency range lower than the electron cyclotron frequency (f_{ce}), satisfy the cyclotron resonance condition with electrons in a wide energy range from keV to MeV. Whistler mode chorus emissions outside the plasmapause from the nightside to the morning side play an important role. Chorus emissions are observed in the frequency range from 0.2 to 0.8 f_{ce}, and a gap at 0.5 f_{ce} divides chorus emissions into lower- and upper-band emissions. Chorus emissions are generated by an instability driven by the temperature anisotropy of energetic electrons and subsequent nonlinear wave–particle interaction. The temperature anisotropy is developed during the adiabatic transport of hot electrons from the plasma sheet to the inner magnetosphere by injection and the convection electric field, as described above.

Chorus emissions are a group of coherent wave elements characterized by their spectral fine structure showing a frequency change with time; rising tones are typical, and falling tones or hooks are also observed. Theoretical and simulation studies of nonlinear wave–particle interactions in nonadiabatic processes have made significant progress since the 2000s. Omura et al. (2008, 2009) theoretically clarified the physical processes of the generation mechanism of rising-tone chorus emissions. For details of the theory of the chorus generation process, see Omura (2021). Omura and Summers (2006) showed that a part of resonant electrons is trapped by waves, which significantly changes their energies. Nonlinear effects in nonadiabatic acceleration processes have been presented by test particle simulations precisely computing trajectories of resonant electrons moving along a magnetic field line under the presence of a coherent whistler mode wave. Relativistic turning acceleration (RTA; Omura et al. 2007) and ultra-relativistic turning acceleration (URA; Summers and Omura 2007) have been proposed. In RTA, the direction of motion of accelerated electrons changes during the acceleration process, and in URA, electrons bouncing

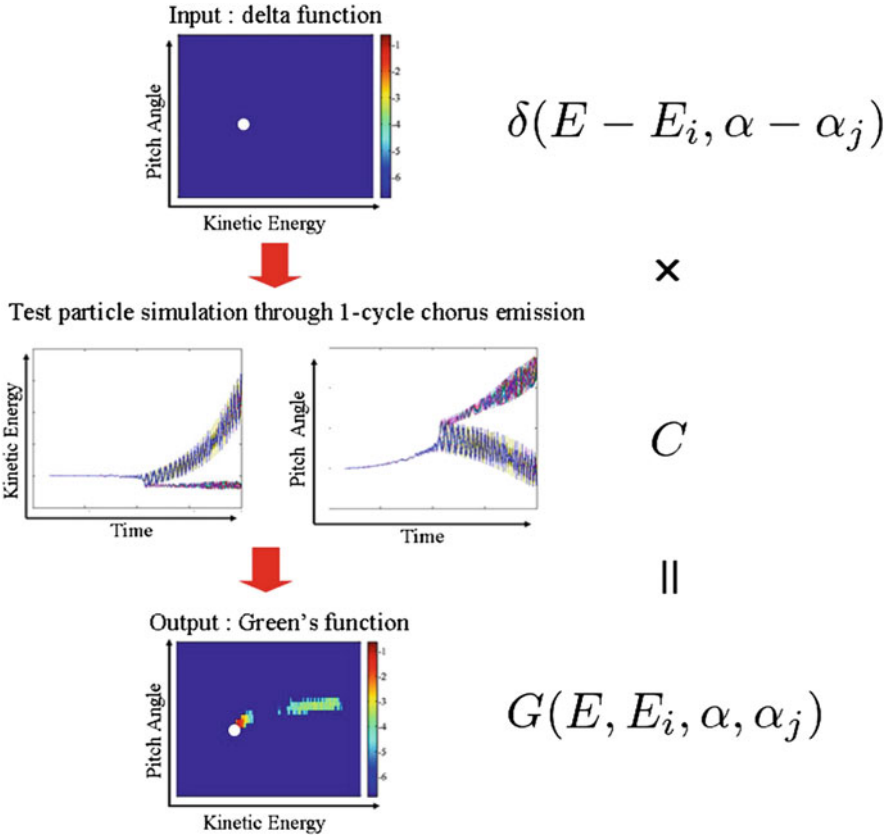

Fig. 5.3 Conceptual diagram of numerical Green's function method (Omura et al. 2015). Test particles with initial values in specific energy and pitch angle ranges are placed in simulation space. Test particle analysis of scattering process by chorus radiation propagating along magnetic field lines is performed to obtain final energy and pitch angle distributions. Initial value of test particle is treated as delta function, and final distribution is used to numerically construct Green's function as response to one cycle of interaction with chorus radiation. Using constructed Green's function, we obtain change in energy and pitch angle distributions of energetic electrons that occur on time scale of few minutes as consequence of interaction corresponding to many cycles

between their mirror points are trapped multiple times by waves and accelerated continuously. By performing numerous test particle analyses for electrons in a wide range of energies and pitch angles, Omura et al. (2015) proposed a numerical Green's function method to construct a response function from the distribution function of high-energy electrons based on RTA and URA. Acceleration of electrons from keV to MeV energy range on a time scale of several minutes has been demonstrated using the numerical Green's function method (Fig. 5.3). The above numerical Green's function method can be regarded as a solution to the unsolved difficult problem of describing the long-term evolution of the distribution function of the radiation belt electrons while considering the nonlinear effects in the time scales

of milliseconds or less. To understand when and how a nonadiabatic acceleration process occurs, it is necessary to develop a method to quantitatively evaluate the conditions for plasma wave generation in the inner magnetosphere. The linear theory is extensively used to obtain the wave growth rate from the distribution function of energetic electrons, which are the energy source of plasma wave generation. In contrast, the chorus generation process is fully nonlinear and, therefore, requires self-consistent or first-principles simulations. Katoh and Omura (2016) reproduced the generation process of chorus emissions using a realistic spatial distribution of the background magnetic field intensity at $L = 4$. Saito et al. (2016) conducted a test particle simulation to reproduce the acceleration of high-energy electrons by chorus emissions and showed that the effect of nonlinear wave trapping on the relativistic electron energy flux is significant (Fig. 5.4). Katoh et al. (2018b) presented the effects of temperature anisotropy and number density of energetic electrons on the conditions for chorus generation by survey calculations using a first-principles electron hybrid code.

Miyoshi et al. (2018) proposed that cross-energy coupling, in which plasma particles of different energy levels ranging from electron volts to MeV are dynamically coupled by wave–particle interactions, is essential for the acceleration of the radiation belt electrons. This is because the energy range responsible for chorus generation is different from that for those accelerated by chorus emissions. Moreover, the background thermal plasma also plays a role in changing the properties of plasma waves. Fig. 5.5 shows a diagram representing cross-energy coupling (Miyoshi et al. 2018), where the horizontal axis represents the distance from the Earth (L value) and the vertical axis is for the kinetic energy. In a collisionless plasma system, particles with different energies interact via plasma waves. Wave–particle interaction is important not only for electron acceleration but also for scattering and ion/electron coupling processes. Cross-energy coupling is a key concept for understanding the dynamics of the inner magnetosphere.

There have also been attempts to directly measure the nonlinear wave–particle interactions between chorus emissions and energetic electrons. A software-type wave–particle interaction analyzer (WPIA) on board the Arase satellite has been used to directly measure the amount of energy exchange between chorus emissions and energetic electrons (Katoh et al. 2018a). Shoji et al. (2017) and Kitamura et al. (2018) demonstrated the measurement of energy exchange between electromagnetic ion cyclotron (EMIC) waves and high-energy ions using the WPIA method. Furthermore, Kitahara and Katoh (2016) proposed a method to measure the momentum change of energetic particles using the WPIA method to directly measure the pitch angle scattering process caused by waves. Kurita et al. (2018a) revealed that the distribution function of 17–30 keV electrons fluctuates with upper-band chorus emissions on a short time scale of less than 30 s, which cannot be explained by the quasi-linear theory. Saito et al. (2021) performed data-driven simulations of the events reported by Kurita et al. (2018a) and demonstrated that a nonlinear process called phase trapping causes acceleration.

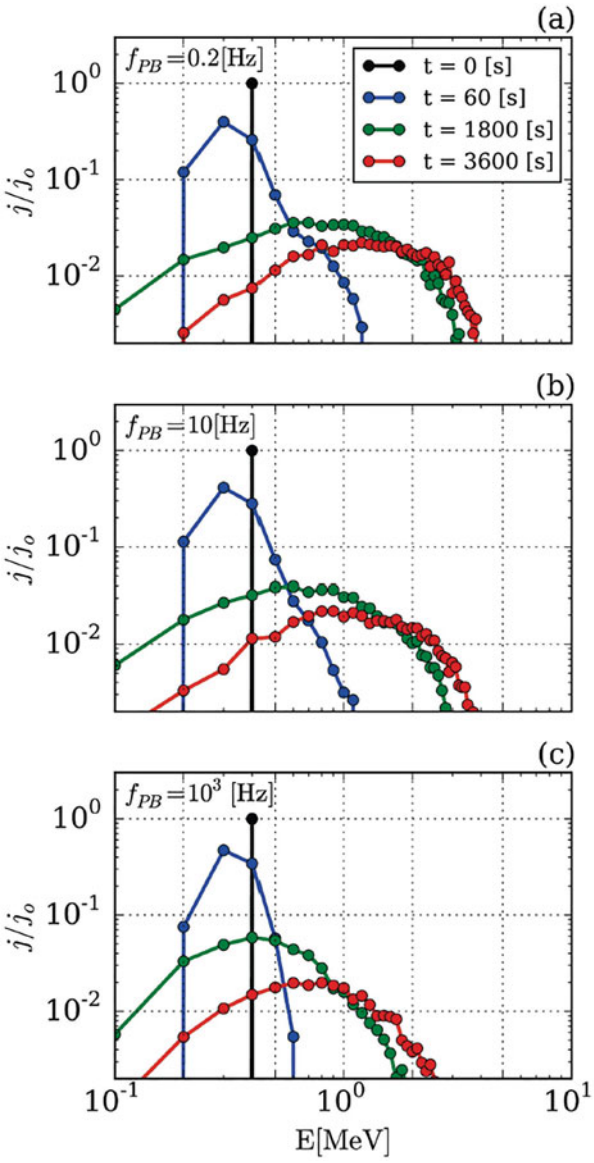

Fig. 5.4 Different changes in energy spectrum of relativistic electrons under conditions in which wave phase is modulated at frequencies of (**a**) 0.2 Hz, (**b**) 10 Hz, and (**c**) 1000 Hz, corresponding to artificial weakening of nonlinear effect in order (Result in (**a**) shows increase in the flux to highest energy, indicating that wave trapping has significant effect on amount of increase in energy flux of relativistic electrons (Saito et al. 2016))

5.4 Loss Mechanisms

5.4.1 Magnetopause Boundary Loss

Magnetopause shadowing (MPS) is a mechanism for the loss of the radiation belt electrons. The basic concept is that electrons in the outer radiation belt escape from

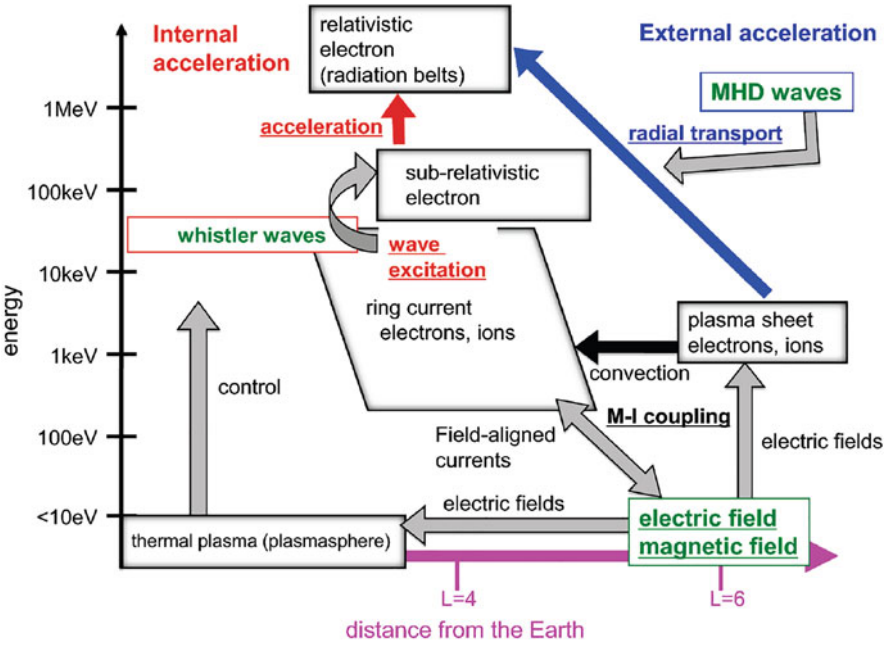

Fig. 5.5 Conceptual diagram of radiation belt electron acceleration via cross-energy coupling. Horizontal axis: distance from the Earth; vertical axis, energy. Blue line shows adiabatic acceleration process due to drift resonance between MHD wave and charged particles, and red line represents nonadiabatic acceleration via cyclotron resonance. Energy band of electrons responsible for excitation of chorus radiation is from several keV to several tens of keV, and energy of accelerated electrons, which are accelerated by excited chorus radiation, is several hundred keV, resulting in formation of relativistic electrons. Moreover, thermal plasmas below few eV also change characteristics of wave–particle interactions as an ambient media. This process of dynamical coupling between plasma particles of different energy hierarchies is called cross-energy coupling (Miyoshi et al. 2018)

the magnetosphere through the magnetopause boundary. Here, the magnetopause boundary is the boundary between the interplanetary space and the magnetosphere. The drift orbit of electrons to reach the magnetopause boundary is mainly controlled by the drift motion, which is perpendicular to the magnetic field vector and its intensity gradient. When an electron with no velocity parallel to the magnetic field lines (pitch angle of 90°) drifts in the magnetosphere of the Earth, it moves along the contour line of the magnetic field intensity on the magnetic equatorial plane. An electron moving on the contour line can flow out to the interplanetary space when the contour line touches the magnetopause boundary, which is the MPS loss process. The magnetic field configuration in the magnetosphere of the Earth depends on the solar wind environment, such as the solar wind dynamic pressure and the IMF. Figure 5.5 shows the magnetic field intensity and its contour lines on the magnetic equatorial plane under different solar wind conditions. The left and right figures show the equatorial magnetic field under slow and fast solar winds, respectively. The

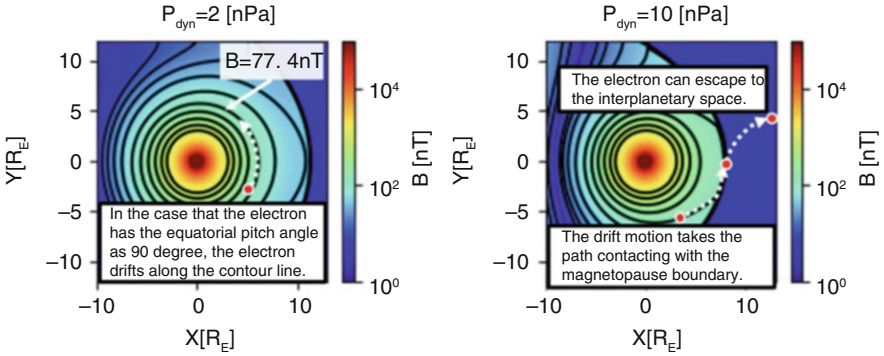

Fig. 5.6 Magnetic field strength in equatorial plane calculated from Tsyganenko model (TS05: Tsyganenko and Sitnov 2005) for solar wind dynamic pressure of (left) 2 [nPa] and (right) 10 [nPa]. Black lines represent contours of magnetic field intensity. Magnetic axis is almost perpendicular to *xy* plane (*z* direction), *x*-axis is from the Earth to Sun, and *y*-axis is perpendicular to both x and z. Southward magnetic field of solar wind is assumed to be 10 nT, and all W parameters are set as zero for simplicity

dense contour lines on the right side of both figures indicate the magnetopause boundary. From the comparison of the two figures, a high dynamic pressure of the solar wind implies a strong magnetic field intensity near the magnetopause boundary. As an example, red dots are placed on the contour lines of the same intensity (77.4 nT) in both figures. When the solar wind dynamic pressure is low (left), the contours (equivalent to the drift orbit) are almost concentrically closed in the magnetosphere. In contrast, when the solar wind dynamic pressure is high (right), the contour line of the same intensity touches the magnetopause boundary on the dayside. Subsequently, electrons in this orbit can flow out to the interplanetary space. The dependence of this flowing out on the solar wind dynamic pressure was investigated by Ohtani et al. (2009), who pointed out that relativistic electrons in GEO tend to be lost when the dynamic pressure is high. This suggests that MPS may dominantly occur. However, Green et al. (2004) argued that the contribution of MPS to the loss of the radiation belt electrons is small because the flux depletion occurs quite inside the magnetopause boundary. It has been pointed out that the loss of the radiation belt electrons inside the magnetopause boundary is related to radial diffusion (electron transport in the *L* direction) (Miyoshi et al. 2003; Turner et al. 2012).

The loss scenario via radial diffusion is as follows: (1) The high dynamic pressure of the solar wind caused by a CME and a CIR compresses the magnetosphere of the Earth, and the magnetopause boundary moves inward (changes from left to right in Fig. 5.6). (2) The electrons near the boundary flow out into interplanetary space, and a sharp density gradient of the radiation belt electrons appears close to the boundary. (3) When the dynamic pressure of the solar wind decreases and the magnetopause boundary expands (change from the right to the left panel in Fig. 5.6), the steep gradient of the spatial density formed by MPS relaxes through radial diffusion.

Consequently, the radiation belt electrons are transported from the inside to the outside. Turner et al. (2014) analyzed the loss event observed at the end of September 2012 and demonstrated this loss scenario based on simultaneous observations by multiple satellites. The observations showed that as the model-estimated maximum L (last closed drift shell) moves inward, the electron phase space density of the outside portion of the outer belt is reduced. Subsequently, as the maximum L moves back outward, the phase space density expands from small to large L, and the phase space density inside is decreased by the electron transport by radial diffusion (e.g., Selesnick and Blake 2000; Miyoshi et al. 2003). This series of processes has been reproduced by numerical simulations (Hudson et al. 2014), and thus, MPS is considered to play an important role as a mechanism for the loss of the electrons in the outer radiation belt. Although an actual outflow of high-energy electrons escaping from the magnetosphere has not been observed, recent MMS satellite observations suggest that high-energy electrons flow into the magnetosheath region (Cohen et al. 2016, 2017). The physical mechanism of the radiation belt loss induced by MPS is being clarified.

5.4.2 Loss Due to Wave–Particle Interactions

The wave–particle interaction with EMIC waves and whistler mode chorus waves has been pointed out as a physical process responsible for the loss of high-energy electrons in the outer radiation belt.

Because EMIC waves satisfy the cyclotron resonance condition with electrons having several hundred keV to several MeV of energy (e.g., Miyoshi et al. 2008), their contribution to the loss of high-energy electrons has been discussed. Omura and Zhao (2012) pointed out the importance of the nonlinear trapping of relativistic electrons by coherent EMIC waves and showed that the pitch angle scattering of high-energy electrons is highly efficient. Omura and Zhao (2013) showed that MeV electrons with an initial pitch angle of more than 60° undergo pitch angle scattering near the loss cone on a time scale of less than 1 s by coherent EMIC waves. In addition, Kubota et al. (2015) showed that the most efficient pitch angle scattering occurs in the energy range of 1.1–3.0 MeV and that 50% electrons can be lost in approximately 2 min. Nakamura et al. (2019) found from Van Allen Probes satellite observations that the MeV electron flux decreases by more than 30% in 1 min with the onset of an EMIC wave. Ground-based observations were also conducted to demonstrate the relationship between relativistic electron loss and EMIC waves. Hirai et al. (2018) showed a good correlation between relativistic electron precipitation detected by VLF/LF radio waves with frequencies between 10 and 100 kHz and EMIC waves observed from induction magnetometers. By an integrated study using observations by the Arase satellite, the Van Allen Probes satellite, and a ground-based observation network, Kurita et al. (2018c) showed that the MeV electron flux decreases significantly on a time scale of several tens of minutes in association with the EMIC wave activity. They succeeded in constraining the time

and spatial scales of the MeV electron loss process from ground-based observations of EMIC waves and observations of adjacent regions by multiple satellites at different timings.

Regarding the role of whistler mode chorus emissions for the loss process of high-energy electrons, it has been pointed out that they are related to microbursts. In these phenomena, electrons from several hundred keV to several MeV are intermittently precipitated into the Earth's atmosphere on time scale of less than 1 s. Saito et al. (2012) performed test particle simulations of the pitch angle scattering of high-energy electrons by whistler mode chorus waves propagating along a magnetic field line. They showed that relativistic electrons are scattered by the chorus waves at high latitudes and precipitated into the atmosphere and that the time scale of the intermittent precipitation is related to the spectral fine structure of the chorus emission. In contrast, Kurita et al. (2018b) focused on the fact that microbursts are frequently observed during the period when fast solar wind reaches the magnetosphere of the Earth. They statistically analyzed observations of the SAMPEX and POES satellites and found that the high occurrence frequency of microbursts corresponds to an increase in the flux of the radiation belt electrons during the period of fast solar wind. These results are good examples to show that a whistler mode chorus emission is deeply related to both the acceleration and loss of the radiation belt electrons.

In addition, Miyoshi et al. (2015, 2021) analyzed observation data obtained from European Incoherent Scatter (EISCAT) radars and showed that the electron density increased up to an altitude below 70 km associated with a pulsating aurora through conjugate observations with Van Allen Probes and Arase. Numerical simulations concluded that precipitation of sub-relativistic and relativistic electrons causes such enhancements. It was considered that the energetic electron precipitation was caused by whistler mode chorus waves, which were simultaneously observed by the satellites. Furthermore, Miyoshi et al. (2020) proposed a theory that relativistic electron microbursts have the same origin as pulsating auroral electrons. They have been advancing research aiming at observational verification. A pulsating aurora is believed to be produced by electrons with energies of keV to a few tens of keV. These electrons are scattered in the pitch angle by whistler mode chorus waves and are precipitated into the atmosphere of the Earth. Based on the results of measurements by the Arase satellite, Kasahara et al. (2018) showed for the first time that the modulation of the electron flux in a loss cone highly correlates with the amplitude modulation of the whistler chorus emission. Kitahara and Katoh (2019) revealed the existence of a nonlinear interaction between coherent whistler mode waves and energetic electrons in a loss cone, in which the pitch angle and energy of electrons nonlinearly trapped by the waves are significantly increased. These results suggest the importance of nonlinear scattering processes in addition to the conventional quasi-linear theory. To understand wave–particle interactions that are important for electron acceleration in the radiation belt, further investigation of the elementary processes involved in nonlinear physics is required.

5.5 Forecast of Radiation Belt Electrons

Modern society uses satellite-based information technology in various fields such as positioning systems using the Global Positioning System satellites and weather observations using meteorological satellites. The orbits of such satellites are in the magnetosphere of the Earth, and many of them are in low-Earth orbits (LEOs) with an altitude of less than 1000 km and GEO with an altitude of approximately 35,000 km. Satellites in LEOs and GEOs are exposed to electrons with energies ranging from a few keV to MeV. In such an environment, satellites are electrically charged by the electrons and are known to produce electrostatic discharges (ESDs), which may cause serious damage to electrical and electronic equipment. To avoid anomalies related to ESDs, it is necessary to develop satellites with robust structures and mechanisms for charging and to operate them in such a manner that fatal enhancement in the radiation belt electron flux can be predicted in advance and failures avoided. Space-based information technology provides us with conveniences for our daily life, and its disruption would have a significant impact on society. Understanding and forecasting the acceleration and loss mechanisms of the radiation belt electrons in the magnetosphere of the Earth play important roles in protecting our social infrastructure.

Several institutes/organizations have developed and operated models for predicting the electron flux in a GEO. The National Institute of Information and Communications Technology (NICT), which is in charge of space weather operations in Japan, has been operating a model to forecast electron fluence above 2 MeV (https://aer-nc-web.nict.go.jp/radi/). This forecast is calculated using a multivariate autoregressive model and a Kalman filter from the solar wind environment data available in real time (Sakaguchi et al. 2013, 2015). The NICT forecasts the electron fluence divided into four levels for the following 1–3 days. Similarly, the Space Weather Prediction Center of the National Oceanic and Atmospheric Administration (NOAA) has operated the relativistic electron forecast model (https://www.swpc.noaa.gov/products/relativistic-electron-forecast-model) using the solar wind data of the past 30 days. In contrast, a group at the University of Colorado, USA, has forecasted variability in the radiation belt electron flux from a quasi-linear diffusion model (http://lasp.colorado.edu/space_weather/xlf3/xlf3.html). This method predicts the flux variations in a GEO by estimating the radial diffusion coefficient from the solar wind data. In Japan, ISEE, Nagoya University, operates a numerical space weather model system (Space-weather-forecast-Usable System Anchored by Numerical Operations and Observations (SUSANOO), http://cidas.isee.nagoya-u.ac.jp/susanoo/index.html), which uses a code coupling of solar wind simulation (Shiota et al. 2014) and the radial diffusion simulation (Miyoshi et al. 2004). Here, the hazard level is evaluated in terms of the probability that the electron flux in the GEO exceeds a certain threshold based on Miyoshi and Kataoka (2008b). The major difference between the forecast model operating at Nagoya University and other models is that it can forecast up to 1 week ahead using the solar wind data predicted

by SUSANOO, whereas other forecast models use real-time solar wind observation data.

Typical forecast methods are mainly based on the solar wind data and radiation belt electron flux observed in a GEO. In addition, a forecasting model of the radiation belt electron variability using satellite data in an LEO has recently been proposed (Chen et al. 2016). When magnetic storms occur, the radiation belt electrons may be scattered by whistler mode waves and ion cyclotron waves and may be precipitated into an LEO as well as the magnetopause shadowing. Precipitation is generally considered to be the cause of radiation belt loss. However, in fact, electron acceleration and loss occur simultaneously. The balance between them needs to be considered to determine the variations in the radiation belt electron flux (e.g., Reeves et al. 2003). Recent observations have indicated that the flux of high-energy electrons tends to increase when relativistic electron microbursts (abrupt relativistic electron precipitations) occur frequently (Kurita et al. 2016). A model proposed by Chen et al. (2016) predicts how much an energetic electron flux will increase in the future by observing variations in the flux of high-energy electrons reaching an LEO. Subsequently, they improved this model to predict for the entire region of the outer radiation belt (Chen et al. 2019) using satellite data from LEOs and GEOs. In addition, an attempt to realize prediction by machine learning based on this model has been started (Pires de Lima et al. 2020).

Most of the radiation belt prediction models currently in use are empirical models based on correlations between observed physical quantities or physical models based on quasi-linear diffusion models (e.g., Li et al. 2001). In the future, it is expected that these models will be further developed to include prediction models based on machine learning and artificial intelligence-based methods, as in Pires de Lima et al. (2020). Although empirical models are effective in tracking phenomena that have been "experienced" frequently, it is difficult to forecast changes that rarely occur. Specifically, it is difficult to reproduce or forecast extreme events. In addition, a quasi-linear diffusion model cannot adequately reproduce nonlinear processes, which might be significant contribution for extreme events. Recently, it has been pointed out that a nonlinear scattering process plays an important role in the rapid increase in the radiation belt electron flux as shown in this chapter. The appropriate incorporation of nonlinear processes is one of the important issues for accurate prediction including extreme events.

To overcome this problem, it is important to reproduce the variations in energetic electron fluxes, including the nonlinear acceleration and loss mechanisms, by numerical simulations demonstrating actual physical phenomena. The motion of the radiation belt electrons, particularly high-energy electrons (>MeV) called killer electrons, can be described as test particles (e.g., Saito et al. 2010), because their contribution to the electromagnetic field is much smaller than that of background plasmas. Here, a test particle model is a model in which the electric current produced by charged particles is not fed back to the electromagnetic field. The behavior of the radiation belt electrons can be treated by modeling the magnetospheric electromagnetic field and the plasma waves (e.g., Saito et al. 2012). Reproduction of such an electron scattering by electromagnetic waves in the entire magnetosphere is one of

the methods to reproduce the temporal variations in the radiation belt electron flux, including nonlinear scattering.

The magnetosphere of the Earth is not a closed physical system but governed by the external boundary conditions of the solar wind, CMEs, and CIRs. These have significant impacts on the physical system of the magnetosphere (e.g., storms, substorms, etc.). Storms continue more than a day from their arrival to reach the peak of the radiation belt electron flux enhancement (Baker et al. 1994). Therefore, a realistic scenario for the forecast of the radiation belt electron flux is to identify events that cause disturbances in the magnetosphere, such as CMEs and CIRs, a few days ahead (Miyoshi and Kataoka 2005, 2008b). To forecast the increase in radiation belt electron flux with a longer lead time, either empirical assumption of the solar wind or a solar wind forecast model such as SUSANOO is required. Particularly, because the radiation belt electrons are strongly affected by the southward IMF, the forecast of the IMF is important (Miyoshi and Kataoka 2008a; Miyoshi et al. 2013, 2016). Improving the forecast accuracy for the IMF is one of the most important issues to improve the forecast of the radiation belt electron. To forecast long-term variability over several days, it is necessary not only to improve the accuracy of radiation belt physics models but also of solar wind forecast models.

References

Baker, D.N., McPherron, R.L., Cayton, T.E., Klebesadel, R.W.: Linear prediction filter analysis of relativistic electron properties at 6.6 RE. J. Geophys. Res. **95**(A9), 15133–15140 (1990a). https://doi.org/10.1029/JA095iA09p15133

Baker, D.N., McPherron, R.L., Cayton, T.E., Klebesadel, R.W.: Linear prediction filter analysis of relativistic electron properties at 6.6 RE. J. Geophys. Res. Space Phys. **95**(A9), 15133–15140 (1990b). https://doi.org/10.1029/JA095iA09p15133

Baker, D.N., Blake, J.B., Callis, L.B., Cummings, J.R., Hovestadt, D., Kanekal, S., Klecker, B., Mewaldt, R.A., Zwickl, R.D.: Relativistic electron acceleration and decay time scales in the inner and outer radiation belts: SAMPEX. Geophys. Res. Lett. **21**(6), 409–412 (1994)

Baker, D.N., Kanekal, S.G., Pulkkinen, T.I., Blake, J.B.: Equinoctial and solstitial averages of magnetospheric relativistic electrons: a strong semiannual modulation. Geophys. Res. Lett. **26**, 3193–3196 (1999)

Baker, D., Jaynes, A., Hoxie, V., et al.: An impenetrable barrier to ultra relativistic electrons in the Van Allen radiation belts. Nature. **515**, 531–534 (2014). https://doi.org/10.1038/nature13956

Boyd, A.J., Turner, D.L., Reeves, G.D., Spence, H.E., Baker, D.N., Blake, J.B.: What causes radiation belt enhancements: a survey of the Van Allen probes era. Geophys. Res. Lett. **45**, 5253–5259 (2018). https://doi.org/10.1029/2018GL077699

Chen, Y., Reeves, G.D., Cunningham, G.S., Redmon, R.J., Hendersion, M.G.: Forecasting and remote sensing outer belt relativistic electrons form low earth orbit. Geophys. Res. Lett. **43**, 1031–1038 (2016). https://doi.org/10.1002/2015GL067481

Chen, Y., Reeves, G.D., Fu, X., Henderson, M.: PreMevE: new predictive model for megaelectron-volt electrons inside Earth's outer radiation belt. Space Weather. **17**, 438–454 (2019). https://doi.org/10.1029/2018SW002095

Claudepierre, S.G., O'Brien, T.P., Looper, M.D., Blake, J.B., Fennell, J.F., Roeder, J.L., et al.: A revised look at relativistic electrons in the Earth's inner radiation zone and slot

region. J. Geophys. Res. Space Physics. **124**, 934–951 (2019). https://doi.org/10.1029/2018JA026349

Cohen, I.J., et al.: Observations of energetic particle escape at the magnetopause: early results from the MMS energetic ion spectrometer (EIS). Geophys. Res. Lett. **43**, 5960–5968 (2016). https://doi.org/10.1002/2016GL068689

Cohen, I.J., et al.: Statistical analysis of MMS observations of energetic electron escape observed at/beyond the dayside magnetopause. J. Geophys. Res. Space Physics. **122**, 9440–9463 (2017). https://doi.org/10.1002/2017JA024401

Fennell, J.F., Claudepierre, S.G., Blake, J.B., O'Brien, T.P., Clemmons, J.H., Baker, D.N., Spence, H.E., Reeves, G.D.: Van Allen probes show the inner radiation zone contains no MeV electrons: ECT/MagEIS data. Geophys. Res. Lett. **42**(5), 1283–1289 (2015). https://doi.org/10.1002/2014GL062874

Foster, J.C., Wygant, J.R., Hudson, M.K., Boyd, A.J., Baker, D.N., Erickson, P.J., Spence, H.E.: Shock-induced prompt relativistic electron acceleration in the inner magnetosphere. J. Geophys. Res. Space Physics. **120**, 1661–1674 (2015). https://doi.org/10.1002/2014JA020642

Green, J.C., Onsager, T.G., O'Brien, T.P., Baker, D.N.: Testing loss mechanisms capable of rapidly depleting relativistic electron flux in the Earth's outer radiation belt. J. Geophys. Res. **109**, A12211 (2004). https://doi.org/10.1029/2004JA010579

Hirai, A., Tsuchiya, F., Obara, T., Kasaba, Y., Katoh, Y., Misawa, H., Shiokawa, K., Miyoshi, Y., Kurita, S., Matsuda, S., Connors, M., Nagatsuma, T., Sakaguchi, K., Kasahara, Y., Kumamoto, A., Matsuoka, A., Shoji, M., Shionohara, I., Albert, J.M.: Temporal and spatial correspondence of Pc1/EMIC waves and relativistic electron precipitations observed with ground-based multi-instruments on 27 March 2017. Geophys. Res. Lett. **45**, 24 (2018). https://doi.org/10.1029/2018GL080126

Horne, R.B., Thorne, R.M., Glauert, S.A., Albert, J.M., Meredith, N.P., Anderson, R.R.: Timescale for radiation belt electron acceleration by whistler mode chorus waves. J. Geophys. Res. **110**(A3) (2005). https://doi.org/10.1029/2004JA010844

Hudson, M.K., Baker, D.N., Goldstein, J., Kress, B.T., Paral, J., Toffoletto, F.R., Wiltberger, M.: Simulated magnetopause losses and Van Allen probe flux dropouts. Geophys. Res. Lett. **41**, 1113–1118 (2014). https://doi.org/10.1002/2014GL059222

Jaynes, A.N., et al.: Source and seed populations for relativistic electrons: their roles in radiation belt changes. J. Geophys. Res. Space Physics. **120**, 7240–7254 (2015). https://doi.org/10.1002/2015JA021234

Kanekal, S., Miyoshi, Y.: Dynamics of the terrestrial radiation belts: a review of recent results during the VarSITI (variability of the sun and its terrestrial impact) era, 2014–2018. Prog Earth Planet Sci. **8**, 35 (2021). https://doi.org/10.1186/s40645-021-00413-6

Kasahara, S., Miyoshi, Y., Yokota, S., Mitani, T., Kasahara, Y., Matsuda, S., Kumamoto, A., Matsuoka, A., Kazama, Y., Frey, H.U., Angelopoulos, V., Kurita, S., Keika, K., Seki, K., Shinohara, I.: Pulsating aurora from electron scattering by chorus waves. Nature. **554**, 337–340 (2018). https://doi.org/10.1038/nature25505

Katoh, Y., Omura, Y.: Electron hybrid code simulation of whistler-mode chorus generation with real parameters in the Earth's inner magnetosphere. Earth Planets Space. **68**, 192 (2016). https://doi.org/10.1186/s40623-016-0568-0

Katoh, Y., Kojima, H., Hikishima, M., Takashima, T., Asamura, K., Miyoshi, Y., Kasahara, Y., Kasahara, S., Mitani, T., Higashio, N., Matsuoka, A., Ozaki, M., Yagitani, S., Yokota, S., Matsuda, S., Kitahara, M., Shinohara, I.: Software-type wave-particle interaction analyzer on board the Arase satellite. Earth Planets Space. **70**, 4 (2018a). https://doi.org/10.1186/s40623-017-0771-7

Katoh, Y., Omura, Y., Miyake, Y., Usui, H., Nakashima, H.: Dependence of generation of whistler mode chorus emissions on the temperature anisotropy and density of energetic electrons in the Earth's inner magnetosphere. J. Geophys. Res. Space Physics. **123**, 1165–1177 (2018b). https://doi.org/10.1002/2017JA024801

Kim, H.-J., Chan, A.A.: Fully adiabatic changes in storm time relativistic electron fluxes. J. Geophys. Res. **102**, 22107–22116 (1997). https://doi.org/10.1029/97JA01814

Kitahara, M., Katoh, Y.: Method for direct detection of pitch angle scattering of energetic electrons caused by whistler mode chorus emissions. J. Geophys. Res. Space Physics. **121**(6), 5137–5148 (2016). https://doi.org/10.1002/2015JA021902

Kitahara, M., Katoh, Y.: Anomalous trapping of low pitch angle electrons by coherent whistler mode waves. J. Geophys. Res. Space Physics. **124**, 5568–5583 (2019). https://doi.org/10.1029/2019JA026493

Kitamura, N., Kitahara, M., Shoji, M., Miyoshi, Y., Hasegawa, H., Nakamura, S., Katoh, Y., Saito, Y., Yokota, S., Gershman, D.J., Vinas, A.F., Giles, B.L., Moore, T.E., Paterson, W.R., Pollock, C.J., Russell, C.T., Strangeway, R.J., Fuselier, S.A., Burch, J.L.: Direct measurements of two-way wave-particle energy transfer in a collisionless space plasma. Science. **361**, 1000–1003 (2018). https://doi.org/10.1126/science.aap8730

Kubota, Y., Omura, Y., Summers, D.: Relativistic electron precipitation induced by EMIC-triggered emissions in a dipole magnetosphere. J. Geophys. Res. Space Physics. **120**, 4384–4399 (2015). https://doi.org/10.1002/2015JA021017

Kurita, S., Miyoshi, Y., Blake, J.B., Reeves, G.D., Kletzing, C.A.: Relativistic electron microbursts and variations in trapped MeV electron fluxes during the 8–9 October 2012 storm: SAMPEX and Van Allen probes observations. Geophys. Res. Lett. **43**, 3017–3025 (2016). https://doi.org/10.1002/2016GL068260

Kurita, S., Miyoshi, Y., Kasahara, S., Yokota, S., Kasahara, Y., Matsuda, S., et al.: Deformation of electron pitch angle distributions caused by upper band chorus observed by the Arase satellite. Geophys. Res. Lett. **45**, 7996–8004 (2018a). https://doi.org/10.1029/2018GL079104

Kurita, S., Miyoshi, Y., Blake, J.B., Friedel, R.H.W.: Response of relativistic electron microbursts to the arrival of high-speed solar wind streams and its relation to flux variation of trapped radiation belt electrons. J. Geophys. Res. Space Physics. **123**, 7452–7461 (2018b). https://doi.org/10.1029/2018JA025675

Kurita, S., Miyoshi, Y., Shiokawa, K., Higashio, N., Mitani, T., Takashima, T., et al.: Rapid loss of relativistic electrons by EMIC waves in the outer radiation belt observed by Arase, Van Allen probes, and the PWING ground stations. Geophys. Res. Lett. **45**, 12720–12729 (2018c). https://doi.org/10.1029/2018GL080262

Li, W., Hudson, M.K.: Earth's Van Allen radiation belts: from discovery to the Van Allen Probes Era. J. Geophys. Res. **124**(11), 8319–8351 (2019). https://doi.org/10.1029/2018JA025940

Li, X., Roth, I., Temerin, M., Wygant, J.R., Hudson, M.K., Blake, J.B.: Simulation of the prompt energization and transport of radiation belt particles during the March 24, 1991 SSC. Geophys. Res. Lett. **20**(22), 2423–2426 (1993). https://doi.org/10.1029/93GL02701

Li, X., Temerin, M., Baker, D.N., Reeves, G.D., Larson, D.: Quantitative prediction of radiation belt electrons at geostationary orbit based on solar wind measurements. Geophys. Res. Lett. **28**(9), 1887–1890 (2001). https://doi.org/10.1029/2000GL012681

Li, X., Temerin, M., Baker, D.N., Reeves, G.D.: Behavior of MeV electrons at geosynchronous orbit during last two solar cycles. J. Geophys. Res. **116**, A11207 (2011). https://doi.org/10.1029/2011JA016934

Li, X., Selesnick, R.S., Baker, D.N., Jaynes, A.N., Kanekal, S.G., Schiller, Q., Blum, L., Fennell, J., Blake, J.B.: Upper limit on the inner radiation belt MeV electron intensity. J. Geophys. Res. Space Physics. **120**, 1215–1228 (2015). https://doi.org/10.1002/2014JA020777

Li, X., Selesnick, R., Schiller, Q., et al.: Measurement of electrons from albedo neutron decay and neutron density in near-earth space. Nature. **552**, 382–385 (2017). https://doi.org/10.1038/nature24642

Miyoshi, Y., Kataoka, R.: Ring current ions and radiation belt electrons during geomagnetic storms driven by coronal mass ejections and corotating interaction regions. Geophys. Res. Lett. **32**, L21105 (2005). https://doi.org/10.1029/2005GL024590

Miyoshi, Y., Kataoka, R.: Flux enhancement of the outer radiation belt electrons after the arrival of stream interaction regions. J. Geophys. Res. **113**, A03S09 (2008a). https://doi.org/10.1029/2007JA012506

Miyoshi, Y., Kataoka, R.: Probabilistic space weather forecast of the relativistic electron flux enhancement at geosynchronous orbit. J. Atm. Solar-Terr. Phys. **70**, 475–481 (2008b)

Miyoshi, Y., Kataoka, R.: Solar cycle variations of outer radiation belt and its relationship to solar wind structure dependences. J. Atm. Solar-Terr. Phys. **73**, 77–87 (2011). https://doi.org/10.1016/j.jastp.2010.09.031

Miyoshi, Y., Morioka, A., Obara, T., Misawa, H., Nagai, T., Kasahara, Y.: Rebuilding process of the outer radiation belt during the November 3, 1993, magnetic storm - NOAA and EXOS-D observations. J. Geophys. Res. **108**(A1), 1004 (2003). https://doi.org/10.1029/2001JA007542

Miyoshi, Y.S., Jordanova, V.K., Morioka, A., Evans, D.S.: Solar cycle variations of the electron radiation belts: observations and radial diffusion simulation, space. Weather. **2**, S10S02 (2004). https://doi.org/10.1029/2004SW000070

Miyoshi, Y., Sakaguchi, K., Shiokawa, K., Evans, D., Albert, J., Conners, M., Jordanova, V.: Precipitation of radiation belt electrons by EMIC waves, observed from ground and space. Geophys. Res. Lett. **35**, L23101 (2008). https://doi.org/10.1029/2008GL035727

Miyoshi, Y., Kataoka, R., Kasahara, Y., Kumamoto, A., Nagai, T., Thomsen, M.F.: High-speed solar wind with southward interplanetary magnetic field causes relativistic electron flux enhancement of the outer radiation belt via enhanced condition of whistler waves. Geophys. Res. Lett. **40** (2013). https://doi.org/10.1002/grl.50916

Miyoshi, Y., et al.: Energetic electron precipitation associated with pulsating aurora: EISCAT and Van Allen probe observations. J. Geophys. Res. Space Physics. **120**, 2754–2766 (2015). https://doi.org/10.1002/2014JA020690

Miyoshi, Y., Kataoka, R., Ebihara, Y.: Flux enhancement of relativistic electrons associated with substorms. In: Balasis, G., Daglis, I.A., Mann, I.R. (eds.) Waves, Particles, and Storms in Geospace, vol. 333-353. Oxford Press (2016)

Miyoshi, Y., Shinohara, I., Takashima, T., Asamura, K., Higashio, N., Mitani, T., Kasahara, S., Yokota, S., Kazama, Y., Wang, S.-Y., Tam, S.W., Ho, P.T.P., Kasahara, Y., Kasaba, Y., Yagitani, S., Matsuoka, A., Kojima, H., Katoh, H., Shiokawa, K., Seki, K.: Geospace exploration project ERG. Earth Planets Space. **70**, 101 (2018). https://doi.org/10.1186/s40623-018-0862-0

Miyoshi, Y., Saito, S., Kurita, S., Asamura, K., Hosokawa, K., Sakanoi, T., Mitani, T., Ogawa, Y., Oyama, S., Tsuchiya, F., Jones, S.L., Jaynes, A.N., Blake, J.B.: Relativistic electron microbursts as high energy tail of pulsating Aurora electrons. Geophys. Res. Lett. **47** (2020). https://doi.org/10.1029/2020GL090360

Miyoshi, Y., Hosokawa, K., Kurita, S., Oyama, S.-I., Ogawa, Y., Saito, S., Shinohara, I., Kero, A., Turunen, E., Verronen, P.T., Kasahara, S., Yokota, S., Mitani, T., Takashima, T., Higashio, N., Kasahara, Y., Masuda, S., Tsuchiya, F., Kumamoto, A., Matsuoka, A., Hori, T., Keika, K., Shoji, M., Teramoto, M., Imajo, S., Jun, C., Nakamura, S.: Penetration of MeV electrons into the mesosphere accompanying pulsating aurorae. Sci. Rep. **11**, 13724 (2021). https://doi.org/10.1038/s41598-021-92611-3

Nakamura, S., Omura, Y., Kletzing, C., Baker, D.N.: Rapid precipitation of relativistic electron by EMIC rising-tone emissions observed by the Van Allen probes. J. Geophys. Res. Space Physics. **124**, 6701–6714 (2019). https://doi.org/10.1029/2019JA026772

Ohtani, S., Miyoshi, Y., Singer, H.J., Weygand, J.M.: On the loss of relativistic electrons at geosynchronous altitude: its dependence on magnetic configurations and external conditions. J. Geophys. Res. **114**, A01202 (2009). https://doi.org/10.1029/2008JA013391

Omura, Y.: Nonlinear wave growth theory of whistler-mode chorus and hiss emissions in the magnetosphere. Earth Planets Space. **73**, 95 (2021). https://doi.org/10.1186/s40623-021-01380-w

Omura, Y., Summers, D.: Dynamics of high-energy electrons interacting with whistler mode chorus emissions in the magnetosphere. J. Geophys. Res. **111** (2006). https://doi.org/10.1029/2006JA011600

Omura, Y., Zhao, Q.: Nonlinear pitch angle scattering of relativistic electrons by EMIC waves in the inner magnetosphere. J. Geophys. Res. **117**, A08227 (2012). https://doi.org/10.1029/2012JA017943

Omura, Y., Zhao, Q.: Relativistic electron microbursts due to nonlinear pitch angle scattering by EMIC triggered emissions. J. Geophys. Res. Space Physics. **118**, 5008–5020 (2013). https://doi.org/10.1002/jgra.50477

Omura, Y., Furuya, N., Summers, D.: Relativistic turning acceleration of resonant electrons by coherent whistler mode waves in a dipole magnetic field. J. Geophys. Res. **112**, A06236 (2007). https://doi.org/10.1029/2006JA012243

Omura, Y., Katoh, Y., Summers, D.: Theory and simulation of the generation of whistler-mode chorus. J. Geophys. Res. **113**, A04223 (2008). https://doi.org/10.1029/2007JA012622

Omura, Y., Hikishima, M., Katoh, Y., Summers, D., Yagitani, S.: Nonlinear mechanisms of lower-band and upper-band VLF chorus emissions in the magnetosphere. J. Geophys. Res. **114**, A07217 (2009). https://doi.org/10.1029/2009JA014206

Omura, Y., Miyashita, Y., Yoshikawa, M., Summers, D., Hikishima, M., Ebihara, Y., Kubota, Y.: Formation process of relativistic electron flux through interaction with chorus emissions in the Earth's inner magnetosphere. J. Geophys. Res. Space Physics. **120**, 9545–9562 (2015). https://doi.org/10.1002/2015JA021563

Paulikas, G.A., Blake, J.B.: Effects of the solar wind on magnetospheric dynamics: energetic electrons at the synchronous orbit, In Quantitative Modeling of Magnetospheric Processes, Geophys. Monogr. Ser., vol. 21, pp. 180–202. AGU, Washington, DC (1979)

Pires de Lima, R., Chen, Y., Lin, Y.: Forecasting megaelectron-volt electrons inside Earth's outer radiation belt: PreMevE 2.0 based on supervised machine learning algorithms. Space Weather. **18**, e2019SW002399 (2020). https://doi.org/10.1029/2019SW002399

Reeves, G.D., McAdams, K.L., Friedel, R.H.W., O'Brien, T.P.: Acceleration and loss of relativistic electrons during geomagnetic storms. Geophys. Res. Lett. **30**(10), 1529 (2003). https://doi.org/10.1029/2002GL016513/

Reeves, G.D., Morley, S.K., Friedel, R.H.W., Henderson, M.G., Cayton, T.E., Cunningham, G., Blake, J.B., Christensen, R.A., Thomsen, D.: On the relationship between relativistic electron flux and solar wind velocity: Paulikas and Blake revisited. J. Geophys. Res. **116**, A02213 (2011). https://doi.org/10.1029/2010JA015735

Russell, C.T., McPherron, R.L.: Semiannual variation of geomagnetic activity. J. Geophys. Res. **78**, 92–108 (1973)

Saito, S., Miyoshi, Y., Seki, K.: A split in the outer radiation belt by magnetopause shadowing: test particle simulations. J. Geophys. Res. **115**, A08210 (2010). https://doi.org/10.1029/2009JA014738

Saito, S., Miyoshi, Y., Seki, K.: Relativistic electron microbursts associated with whistler chorus rising tone elements: GEMSIS-RBW simulations. J. Geophys. Res. **117**, A10206 (2012). https://doi.org/10.1029/2012JA018020

Saito, S., Miyoshi, Y., Seki, K.: Rapid increase in relativistic electron flux controlled by nonlinear phase trapping of whistler chorus elements. J. Geophys. Res. Space Physics. **121**, 6573–6589 (2016). https://doi.org/10.1002/2016JA022696

Saito, S., Kurita, S., Miyoshi, Y., Kasahara, S., Yokota, S., Keika, K., Hori, T., Kasahara, Y., Matsuda, S., Shoji, M., Nakamura, S., Matsuoka, A., Imajo, S., Shinohara, I.: Data-driven simulation of rapid flux enhancement of energetic electrons with an upper-band whistler burst. J. Geophys. Res. Space Phys. **126**, e2020JA028979 (2021). https://doi.org/10.1029/2020JA028979

Sakaguchi, K., Miyoshi, Y., Saito, S., Nagatsuma, T., Seki, K., Murata, K.T.: Relativistic electron flux forecast at geostationary orbit using Kalman filter based on multivariate autoregressive model. Space Weather. **11**, 79–89 (2013). https://doi.org/10.1002/swe.20020

Sakaguchi, K., Nagatsuma, T., Reeves, G.D., Spence, H.E.: Prediction of MeV electron fluxes throughout the outer radiation belt using multivariate autoregressive models. Space Weather. **13**, 853–867 (2015). https://doi.org/10.1002/2015SW001254

Selesnick, R.S., Blake, J.B.: On the source location of radiation belt relativistic electrons. J. Geophys. Res. **105** (2000). https://doi.org/10.1029/1999JA900445

Shiota, D., Kataoka, R., Miyoshi, Y., Hara, T., Tao, C., Masunaga, K., Futaana, Y., Terada, N.: Inner heliosphere MHD modeling system applicable to space weather forecasting for the other planets. Space Weather. **12**, 187–204 (2014). https://doi.org/10.1002/2013SW000989

Shoji, M., Miyoshi, Y., Katoh, Y., Keika, K., Angelopoulos, V., Kasahara, S., Asamura, K., Nakamura, S., Omura, Y.: Ion hole formation and nonlinear generation of electromagnetic ion cyclotron waves: THEMIS observations. Geophys. Res. Lett. **44**, 8730–8738 (2017). https://doi.org/10.1023/2017GL074254

Summers, D., Omura, Y.: Ultra-relativistic acceleration of electrons in planetary magnetospheres. Geophys. Res. Lett. **34**, L24205 (2007). https://doi.org/10.1029/2007GL032226

Summers, D., Thorne, R.M., Xiao, F.: Relativistic theory of wave-particle resonant diffusion with application to electron acceleration in the magnetosphere. J. Geophys. Res. Space Phys. **103**(A9), 7–20500 (1998). https://doi.org/10.1029/98JA01740

Tsurutani, T., Gonzalez, W.D.: The cause of high-intensity long-duration continuous AE activity (HILDCAAs): Interplanetary Alfvén wave trains. Planet Space Sci. **35**(4), 405–412 (1988). https://doi.org/10.1016/0032-0633(87)90097-3

Tsyganenko, N.A., Sitnov, M.I.: Modeling the dynamics of the inner magnetosphere during strong geomagnetic storms. J. Geophys. Res. **110**, A03208 (2005). https://doi.org/10.1029/2004JA010798

Turner, D., Shprits, Y., Hartinger, M., et al.: Explaining sudden losses of outer radiation belt electrons during geomagnetic storms. Nat. Phys. **8**, 208–212 (2012). https://doi.org/10.1038/nphys2185

Turner, D.L., et al.: On the cause and extent of outer radiation belt losses during the 30 September 2012 dropout event. J. Geophys. Res. Space Physics. **119**, 1530–1540 (2014). https://doi.org/10.1002/2013JA019446

Turner, D.L., Kilpua, E.K.J., Hietala, H., Claudepierre, S.G., O'Brien, T.P., Fennell, J.F., et al.: The response of Earth's electron radiation belts to geomagnetic storms: statistics from the Van Allen probes era including effects from different storm drivers. J. Geophys. Res. Space Physics. **124**, 1013–1034 (2019)

Xiang, Z., Li, X., Selesnick, R., Temerin, M.A., Ni, B., Zhan, H., Zhang, K., Khoo, L.Y.: Modeling the quasi-trapped electron fluxes from cosmic ray albedo neutron decay (CRAND). Geophys. Res. Lett. (2019). https://doi.org/10.1029/2018GL081730

Chapter 6
Geomagnetic Variability and GIC

Yusuke Ebihara, Satoko Nakamura, Tada-nori Goto, Shinichi Watari, and Takashi Kikuchi

6.1 Mechanism of GIC Generation

The electric currents flowing in the near-Earth space cause the magnetic disturbances and geomagnetically induced currents (GIC). In particular, the magnetopause current, the ring current, the field-aligned current, and the ionospheric current are known to cause the large-amplitude GIC. The magnitudes of these currents change rapidly and significantly in response to the solar wind conditions. Typical changes in the geomagnetic field caused by the above current systems are discussed in detail subsequently in Sect. 6.3. The change in the currents flowing in the near-Earth space induces an electric field on and below the surface of the Earth, which is called geomagnetically induced electric field (GIE). In Fig. 6.1, we show an example of the natural fluctuations of the geomagnetic field and GIE recorded at the Kakioka Magnetic Observatory of the Japan Meteorological Agency (36.232° N, 140.186° E).

The generation process of GIC through a power grid is schematically shown in Fig. 6.2. Part of the induced currents flowing underground goes into and out of a power grid through the grounding points of power plants and substations. For

Y. Ebihara (✉)
Research Institute for Sustainable Humanosphere, Kyoto University, Kyoto, Japan
e-mail: ebihara@rish.kyoto-u.ac.jp

S. Nakamura · T. Kikuchi
Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Aichi, Japan

T.-n. Goto
University of Hyogo, Himeji, Hyogo, Japan

S. Watari
National Institute of Information and Communications Technology, Tokyo, Japan

K. Kusano (ed.), *Solar-Terrestrial Environmental Prediction*,
https://doi.org/10.1007/978-981-19-7765-7_6

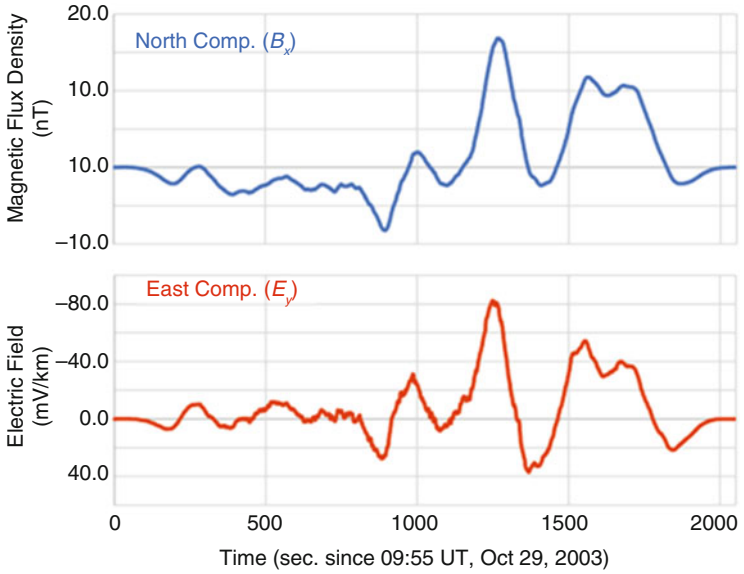

Fig. 6.1 Geomagnetic variations (blue, northward component) and GIE at surface (red, eastward component) recorded at Kakioka, Ishioka, Ibaraki, Japan. Both data are after removal of mean and application of Hanning window to original record of 2048-s values per second. Public data from website of Kakioka Magnetic Observatory of Japan Meteorological Agency are used

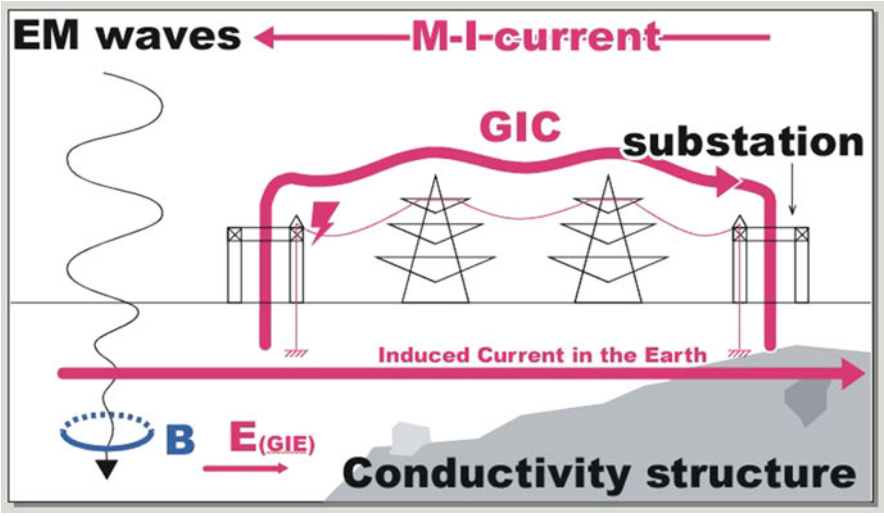

Fig. 6.2 Conceptual diagram of mechanism of GIC generation. GIC are induced currents in power grids caused by magnetospheric ionospheric currents generated by solar wind

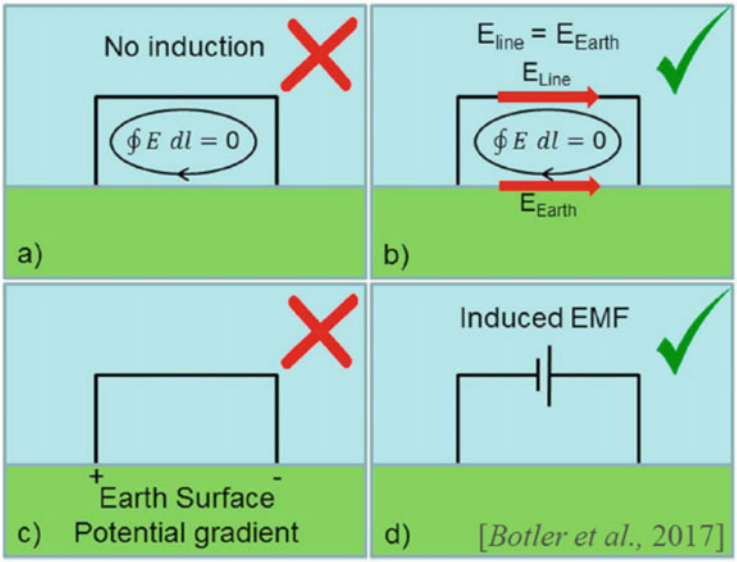

Fig. 6.3 Conceptual diagram of physical mechanism of GIC. See text for detailed explanation (Boteler and Pirjola 2017)

example, the neutral point of a transformer is grounded for safety. Large-amplitude GIC are known to cause severe heating of the transformer and malfunction of the protective relay of the power grid.

Boteler and Pirjola (2017) summarized questions that have arisen regarding the physical driver of GIC. Can we use Faraday's law to derive the electromotive force integrated with the time derivative of the magnetic flux around the loop formed by power lines and the surface of the Earth (Fig. 6.3a)? Can we consider the potential difference between the two points on the ground to cause the current flowing into and out of the power lines (Fig. 6.3c)? The answers to these questions are no. Geomagnetically disturbances caused by near-Earth space currents have so large spatial scales greater than 100 km that the area encircled by the power line and the ground surface is too small to induce the electric field integrated along the loop formed by power lines and the surface of the Earth. The loop should extend down to the interior of the Earth. For the vertically elongated loop, the electric field induced in the power lines and the ground surface can be considered to be almost the same (Fig. 6.3b). The idea of the potential difference comes from the facts that the current flows in the ground and that the underground has finite conductivity. The corresponding potential drop arising from the current and the finite conductivity seems to appear on the ground surface. However, the electric field is purely inductive, so that the idea of the potential differences leads to confusion. After discussing the above ideas, Boteler and Pirjola (2017) suggested that the power lines act as extra voltage sources and drive the GIC.

The direct driver of GIC is the GIE associated with geomagnetic variations. For a uniform surface impedance $\widetilde{Z}$, GIE $\widetilde{E}$ is related to the horizontal component of the geomagnetic field, $\widetilde{B}$, in the frequency domain (Pulkkinen et al. 2007) as

$$\widetilde{E}(\omega) = \pm \frac{\widetilde{Z}}{\mu_0} \widetilde{B}(\omega), \tag{6.1}$$

where ω is the angular frequency. Thus, the temporal variations in the geomagnetic field can be used as a measure of the magnitude of GIE. Figure 6.4 shows the latitudinal distributions of the maximum calculated GIE, maximum time derivative of the horizontal component of the geomagnetic field, and maximum amplitude of the horizontal magnetic field during the large magnetic storms that occurred on March 13–15, 1989, and October 29–31, 2003. The maximum GIE is shown to be large at high latitudes because of the significant contribution from the ionospheric currents, including auroral electrojets (AEJs). The maximum GIE is also high at the geomagnetic equator, most probably owing to equatorial electrojets (EEJs). The GIE is relatively low at geomagnetic latitudes lower than 50°, indicating that the potential risk for power grids at low latitudes is relatively low. However, as mentioned in Sect. 6.2, large-amplitude GIC have been observed at low latitudes during extreme magnetic storms. The observation results suggest that power grids at low latitudes are not typically safe against GIC.

The electric field induced by the magnetospheric and ionospheric currents is modulated by the current flow in the subsurface. The current flowing in the magnetosphere and the ionosphere has a very large spatial scale, from several hundred kilometers to several times the radius of the Earth (Re). It depends not only on the local electrical conductivity but also on the three-dimensional structure of the electrical conductivity because of the continuity of the current. The depth, d, at which electromagnetic energy can penetrate the ground is called the skin depth. In the case of uniform electrical conductivity, the skin depth is expressed as

$$d = \sqrt{\frac{2}{\omega \mu_0 \sigma}} \approx 503 \sqrt{\frac{T}{\sigma}}, \tag{6.2}$$

where μ_0, σ, and T are the magnetic constant, electrical conductivity, and period, respectively. It is noticeable that the skin depth depends on the electrical conductivity of the ground and the period of fluctuations. For example, for geomagnetic fluctuations with periods of 100, 1000, and 10,000 s and ground conductivity of 0.01 S/m, the skin depths, d, are ~50, ~300, and ~ 500 km, respectively. These periods correspond to the characteristic timescales for sudden commencements (SCs), substorms, and magnetic storms, respectively. Because of the dependence on the conductivity, information on the distribution of electrical conductivity deep underground is necessary.

Specific theoretical calculations have confirmed that the electrical conductivity structure of the mantle of the Earth can affect GIE and GIC and that large

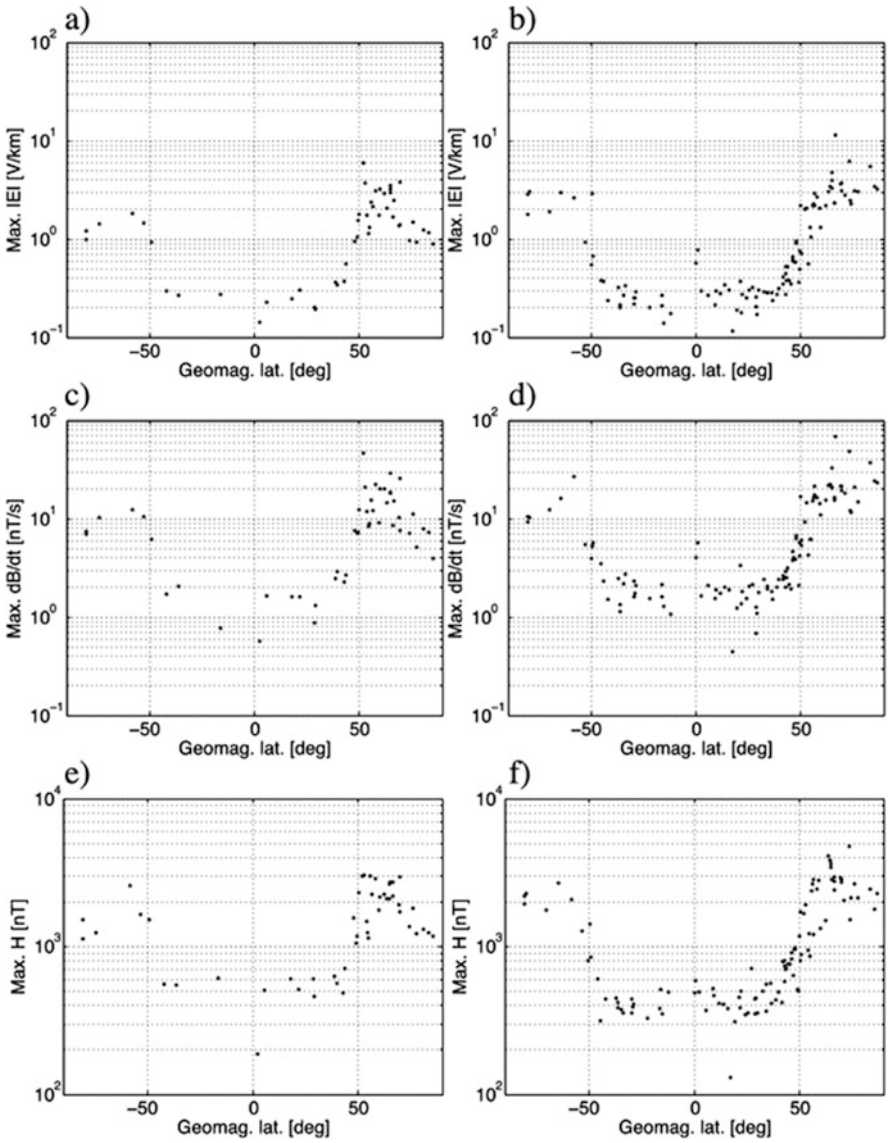

Fig. 6.4 GIE (calculated), dB/dt (observed), and H (observed) values for magnetic storms of March 13–14, 1989 (**a**, **c**, **e**), and October 29–31, 2003 (**b**, **d**, **f**) (Pulkkinen et al. 2012)

geomagnetic fluctuations can produce large GIE and GIC. Figure 6.5 shows the theoretical GIE generated by geomagnetic fluctuations under the assumption of a horizontally stratified Earth (Goto 2015). Two models of the subsurface conductivity structure are considered here. In both models, the conductivity is 0.01 S/m from the surface to 10-km depth and 0.001 S/m from 10- to 40-km depth. At depths exceeding

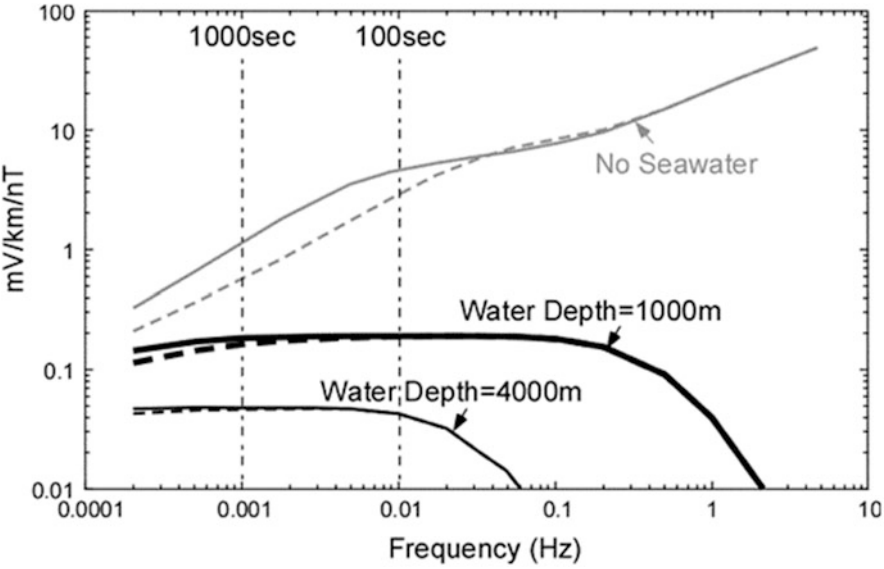

Fig. 6.5 Examples of estimated induced electric fields (GIE) produced by geomagnetic variations of various frequencies (Goto 2015). Solid and dashed lines are from different subsurface electrical conductivity models (see text). Gray represents GIE at surface on land, and solid line denotes GIE at seafloor surface when layer of seawater is placed over this land, for depths of 1000 and 4000 m, respectively

40 km, the conductivity differs for the two models, with one model showing 0.001 S/m at depths between 40 and 150 km and 0.033 S/m at depths greater than 150 km (solid line in Fig. 6.5). In the other model, the value is 0.033 S/m at depths deeper than 40 km (dashed line in Fig. 6.5). For a period of 1000 s, the difference in the GIE is approximately a factor of two owing to the difference in mantle structure at depths deeper than 40 km. In these models, if a geomagnetic fluctuation with an amplitude of 2000 nT (period 1000 s) occurs over the Earth, the predicted GIE is 1.1–2.3 V/km. This GIE can generate a voltage fluctuation of approximately 1.1–2.3 kV between two points 1000 km apart. If a geomagnetic variation with a period of 100 s occurs with the same amplitude, the voltage variation between these two points can reach approximately 5.7–9.3 kV. Such large GIE is expected to produce large GIC. Figure 6.5 also shows the predicted values of the GIE at the seafloor. Because it is known that the electromagnetic field decays rapidly in seawater, the effect of GIC is reduced compared to that on land. However, in the case of shallow water and long-period geomagnetic fluctuations, the GIE are expected to be a fraction of those on land. Long infrastructure facilities (submarine cables and pipelines) are built on the seafloor; therefore, caution should be exercised in designing systems that ignore GIC.

For GIE prediction, three-dimensional subsurface electrical conductivity distributions over a wide area are required. Currently, modeling of subsurface structures is being conducted in many countries and regions. For example, in the case of the

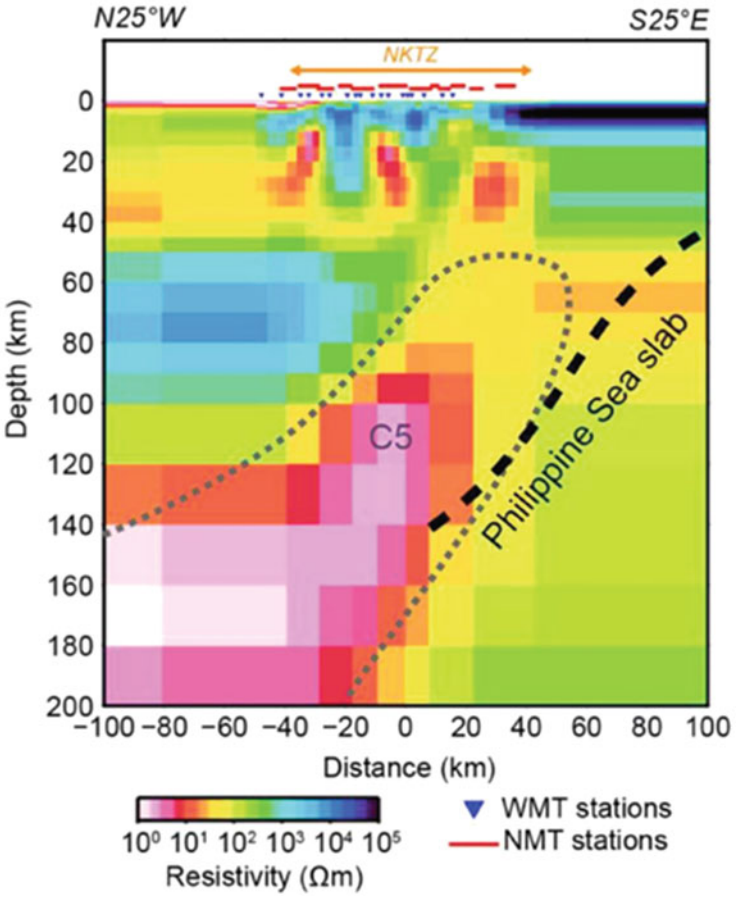

Fig. 6.6 Cross section of subsurface electrical conductivity structure in Central Japan (leftmost, Ishikawa Prefecture; rightmost, Shizuoka Prefecture) (Usui et al. 2021). Many heterogeneous structures are found in crust located in the Kobe/Niigata strain concentration zone (NKTZ). WMT, wideband MT survey stations; NMT, long-period network MT stations. Data are expressed in resistivity (Ωm), which is reciprocal of electrical conductivity

Japanese islands, a large-scale crustal and mantle electrical conductivity structure, as shown in Fig. 6.6, has been found in observations. In fact, GIE have been used for such subsurface visualization. Small-scale geomagnetic disturbances and the resulting weak GIE at the surface occur regularly and do not produce large GIC. If the geomagnetic and electric field fluctuations are measured at multiple locations on the surface of the Earth, the strength of the induced electric field relative to the unit magnetic field fluctuation can be determined over a wide frequency range (from 10 kHz to near DC) (this is called the impedance of the Earth). Because electromagnetic field fluctuations at high frequencies are attenuated in a shallow subsurface, the impedance at high frequencies has information about its electrical conductivity.

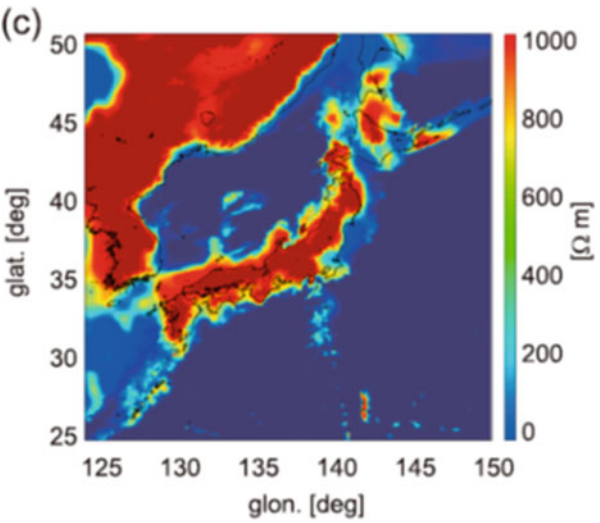

Fig. 6.7 Electrical conductivity structure model for 2 km of surface around Japan (Nakamura et al. 2018). Color represents resistivity (inverse of electrical conductivity)

Electromagnetic field fluctuations at low frequencies do not decay easily; therefore, they contain information on the electrical conductivity deep underground. A geophysical subsurface exploration method using such natural electromagnetic field fluctuations is called the magnetotelluric (MT) method (Chave and Jones 2012). The MT method is commonly used on land and at the seafloor worldwide for the exploration of underground resources, active faults and volcanoes, and the mantle structure. In recent years, there have also been attempts to obtain the three-dimensional distribution of the electrical conductivity of the mantle of the Earth using data of magnetometers aboard satellites and data from ground-based geomagnetic observatories (e.g., Kuvshinov et al. 2021). However, a wide-area three-dimensional electrical conductivity distribution in the crust and mantle of the Earth based on observation data has not been completed for the Japanese islands, and it is still under construction for other countries globally.

Therefore, there have been several efforts to build a simple subsurface electrical conductivity model incorporating known shallow subsurface information and conduct first-principles studies on the GIE distribution and the generation of GIC in a power grid. As an example, Fig. 6.7 shows a model of the subsurface resistivity structure (inverse of the electrical conductivity) around Japan. Here, seawater distribution (seafloor topography) and known sedimentary layer distribution are incorporated in a three-dimensional structure. From the broad structure, it can be seen that the land has a higher specific resistance than the sea. This is because seawater has a higher electrical conductivity (approximately 3–4 S/m) than the sediments and rocks of the crust and the mantle. Comparing the sea areas surrounding Japan, the resistivity is high west of Kyushu. This is owing to the existence of a continental shelf and the shallow water depth of the East China Sea. The resistivity distribution around the land/sea boundary has a very complex structure, and there are areas in which low-resistivity regions similar to that of the sea extend inland from the

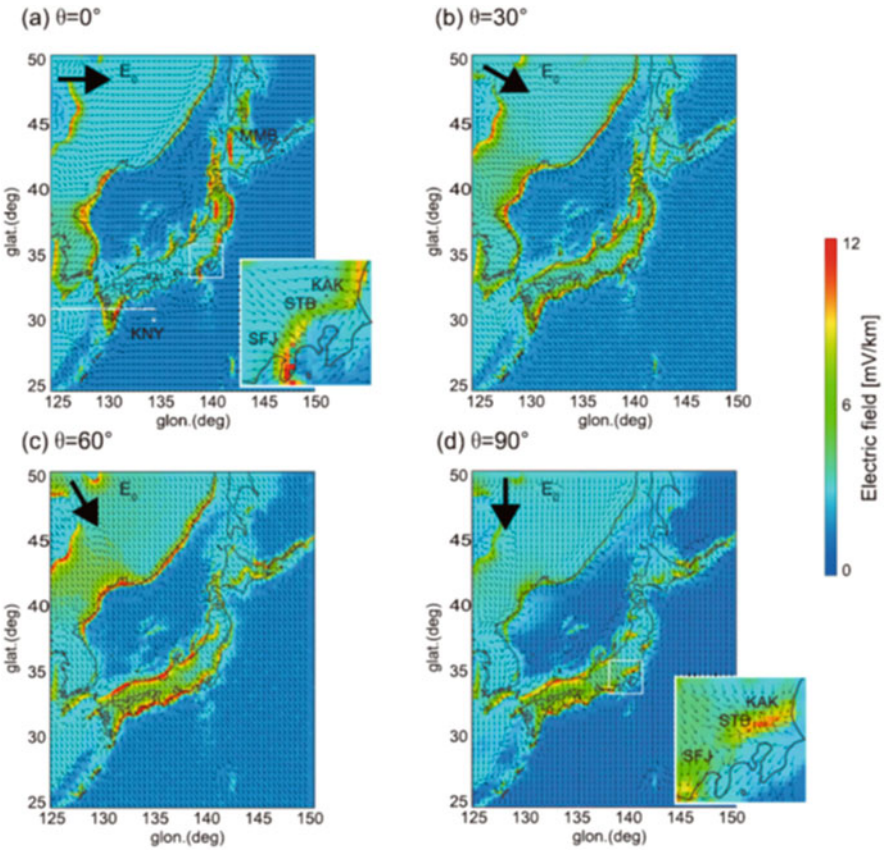

Fig. 6.8 Simulation results of GIE generated for 100-s electromagnetic field variation (Nakamura et al. 2018)

actual coastline. In particular, these low-resistivity regions are localized in the inland areas of the Hokkaido, Tohoku, and Kanto regions. These are locations with differences in land geology and where sedimentary layers with high electrical conductivity are modeled. In areas where sediments are thickly distributed, such as sedimentary plains and sedimentary basins, the resistivity of a model tends to be low. Note that the model in Fig. 6.7 does not incorporate deep crustal and mantle heterogeneities as that in Fig. 6.6.

Using a subsurface electrical conductivity model, GIE generated by external geomagnetic disturbances can be calculated. The numerical results using the model in Fig. 6.7 are shown in Fig. 6.8. The calculations are performed assuming a spatially uniform and predominant external geomagnetic disturbance in a certain direction. For the spatially uniform external disturbance, it can be seen that the GIE has a spatial structure at the surface and that there are regions with very large GIE values. As the predominant direction of the external disturbance (current direction of

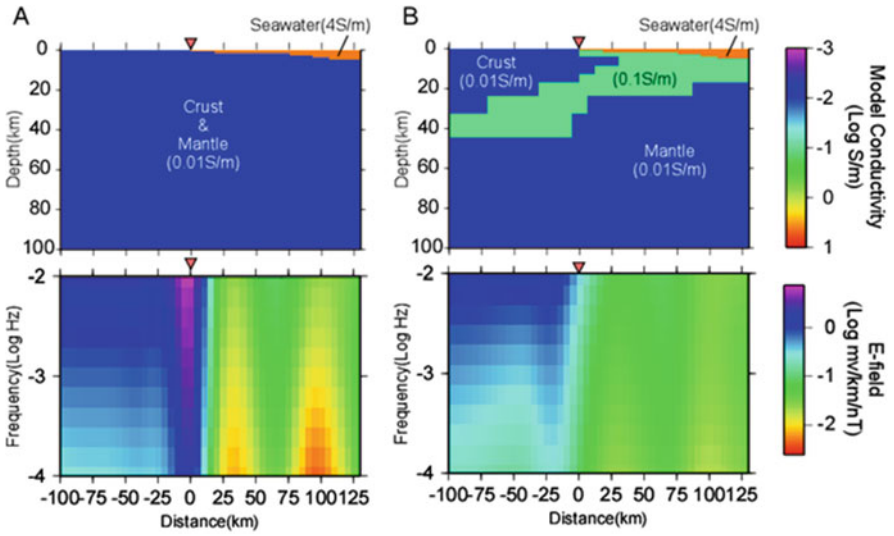

Fig. 6.9 Numerically calculated GIE under assumption of two-dimensional subsurface electrical conductivity (Goto 2015). (**a**) Model with uniform subsurface structure and only contrast with seawater layer. (**b**) Model with high conductivity (0.1 S/m) spanning sea–land boundary. Model assumes high-conductivity body around upper surface of Philippine Sea Plate. Below are calculated GIE for both models with periods of 100 s–10000 s (frequencies of 10^{-2}–10^{-4} Hz)

the current source) changes, the location of the GIE anomaly also changes, and particularly near the coast where the direction of the GIE and the coastline are orthogonal, the GIE tends to be large. This is because the presence of a spatial gradient in the electrical conductivity charges the discontinuity boundary of the electrical conductivity owing to the continuity of the current, and a secondary electric field is generated. To examine this effect in more detail, we compared GIE obtained using a simple model of a subsurface conductivity structure, as shown in Fig. 6.9. In the case of a uniform subsurface conductivity structure (Fig. 6.9a), large GIE are generated on the landward side of the boundary between the seawater and the crust, which is a good conductor. At this time, the GIE decreases on the seafloor side. In contrast, when a low-resistivity layer is distributed in the ground in a manner that intersects the coastline (Fig. 6.9b), the difference in the GIE near the sea–land boundary becomes small, and the amplification effect of the GIE from the coastline to the landside is suppressed. Such local amplification and attenuation of GIE can also occur in areas of rapid subsurface structural changes on land. For example, Fig. 6.8 shows that the GIE are abnormally large in the sedimentary stratum regions of Hokkaido, Tohoku, and Kanto as well as at the sea–land boundary. As described above, the subsurface electrical conductivity structure plays a very important role in the generation of GIE.

When the distribution of GIE is provided, the magnitude of the GIC can be determined by the topology and electrical parameters of the power grid, such as the resistance values of the ground and the transmission line. Lehtinen and Pirjola (1985) suggested a scheme to calculate the GIC flowing in an entire grid for a

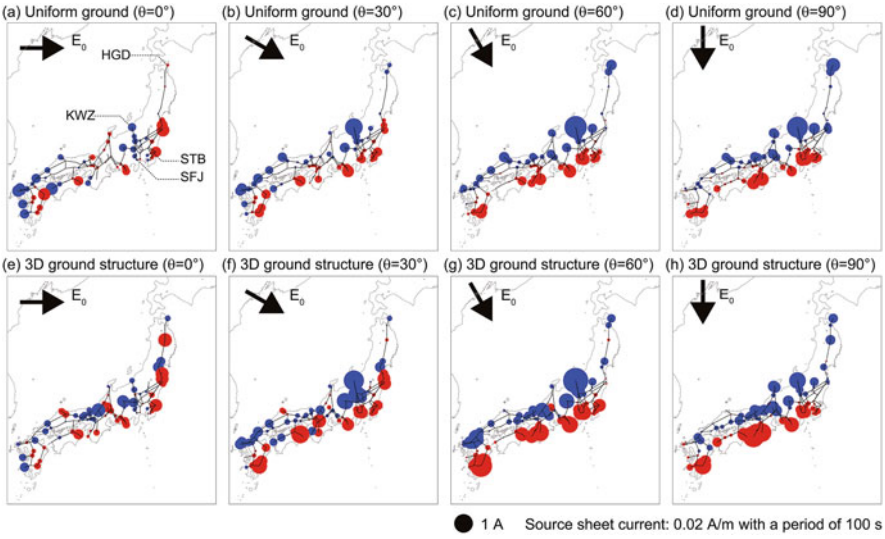

Fig. 6.10 Simulation results of GIC generated at substations in high-voltage power grid for geomagnetic variations with period of 100 s. The electric field directing (a,e) 0°, (b,f) 30°, (c,g) 60°, (d,h) 90° from geographical east is emposed. The top panels show the case in which the ground conductivity is uniform, whereas the bottom ones show the case in which inhomogeneous ground conductivity is introduced (Nakamura et al. 2018)

given GIE and parameters of the power grid. Figure 6.10 shows examples of GIC calculated for a 500-kV transmission network in Japan using the GIE distribution shown in Fig. 6.8. From left to the right, electric fields directing 0°, 30°, 60°, and 90° from geographical east are imposed on the Earth. The magnitude of the GIC flowing at a substation is indicated by the radius of the circle, and the polarity of the GIC (flowing into or out of the grid) is indicated by color. It is noticeable from the numerical simulations that large-amplitude GIC tend to flow at substations near the coast. In addition, particularly large GIC are found in the inland areas where the gradient of the conductivity is high. This is a notable example showing that the inhomogeneity of the conductivity and the topology of the power grid are significant in the generation of GIC.

Figure 6.11 shows examples of the temporal variations in the GIC at two substations near Tokyo during the small magnetic storm on May 27–28, 2017. The black line indicates the observed GIC, and the red one indicates the calculated one using the same method to obtain Fig. 6.10. The calculated GIC are similar to the observations. A good agreement is found at STB for the first half of the magnetic storm. The blue line indicates the GIC estimated based on the GIE observed at the Kakioka Magnetic Observatory, which is close to the Shin-Tsukuba Observatory (STB). The GIE are assumed to be uniform. The blue line is very close to the observations at the STB, whereas it significantly deviates from those at the Shin-Fuji Observatory (SFJ). This discrepancy can be explained in terms of the inhomogeneity of the conductivity.

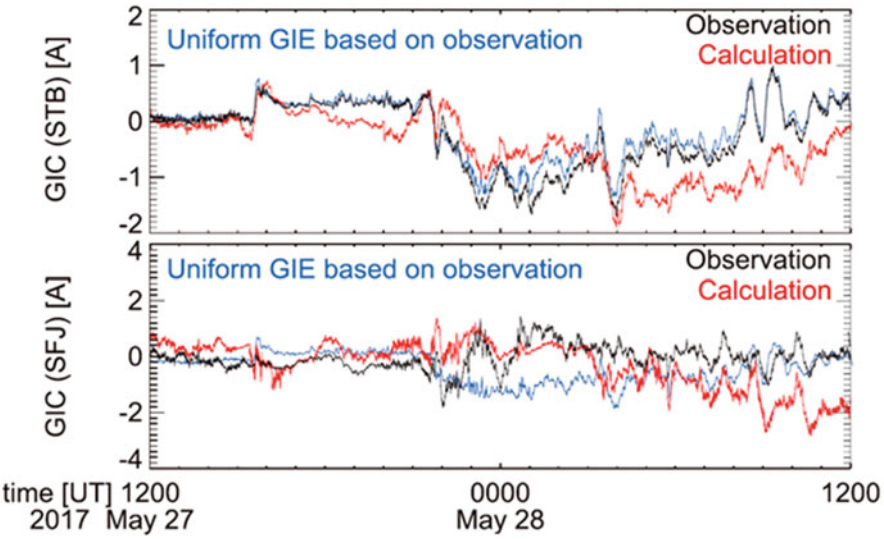

Fig. 6.11 Comparison of actually observed and numerically calculated GIC (Nakamura et al. 2018). Black, observed; red, calculated (using three-dimensional model; based on Fig. 6.10); blue, calculated (using uniform structure model); STB, Shin-Tsukuba Observatory; SFJ, Shin-Fuji Observatory

As described above, the mechanism of GIC has been clarified to some extent by studies in the last 10 years; however, there are still some problems to be solved for the reproduction of actual GIC and the estimation of possible GIC in the future. Further studies are required to improve the accuracy of GIC reproduction and the prediction capability of GIC against extremely large geomagnetic storms. Another issue that has not been discussed well this far is the effect of the complex magnetospheric ionospheric current system that emerges during an extremely large geomagnetic storm. In addition, the interaction between the ground and transmission lines (i.e., electromagnetic coupling between ground and transmission lines) has not been fully discussed. Field and numerical experiments will help understand the coupling and evaluate the generation of GIC more precisely.

6.2 GIC Observations

GIC observations are necessary to evaluate the impact of GIC on electrical power grids. It is also important to verify the results of GIC simulations with actual observation data (Viljanen 2011; Sun and Blach 2019; Nakamura et al. 2018). Besides direct observations, GIC can also be estimated from geomagnetic or geoelectric field measurements considering the configurations of power grids and

Fig. 6.12 Example of a clamp-type current probe used for GIC observations

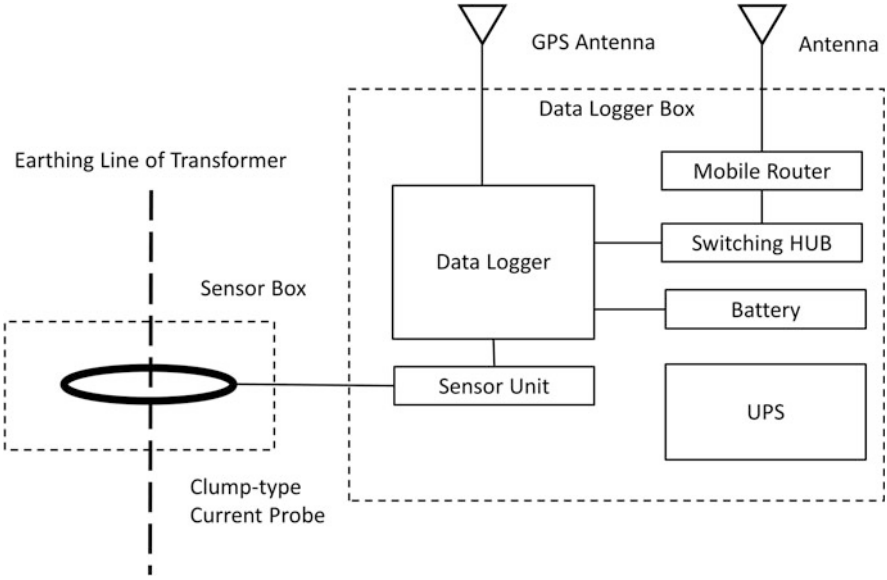

Fig. 6.13 Example of a block diagram of a GIC observation equipment

underground conductivity structures (Pulkkinen et al. 2007; Boteler and Pirjola 2017).

As a method to directly observe GIC, it is common to attach a clamp-type current probe using a Hall element or a fluxgate magnetic sensor to the earthing line of a power transformer without stopping its operation. A clamp-type current probe used for GIC observations and a block diagram of an observation equipment are shown in Figs. 6.12 and 6.13, respectively.

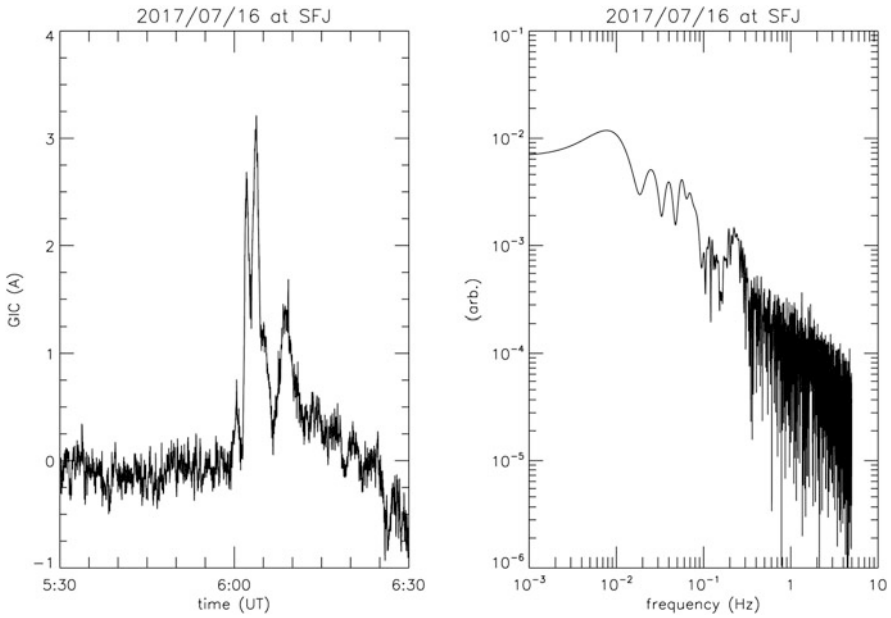

Fig. 6.14 GIC and its frequency spectrum observed at SFJ around Tokyo, Japan, associated with SSC on July 16, 2017

The following points should be considered for GIC observations (Watari et al. 2021). It is desirable to sample data as fast as possible to capture the fine variations in GIC. It is known that large GIC are observed in association with rapid geomagnetic fluctuations called storm sudden commencements (SSCs) or sudden impulses (SIs) caused by the arrival of interplanetary shocks (Kappenman 2003). Figure 6.14 shows the GIC variations associated with an SSC that occurred on July 16, 2017, and its frequency spectrum. Although there is a trade-off between the amount of data and the sampling rate owing to the data storage capacity, a sampling rate of 1 s or more is desirable for GIC observations.

To evaluate the effect of the configuration of a power grid on GIC, it is necessary to conduct GIC measurements to the maximum extent in its adjacent substations. To evaluate the inflow and outflow of GIC at each substation, it is also necessary to observe the polarity of the GIC. Large GIC can be observed at substations which are located at the end points of a power grid where the inflows or outflows from the adjacent substations concentrate.

To compare GIC data with geomagnetic or geoelectric field data observed with high time resolution, high precision time information of the data is required to avoid the influence of the time deviation between these data. The time information from the Global Navigation Satellite System, such as the Global Positioning System (GPS), is suitable for this. It is necessary to select a suitable method to collect the recorded data because a substation where an observation equipment is installed is typically located in the remote countryside. One method is to use the Internet by a cellular network.

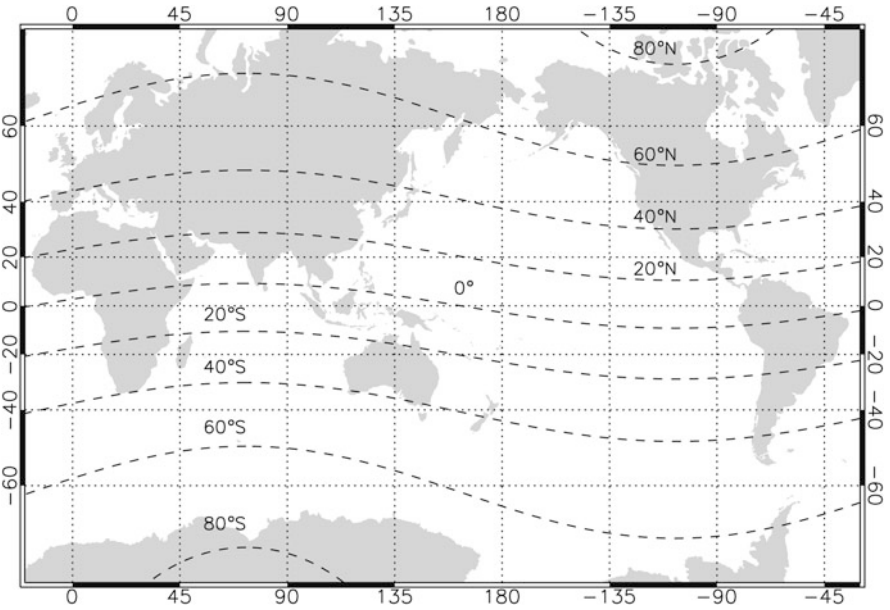

Fig. 6.15 Global map of geographic and geomagnetic latitudes

The Internet enables obtaining the above data in quasi-real time, which reduces the restriction of the sample rate due to the data storage capacity of an equipment. It also allows remotely obtaining the status information of the equipment and restarting it when a problem occurs in it. A GIC observation equipment needs electricity to operate. It is necessary to consider the use of solar panels if it is difficult to obtain electricity at the installation sites. To capture the GIC variations in response to various geomagnetic disturbances such as large magnetic storms, it is necessary to conduct long-duration observations.The impacts of GIC on power grids were previously believed to occur in regions of high geomagnetic latitudes, such as northern Europe and North America. However, a power transformer failure occurrence in New Zealand associated with the November 6, 2001, magnetic storm (Marshall et al. 2012) and thermal damage of a transformer in South Africa following the major magnetic storm at the end of October 2003 called the Halloween event (Gaunt and Coetzee 2007) have been reported. These have led the impacts of GIC on power grids in mid- and low-latitude countries to become concerning; consequently, observations of GIC have been conducted in such countries. Furthermore, it has been reported that large GIC may occur near the equator with the arrival of interplanetary shocks, based on analysis using geomagnetic data (Carter 2015). Figure 6.15 shows a global map of the geographic and geomagnetic latitudes. Examples of GIC observations worldwide are summarized below.

Japan. Observations were conducted at the Memanbetsu (MMB) substation in Hokkaido from December 2005 to March 2008 (Watari et al. 2009; Watari 2015). Although it is generally believed that GIC fluctuations correspond to the time

derivative of geomagnetic field fluctuations, the above observations showed GIC fluctuations that were similar to geomagnetic field fluctuations. Since February 2017, observations have been conducted at three substations around Tokyo and at a substation in the Yamaguchi Prefecture in Japan (Nakamura et al. 2018; Watari et al. 2021). Figure 6.16 shows the GIC observation data during the May 5, 2018, magnetic storm at the Shin-Fuji (SFJ), Shin-Tsukuba (STB), and Shin-Fukushima (SFS) substations around Tokyo with geomagnetic and geoelectric field data observed at the Kakioka Magnetic Observatory.

China. GIC measurements were started in 2004 at the Ling'ao Power Plant near Hong Kong. Specifically, a GIC monitoring device automatically starts recording data when the output of the Hall current sensor exceeds a predetermined value. A GIC value of 55.8 A was observed during the intense magnetic storm on November 9–10, 2004 (Liu et al. 2009).

New Zealand. The current flowing through the earthing lines of power transformers has been measured to monitor the return currents of a high-voltage DC (HVDC) link between the south and north islands of New Zealand. A transformer failure due to the operation of protective relays at approximately 2 universal time (UT) on November 6, 2001, was analyzed using the above data. It was clarified that this failure was caused by rapid fluctuations of the geomagnetic field due to the arrival of an interplanetary shock (Marshall et al. 2012).

Australia. The Hall-effect current sensors (LEM model LT505-S) have been installed at several substations to observe GIC with 1-minute sampling. GIC of approximately 4–5 A were observed during the magnetic storm on July 14–15, 2012 (Marshall et al. 2013).

Brazil. The Brazilian National Institute for Space Research has observed GIC since 2004 in cooperation with the power company. Because the geoelectric field tends to be more dominant in the east–west direction than in the north–south direction, the GIC observations have been conducted in the 150–200-kilometer-long 500-kV power lines between Itumbiara (18.4°S, 49.2°W) and São Simão (19.0° S, 50.5°W) and between Pimenta (20.5°S, 45.8°W) and Barreiro (20.0°S, 44.0°W). GIC of 15–20 A were observed during the magnetic storm on November 7–10, 2004 (Trivedi et al. 2007).

South Africa. In South Africa, GIC observations have been conducted at Grassridge (33.7°S, 25.6°E) and Hartebeesthoek (27.7°E, 25.9°S) substations since 2001 as part of the SUNBURST project of the Electrical Power Research Institute (EPRI) (Koen and Gaunt 2003; Ngwira et al. 2008). The SUNBURST project is a collaborative project by the EPRI for monitoring GIC and their impact on power grids in near real time (http://sunburstproject.net). At the time of a regular maintenance after the major magnetic storm in the end of October 2003 (called the Halloween event), increments in the dissolved gas in the insulation oils of transformers were found. Some of them were found to be thermally damaged (Gaunt and Coetzee 2007).

Spain. GIC were measured at Vandellos in north Spain at a sample rate of 1 kHz by a Hall-effect current sensor. The data were recorded with the time information obtained from the GPS and transmitted via a cellular network (Torta et al. 2012).

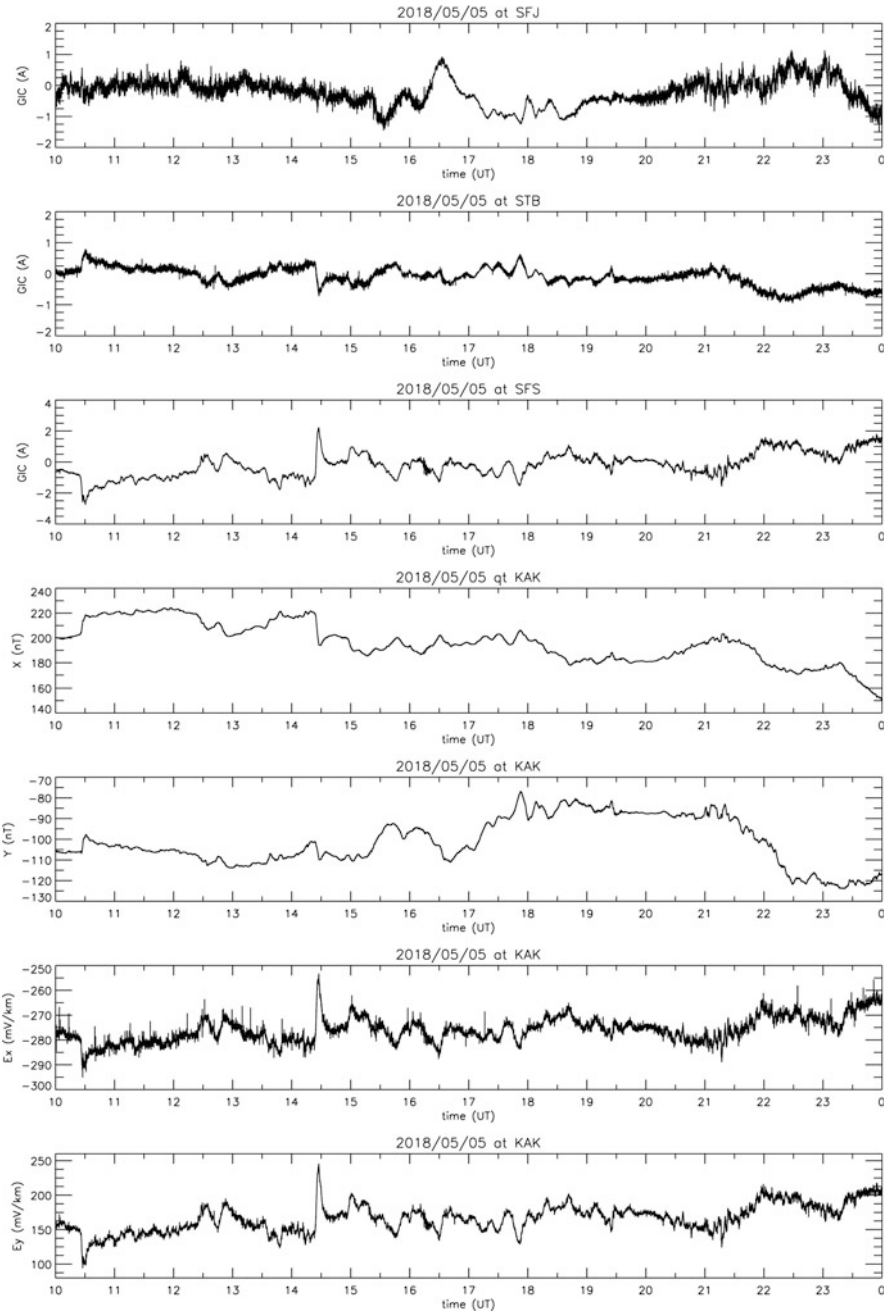

Fig. 6.16 GIC observed at substations (SFJ, STB, and SFS) around Tokyo, Japan, during May 5, 2018, magnetic storm and X- and Y-components of geomagnetic and geoelectric field data at Kakioka Magnetic Observatory

United Kingdom. ScottishPower has installed GIC observation equipment using a Hall-effect sensor with a low-pass filter on the input to remove high frequencies at three locations in a 400-kV network. When the magnitude of GIC exceeds ±5 A, the system voltage and current are recorded by a high-speed sampling rate of 0.1 s or 6.4 kHz. In addition, a monitor for the dissolved gas in the insulation oils of transformers has been installed (Breckenridge et al. 2001). The British Geological Survey provides geomagnetic observation data to the power company.

Sweden. A power blackout occurred in Malmö, Sweden, on October 30, 2003, during the Halloween event (Pulkkinen et al. 2005). One-minute GIC values have been observed at a Simpevarp 400-kV transformer. A GIC value of 150 A was observed on October 29, 2013 (Wik et al. 2009).

Finland. The Finnish Meteorological Institute, in cooperation with power companies, has conducted GIC observations on a 400-kV network for approximately 30 years. The largest GIC value until now has been 201 A (1-min value), which was observed at the Rauma transformer during the magnetic storm on March 24, 1991 (Pirjola et al. 2007).

Russia. GIC observations have been conducted on 110–330-kV networks in the Kola Peninsula, Russia (Sakharov et al. 2007). Using 10-s GIC data, Belakhovsky et al. (2019) showed several examples of large GIC due to locally occurring rapid geomagnetic variations at high latitudes, such as daytime traveling convection vortices.

Canada. During the magnetic storm on March 13, 1989, GIC caused a disruption in the power grid of the Canadian power company, Hydro-Québec, resulting in a power blackout in Quebec. GIC larger than 110 A were observed at a Radisson transformer during the March 24, 1991, magnetic storm (Bolduc 2002). The Canadian power company, Hydro One, has installed GIC observation equipment at 17 locations on its power networks and magnetometers at six locations (Marti and Yiu 2015).

United States of America. Dominion Virginia Power, an electric power company in the United States of America (USA), participated in the SUNBURST project of the EPRI to observe GIC on its 500-kV network 2 years after a power blackout was caused by GIC in Canada. In addition to GIC monitors, harmonics, dissolved gas, and temperature monitors have been installed on its transformers (Sun et al. 2015).

6.3 Magnetospheric Ionospheric Current Systems and Geomagnetic Field Fluctuations

A dynamo is activated in the outer magnetosphere by the interactions between the solar wind and the magnetosphere, and it supplies energy to the inner magnetosphere and the ionosphere extending from the polar to equatorial latitudes, generating magnetic storms and polar magnetic storms (substorms). A dynamo developed by the southward interplanetary magnetic field (IMF) generates dawn-to-dusk

convection electric fields and the Region 1 field-aligned currents (FACs) (R1FACs). The convection electric field develops ring currents (RCs) around the Earth, causing magnetic field decreases characteristic of magnetic storms at mid- and low latitudes. The partial ring current (PRC) generates dusk-to-dawn shielding electric fields and the Region 2 FACs (R2FACs) as the inner magnetospheric dynamo. The R1FACs and R2FACs flow into the high- and low-latitude sides of the auroral zone, respectively. The resulting electric fields reinforce each other in the auroral zone, producing AEJs composed of eastward Hall currents in the afternoon and westward Hall currents from midnight to morning. This produces strong magnetic field variations (500–3000 nT) on the ground. In the main phase of a geomagnetic storm and the growth phase of a substorm, the southward IMF enhances the convection electric field, and two-vortex Hall currents (DP2 currents) flow in the high- to mid-latitude ionospheric E layer surrounding the R1FAC. This R1FAC flows through the mid-latitude E layer and produces jet currents (EEJs) in the equatorial E layer where the electrical conductivity is anomalously large. In the expansion phase of a substorm and the recovery phase of a magnetic storm initiated by the IMF turning northward, the DP2 current is weakened (shielding) owing to the predominance of the R2FAC. Overshielding frequently occurs when the shielding electric field exceeds the convection electric field, and the direction of the electric field and current from the mid-latitudes to the equator is reversed. At the equator, the EEJ becomes a counterelectrojet (CEJ). The ionospheric currents between high latitudes and the equator are transmitted at the speed of light by electromagnetic waves in the Earth–ionosphere waveguide/transmission line, which is composed of the ionospheric E layer and the ground (Kikuchi 2014). Consequently, similar geomagnetic variations occur simultaneously at high latitudes and the equator. In the mid- and low latitudes where Japan is located, magnetic field fluctuations due to magnetopause currents, RCs, and FACs are dominant because the magnetic fields due to the ionospheric currents are small. Ground magnetic field variations include magnetic storms and substorms, which are mainly caused by the IMF variations, geomagnetic SCs (1–10 min) due to the solar wind pressure variations, and geomagnetic pulsations (1–10 min) due to the solar wind pressure variations and magnetospheric plasma instability. Because of the contributions of currents from these different sources, the temporal variations and magnitudes of the ground magnetic variations are highly dependent on the latitude and the local time. After reaching the ground, geomagnetic variations propagate into the interior of the Earth, inducing electric field and current. The induced electric field makes the alternating currents (GIC) flow to power transmission lines through grounded transformers. Consequently, the electromagnetic energy emitted from the magnetospheric dynamo is transported to the ground. The electric field moves the ionospheric F-layer plasma over the globe from high latitudes to the equator, causing increases in the total electron content and plasma instabilities. Therefore, understanding the magnetosphere–ionosphere current system helps understand the mid- and low-latitude ionospheric variability caused by magnetic storms and substorms.

For details of this paper, please refer to AGU Geophysical Monograph 260, Chap. 14 (Kikuchi 2021).

6.3.1 Convection Electric Field, R1FAC, and Shielding Electric Field, R2FAC

Figure 6.17a shows the averaged distribution of FACs observed by low-altitude satellites passing through the polar ionosphere. When the IMF is southward, two pairs of FACs appear at auroral latitudes. The poleward pair (downward in the morning and upward in the afternoon) is R1FACs, and the low-latitude pair (upward in the morning and downward in the afternoon) is R2FACs. Figure 6.17b shows the averaged distribution of the ionospheric potential observed by the Super Dual Auroral Radar Network (SuperDARN) radars. Two vortex patterns—high potential in the morning and low potential in the afternoon—are noticeable when the IMF is southward. When the IMF turns northward after a long period of southward IMF, the convection electric field decreases, and the R2FACs become relatively strong, resulting in an overshielding electric field opposite to the convection electric field. Consequently, reverse flow vortices are generated on the low-latitude sides of the two vortices, and a pattern of four vortices is observed (Kikuchi et al. 2010). The ionospheric convection pattern is consistent with the Hall current pattern because it represents the electron motion in the ionospheric E layer. This result is consistent with the ionospheric equivalent current pattern of quasiperiodic DP2 magnetic field fluctuations (Nishida 1968).

6.3.2 Quasiperiodic DP2 Magnetic Field Fluctuations and Substorms

Figure 6.18 shows quasiperiodic DP2 fluctuations (12–14 UT) observed from the auroral zone (Kilpisjärvi) to the equator (Mokolo) in the afternoon and the magnetic field variations caused by the subsequent substorm (14–16 UT) (Kikuchi et al. 1996). The DP2 fluctuation at Kilpisjärvi correlates with the east–west Hall current derived from the electric field and the conductivity observed by the European Incoherent Scatter (EISCAT) radars. The DP2 fluctuation also correlates well with those observed at the subauroral zone (Nurmijärvi) and the equator (Mokolo). However, DP2 fluctuations barely appear at low latitudes (Tamanrasset). The latitudinal decrease of the amplitude is due to the geometrical decay of the ionospheric currents spreading two-dimensionally from the polar region. The high correlation coefficients (0.9 with a 25 s accuracy) between the polar and equatorial DP2 fluctuations indicate that the electric field propagates almost instantaneously from high latitudes to the equator. The increase in the equatorial current is caused by the magnetic field of the Earth being horizontal. Consequently, the vertical Hall current due to the east–west electric field creates a secondary electric field, which causes the Hall current to flow in the same direction as the original electric field. This effect is called the Cowling effect (Hirono 1952).

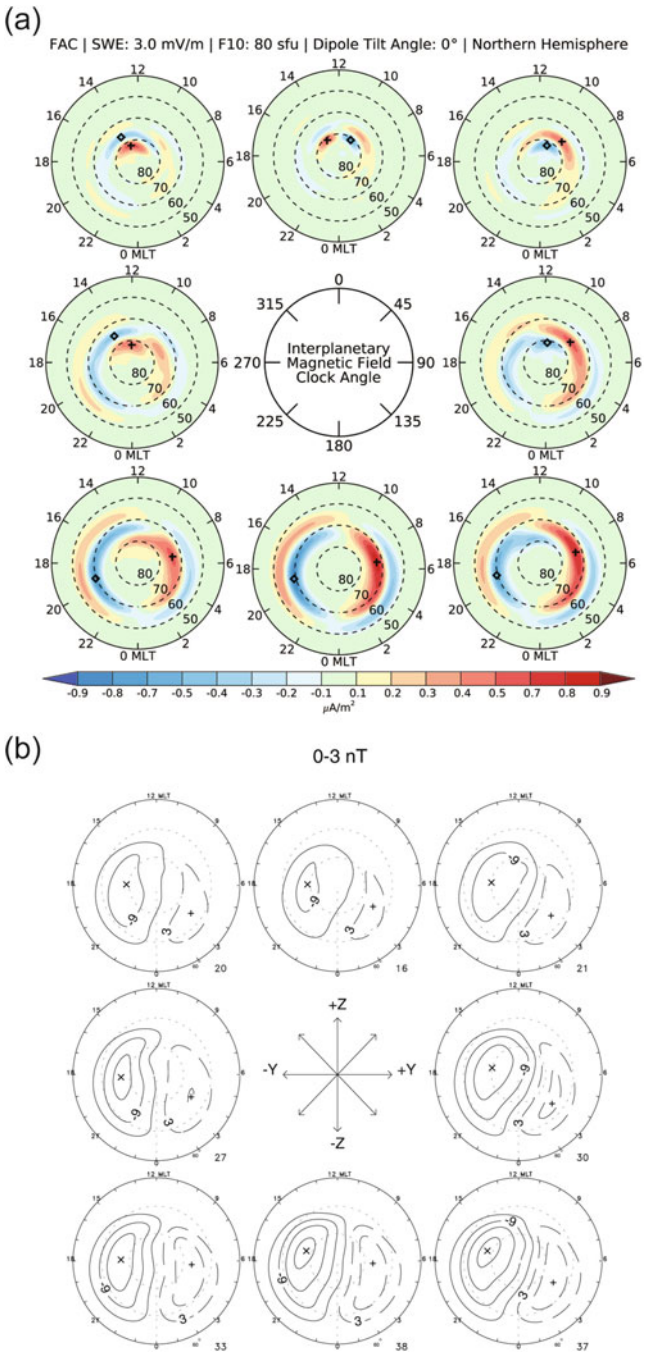

Fig. 6.17 (**a**) FACs observed by low-altitude satellites (Edwards et al. 2020) and (**b**) distribution of polar ionospheric potential observed by SuperDARN radars in Northern Hemisphere (Ruohoniemi and Greenwald 2005). Dashed and solid lines represent positive and negative potentials, respectively

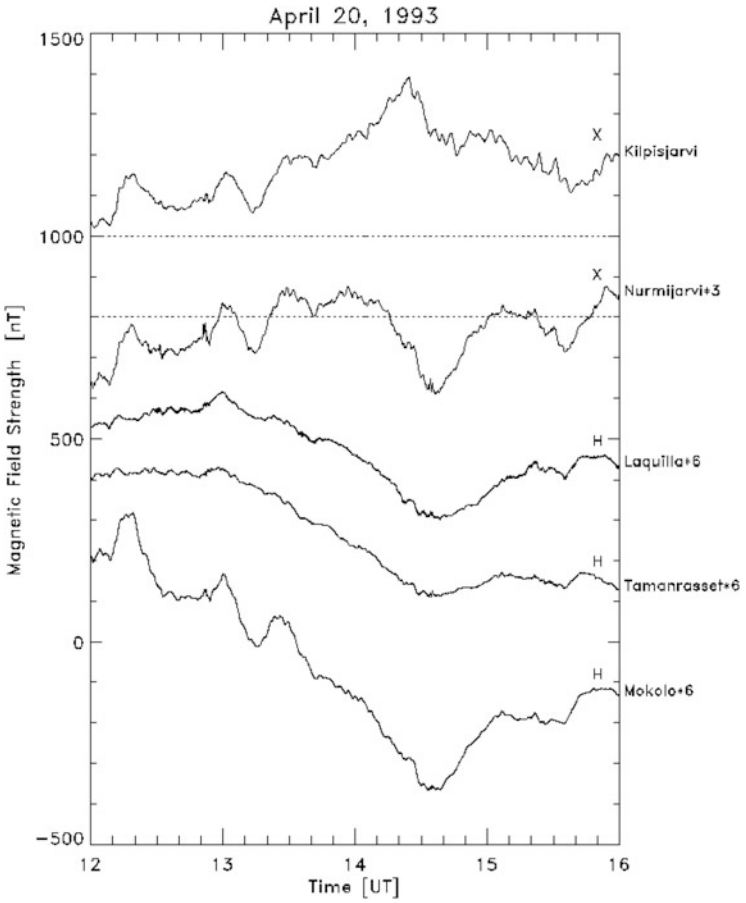

Fig. 6.18 Quasiperiodic DP2 (12–14 UT) and substorm (14–16 UT) observed in auroral zone (Kilpisjärvi), subauroral zone (Nurmijärvi), low latitude (Tamanrasset), and equator (Mokolo) (Fig. 6.2 of Kikuchi et al. (1996))

In the growth phase of substorm, magnetospheric convection, PRCs, and ionospheric DP2 currents are developed by the southward IMF. In the expansion phase, the R1FAC (substorm current wedge (SCW)) develops mainly on the nightside, and the R2FAC develops from evening to afternoon. Therefore, in Fig. 6.18, X (Kilpisjärvi) increases, whereas X (Nurmijärvi) decreases in the substorm expansion phase during 14–16 UT. This breaking of the correlation is due to the inflow of the R2FAC between the two stations. The latitudinal distribution of the magnetic field variations allows separating the convection electric field, E1, and the shielding electric field, E2, in the auroral (upper panel of Fig. 6.19a) and subauroral latitudes (lower panel) (Kikuchi et al. 2000). The R2FAC enhancement during substorms is due to the magnetic dipolarization in the magnetotail, which accumulates magnetized plasma in the near-Earth plasma sheet and generates PRCs (Tanaka et al. 2010).

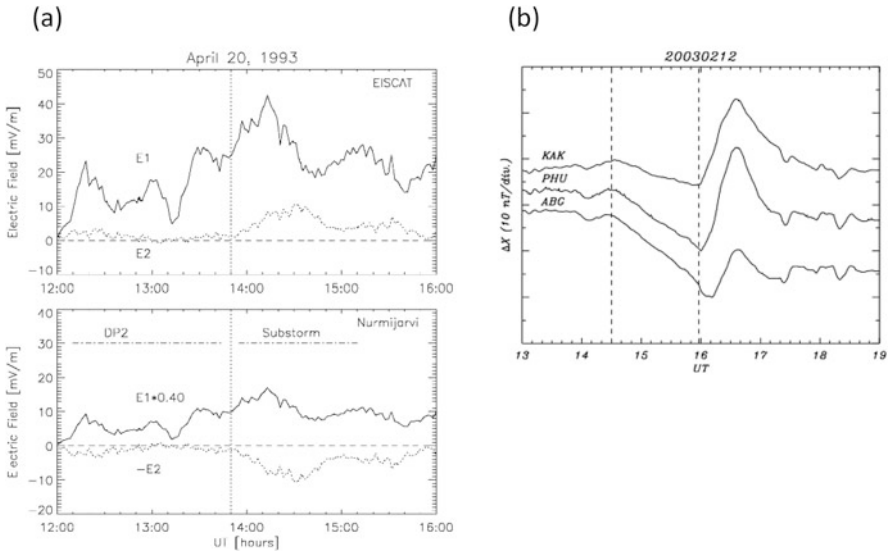

Fig. 6.19 (**a**) Convection electric field E1 and shielding electric field E2 estimated in auroral zone (EISCAT) and subauroral zone (Nurmijärvi) (Fig. 6.8 of Kikuchi et al. (2000)). (**b**) Example of substorm positive bay observed at mid-to-low latitudes at midnight (Fig. 6.9 of Hashimoto et al. (2011))

In the mid-latitudes, in the growth phase, the negative bay is caused by PRCs, which is followed by the positive bay due to the SCW in the expansion phase (Fig. 6.19b) (Hashimoto et al. 2011). The R2FAC enhancement leads to overshielding, and at the afternoon equator (Mokolo), an amplitude enhancement of the negative bay is due to the CEJ (Fig. 6.18).

Figure 6.20a schematically shows the current system that develops when the R1FAC prevails during the DP2 fluctuations and substorm growth phase and the storm main phase. The dawn-to-dusk electric field drives two-vortex Hall currents (DP2 currents) around the R1FAC at mid- and high latitudes, and the R1FAC drives the EEJ in the daytime via the mid-latitude Pedersen current (Kikuchi et al. 1996). Figure 6.20b shows the current system achieved when the PRC develops in the inner magnetosphere (Kikuchi et al. 2003). When the R2FAC prevails during the period of the IMF Bz > 0 or during the substorm expansion phase, the Hall currents surrounding the R2FAC reverse the electric field and currents at mid- and low latitudes, i.e., overshielding occurs. This results in the superimposition of the magnetic field reduction due to the westward CEJ at the dayside equator (Rastogi and Patel 1975). In contrast, the shielding electric field enhances the electric field in the auroral zone in the afternoon and evening, producing the subauroral polarization stream (SAPS) (Foster and Burke 2002). Although the duration of the shielding electric field is approximately 20–30 min (Somayajulu et al. 1987; Senior and Blanc 1984), during magnetic storms, it increases, and overshielding occurs for more than 2 h (Kikuchi et al. 2008).

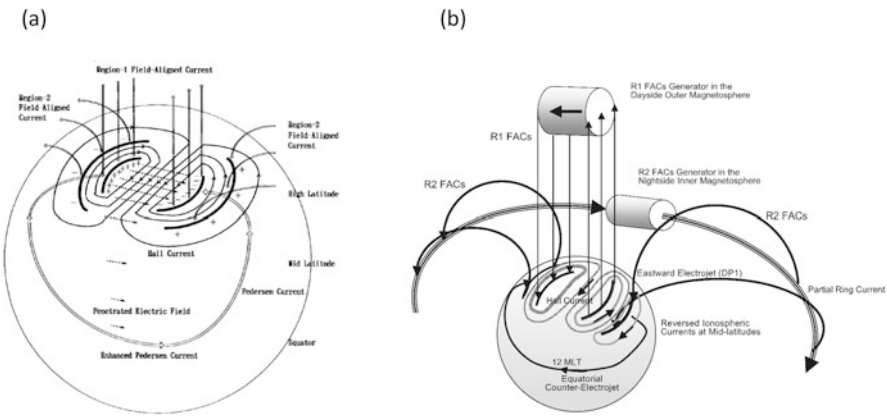

Fig. 6.20 (**a**) Schematic of current system developed during period of IMF Bz < 0. Two-vortex Hall currents around R1FAC are driven by convection electric field, and EEJ is supplied by R1FAC via mid-latitudes (Fig. 6.9 of Kikuchi et al. (1996)). (**b**) Schematic diagram of current system that develops during substorm. Reverse Hall currents are driven by R2FAC, at lower latitudes than R1FAC. R2FAC flows into equatorial ionosphere, and CEJ flows (Fig. 6.11 of Kikuchi et al. (2003))

6.3.3 Storm Main Phase Electric Field and DP2 Currents

In magnetic storms, the magnetic field increase (SC) caused by the solar wind shock is followed by a RC development by the strong southward IMF, which decreases the magnetic field at mid- and low latitudes (Ebihara and Ejiri 2000). Figure 6.21 shows a magnetic storm in which the magnetic field at low latitudes (OKI) decreases rapidly in the main phase and recovers gradually in the recovery phase (Kikuchi et al. 2008). At the magnetic equator (YAP), the amplitude is 2.7 times that of the OKI magnetic storm. As shown in Fig. 6.22a, a strong westward AEJ of 2000 nT flows in the main phase at mid-latitudes (TAR) in the morning, which moves into the auroral zone (KIL) in the recovery phase. The EEJ calculated as the difference between H(YAP) and H(OKI) in Fig. 6.21 is positive (EEJ) in the main phase and negative (CEJ) in the recovery phase (lower panel of Fig. 6.22b). The amplitude of the magnetic storm increases at the equator owing to the superposition of the main phase EEJ and the recovery phase CEJ. As shown in Fig. 6.22b (upper panel), the latitude of the AEJ rapidly moves from mid-latitudes (55° CGM, where CGM denotes corrected geomagnetic coordinates) to an auroral latitude (65° CGM), indicating that a substorm overshielding occurred in the early recovery phase. The propagation of the storm electric fields to lower latitudes is observed by the magnetometer, as shown in Fig. 6.22b, as well as by the Jicamarca incoherent radar (Fejer et al. 2007) and low-orbit satellites DMSP, C/NOFS, and ROCSAT-1 (Burke et al. 2000; Huang 2015).

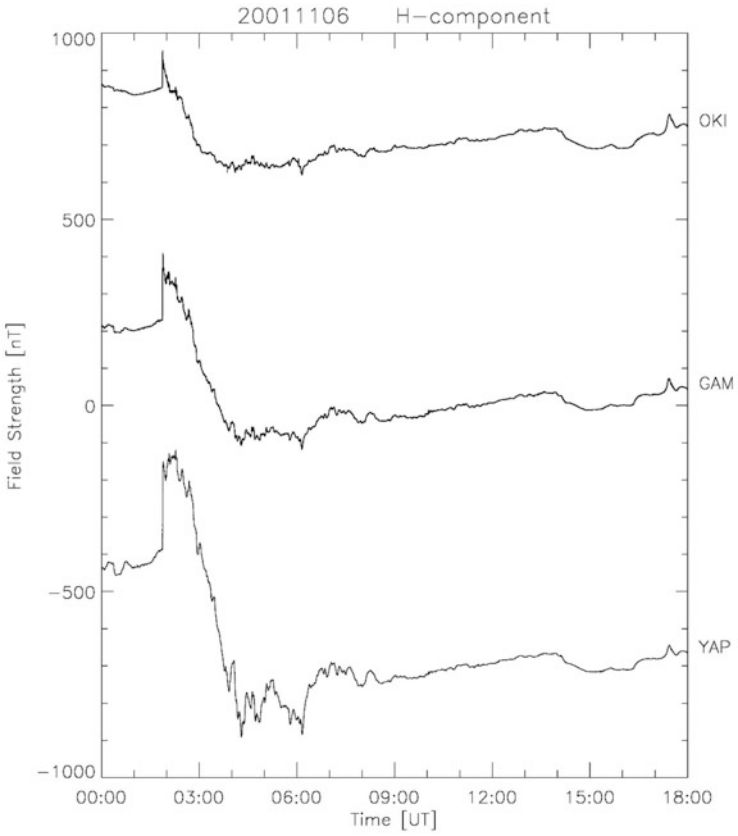

Fig. 6.21 Geomagnetic storm at low latitude (OKI, Okinawa), near magnetic equator (GAM, Guam), and magnetic equator (YAP, Yap) (Fig. 6.6 of Kikuchi et al. (2008)). Amplitude of storm is larger at YAP than that at OKI by 2.7

6.3.4 Wave Phenomena in the Magnetosphere

During magnetic storms, waves of geomagnetic pulsations Pc5 (period 150–600 s), Pi2 (period 40–150 s), and SCs (1–10 min impulse) are generated. Magnetic field fluctuations parallel to the magnetic field lines propagate perpendicular to them and are observed at mid- and low latitudes. In parallel, FACs flow into the polar ionosphere and drive DP2-type ionospheric currents. Consequently, magnetic field fluctuations occur simultaneously in the polar regions and at the equator.

Pc5 Geomagnetic Pulsation. Pc5 is observed in the auroral zone and also at mid-latitudes and the equator during magnetic storms. In the auroral zone, the phase changes 180° at a certain latitude and reverses near the noon (Samson et al. 1971). This property is attributed to the field line resonance, Kelvin–Helmholtz instability at the magnetospheric boundary (Chen and Hasegawa 1974), and magnetospheric

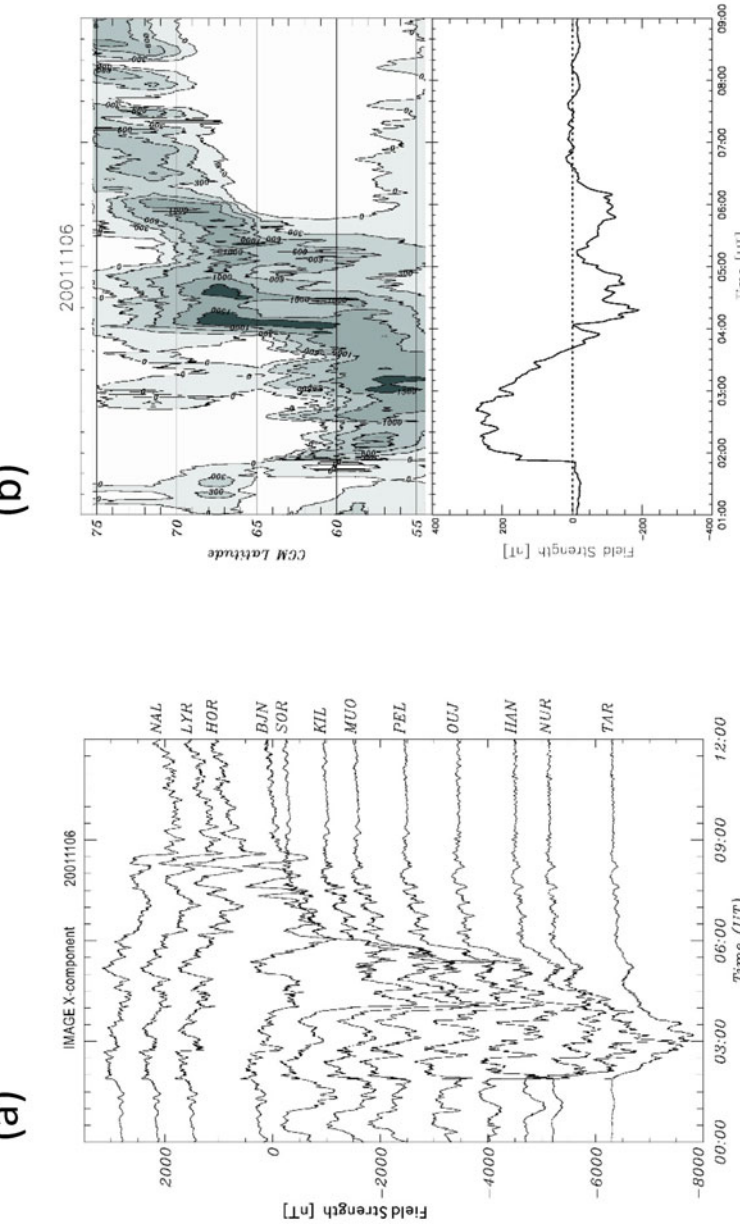

Fig. 6.22 (**a**) Magnetic field reduction due to westward AEJ observed by IMAGE magnetometer chain (Fig. 6.5 of Kikuchi et al. (2008)). (**b**) (Upper) Latitude distribution of AEJ intensity. AEJ flows in mid-latitudes in main phase and in auroral zone in recovery phase. (Lower) EEJ in main phase and CEJ in recovery phase (Fig. 6.8 of Kikuchi et al. (2008))

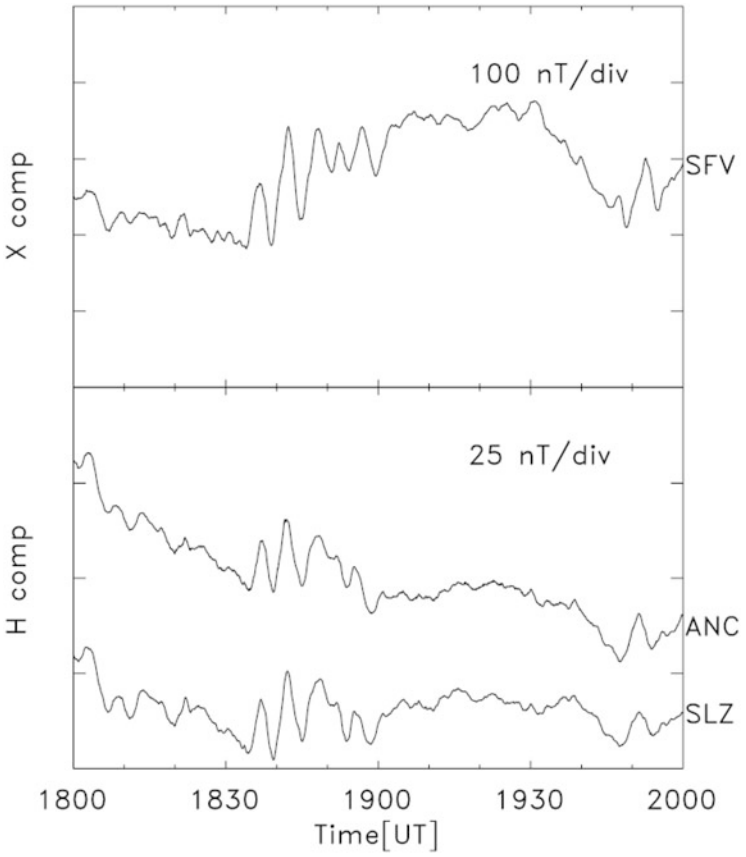

Fig. 6.23 Pc5 geomagnetic pulsations observed simultaneously at high latitudes (SFV) and equator (ANC, SLZ) (Fig. 6.2 of Motoba et al. (2002))

cavity resonance (Kivelson and Southwood 1985). Stormtime Pc5 propagates to low latitudes as compressional magnetohydrodynamic (MHD) waves excited by the solar wind pressure, with amplitude enhancement at the daytime equator (Reddy et al. 1994). In the case shown in Fig. 6.23, DP2-type ionospheric currents develop, and Pc5 occurs without time difference at high latitudes (SFV) and the dayside equator (ANC, SLZ) (Motoba et al. 2002). This suggests that the ionospheric currents were transmitted from the high latitudes to the equator instantaneously.

Geomagnetic Impulse. The magnetic field increase generated by the solar wind shock impacting the magnetosphere propagates as compressional MHD waves and is observed as a geomagnetic SC/SI at a low latitude. A typical low-latitude SC is a step-like increase, as shown in Fig. 6.24 (OKI, Okinawa) (Kikuchi et al. 2001). High and mid-latitude SCs consist of preliminary impulses (PIs, 1 min) and main impulses (MIs, 5–10 min), where the PIs are positive in the morning and negative in the afternoon and the MIs are the opposite to the PIs (Araki 1994). A typical mid-latitude

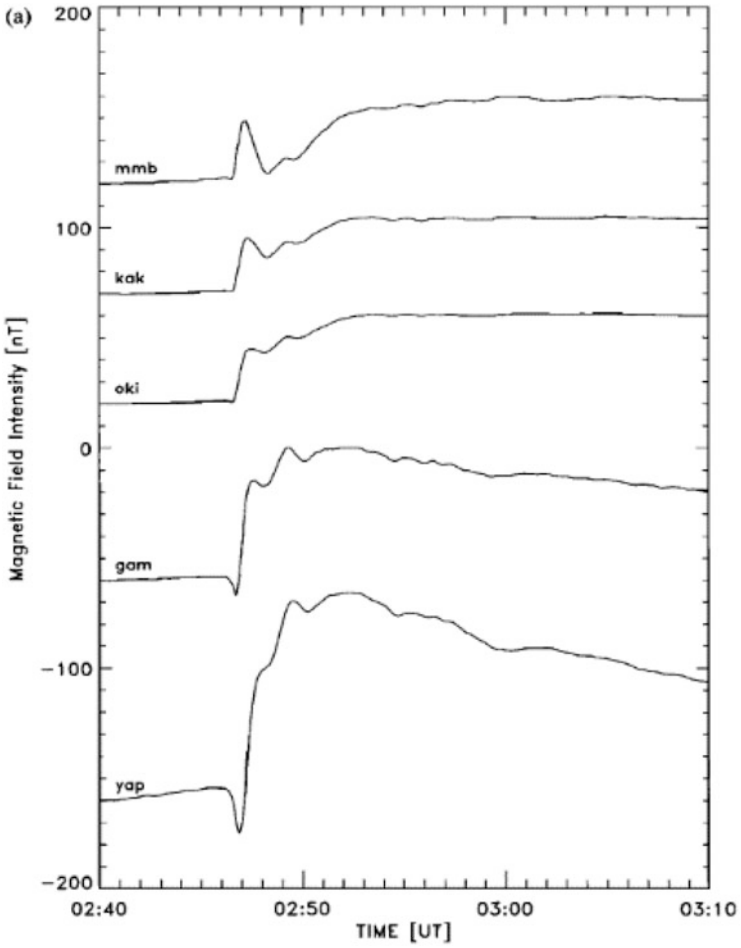

Fig. 6.24 SCs observed at mid-latitude (MMB), low latitude (OKI), and equator (YAP) on same meridian (Fig. 6.3 of Kikuchi et al. (2001))

SC observed in the morning is shown in Fig. 6.24 (Memambetsu (MMB)). At the magnetic equator during the daytime, negative PIs and positive MIs are superimposed on the low-latitude magnetic field increase, as observed at YAP in Fig. 6.24. The PIs and MIs are generated by DP2-type ionospheric currents due to dusk-to-dawn and dawn-to-dusk electric fields, respectively (Kikuchi et al. 2016, 2021a, b). The PI and MI currents are supplied by FACs, as reproduced by global MHD simulations (Slinker et al. 1999; Fujita et al. 2003). PIs provide indicators of the propagation mechanism of the global electric field and currents, because their onset times can be identified with an accuracy of less than 10 s (Araki 1994).

6.3.5 Geomagnetically Induced Currents

The magnetic field propagates from the magnetosphere and the ionosphere to the ground, and in the process of further propagation into the interior of the Earth, induced currents flow. Because the current flowing on the surface of the Earth creates a potential difference between the transformers in a power transmission line network, a quasi-DC current (called GIC) flows in the power transmission line. The GIC observed at MMB, Hokkaido, Japan, correlates well with the geomagnetic east–west component, B_y (Watari et al. 2009), as shown in Fig. 6.25 from Kikuchi et al. (2021a, b). Figure 6.25 shows geomagnetic fields B_x and B_y (upper panel) and the observed GIC (lower panel, solid line). The correlation coefficients of the GIC with B_x and B_y are -0.04 and 0.8, respectively, indicating a strong correlation with B_y. The reason for the strong correlation with B (not dB/dt) is that when the conductivity distribution is large in the upper layer and small in the lower layer, the induced current is concentrated in the upper layer (Pirjola 2010).

In a two-layer conductivity model (upper layer σ_1 and lower layer σ_2), the electric field, E_x, induced by B_y is given by the convolution of dB_y(t)/dt and the step response function of the two-layer conductor, $G(t)$ (Eq. (6.3)), as expressed in Eq. (6.4) (Kikuchi et al. 2021a, b).

$$G(t) = \frac{1}{\sqrt{\pi\mu\sigma_1}} \frac{1}{\sqrt{t}} \left(1 + 2 \sum_{j=1}^{n} \left(\frac{\sqrt{\sigma_1} - \sqrt{\sigma_2}}{\sqrt{\sigma_1} + \sqrt{\sigma_2}} \right)^j e^{-\frac{(jd)^2 \mu\sigma_1}{t}} \right) U(t) \tag{6.3}$$

$$E_x(t) = \int_0^t G(t-\tau) \frac{dB_y(\tau)}{d\tau} d\tau \tag{6.4}$$

Using the above equations, the correlation coefficient between the GIC and E_x is obtained as 0.98. When plotted over GIC, they are in close agreement (lower part of Fig. 6.25, dashed line). Figure 6.26 shows GIC observed during magnetic impulses, substorms, and magnetic storms (solid line) and the E_x calculated by the above two-layer model (dotted line). The correlation is high (correlation coefficient > 0.95) regardless of the type of disturbance.

6.3.6 Mechanism of Current Flow from Magnetosphere to Low-Latitude and Equatorial Ionosphere

It is assumed that the external force ($\boldsymbol{F}$) caused by the interaction between the solar wind and the magnetosphere produces a current, $\boldsymbol{F} \times \boldsymbol{B}/B^2$, where $\boldsymbol{B}$ is the magnetic field. The vector potential, $\boldsymbol{A}_\perp$, generated by this current propagates perpendicular to the magnetic field lines as a compressional MHD wave and propagates parallel to the

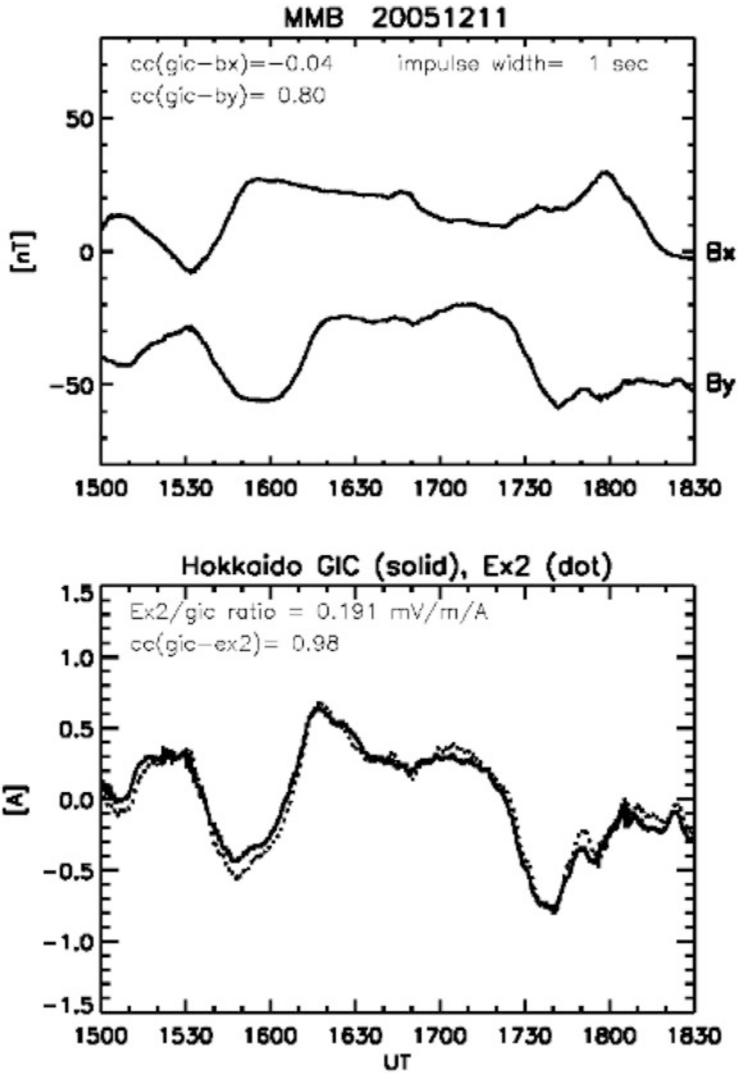

Fig. 6.25 (Upper) Geomagnetic Bx and By observed at Memambetsu (MMB). (Lower) Observed GIC (solid line) and electric field Ex2 (dotted line) induced by B_y in two-layer model

magnetic field lines as an Alfvén wave. Their propagations are expressed by Eqs. (6.5) and (6.6), respectively (Tamao 1964).

$$\left(\nabla^2 - \frac{1}{{V_A}^2}\frac{\partial^2}{\partial t^2}\right)\nabla_z \times \mathbf{A}_\perp = -\mu_0 \nabla_z \times \left(\frac{\mathbf{F}\times\mathbf{B}}{B^2}\right) \tag{6.5}$$

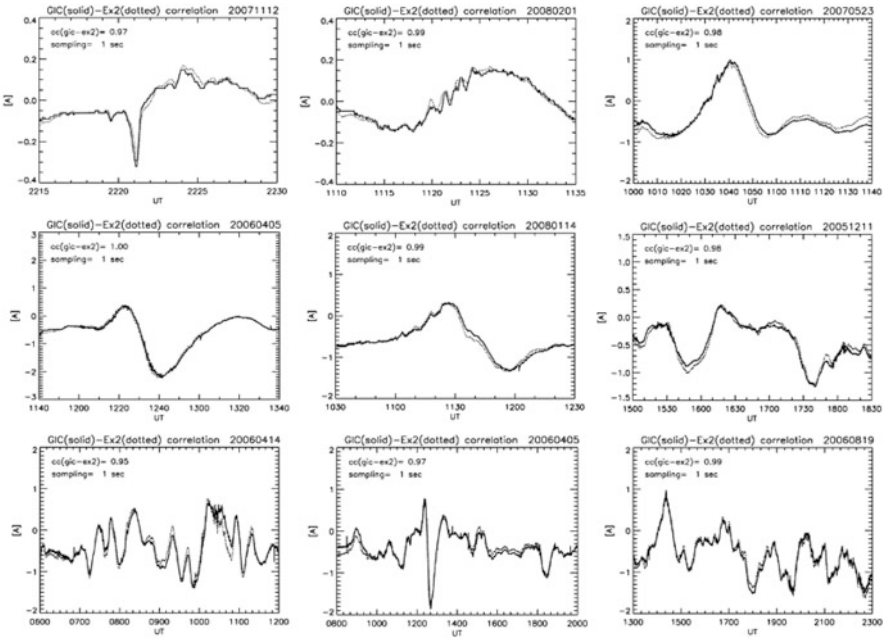

Fig. 6.26 GIC (solid line) observed during various geomagnetic disturbances and induced electric field Ex2 (dashed line) calculated with two-layer model. Cc (gic-ex2) in panel represents correlation coefficient between GIC and Ex2 (Fig. 6.7 of Kikuchi et al. (2021b))

$$\left(\frac{\partial^2}{\partial z^2} - \frac{1}{{V_A}^2}\frac{\partial^2}{\partial t^2}\right)\nabla \cdot \mathbf{A}_\perp = -\mu_0 \nabla \cdot \left(\frac{\mathbf{F}\times\mathbf{B}}{B^2}\right) \tag{6.6}$$

where V_A $(= B/(\mu_0\rho)^{1/2}$, where ρ is the mass density of the plasma, is the Alfvén speed. $\nabla_z \times \mathbf{A}_\perp$ represents the change in the magnetic field in the direction parallel to the background magnetic field, z, which causes magnetic field changes on the low-latitude ground during an SC. $\nabla \cdot \mathbf{A}_\perp$ represents the electric potential, V, created by the dynamo current due to the Lorentz transformation, $\nabla \cdot \mathbf{A}_\perp + 1/V_A^2 \cdot dV/dt = 0$, which propagates to the polar ionosphere with FACs, and produces the polar cap potential. The polar ionospheric potential generates a vertical electric field between the ground and the ionosphere, which is propagated by zeroth-order transverse magnetic (TM) mode electromagnetic waves at the speed of light to the equator (Fig. 6.27). This results in the formation of a current circuit in the ionosphere between the polar and equatorial latitudes (Fig. 6.28a). Assuming that the height of the ionosphere is h, the height-integrated electrical conductivity is Σ, and the propagation distance is d, the ionospheric currents flowing at distance x are obtained from (Kikuchi 2014)

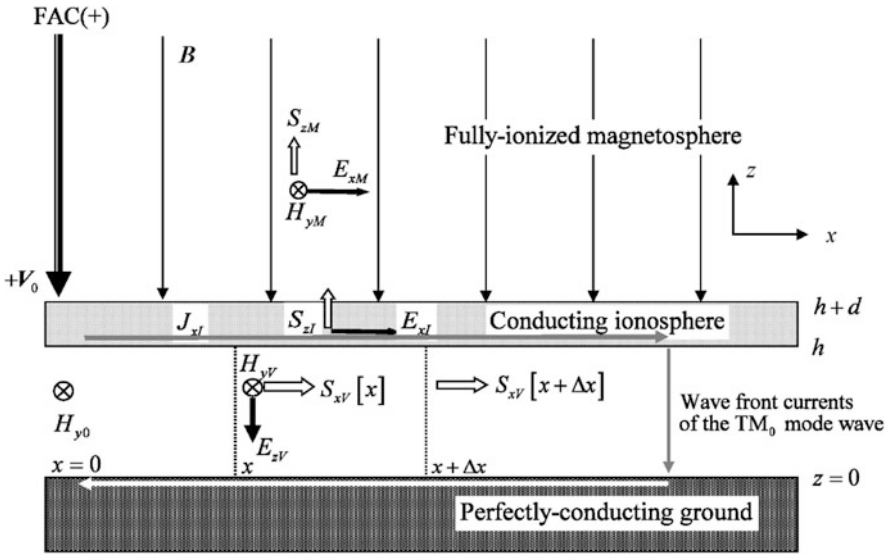

Fig. 6.27 Earth–ionosphere waveguide model, where ionospheric potential V_0 given by FAC is transmitted to equator at speed of light by TM_0 mode wave (Fig. 6.1 of Kikuchi (2014))

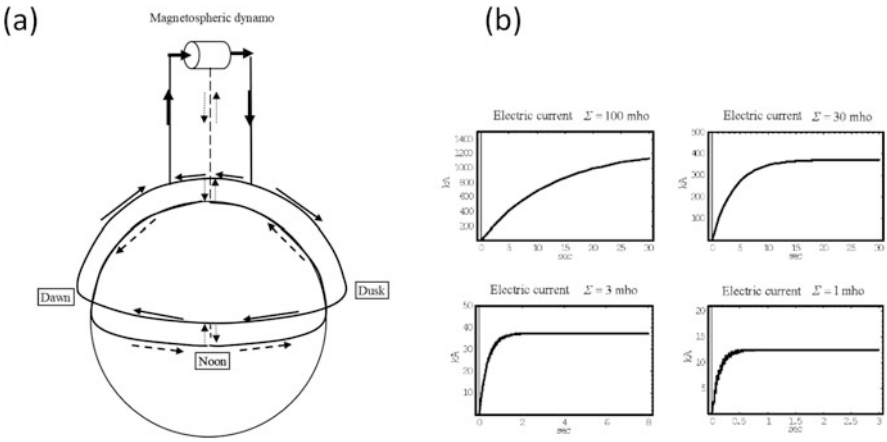

Fig. 6.28 (**a**) Magnetosphere–ionosphere–ground transmission line model in which a pair of FACs are coupled to Earth–ionosphere waveguide. Part of FACs forms circuit with polar ionosphere and another part with EEJ via mid-to-low-latitude ionosphere (Fig. 6.12 of Kikuchi (2014)). (**b**) Ionospheric currents calculated for ionospheric conductivity of 100, 30, 3, and 1 mho. Ionospheric currents increase with time constants of several to 20 s, which depends on electrical conductivity (Fig. 6.10 of Kikuchi (2014))

$$
\begin{aligned}
I(x,t) &= V_0 \frac{1}{Z} e^{-\frac{\alpha}{2}t} \left\{ \sum_{n=1}^{\infty} U(t-t_{n1}) I_0\left(\frac{\alpha}{2}\sqrt{t^2-{t_{n1}}^2}\right) \right. \\
&\quad \left. + \sum_{n=1}^{\infty} U(t-t_{n2}) I_0\left(\frac{\alpha}{2}\sqrt{t^2-{t_{n2}}^2}\right) \right\} \quad t_{n1} \\
&= \frac{2(n-1)d_2 + x}{c} t_{n2} = \frac{2nd_2 - x}{c} \alpha = \frac{1}{\mu_0 \Sigma \cdot h}
\end{aligned}
\tag{6.7}
$$

where I_0 is the modified Bessel function of the first kind of the zeroth order and Z is the characteristic impedance, which depends on the geometry of the Earth–ionosphere waveguide and is approximately 18 Ω. Figure 6.28b shows the ionospheric currents for ionospheric electrical conductivities of 1, 3, 30, and 100 S (mho). The time constant to reach a quasi-steady-state value increases with increasing conductivity and ranges from a few seconds to 20 s. The equatorial PI starts simultaneously with that in the polar regions; however, there is a delay of approximately 20 s in the peak (Takahashi et al. 2015) caused by the time constant in the growth of the current. The horizontal electric field, which is the ionospheric current divided by the electrical conductivity, propagates from the E layer to the F layer along the magnetic field lines and drives the ionospheric plasma motion. The ionospheric plasma motion is observed by incoherent scatter radars and HF Doppler sounders (Kikuchi et al. 2021a, b).

Acknowledgments We thank the Kakioka Magnetic Observatory of the Japan Meteorological Agency for providing geomagnetic and geoelectric field data.

References

Araki, T.: A physical model of the geomagnetic sudden commencement. In: Solar Wind Sources of Magnetospheric Ultra-Low-Frequency Waves, Geophysical Monograph, vol. 81, pp. 183–200. American Geophysical Union (1994)

Belakhovsky, V., et al.: Impulsive disturbances of geomagnetic field as a cause of induced currents of electric power lines. J. Space Weather Space Clim. **9**, A18 (2019). https://doi.org/10.1051/swsc/2019015

Bolduc, L.: GIC observations and studies in the hydro-Quebec power system. J. Atmos. Solar Terrestr. Phys. **64**(16), 1793–1082 (2002). https://doi.org/10.1016/S1364-6826(02)00128-1

Boteler, D.H., Pirjola, R.J.: Modeling geomagnetically induced currents. Space Weather. **15**, 258–276 (2017). https://doi.org/10.1002/2016SW001499

Breckenridge, T. H., Cumming, T., Merron, J.: Geomagnetic induced current determination and monitoring, 2001 Seventh International Conference on Developments in Power System Protection (IEE), 9–12 April 2001, Amsterdam, Netherlands (2001) https://doi.org/10.1049/cp:20010147

Burke, W.J., Rubin, A.G., Maynard, N.C., Gentile, L.C., Sultan, P.J., Rich, F.J., de La Beaujardiere, O., Huang, C.Y., Wilson, G.R.: Ionospheric disturbances observed by DMSP at middle to low latitudes during the magnetic storm of June 4-6, 1991. J. Geophys. Res. **105**, 18391–18405 (2000)

Carter, B.A.: Interplanetary shocks and the resulting geomagnetically induced currents at the equator. Geophys. Res. Lett. **42**(16), 6554–6559 (2015). https://doi.org/10.1002/2015GL065060

Chave, A.D., Jones, A.G.: The Magnetotelluric Method: Theory and Practice, p. 604. Cambridge University Press, New York (2012)

Chen, L., Hasegawa, A.: A theory of long-period magnetic pulsations, 1. Steady state excitation of field line resonance. J. Geophys. Res. **79**, 1024 (1974)

Ebihara, Y., Ejiri, M.: Simulation study on fundamental properties of the storm-time ring current. J. Geophys. Res. **105**, 15843–15859 (2000)

Edwards, T.R., Weimer, D.R., Olsen, N., Lühr, H., Tobiska, W.K., Anderson, B.J.: A third generation field-aligned current model. Journal of geophysical research: space. Physics. **125**, e2019JA027249 (2020). https://doi.org/10.1029/2019JA027249

Fejer, B.G., Jensen, J.W., Kikuchi, T., Abdu, M.A., Chau, J.L.: Equatorial ionospheric electric fields during the November 2004 magnetic storm. J. Geophys. Res. **112**, A10304 (2007). https://doi.org/10.1029/2007JA012376

Foster, J.C., Burke, W.J.: SAPS: a new characterization for sub-auroral electric fields. Eos. **83**, 393–394 (2002)

Fujita, S., Tanaka, T., Kikuchi, T., Fujimoto, K., Itonaga, M.: A numerical simulation of the geomagnetic sudden commencement: 2. Plasma processes in the Main impulse. J. Geophys. Res. **108**(A12), 1417 (2003). https://doi.org/10.1029/2002JA009763

Gaunt, C.T. and Coetzee, G.: Transformer failures in regions incorrectly considered to have low GIC-risk, 2007 IEEE Lausanne PowerTech Conference, 1-5 July 2007, Lausanne, Switzerland (2007) https://doi.org/10.1109/PCT2007.4538419

Goto, T.: Numerical studies of geomagnetically induced electric field on seafloor and near coastal zones incorporated with heterogeneous conductivity distributions. Earth Planets Space. **67**(1), 1–9 (2015). https://doi.org/10.1186/s40623-015-0356-2

Hashimoto, K.K., Kikuchi, T., Watari, S., Abdu, M.A.: Polar-equatorial ionospheric currents driven by the region 2 field aligned currents at the onset of substorms. J. Geophys. Res. **116**, A09217 (2011). https://doi.org/10.1029/2011JA016442

Hirono, M.: A theory of diurnal magnetic variations in equatorial regions and conductivity of the ionosphere E region. J. Geomag. Geoelectr. Kyoto. **4**, 7–21 (1952)

Huang, C.-S.: Storm-to-storm main phase repeatability of the local time variation of disturbed low-latitude vertical ion drifts. Geophys. Res. Lett. **42**, 5694–5701 (2015). https://doi.org/10.1002/2015GL064674

Kappenman, J.G.: Storm sudden commencement events and the associated geomagnetically induced current risks to ground-based systems at low-latitude and midlatitude locations. Space Weather. **1**(3), 1016 (2003). https://doi.org/10.1029/2003SW000009

Kikuchi, T.: Transmission line model for the near-instantaneous transmission of the ionospheric electric field and currents to the equator. J. Geophys. Res. Space Physics. **119** (2014). https://doi.org/10.1002/2013JA019515

Kikuchi, T.: Penetration of the magnetospheric electric fields to the low latitude ionosphere. In: Huang, C., Lu, G. (eds.) Space Physics and Aeronomy Collection Volume 3: Ionosphere Dynamics and Applications, Geophysical Monograph 260. © 2021 American Geophysical Union. Published by John Wiley & Sons, Inc. (2021). https://doi.org/10.1002/9781119507512.ch14

Kikuchi, T., Lühr, H., Kitamura, T., Saka, O., Schlegel, K.: Direct penetration of the polar electric field to the equator during a DP2 event as detected by the auroral and equatorial magnetometer chains and the EISCAT radar. J. Geophys. Res. **101**, 17161–17173 (1996)

Kikuchi, T., Lühr, H., Schlegel, K., Tachihara, H., Shinohara, M., Kitamura, T.-I.: Penetration of auroral electric fields to the equator during a substorm. J. Geophys. Res. **105**, 23251–23261 (2000)

Kikuchi, T., Tsunomura, S., Hashimoto, K., Nozaki, K.: Field-aligned current effects on midlatitude geomagnetic sudden commencements. J. Geophys. Res. **106**, 15555–15565 (2001)

Kikuchi, T., Hashimoto, K.K., Kitamura, T.-I., Tachihara, H., Fejer, B.: Equatorial counterelectrojets during substorms. J. Geophys. Res. **108**(A11), 1406 (2003). https://doi.org/10.1029/2003JA009915

Kikuchi, T., Hashimoto, K.K., Nozaki, K.: Penetration of magnetospheric electric fields to the equator during a geomagnetic storm. J. Geophys. Res. **113**, A06214 (2008). https://doi.org/10.1029/2007JA012628

Kikuchi, T., Ebihara, Y., Hashimoto, K.K., Kataoka, R., Hori, T., Watari, S., Nishitani, N.: Penetration of the convection and overshielding electric fields to the equatorial ionosphere during a quasiperiodic DP 2 geomagnetic fluctuation event. J. Geophys. Res. **115**, A05209 (2010). https://doi.org/10.1029/2008JA013948

Kikuchi, T., Hashimoto, K.K., Tomizawa, I., Ebihara, Y., Nishimura, Y., Araki, T., Shinbori, A., Veenadhari, B., Tanaka, T., Nagatsuma, T.: Response of the incompressible ionosphere to the compression of the magnetosphere during the geomagnetic sudden commencements. J. Geophys. Res. Space Physics. **121** (2016). https://doi.org/10.1002/2015JA022166

Kikuchi, T., Chum, J., Tomizawa, I., Hashimoto, K.K., Hosokawa, K., Ebihara, Y., Hozumi, K., Supnithi, P.: Penetration of the electric fields of the geomagnetic sudden commencement over the globe as observed with the HF Doppler sounders and magnetometers. Earth Planets Space. **73**(1), 1–3 (2021a). https://doi.org/10.1186/s40623-020-01350-8

Kikuchi, T., Ebihara, Y., Hashimoto, K.K., Kitamura, K., Watari, S.-I.: Reproducibility of the geomagnetically induced currents at middle latitudes during space weather disturbances. Front. Astron. Space Sci. **8**, 759431 (2021b). https://doi.org/10.3389/fspas.2021.759431

Kivelson, M.G., Southwood, D.J.: Resonant ULF waves: a new interpretation. Geophys. Res. Lett. **12**(1), 49–52 (1985)

Koen, J., Gaunt, T.: Geomagnetically induced currents in the southern Africa electricity transmission network, 2003 IEEE Bologna PowerTech Conference, 23-26 June 2003, Bologna, Italy (2003) https://doi.org/10.1109/PTC.2003.1304165

Kuvshinov, A., Grayver, A., Toffner-Clausen, L., Olsen, N.: Probing 3-D electrical conductivity of the mantle using 6 years of Swarm, CryoSat-2 and observatory magnetic data and exploiting matrix Q-responses approach. Earth Planets Space. **73**(1), 1–26 (2021). https://doi.org/10.1186/s40623-020-01341-9

Lehtinen, M., Pirjola, R.: Currents produced in earthed conductor networks by geomagnetically-induced electric fields. Ann. Geophys. **3**(4), 479–484 (1985)

Liu, C., Liu, L., Yang, Y.: Monitoring and modeling geomagnetically induced currents in power grids of China, 2009 Asia-Pacific Power and Engineering Conference, 27-31 March 2009, Wuhan, China (2009) https://doi.org/10.1109/APPEEC.2009.4918502

Marshall, R.A., et al.: Geomagnetically induced currents in the New Zealand power network. Space Weather. **10**(8), S08003 (2012). https://doi.org/10.1029/2012SW000806

Marshall, R.A., et al.: Observations of geomagnetically induced currents in the Australian power network. Space Weather. **11**(1), 6–16 (2013). https://doi.org/10.1029/2012SW000849

Marti, L., Yiu, C.: Real-time management of geomagnetic disturbances: hydro One's eXtreme space weather control room tools. IEEE Electrific Magaz. **3**(4), 46–51 (2015). https://doi.org/10.1109/MELE.2015.2480637

Motoba, T., Kikuchi, T., Lühr, H., Tachihara, H., Kitamura, T.-I., Hayashi, K., Okuzawa, T.: Global Pc5 caused by a DP2-type ionospheric current system. J. Geophys. Res. **107**, A2 (2002). https://doi.org/10.1029/2001JA900156

Nakamura, S., Ebihara, Y., Fujita, S., Goto, T., Yamada, N., Watari, S., Omura, Y.: Time domain simulation of geomagnetically induced current (GIC) flowing in 500-kV power grid in Japan including a three-dimensional ground inhomogeneity. Space Weather. **16**, 1946–1959 (2018). https://doi.org/10.1029/2018SW002004

Ngwira, C.M., et al.: Improved modeling of geomagnetically induced currents in the South African power network. Space Weather. **6**(11), S11004 (2008). https://doi.org/10.1029/2008SW000408

Nishida, A.: Coherence of geomagnetic DP2 magnetic fluctuations with interplanetary magnetic variations. J. Geophys. Res. **73**, 5549–5559 (1968)

Pirjola, R.: Derivation of characteristics of the relation between geomagnetic and geoelectric variation fields from the surface impedance for a two-layer earth. Earth Planets Space. **62**, 287–295 (2010)

Pirjola R.J., Viljanen, A.T., and Pulkkinen, A.A.: Research of Geomagnetically Induced Currents (GIC) in Finland, 2007 Seventh International Symposium on Electromagnetic Compatibility and Electromagnetic Ecology, 26–29 June 2007, St. Peters-burg, Russia (2007) https://doi.org/10.1109/EMCECO.2007.4371707

Pulkkinen, A., et al.: Geomagnetic storm of 29-31 October 2003: geomagnetically induced currents and their relation to problems in the Swedish high-voltage power transmission system. Space Weather. **3**(8), S08C03 (2005). https://doi.org/10.1029/2004SW000123

Pulkkinen, A., Pirjola, R., Viljanen, A.: Determination of ground conductivity and system parameters for optimal modeling of geomagnetically induced current flow in technological systems. Earth Planets Space. **59**, 999–1006 (2007). https://doi.org/10.1186/BF03352040

Pulkkinen, A., Bernabeu, E., Eichner, J., Beggan, C., Thomson, A.W.P.: Generation of 100-year geomagnetically induced current scenarios. Space Weather. **10**, S04003 (2012). https://doi.org/10.1029/2011SW000750

Rastogi, R.G., Patel, V.L.: Effect of interplanetary magnetic field on ionosphere over the magnetic equator. Proc. Indian Acad. Sci. **82**, 121–141 (1975)

Reddy, C.A., Sudha Ravindran, K.S., Viswanathan, B.V., Krishna Murthy, D.R.K., Rao, T.A.: Observations of Pc5 micropulsation-related electric field oscillations in the equatorial ionosphere. Ann. Geophys. **12**(6), 565–573 (1994)

Ruohoniemi, J.M., Greenwald, R.A.: Dependencies of high-latitude plasma convection: consideration of interplanetary magnetic field, seasonal, and universal time factors in statistical patterns. J. Geophys. Res. **110**, A09204 (2005). https://doi.org/10.1029/2004JA010815

Sakharov, YA.A., Danilin, A.N., and Ostafiychuk, R.M.: Registration of GIC in Power Systems of the Kola Peninsula, 2007 Seventh International Symposium on Electromagnetic Compatibility and Electromagnetic Ecology, 26–29 June 2007, St. Petersburg, Russia (2007) https://doi.org/10.1109/EMCECO.2007.4371714

Samson, J.C., Jacobs, J.A., Rostoker, G.: Latitude dependent characteristics of long-period geomagnetic pulsations. J. Geophys. Res. **76**, 3675 (1971)

Senior, C., Blanc, M.: On the control of magnetospheric convection by the spatial distribution of ionospheric conductivities. J. Geophys. Res. **89**, 261–284 (1984)

Slinker, S.P., Fedder, J.A., Hughes, W.J., Lyon, J.G.: Response of the ionosphere to a density pulse in the solar wind: simulation of traveling convection vortices. Geophys. Res. Lett. **26**, 3549–3552 (1999)

Somayajulu, V.V., Reddy, C.A., Viswanathan, K.S.: Penetration of magnetospheric convective electric field to the equatorial ionosphere during the substorm of March 22, 1979. Geophys. Res. Lett. **14**, 876–879 (1987)

Sun, R., Blach, C.: Comparison between 1-D and 3-D geoelectric field methods to calculate geomagnetically induced cur-rents: a case study. IEEE Trans. Power Del. **34**(6), 2163–2172 (2019). https://doi.org/10.1109/TPWRD.2019.2905532

Sun, R., et al.: Mitigating geomagnetic disturbances: a summary of dominion Virginia power's efforts. IEEE Electrific Magaz. **3**(4), 34–45 (2015). https://doi.org/10.1109/MELE.2015.2480636

Takahashi, N., Kasaba, Y., Shinbori, A., Nishimura, Y., Kikuchi, T., Ebihara, Y., Nagatsuma, T.: Response of ionospheric electric fields at mid-low latitudes during sudden commencements. J. Geophys. Res. Space Physics. **120**, 4849–4862 (2015). https://doi.org/10.1002/2015JA021309

Tamao, T.: The structure of three-dimensional hydromagnetic waves in a uniform cold plasma. J. Geomagn. Geoelectr. **48**, 89–114 (1964)

Tanaka, T., Nakamizo, A., Yoshikawa, A., Fujita, S., Shinagawa, H., Shimazu, H., Kikuchi, T., Hashimoto, K.K.: Substorm convection and current system deduced from the global simulation. J. Geophys. Res. **115**, A05220 (2010). https://doi.org/10.1029/2009JA014676

Torta, J.M., et al.: Geomagnetically induced currents in a power grid of northern Spain. Space Weather. **10**(6), S06002 (2012). https://doi.org/10.1029/2012SW000793

Trivedi, N.B., et al.: Geomagnetically induced currents in an electric power transmission system at low latitudes in Brazil: a case study. Space Weather. **5**(4), S04004 (2007). https://doi.org/10.1029/2006SW000282

Usui, Y., Uyeshima, M., Ogawa, T., Yoshimura, R., Oshiman, N., Yamaguchi, S., et al.: Electrical resistivity structure around the Atotsugawa fault, Central Japan, revealed by a new 2-D inversion method combining wideband-MT and network-MT data sets. J. Geophys. Res. Solid Earth. **126**(4), e2020JB020904 (2021)

Viljanen, A.: European project to improve models of geomagnetically induced currents. Space Weather. **9**(7), S07007 (2011). https://doi.org/10.1029/2011SW000680

Watari, S.: Estimation of geomagnetically induced currents based on the measurement data of a transformer in a Japanese power network and geoelectric field observations. Earth Planets Space. **77**, 67 (2015). https://doi.org/10.1186/s40623-015-0253-8

Watari, S., et al.: Measurements of geomagnetically induced current in a power grid in Hokkaido, Japan. Space Weather. **7**, S03002 (2009). https://doi.org/10.1029/2008SW000417

Watari, S., Nakamura, S., Ebihara, Y.: Measurement of geomagnetically induced current (GIC) around Tokyo, Japan. Earth Planets Space. **73**, 102 (2021). https://doi.org/10.1186/s40623-021-01422-3

Wik, M., et al.: Space weather events in July 1982 and October 2003 and the effects of geomagnetically induced currents on Swedish technical systems. Ann. Geophys. **27**, 1775–1787 (2009). https://doi.org/10.5194/angeo-27-1775-2009

Chapter 7
Ionospheric Variability

Yuichi Otsuka, Hidekatsu Jin, Hiroyuki Shinagawa, Keisuke Hosokawa, and Takuo Tsuda

7.1 Propagation of Radio Waves in Ionosphere

7.1.1 Introduction

Radio waves in the shortwave band (frequencies of 3–30 MHz), which propagate over long distances by reflecting between the ionosphere and the ground, have been traditionally used for long-distance broadcasting and communication. Before the ionosphere was discovered, it had been recognized as a layer that reflects radio waves. In recent years, high-frequency radio waves that are transmitted from satellites and pass through the ionosphere are used for satellite communication and broadcasting as well as satellite positioning, such as the Global Navigation Satellite System (GNSS) including the Global Positioning System (GPS). Radio waves, which are currently indispensable in our daily life, are affected by the ionosphere in various ways. In this section, we describe the effects of the ionosphere on radio wave propagation.

Y. Otsuka (✉)
Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Aichi, Japan
e-mail: otsuka@isee.nagoya-u.ac.jp

H. Jin · H. Shinagawa
National Institute of Information and Communications Technology, Tokyo, Japan

K. Hosokawa · T. Tsuda
Department of Computer and Network Engineering, The University of Electro-Communications, Tokyo, Japan

K. Kusano (ed.), *Solar-Terrestrial Environmental Prediction*,
https://doi.org/10.1007/978-981-19-7765-7_7

7.1.2 *Reflection and Refraction of Radio Waves in Ionosphere*

Because of the presence of plasma in the ionosphere, radio waves propagating in the ionosphere are affected by their interaction with electrons and have different velocities than those propagating in vacuum. This effect is expressed by the refractive index, μ, and is known as the Appleton–Hartree equation (e.g., Hunsucker and Hargreaves 2002). The frequency of radio waves, f, is sufficiently higher than the gyrofrequency of an electron; therefore, μ is

$$\mu = \sqrt{1 - \frac{f_p^2}{f^2}} \qquad f_p = \frac{1}{2\pi}\sqrt{\frac{e^2 n_e}{\varepsilon_0 m_e}} \tag{7.1}$$

where f_p is the plasma frequency, which is a function of the electron density, n_e. e is the elementary charge, ε_0 is the permittivity in vacuum, and m_e is the mass of an electron. In vacuum, because $n_e = 0$, $\mu = 1$. As the electron density increases, μ becomes smaller. When the radio wave frequency coincides with the plasma frequency, $\mu = 0$, and thus, that radio wave is reflected.

In the ionosphere, a radio wave is refracted according to Snell's law (Fig. 7.1). If a radio wave is incident from a medium with refractive index μ_1 at an angle of incidence i_1 to a medium with refractive index μ_2, the angle of refraction,i_2, can be written as

$$\mu_1 \sin i_1 = \mu_2 \sin i_2 \tag{7.2}$$

In the lower part of the ionosphere, the electron density increases with the altitude, whereas the refractive index decreases with it. Therefore, the refraction angle of a radio wave increases with the altitude. When the angle of refraction reaches 90°, the radio wave is reflected. By setting the refractive index on the ground as 1 in Eq. (7.2), the condition under with a radio wave transmitted from the ground at the zenith angle, i_0, is reflected in the ionosphere which can be expressed as

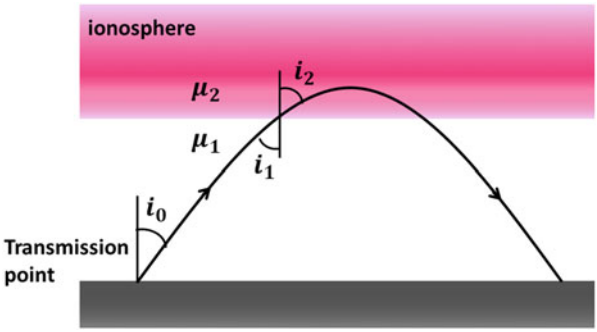

Fig. 7.1 Reflection of radio waves by ionosphere

$$f = \frac{f_p}{\cos i_0} \tag{7.3}$$

Therefore, as the zenith angle of the radio waves transmitted from the ground increases, the radio waves at higher frequencies are reflected. Thus, radio waves in the very high-frequency (VHF) band (frequencies 30–300 MHz) are also reflected by a sporadic *E* (Es) layer and can propagate for a long distance (discussed in detail in Sect. 7.7). At a certain distance, the maximum frequency at which radio waves can be reflected by the ionosphere is called the maximum usable frequency (MUF). It is used as an index for long-distance communication by radio waves.

7.1.3 Propagation Delay of Radio Waves in Ionosphere

For satellite communication and broadcasting and for satellite positioning such as the GPS, the radio waves having sufficiently higher than the plasma frequency in the ionosphere are used to ensure that they can pass through the ionosphere. If $f_p \gg f$, the refractive index, μ, can be approximated as

$$\mu = 1 - \frac{e^2 n_e}{8\pi^2 \varepsilon_0 m_e f^2} \tag{7.4}$$

The phase velocity of a radio wave is obtained as c/μ, where c is the speed of light in vacuum. Because μ is smaller than 1 in the ionosphere, the phase velocity of a radio wave is larger than c. Concurrently, the group velocity of radio waves is obtained as c/μ', where μ' is the group refractive index.

$$\mu' = 1 + \frac{e^2 n_e}{8\pi^2 \varepsilon_0 m_e f^2} \tag{7.5}$$

In satellite positioning, such as the GPS, and in the measurement methods of the altitude of the sea surface from a satellite, the time taken for a radio wave transmitted from the satellite to reach the receiver on the surface of the Earth is measured. Thus, the distance is calculated by multiplying this time with the speed of the radio wave (Misra and Enge 2006). As mentioned above, in the ionosphere, the speed of a radio wave is not constant; in fact, it depends on the electron density. When a radio wave transmitted from a satellite penetrates the ionosphere and is received on the ground, the phase delay, Δr_p, and the group delay, Δr_g, of the radio wave in distance are given as

$$\Delta r_g = -\Delta r_p = \frac{e^2}{8\pi^2 \varepsilon_0 m_e f^2} T \tag{7.6}$$

where T is the total electron content (TEC), which is the integration of the electron density along the propagation path of the radio wave. Δr_g and Δr_p are equal in magnitude but opposite in sign. The effects of the ionosphere cause Δr_g to increase whereas Δr_p to decrease. Both delays are inversely proportional to the square of the frequency of the radio wave and proportional to the TEC. Therefore, a low frequency induces a large delay. For ionospheric studies, TECU is used as a unit of the TEC, where 1 TECU = 10^{16} electron/m^2. For a radio wave with a frequency of 1.57542 MHz, which is used by the GPS as the L1 band, a TEC with a magnitude of 1 TECU causes a delay distance of approximately 16 cm. The TEC significantly varies depending on the location, time of day, and solar activity, and it ranges from 1 to 100 TECU. Therefore, the ionosphere can cause GPS positioning errors of approximately 10 m.

7.1.4 Absorption of Radio Waves in Ionosphere

When radio waves propagate through the ionosphere, electrons in the ionosphere oscillate because radio waves are accompanied by fluctuations in the electric field. In the *D* region (60–90 km altitude), which is located in the lower part of the ionosphere, electrons collide with neutral particles and lose their energy. Therefore, when the electron density in the *D* region increases, radio waves are attenuated in the ionosphere. In particular, when a solar flare occurs, ionization in the D region is increased by the rapid increase in X-rays, and shortwave radios, which typically propagate over long distances by reflecting between the ionosphere and the surface of the Earth, are absorbed. These result in a blackout that can interrupt long-distance communication for several minutes to several hours. This phenomenon is called the Dellinger phenomenon, named after the discoverer of its connection with solar flares (Dellinger 1937).

In the polar cap regions, the *D* region is ionized by the energetic protons emitted from the Sun during solar flares, and radio absorption occurs. This phenomenon is called "polar cap absorption." The polar cap absorption is discussed in detail in Sect. 7.4.4.

7.1.5 Scintillation

When there are electron density irregularities in the ionosphere, the phase and amplitude of the radio waves passing through the ionosphere may fluctuate. This

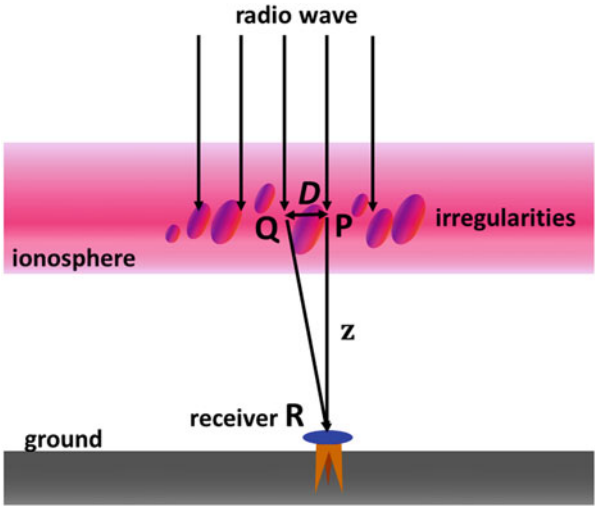

Fig. 7.2 Schematic showing principle of amplitude scintillation

phenomenon is called scintillation. Scintillation can affect satellite broadcasting, communication, and satellite positioning, such as the GPS.

As described in Sect. 7.1.2, the phase speed of the radio waves propagating in the ionosphere increases with increase in the electron density. When the electron density is spatially inhomogeneous, the phase of a radio wave differs depending on its propagation path, and the phase of the received radio wave fluctuates. This phenomenon is called phase scintillation.

In contrast, amplitude scintillation is caused by the diffraction of a radio wave due to the plasma density irregularities in the ionosphere, i.e., inhomogeneity of the refractive index, leading to fluctuations in the amplitude of the received radio wave. As shown in Fig. 7.2, we consider a case in which a radio wave transmitted from a satellite passes through the ionosphere and is received at the receiving point, R. In the case of a satellite flying at a high altitude, such as a geostationary satellite or a GPS satellite, the distance between the satellite and ionosphere is sufficiently long, and thus, its transmitted radio waves can be considered as plane waves. Radio waves passing through paths PR and QR, of which the latter is diffracted by the plasma density irregularities in the ionosphere, are received simultaneously at the receiving point, R. When the difference in the distance between paths PR and QR is an integer multiple of the wavelength, the two radio waves are in phase and, thus, strengthen each other. However, when the difference in the paths is a half of the wavelength, they weaken each other. Therefore, the amplitude of the received signal at point R fluctuates. The size of an irregularity corresponds to the Fresnel scale represented as

$$D = \sqrt{2\lambda z} \tag{7.7}$$

where z is the distance from the receiving point to the irregularity and λ is the wavelength of the radio wave. For the case of the GPS, λ is approximately 19 cm,

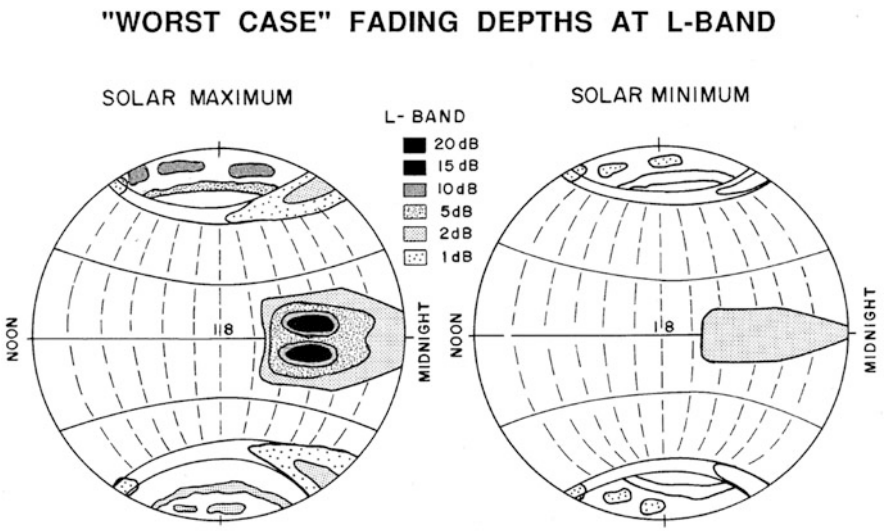

Fig. 7.3 Global distribution of scintillation occurrence during solar activity maxima (left) and minima (right) (Basu et al. 1988)

and thus, D is approximately 300 m. From Eq. (7.7), we can see that the Fresnel zone becomes larger as the wavelength of the radio wave increases. In general, it is known that an electron density irregularity with a long wavelength has large amplitude. Therefore, the scintillation intensity also depends on the wavelength of the radio wave. Intense scintillation tends to occur for low-frequency radio waves.

The global distribution of the amplitude scintillation occurrence rates is shown in Fig. 7.3. The intensity of amplitude scintillation depends on the amplitude of the electron density fluctuation. Therefore, amplitude scintillation occurs most frequently at the crest of equatorial ionization anomaly, where the electron density is the highest. The scintillation occurrence rate near the equatorial region has a remarkable positive correlation with the solar activity because of an irregularity generated inside the equatorial plasma bubble, which is an electron density disturbance in the ionosphere. As shown in the figure, scintillation occurs most frequently immediately after sunset when a plasma bubble is generated (Basu et al. 1988). The scintillation occurrence rate decreases with local time because the irregularities are dissipated by diffusion. Scintillation in the VHF band (frequency 30–300 MHz) occurs throughout night, from sunset to sunrise, whereas in the GPS L1 signal (frequency approximately 1.6 GHz), scintillation rarely occurs after midnight (Secan et al. 1985). Plasma bubbles are discussed in Sect. 7.6. At high latitudes, scintillation occurs in the polar cap regions, auroral oval on the nightside, and cusp region on the dayside. Scintillation in the polar regions is discussed in more detail in Sect. 7.4. Scintillation can be observed during geomagnetic disturbances, although it occurs less frequently at mid-latitudes than in the polar and equatorial regions (Ogawa et al. 1980).

7.2 Ionospheric Changes Caused by Solar Flares

7.2.1 Introduction

Solar radiation is the most important source of ionizing and heating energy for the upper atmosphere (ionosphere and thermosphere) at mid- and low latitudes. During solar flares, the intensity of solar radiation increases over a wide range of wavelengths, and the increase in X-rays to extreme ultraviolet (EUV) radiation enhances the ionization and molecular dissociation of the upper atmospheric components. Correspondingly, ion and electron densities in the wide altitudes of the ionosphere (D to F_2 layers) temporarily increase, causing disturbances in radio propagation, as described in the previous section. In the thermosphere, heating causes the expansion of the atmosphere, which increases the atmospheric drag of satellites. In this section, we outline these phenomena in detail using examples of numerical models developed during the Project for Solar-Terrestrial Environmental Prediction.

7.2.2 Examples of Solar Radiation Variations (X9.3 Flare in September 2017)

Before discussing the ionospheric variations caused by solar flares, we need to understand how the source of the variations in the solar radiation changes at different wavelengths. In this section, we introduce an example of an actual solar flare. For a general description of solar flares, please refer to Chap. 8. Figure 7.4 shows an

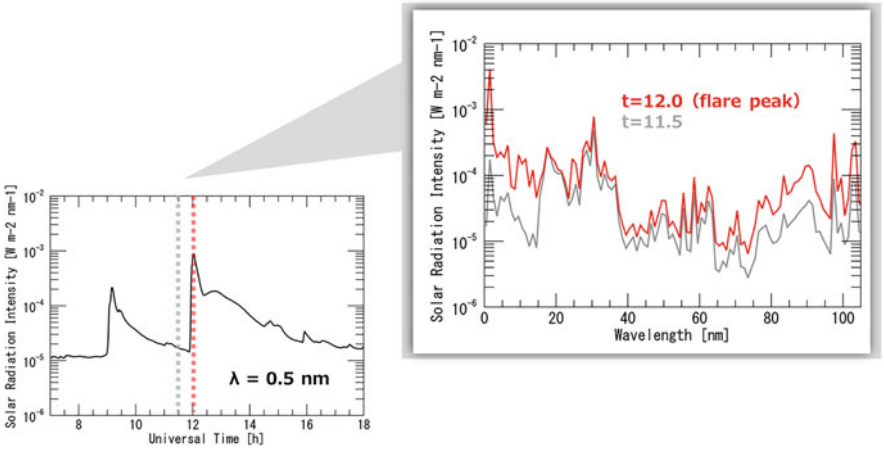

Fig. 7.4 (Left) Time variations in solar X-ray intensity at wavelength of 0.5 nm when X9.3 solar flare occurred on September 6, 2017 (FISM output and wavelength 0.5 nm). (Right) Comparison of solar radiation spectra before and after X9.3 flare (FISM output)

example of a solar flare that occurred on September 6, 2017. The solar radiation intensity shown in this figure is obtained using the Flare Irradiance Spectral Model (FISM) (Chamberlin et al. 2008), an empirical model that uses solar X-ray observations by the Geostationary Operational Environmental Satellite (GOES) as a proxy. In the considered event, two X-class flares occurred (X2.2 at 8:50 universal time (UT) and X9.3 at 11:53 UT). In the same figure (right), the radiation spectra before and after the second flare are compared. Although the changes in the emission spectrum are different for both flares, the emission intensity is increased by a factor of 2–40 in the 0.1–10 nm X-ray band and by a factor of 1.5–4 in the 30–110 nm EUV band.

7.2.3 *Absorption of Solar Radiation by Upper Atmosphere and Change in Ionization Rate*

The main atmospheric components of the thermosphere at altitudes of 100–400 km are nitrogen and oxygen molecules and atoms. The amount of ionization of these molecules and atoms in response to the injecting photon energy is represented by their photoionization cross-section. As shown in Fig. 7.5, the photoionization cross-sections of oxygen atoms/molecules and nitrogen molecules are large in the EUV band with wavelengths of 30–100 nm, and in this band, the solar radiation is absorbed at altitudes of 100 km or more. At shorter wavelengths (0–30 nm), the photoionization cross-sections are fractions to two orders of magnitude lower than at the above wavelengths; thus, the solar radiation cannot be fully absorbed by the thermosphere and reaches the mesosphere at altitudes below 100 km. Although the photoionization cross-sections are small in this wavelength range, the ionization rate due to photoelectron collisions is several times to two orders of magnitude higher than the photoionization rate due to the high energy of the produced photoelectrons (Solomon and Qian 2005). The ultraviolet–infrared solar flux at wavelengths longer than 120 nm is not absorbed in the thermosphere; instead it is absorbed in the atmosphere below the mesosphere or reaches the surface of the Earth.

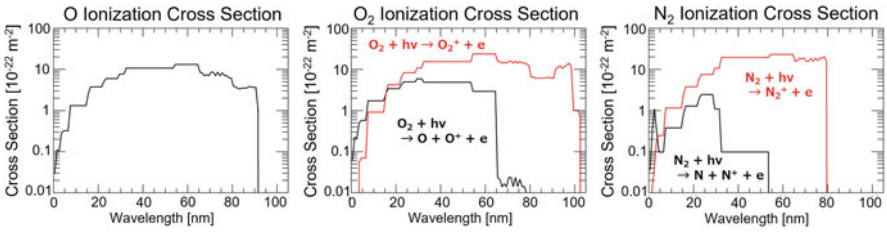

Fig. 7.5 (Left) Photoionization cross-sections of oxygen atoms, (middle) photoionization and photodissociative ionization cross-sections of oxygen molecules, and (right) photoionization and photodissociative ionization cross-sections of nitrogen molecules for each wavelength of input light. Cross-section data are based on Solomon and Qian (2005)

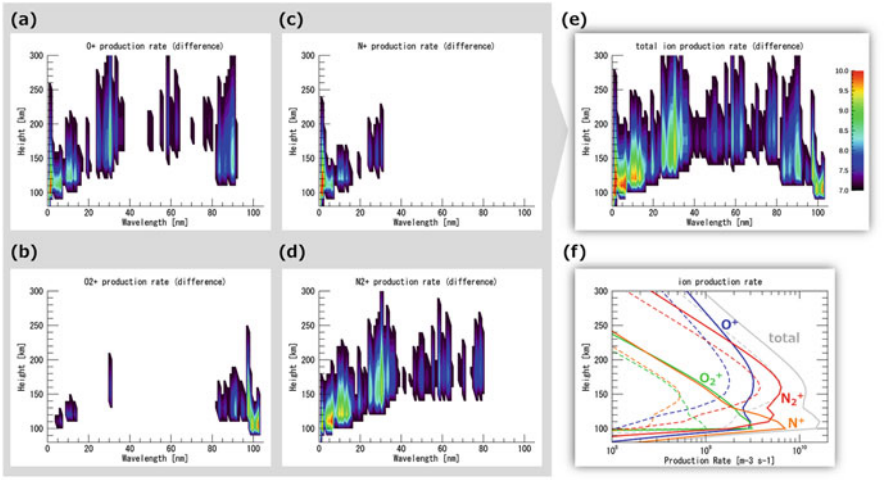

Fig. 7.6 Changes in ionization rate of each ion due to X9.3 flare in September 2017, including ionization by sunlight and secondary photoelectrons. Locations are 0°E, 30°N. (Differences in ionization rates before and after flare for (**a**) O^+, (**b**) O_2^+, (**c**) N^+, (**d**) N_2^+, and (**e**) total ions, respectively, are shown as functions of wavelength of sunlight (in Log $_{10}$(Wm^{-2} nm^{-1}).) (**f**) Changes in altitude distribution of ionization rates of each ion. Each solid line is immediately after flare (12:00 UT), and each dotted line is before the flare (11:30 UT)

Using the aforementioned solar radiation spectrum (Fig. 7.4) and the photoionization cross-section of the atmospheric components (Fig. 7.5), the wavelengths of the solar radiation that enhance ionization and the corresponding altitudes during a flare are calculated, which are shown in Fig. 7.6. The increase in the ionization rate due to solar flares is larger at lower altitudes from the *F* to *E* layers. In the example of the X9.3 flare in September 2017, the ionization rates of all ions increased by factors of 1.8–10 from 150 to 100 km altitude after the flare (Fig. 7.6f). In the F_2 layer, the ionization rates of O^+ and N_2^+ were increased mainly by the contribution of the EUV light in the wavelength range of 25–90 nm (Fig. 7.6a–e). In contrast, in the *E* layer, the ionization rates of N^+, N_2^+, and O^+ by X-rays in the wavelength range of 7 nm or less (more precisely, ionization by photoelectrons) and of O_2^+ by the EUV light in the wavelength range of 95–105 nm increase. Thus, the altitudinal range and components of the atmosphere acting on the solar radiation differ depending on its wavelength. Concurrently, the solar radiation spectra are different for different flares, and they depend on the locations of solar flares. In flares occurring near the solar limb, the X-ray intensity is similar to those in the center of the solar disk, whereas the EUV light is attenuated by the optical thickness of the solar atmosphere. Therefore, the mechanism of its effects on the upper atmosphere can change (Qian et al. 2010).

7.2.4 *Ionospheric Variability*

A case of the solar flare on September 6, 2017, is shown in Fig. 7.7 as an example of ionospheric variations occurring immediately after a solar flare. The figure is obtained by inputting the aforementioned solar radiation spectrum from the FISM (Fig. 7.4) into GAIA, which is an upper atmosphere simulation model (Jin et al. 2011; Matsumura et al. 2018). To extract the effects of the solar flare, two solar radiation spectrum models (with and without solar flares) are input and the calculation results are compared. The figure (right) shows the ion height distribution at the same location and time as the ion ionization rate in Fig. 7.6f. At the *F* layer altitude, the ionizations of O^+ and N_2^+ account for a large fraction of the total ion production, and N_2^+ contributes to the increase in the densities of O^+ and NO by chemical reactions with O and NO^+. In the E layer, the increase in the amounts of N, N, and O production is large, with that in the amounts of O and N also contributing to the increase in the densities of NO^+ and O_2^+. The rate of density increase due to the flare increases as the altitude decreases from the *F* to *E* layer (in this case, approximately 1.5–7 times from 250 to 100 km altitude). At higher altitudes, the time constant of the chemical reaction is reduced, and the effects of motion such as diffusion become more dominant. Figure 7.7 (left) shows the calculated TEC as a function of time in

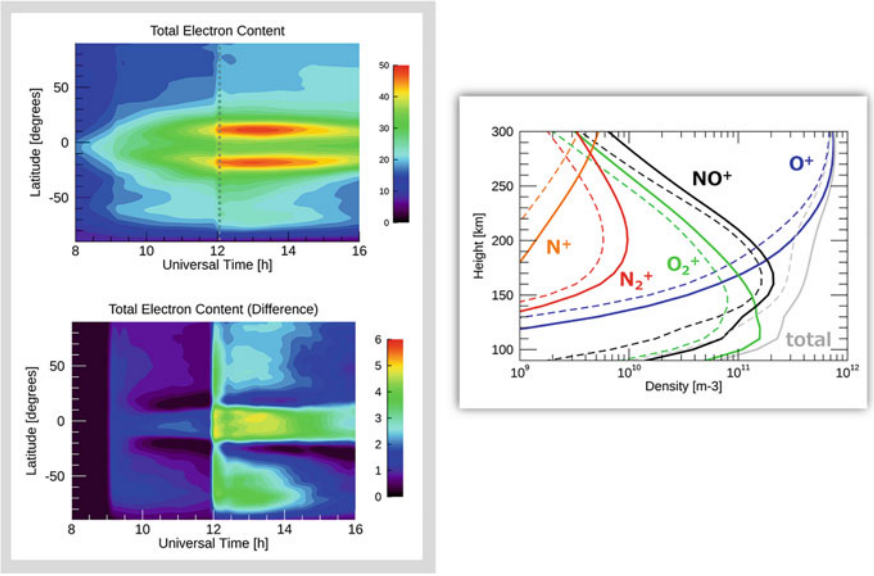

Fig. 7.7 Numerical simulation results of impacts of solar flare on ionosphere on September 6, 2017. (Top left) Variations in total electron count (TEC) (longitude 0°E). Unit is TECU (= 10^{16} electrons m^{-2}). (Lower left) Difference between TECs from calculation results with and without solar radiation due to solar flares. (Right) Comparison of altitude distributions of each ion with (solid line) and without (dotted line) solar radiation input immediately after X9.3 solar flare (12 UT). Location is (0°E, 30°N)

the broad ionosphere, presenting increments in the TEC corresponding to two flares (X2.2 at 8:50 UT and X9.3 at 11:53 UT). The increase in the ionospheric ion and electron densities is considered to depend on the solar radiation intensity of a flare, which is suggested to be correlated with the solar zenith angle observationally (Zhang and Xiao 2005). In the above calculation example, an increase of approximately 5 TECU ($= 5.0 \times 10^{16}$ electrons m^{-2}) (~12% increase) occurred immediately after the flare at the equatorial region. In comparison, observed cases showed the same level (~5 TECU) in the X5.7 flare on July 14, 2000, and an increase of up to 25 TECU (30% increase) in the X17.2 flare on October 28, 2003 (Tsurutani et al. 2006).

7.2.5 *Variations in Thermosphere*

A case of the solar flare on September 6, 2017, is shown in Fig. 7.8 as an example of thermospheric variations occurring immediately after a solar flare. This figure is the result of the GAIA simulations described above. The temperature and mass density at an altitude of 400 km increased over a wide area in response to two flares (X2.2 at 8:50 UT and X9.3 flare at 11:53 UT). In this example, the second X9.3 flare

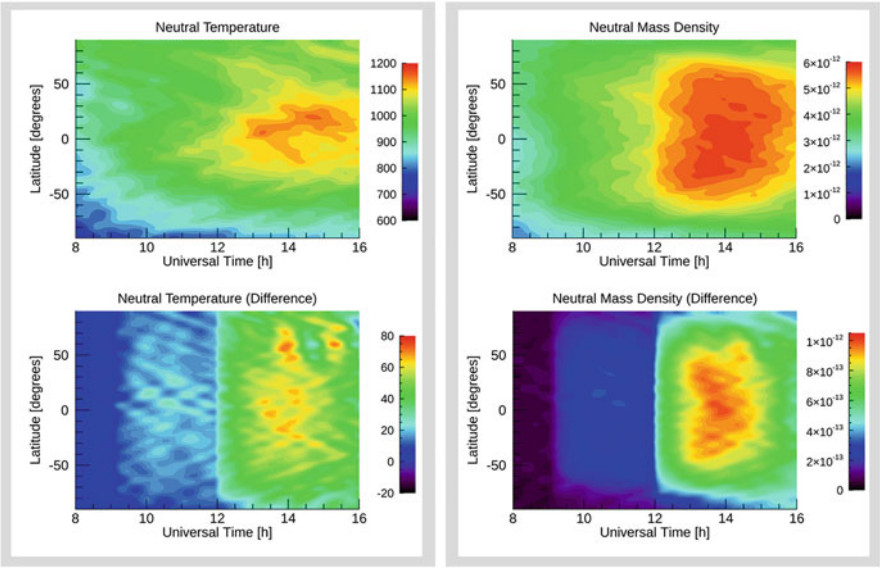

Fig. 7.8 Numerical simulation results of impacts of solar flare on thermosphere on September 6, 2017. (Upper left) Variations in temperature of neutral atmosphere (longitude 0°E, altitude 400 km). (Lower left) Differences between temperatures from calculations with and without solar radiation from solar flare. (Upper right) Variations in mass density of neutral atmosphere (longitude 0°E, altitude 400 km). Unit is kgm^{-3}. (Lower right) Differences between mass density from calculations with and without solar radiation due to solar flares

increased the temperature and mass density by up to 6% and 18%, respectively, approximately 2 h after the peak of the flare. According to the statistics of satellite observation data, the peak increase in the mass density near 400 km altitude for flares above X5 is approximately 10–13% (a maximum of 27% increase in thermospheric mass density occurred for an X17.2 flare on October 28, 2003). Moreover, the delay in the peak of solar X-rays is reported to be within 4 h or less (Le et al. 2012).

As described in Sect. 7.2.3, the solar radiation acts on the ionization and dissociation of different altitudinal and atmospheric components depending on the wavelength as well as on the heating of the thermosphere. EUV light in the 25–90 nm wavelength range mainly contributes to the ionization of the *F* layer above 150 km altitude, whereas the solar radiation in the wavelength range below 15 nm including X-rays and above 95 nm contributes to the ionization of the *E* layer below 130 km altitude. The increase in the ion and electron densities in these altitudes enhances exothermic ion-neutral reactions (e.g., dissociative recombination $O_2^+ + e^- \rightarrow O + O$) and increases the collisions with the ionospheric ions and electrons with higher temperature, resulting in heating of the thermosphere (Smithtro 2008; Roble et al. 1987). Although X-rays contribute significantly to the increase in the incident energy to the thermosphere due to flares, EUV light has a larger contribution to the increase in temperature due to the higher energy input per atmospheric mass. This is because it acts at higher altitudes and, thus, impacts the atmosphere at lower densities (Huang et al. 2014). Therefore, the contribution of the EUV band is also significant in the increase in the thermospheric mass density due to the expansion of the atmosphere. It has been reported that the effects on the thermosphere depend not only on the scale of the flare but also on the location of its occurrence on the solar surface. This is because the intensity of the solar EUV radiation decays as the position of a flare varies from the center to the edge of the solar surface, whereas the intensity of a solar X-ray does not significantly change (Qian et al. 2010; Le et al. 2012).

7.2.6 Conclusion

As already described in this section, the impacts of a solar flare on the ionosphere and the thermosphere depend on the solar radiation spectrum. In terms of the effects on society, X-rays contribute to the density increase of the ionosphere at low altitudes below the *E* layer and cause shortwave attenuation, as explained in Sect. 7.1. In comparison, the EUV band contributes to the ionization of the *F* layer and the heating and expansion of the thermosphere, affecting the maximum usable frequency of the ionospheric propagation, amount of the GNSS positioning delay, and variations in the atmospheric drag of satellites. However, in most cases, the impacts of the subsequent magnetic storms on the ionosphere and the thermosphere are larger than those of the solar flare itself. The effects of magnetic storms are discussed in the next section.

7.3 Ionospheric Variations Associated with Geomagnetic Storms

7.3.1 Introduction

During a geomagnetic storm, the electric field and electric current in the polar upper atmosphere increase, and auroral particles originating from the magnetosphere precipitate into the upper atmosphere, causing drastic changes in the thermosphere and the ionosphere from their normal states. The electric field in the ionosphere causes a drift motion of the ionospheric plasma, and the precipitating particles ionize the neutral atmosphere and increase the electron density and the electrical conductivity. In addition, Joule heating by ionospheric currents and precipitating particles heat the thermosphere and affect the electron density distribution and dynamics in the ionosphere by changes in both the thermospheric wind system and atmospheric composition. These effects extend not only to the polar regions but also to mid- and low latitudes, leading to global thermospheric and ionospheric variations. Here we discuss the effects of geomagnetic storms on the upper atmosphere, focusing on ionospheric storms, large-scale traveling ionospheric disturbances (LSTIDs), thermospheric disturbances, and atmospheric drag.

7.3.2 Ionospheric Storms

The phenomenon of large variations in the electron density in the ionosphere associated with geomagnetic storms is called an "ionospheric storm" and is particularly pronounced in the *F* region of the ionosphere. Ionospheric storms can be classified into two types: (1) "positive storms" in which the electron density significantly increases and (2) "negative storms" in which the electron density significantly decreases (Fig. 7.9). When a positive storm occurs, the delay of radio waves passing through the ionosphere increases, which may increase the positioning errors in the GPS. In a negative storm, radio waves that are typically reflected in the ionosphere are not reflected at certain frequencies, which can affect radio communication.

There are two main causes of positive storms: (1) penetration of the electric field from the magnetosphere and (2) increase in the thermospheric winds due to the heating of the polar thermosphere. The mechanism of (1) is that the electric field applied from the magnetosphere to the polar regions increases with the geomagnetic storm, and the electric field penetrates to lower latitudes (Kikuchi et al. 1996, 2008). The direction of the electric field is eastward on the dayside and westward on the nightside (Fig. 7.10).

When an eastward electric field is applied to the ionosphere, the ionosphere is lifted upward by the $\mathbf{E} \times \mathbf{B}$ drift. In the *F* region of the ionosphere, oxygen ions O^+ are mainly lost by chemical reactions with nitrogen molecules N2; however, the loss

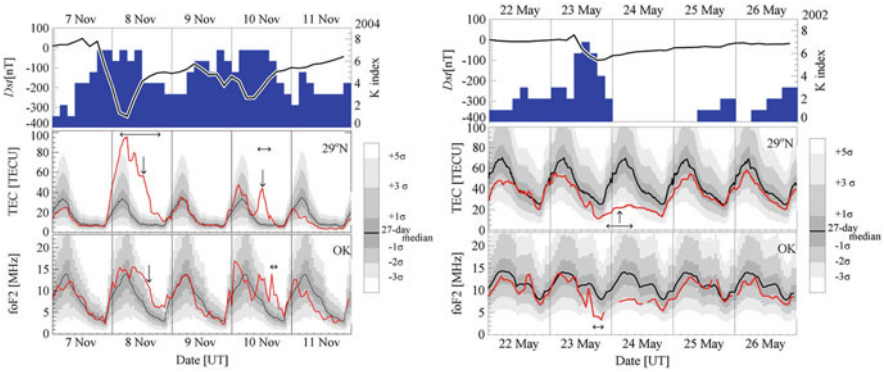

Fig. 7.9 Examples of positive storms (left) and negative storms (right). Top row shows Dst (solid line) and K indices (blue bars). Middle row shows the total electron content (TEC), median of 27 days (black solid line), and observed values (red solid line). Bottom row shows ionospheric critical frequency (foF2) values at Okinawa (Nishioka et al. 2017)

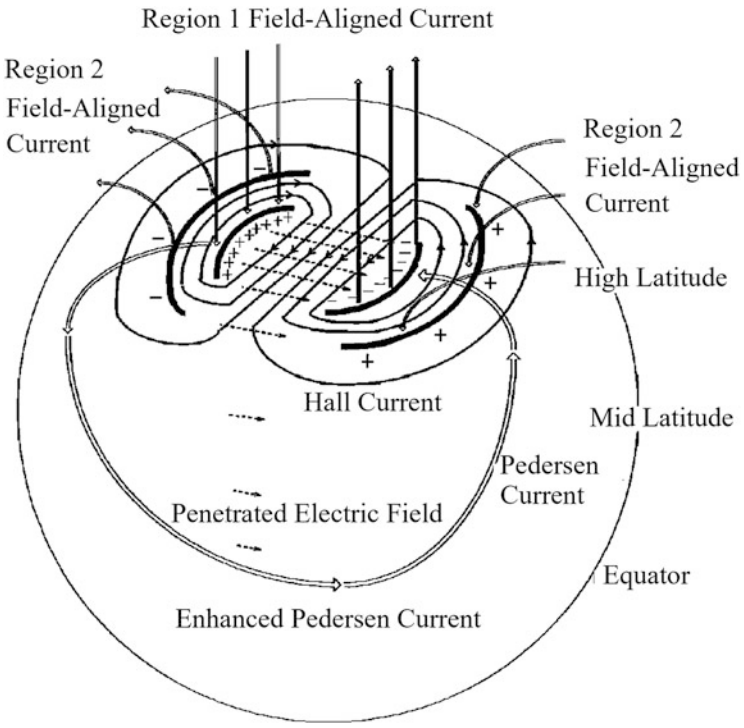

Fig. 7.10 Penetration process of polar cap electric field into low latitudes. Electric field applied to polar region penetrates to mid- and low latitudes and becomes eastward electric field (Kikuchi et al. 1996)

rate of O^+ decreases with increasing altitude due to the decrease in the ratio of the density of nitrogen molecules N2 to the density of oxygen atoms O, i.e., [N2]/[O]. Photoionization production still occurs, and consequently the electron density increases. Because this phenomenon is due to a transient eastward penetrating the electric field, the duration is generally short, approximately 2–3 h. The penetration process of the electric field from the magnetosphere does not typically cause a positive storm, because the direction of the electric field may actually change owing to the interaction process between the magnetosphere, ionosphere, and thermosphere.

In addition to the above mechanisms of positive storms, the changes in the thermospheric neutral winds due to sustained heating of the polar regions can also lead to positive storms. During geomagnetic storms, the polar thermosphere is heated by Joule heating and particle precipitation, resulting in equatorward wind enhancement and traveling atmospheric disturbances (TADs), which are atmospheric gravity waves in the thermosphere driven by polar heating. Consequently, the ionospheric plasma in the *F* layer is lifted upward in the direction of magnetic field lines. Subsequently, the flow of the thermosphere heated and lifted in the polar regions turns downward in the low-latitude region, and the $[N_2]/[O]$ ratio of the thermosphere decreases. This process also enhances the electron density in the ionosphere. The region of decreased $[N_2]/[O]$ gradually moves to low latitudes, whereas that of increased $[N_2]/[O]$ extends to the mid-latitude region. In general, the positive storm effect caused by the descent of the neutral atmosphere with low $[N_2]/[O]$ is more probable to appear in the low-latitude region.

The main cause of negative storms is considered to be the variations in the neutral atmospheric composition. As in the case of positive storms, the origin of negative storms is the injection of energy into the polar regions by geomagnetic storms. However, in contrast to positive storms, which are mainly caused by the direct effects of dynamics, negative storms are caused by upwelling of the neutral atmosphere associated with the thermospheric heating in the polar regions, which increases $[N_2]/[O]$. The atmosphere with high $[N_2]/[O]$ propagates to lower latitudes as equatorward winds and TADs. In its first stage, an equatorward wind strengthens and may be preceded by a positive storm, which pushes the ionospheric plasma along tilted magnetic field lines and causes an increase in the electron density. In actual geomagnetic storms, the variations in the electron density typically depend on the time of occurrence.

Apart from ionospheric storms described above, a phenomenon called SED (storm-enhanced density) occurs in large geomagnetic storms (Fig. 7.11), where the electron density tends to increase significantly mainly on the equatorial side of the afternoon auroral region (Foster 1993). An elongated structure with a high-electron density, called the SED plume, occasionally extends from the daytime cusp region into the polar cap. This prominent structure in the polar cap is called the tongue of ionization (TOI) (Fig. 7.11).

Although the formation mechanisms of SED and SED plumes have not been fully elucidated yet, several mechanisms have been proposed, such as the transport of the plasma due to the increase in the local electric field in the polar region associated

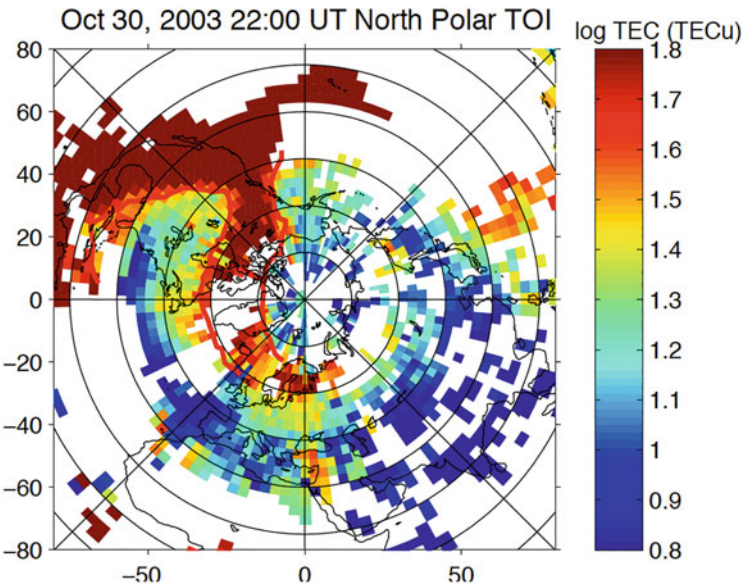

Fig. 7.11 SED that occurred during October 2003 geomagnetic storm, with SED plume extending toward pole (Foster and Rideout 2007)

with geomagnetic storms. Other possible mechanisms are variations in the electron density due to the thermospheric winds caused by the heating of the polar regions and expansion of the equatorial anomaly to high latitudes associated with geomagnetic storms.

7.3.3 Evaluation of Magnitudes of Ionospheric Storms

The electron density significantly varies during ionospheric storms. A factor that makes the evaluation and prediction of the magnitudes of ionospheric storms difficult is that the range of the variations in the electron density significantly depends on the solar activity, season, latitude, and local time even for geomagnetic storm of the same scale. Because of the difficulty in the evaluation of ionospheric storm intensity, the general scale of ionospheric disturbances has not been developed. However, recently, an index called the I-scale has been devised as a standard for universally expressing the extent of disturbances (Nishioka et al. 2017) (Fig. 7.9). For positive storms, the deviations are defined as IP3, IP2, and IP1, starting with the largest in magnitude, and for negative storms with decreasing parameters, the deviations are defined as IN3, IN2, and IN1, starting with the largest in magnitude, and I0 for quiet conditions.

In the space weather forecast by the National Institute of Information and Communications Technology, an ionospheric storm is rated by the I-scale for the

ionospheric TEC and the ionospheric critical frequency (foF2), respectively (https://swc.nict.go.jp/en/trend/ionosphere.html).

7.3.4 Large-Scale Traveling Ionospheric Disturbances

Large-scale traveling ionospheric disturbances (LSTIDs) are ionospheric disturbances associated with geomagnetic storms and substorms. They are considered to be driven by TADs generated by polar heating (Hunsucker 1982). TADs propagate in the equatorial direction with a phase velocity of 600–800 m/s. Equatorward propagation of this wind causes the ionospheric plasma to move up and down along the magnetic field lines, which, in turn, causes height variations in the ionospheric *F* region. When the altitude of the ionospheric *F* region moves up and down, the ion loss rate changes because the $[N_2]/[O]$ ratio changes depending on the altitude, resulting in electron density fluctuations. These TAD-induced large-scale fluctuations in the ionosphere are considered to be LSTIDs. Figure 7.12 shows examples of LSTIDs appearing in the TEC around Japan.

7.3.5 Thermospheric Disturbances and Atmospheric Drag

When a geomagnetic storm occurs, the temperature and density of the neutral atmosphere increase significantly in the polar thermosphere due to Joule heating and heating by precipitating particles. Consequently, a drag force of the atmosphere (atmospheric drag) exerts on satellites orbiting at altitudes below approximately 400 km, which may change their orbits and attitudes. Therefore, in actual satellite operations, it is important to constantly monitor and predict changes in the thermospheric atmospheric density. However, it is difficult to quantitatively estimate the extent of increase in the thermospheric density caused by a given geomagnetic storm, because of the complex relationships between various physical and chemical processes.

To reproduce the variations in the thermosphere, various numerical simulation models have been developed since the 1980s, and the increase in the thermospheric density associated with geomagnetic storms and substorms has been reproduced well qualitatively (Lu et al. 2016, see Fig. 7.13). However, variations in the thermospheric density associated with individual geomagnetic storms have not been adequately reproduced because of uncertainties in the energy input from the magnetosphere and in the physical and chemical processes.

Regarding the atmospheric drag, empirical models have been developed specifically to reproduce the thermospheric mass density. The basic method is to construct an empirical model that reproduces the thermospheric mass density data obtained from satellite observations using the solar EUV parameter, F10.7, solar wind

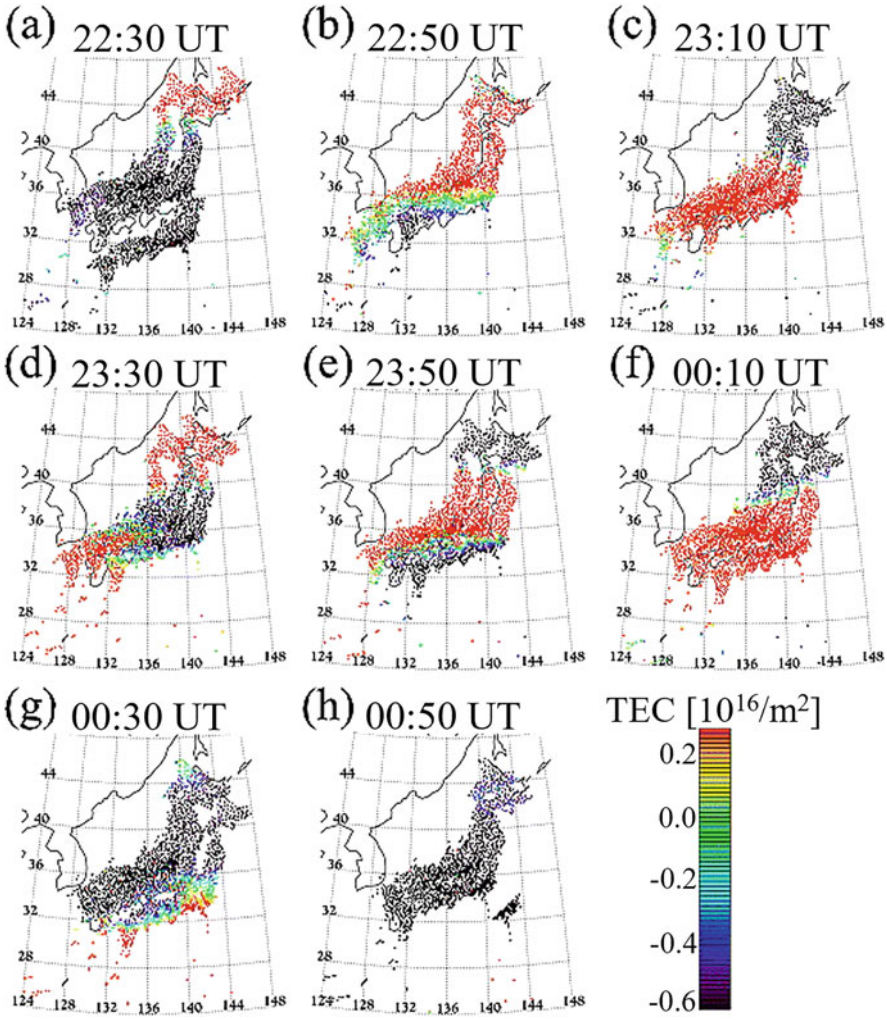

Fig. 7.12 (**a**–**h**) Two-dimensional distributions of variations in TEC around Japan between 22:30 UT on 22 September 1999 and 00:50 UT on the subsequent day with a 20-minute interval. In (**a**–**c**), region with increase in TEC is propagating from north to south, which is LSTID. In (**d**–**f**), another LSTID is moving from north to south around Japan (Tsugawa et al. 2003)

parameters observed by a satellite, and geomagnetic activity indices (such as Kp, ap, and Dst) based on geomagnetic data as input parameters.

Among many empirical models, some of the oldest and best known include the Jacchia-70 model (Jacchia 1970), MSISE-90 (Hedin 1991), and NRLMSISE-00 (Picone et al. 2002). Some models using a comprehensive set of recent satellite data include JB2008 (Bowman et al. 2008) and DTM-2013 (Bruinsma 2015). More recently, predictions using neural network techniques (Chen et al. 2014) and methods combining numerical and empirical models (Fedrizzi et al. 2012) have

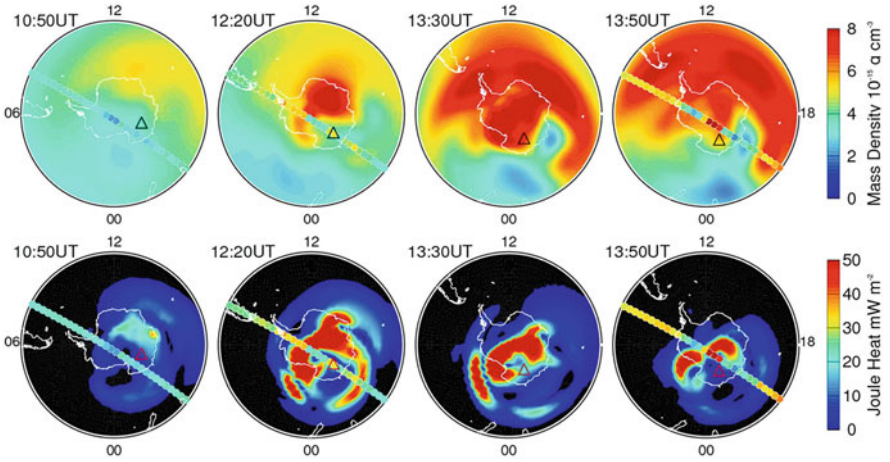

Fig. 7.13 (Top row) Mass density of neutral atmosphere at altitude of 380 km reproduced by TIEGCM (Thermosphere-Ionosphere Electrodynamics General Circulation Model). (Bottom row) Height-integrated Joule heating rate in Southern Hemisphere. Small colored dots indicate mass density of the neutral atmosphere as observed by the CHAMP satellite (Lu et al. 2016)

also been developed. At present, empirical models are more accurate than numerical simulation models for the prediction of the density variations in the thermosphere and are more commonly used in satellite operations.

7.3.6 Conclusion

The prediction of the thermospheric and ionospheric variations associated with geomagnetic storms is one of the most important issues in space weather forecasting. However, it has been almost impossible to predict the variability at a given location and time with sufficient accuracy because many physical and chemical processes interact with each other simultaneously in the thermosphere and the ionosphere. In recent years, numerical simulation models with high accuracy have been steadily developed for the solar wind, magnetosphere, ionosphere, thermosphere, and atmosphere, and the amount of data from satellite and ground-based observations has been rapidly increasing. Furthermore, data assimilation and artificial intelligence-based prediction techniques have also been rapidly developing. Therefore, it is expected that practical prediction of ionospheric and thermospheric disturbances will become possible in the near future.

7.4 Ionospheric Variations in Polar Regions

7.4.1 *Recent Observations of Scintillation Associated with Aurora*

An aurora is a most prominent phenomenon observed in the polar upper atmosphere. Auroras emit light when energetic charged particles (electrons and protons) precipitate from the magnetosphere along the magnetic field lines of the Earth and excite atmospheric atoms and molecules. The precipitation of auroral electrons not only excites but also ionizes the atmosphere and is known to produce significant variations in the ionospheric electron density (mainly at approximately 100 km altitude, corresponding to the E region of the ionosphere). For example, a phase scintillation (see Sect. 7.1.4), which is a fluctuation of the phase of GNSS satellite positioning signals, has been shown to occur in association with an auroral breakup. An auroral breakup is an abrupt start of the expansion phase of an auroral substorm (see Fig. 7.14, which is taken from Hosokawa et al. 2014). However, it is known that such phase scintillation does not frequently lead to GNSS signal lock loss and does not directly cause positioning interruptions. This can be attributed to the fact that the ionization associated with an aurora is mainly in the ionospheric E region, where the absolute value of the electron density variation is smaller than that in the F region. In the following, we present several previous observations of ionospheric phenomena other than auroras in the polar ionospheric F region.

7.4.2 *Mechanism of Polar Cap Patch Generation and its Relation to TOI/SED*

Polar cap patches are patchy regions of increased electron density that occur in the ionospheric F region (at approximately 300 km of altitude) of the polar cap regions, where the magnetic latitude exceeds 80°. A polar cap patch is composed of a high-density plasma transported by the anti-sunward plasma convection in the polar cap driven by the interaction between the magnetosphere and the solar wind when the interplanetary magnetic field (IMF) is directed southward. In the ionospheric F region, where atomic ions (oxygen ions) dominate, it takes a long time (several hours) for the plasma to disappear by recombination even after the plasma is transported to the dark region where the plasma production by the solar EUV radiation is unexpected. Therefore, the high-density plasma in the sunlit region transported to the dark area on the nightside can survive for a long time and is observed as polar cap patches by ionosondes, radars, and airglow imagers.

Although various mechanisms have been proposed for the generation of polar cap patches, they employ the temporal and spatial variations in the strength and shape (pattern) of the high-latitude convection for capturing the dense sunlit plasma on the dayside into smaller-scale structures, i.e., polar cap patches.

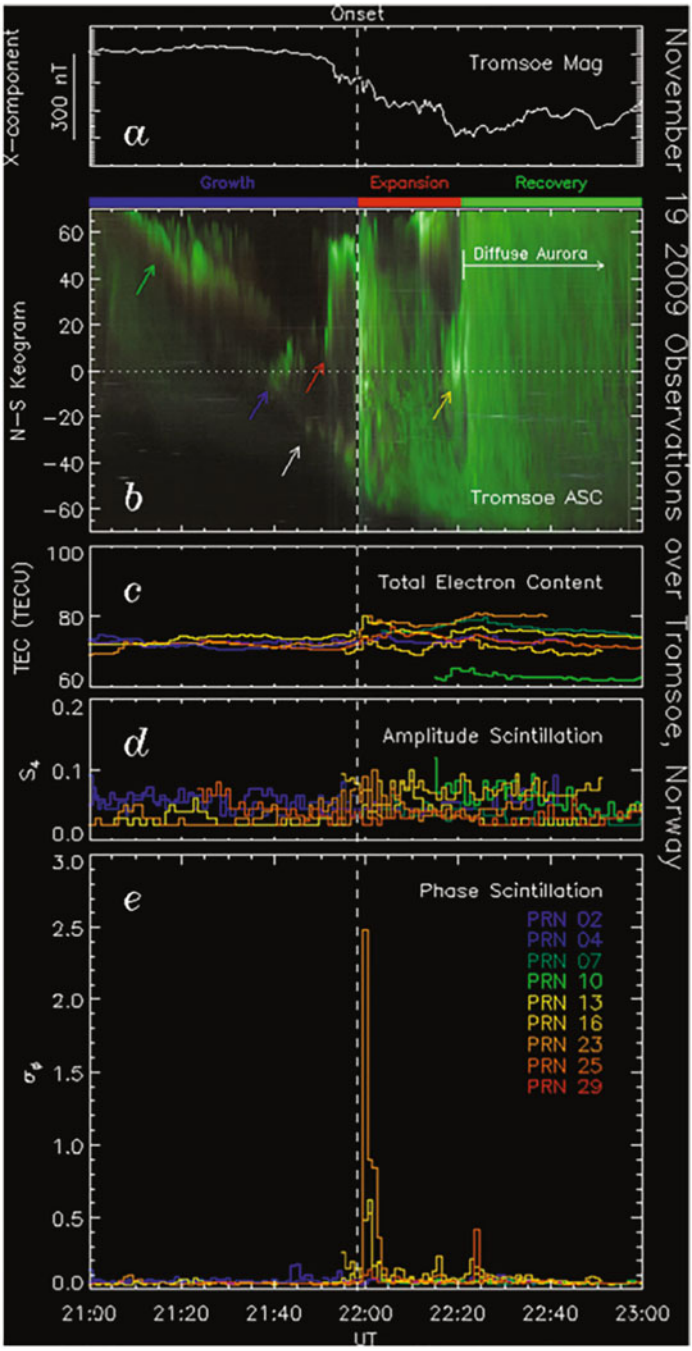

Fig. 7.14 Examples of GPS scintillation associated with an aurora substorm. (**a**) Geomagnetic field decrease indicating the onset of the auroral substorm, (**b**) time series of north–south cross-sections of auroral all-sky images, showing rapid intensification of aurora in middle of interval, (**c**) ionospheric TEC calculated from GPS receiver data, (**d**) amplitude scintillation, (**e**) phase

A polar cap patch is a phenomenon in which regions of high-electron density propagate in a patchy manner with a horizontal scale of 500–1000 km. However, the spatial scale of the high-density plasma in the polar cap is sometimes larger than the above range, and, in such a case, the electron density enhancement is seen to extend across the polar cap regions from the daytime to the nighttime. Such large-scale electron density enhancements are called TOI (tongue of ionization; Foster et al. 2005) and are considered to be formed when dense plasmas created by the solar EUV radiation on the dayside are transported along the polar cap ionospheric convection toward the nightside without being split into patches. Figure 7.15 shows a mapping of a TOI extending from the dayside sunlit region toward the nightside as reported by Foster et al. (2005) using GPS TEC observations. A TOI is frequently generated during coronal mass ejection (CME)-type disturbances such as magnetic storms, because the temporal variation in the IMF associated with CMEs and the resultant fluctuation in the polar cap convection are larger than those of corotating interaction region-type disturbances (Hosokawa et al. 2009). Foster et al. (2005) also proposed the existence of an SED, which is formed when a fast westward plasma stream (subauroral polarization stream: SAPS) appearing in a mid-latitude trough entrains the dense sunlit plasma in the dusk side toward high latitudes. This suggests that an SED, which is a source of GNSS scintillation in mid-latitudes, and a TOI, which produces polar cap scintillation, may be the same phenomenon.

7.4.3 Ionospheric Irregular Structure and Scintillation Associated with Polar Cap Patches

It is known observationally that in a polar cap patch, irregular structures in the ionospheric electron density (ionospheric irregularities) are distributed. Such irregularities are detected as scintillations of the radio waves between a satellite and the ground. These electron density irregularities are known to produce backscatter in a coherent scattering radar. Observations by the Super Dual Auroral Radar Network (SuperDARN) have suggested the formation of irregularities with large amplitudes at the trailing edge of polar cap patches (Milan et al. 2002). It has long been pointed out that such irregularities inside a polar patch are amplified by gradient-drift instability (GDI). Considering the generation of irregularities associated with patches based on GDI, small-scale density perturbations are expected to be amplified where the spatial gradient of the electron density at the trailing edge of a polar cap patch and the background plasma convection are in the same direction. This explains the early SuperDARN radar observations. However, recent studies investigating the spatial relationship between polar cap patches observed by an airglow imager and

Fig. 7.14 (continued) scintillation showing brief and sharp increase at the expansion phase onset of the auroral substorm (Hosokawa et al. 2014)

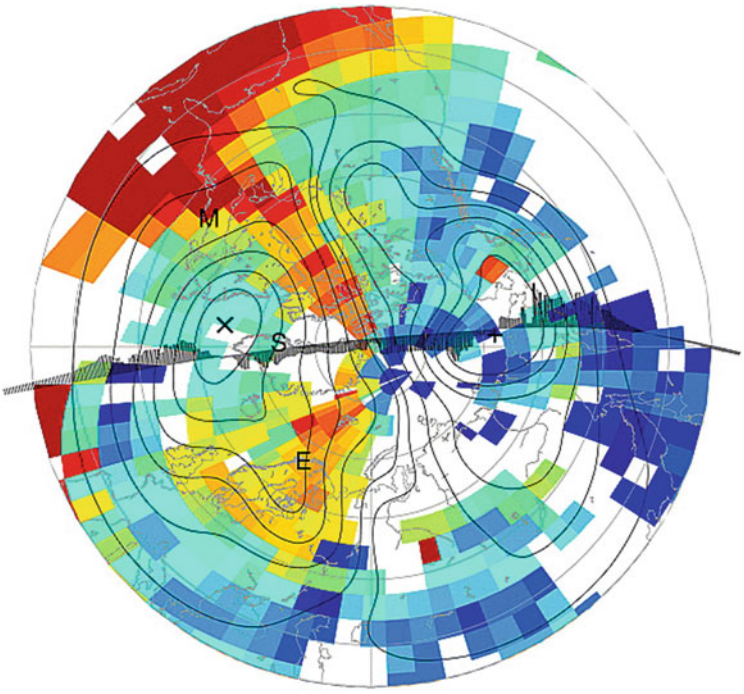

Fig. 7.15 Large-scale structure of TOI visualized by GPS total electron count (TEC) observations. Center of circle is the magnetic pole, and upper part is the dayside. Dense daytime sunlit plasma is continuously transported toward nighttime by the anti-sunward plasma convection (Foster et al. 2005)

the associated irregularities detected by a radar as backscatter echoes have confirmed the presence of irregularities in entire region of polar cap patches (Fig. 7.16 taken from Hosokawa et al. 2009). This suggests that the growth of irregularities by GDI extends over an entire polar cap patch and produces scintillation on a relatively short timescale (several tens of minutes). It is also known that the scintillation associated with a polar cap patch is mainly phase scintillation. This tendency is very different from that of an equatorial plasma bubble. This is considered to be attributed to the rapid change in the phase velocity of radio waves caused by the temporal change in the refractive index along the radio wave path, as a patch propagates at speeds that can reach 1 km/s in the fastest case.

7.4.4 Polar Cap Absorption

Solar proton events (SPEs) originating from solar flares penetrate the atmosphere of the Earth and strongly ionize the lower part of the ionosphere (D region: 50–90 km altitude). The extent of the region ionized by SPEs depends on the energy of the

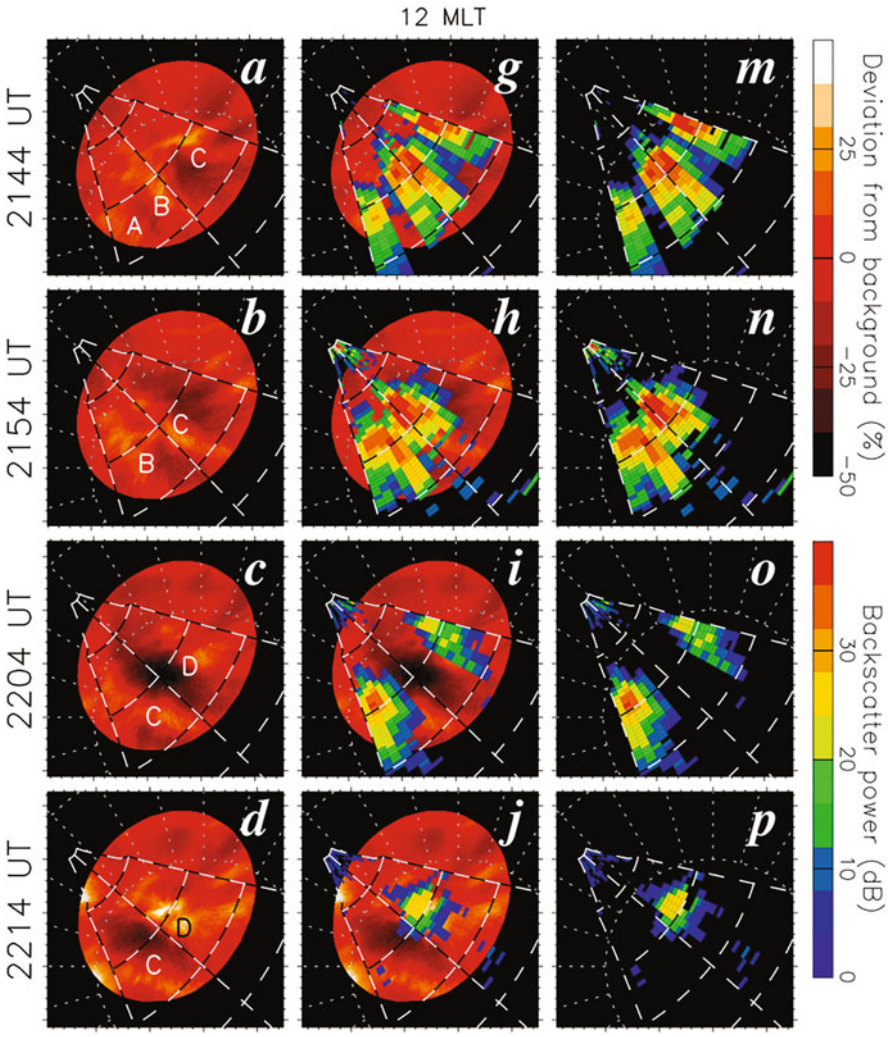

Fig. 7.16 (Left) Polar cap patch visualized by all-sky airglow observations, (right) plasma irregularities associated with polar cap patch observed by SuperDARN radars, and (center) superimposition of airglow data and radar observations. It can be seen that plasma irregularities are present in the entire area of polar cap patch detected with the airglow imager (Hosokawa et al. 2009)

incoming energetic protons and the shape of the magnetic field lines; however, it is typically distributed in the polar regions at 60–90° magnetic latitude (Kress et al. 2010). In particular, the anomalous ionization in the polar cap regions is frequently accompanied by radio absorption, which is known as polar cap absorption (PCA). PCA is frequently observed during large solar flares with a time delay of several minutes to several hours. PCAs have long been investigated because the strong ionization in the D region absorbs high-frequency (HF) and very high-frequency

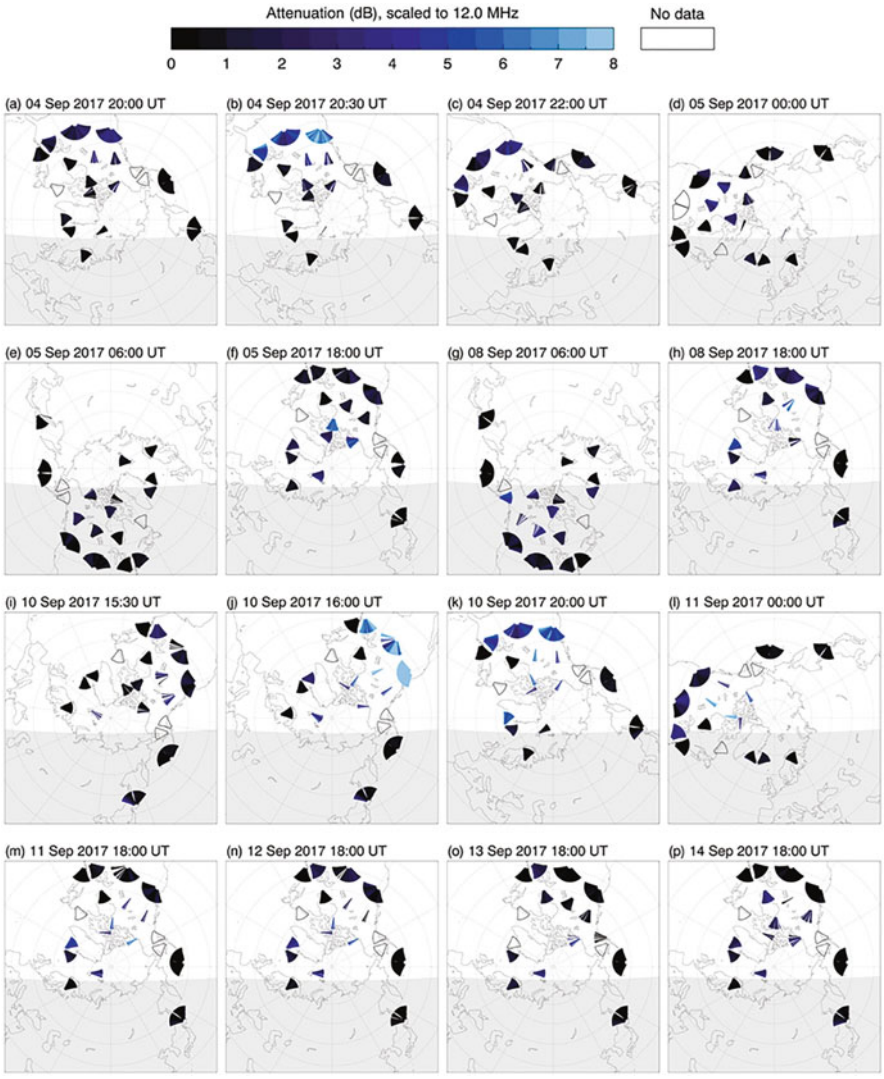

Fig. 7.17 Visualization of extent of region of ionization and radio absorption caused by SEP using SuperDARN data. Radio absorption is strong in light-colored regions, indicating that ionization occurs at low latitudes (Bland et al. 2018)

(VHF) band radio waves used for long-distance communication. PCA has been studied using data obtained from riometers, incoherent scatter radars, VLF radio observations, and GPS TEC observations (Jones et al. 2005). Recently, an attempt has been made to visualize the spatiotemporal evolution of PCAs by SuperDARN, a network of coherent scatter radars using HF radio waves (Bland et al. 2018). In Fig. 7.17, the SuperDARN data, which cover a large area, are used to visualize the

extent of the region where ionization and radio absorption happen due to a SEP. The radio absorption is strong in the light-colored regions, indicating that ionization occurs even at low latitudes.

7.5 Medium-Scale Propagating Ionospheric Disturbances

7.5.1 Introduction

Traveling ionospheric disturbances (TIDs), which are wavy structures of the ionospheric electron density, have been observed and studied extensively (Hunsucker 1982; Hocke and Schlegel 1996). Based on their characteristics, TIDs are classified into two types. The first type is medium-scale TIDs (MSTIDs), which have periods of 15–60 min and horizontal wavelengths of several 100 km, and the second type is large-scale TIDs (LSTIDs), which have periods of 30 min–3 h and horizontal wavelengths of 1000 km or more. Since the study by Hines (1960), TIDs were considered to be caused by atmospheric gravity waves, which are oscillations of the neutral atmosphere. However, in the late 1990s, observations by an all-sky airglow imager and GPS receiver networks began to reveal the horizontal two-dimensional structure of TIDs (Fig. 7.18). It was shown that most MSTIDs observed at night in mid-latitudes propagate southwestward in the Northern Hemisphere and northwestward in the Southern Hemisphere. These results suggest that nighttime MSTIDs may be caused by plasma instabilities in the ionosphere. In contrast, LSTIDs are considered to be caused by atmospheric gravity waves generated in high latitudes by geomagnetic disturbances, because they frequently appear during geomagnetic disturbances and propagate equatorward, as described in Sect. 7.3. In this section,

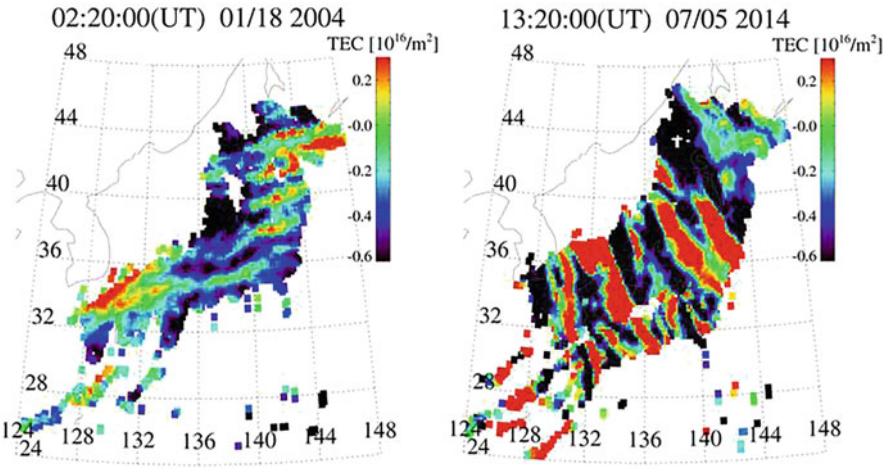

Fig. 7.18 TEC variations over Japan (https://aer-nc-web.nict.go.jp/GPS/DRAWING-TEC/). MSTIDs during (left) daytime and (right) nighttime are observed

we describe MSTIDs, which occur frequently even in geomagnetically quiet conditions. Although the electron density variation due to MSTIDs is only a few percent of the background, it can be a factor of positioning error in satellite positioning such as the GPS when precise measurement is required.

7.5.2 MSTIDs During Daytime

Most of the MSTIDs that appear during daytime propagate equatorward and occur most frequently in winter. This feature is consistent for different longitudinal sectors in both the Northern and Southern Hemispheres. The preference of equatorward propagation is one of the reasons daytime MSTIDs are considered to be caused by atmospheric gravity waves. The electron density perturbations, n_e', caused by neutral wind perturbations due to atmospheric gravity waves can be expressed as follows (Hooke 1968):

$$n_e' = \frac{i}{\omega} u_b' \left\{ \frac{\partial n_e}{\partial z} \sin I - i k_b n_e \right\} \tag{7.8}$$

where u_b' and k_b are the magnetic parallel component of the neutral wind perturbation and the wave number vector, respectively, ω is the angular frequency, n_e is the background electron density, z is the altitude, I is the inclination angle of the magnetic field lines, and i denotes a complex number. Because an ion-neutral collision frequency is much smaller than an ion gyrofrequency in the F region, collision with neutral particles causes ions to move only in the direction parallel to the magnetic field at the same velocity as the neutral particles, u_b'. The first and second terms on the right-hand side of Eq. (7.8) represent the electron density perturbation caused by the vertical motion of ions and by the convergence and divergence of ions, respectively. Although the convergence and divergence of neutral particles due to atmospheric gravity waves do not occur, ions are converged and diverged because they move only in a direction parallel to the magnetic field lines.

Figure 7.19 shows the relationship between the wave front of upward-propagating atmospheric gravity waves and the directions of neutral particle and ion oscillations. In the horizontal and vertical cross-sections, a neutral particle oscillates parallel to the wave front of the gravity waves. The left and right figures in Fig. 7.19 show the cases of equatorward-propagating gravity waves at mid-latitudes and poleward-propagating gravity waves, respectively. Comparing the magnetic field lines of the Earth and the direction of the atmospheric oscillations, the parallel component of neutral particle oscillations is larger when the gravity waves propagate equatorward than that when they propagate poleward. According to Eq. (7.8), when a gravity wave propagates equatorward (poleward), the amplitude of u_b' is larger (smaller), resulting in a larger (smaller) amplitude of n_e'. Because the ion

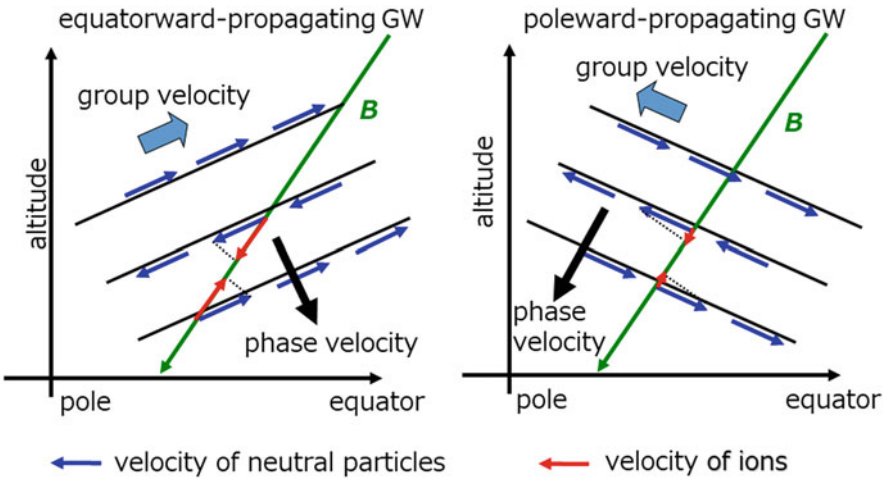

Fig. 7.19 Relation between oscillation directions of neutral particles and ions due to atmospheric gravity waves propagating (left) equatorward and (right) poleward

motion is anisotropic with respect to the magnetic field lines, the electron density perturbation with respect to atmospheric gravity waves is also anisotropic. Even if an atmospheric gravity wave propagates isotropically, the electron density perturbation, i.e., a TID, propagating in the equatorial direction is accompanied by a large electron density perturbation. Therefore, it can be inferred that most of the observed MSTIDs propagate equatorward. In addition, the motion of the neutral atmosphere in the direction of crossing the magnetic field lines is dragged by the collision with ions (ion drag), which attenuates atmospheric gravity waves. Because the ion drag force is smaller in the equatorial direction than those in the other directions, atmospheric gravity waves can propagate over long distances with reduced attenuation. This may be one of the causes of the high occurrence rate of equatorward-propagating MSTIDs.

The vertical structures of MSTID-induced electron density perturbations have been observed by incoherent scatter radars. As shown in Fig. 7.20, the electron density perturbation propagates to lower altitudes with time. Because the directions of the phase velocity and the group velocity of the atmospheric gravity waves are different by 90°, we can deduce that this electron density perturbation is caused by the upward-propagating atmospheric gravity waves. The vertical wavelength of the electron density perturbation tends to be longer at higher altitudes. This is because the viscosity and thermal conductivity of the neutral atmosphere become larger at higher altitudes (Pitteway and Hines 1963). This feature is reproduced by GAIA, which is a coupled atmosphere–ionosphere model (Miyoshi et al. 2018).

An atmospheric gravity wave with vertical wave number m, horizontal wave number k, and horizontal phase velocity c follows the dispersion relation,

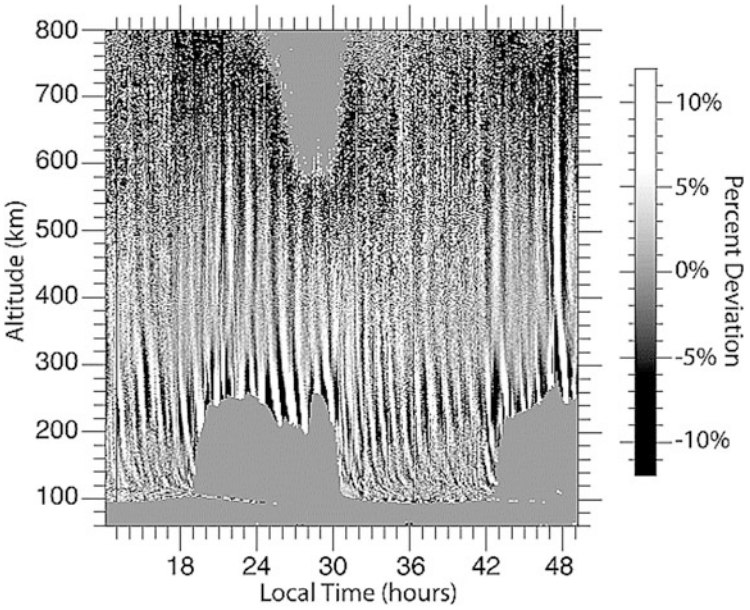

Fig. 7.20 Temporal altitude variation in electron density perturbations with period 1–2 h component observed by incoherent scattering radar at Arecibo, Puerto Rico (Livneh et al. 2007)

$$m^2 = \frac{N^2}{(c-U)^2} - k^2 - \frac{1}{4H^2} \tag{7.9}$$

where U is the background wind speed, N is the Brandt–Vaisala angular frequency, and H is the scale height of the neutral atmosphere. The vertical wave number of the atmospheric gravity wave, m, depends on U. Because of the strong east–west wind in the stratosphere and the mesosphere, U can be close to the phase speed of the gravity wave, c. In this case, $(c - U)^2$ becomes smaller, and thus, m becomes larger (the vertical wavelength becomes shorter). At the altitude where $c = U$, m becomes infinite (the vertical wavelength becomes zero), and the atmospheric gravity wave cannot propagate above this altitude and eventually breaks. In addition, a short vertical wavelength of the atmospheric gravity wave implies a rapid decay due to the viscosity effect. Thus, the background wind velocity for gravity waves propagating from the lower atmosphere to the ionosphere has a significant influence on the propagation of atmospheric gravity waves.

In addition, when atmospheric gravity waves are attenuated or broken during their propagation, secondary atmospheric gravity waves may be excited (Vadas and Liu 2009). These secondary atmospheric gravity waves are probable to be generated in the stratosphere and the mesosphere owing to the increase in the zonal wind speed. The atmospheric gravity waves excited in the lower atmosphere directly propagate to the ionosphere, and the secondary waves generated by their breakup during

propagation are considered to be the cause of daytime MSTIDs. However, it is still unclear where atmospheric gravity waves are excited and how they propagate to the ionosphere, and theoretical and observation studies are underway.

7.5.3 MSTIDs During Nighttime

During nighttime, most MSTIDs propagate southwestward in the Northern Hemisphere and northwestward in the Southern Hemisphere, unlike in the daytime. This is because during nighttime MSTIDs are not caused by atmospheric gravity waves but by electromagnetic dynamics. Figure 7.21 shows that nighttime MSTIDs are accompanied by electric field perturbations (Shiokawa et al. 2003). It presents a two-dimensional distribution of the 630 nm airglow intensity, which is proportional to the electron density in the bottomside of the *F* region. In the figure, MSTIDs with a wave front extending from northwest to southeast can be seen. The electric field perturbations observed by the satellite simultaneously are projected onto an altitude of 300 km along the magnetic field lines. The electric field perturbations, $\delta \boldsymbol{E}$, are southwestward in the region where the airglow intensity increases and northeastward

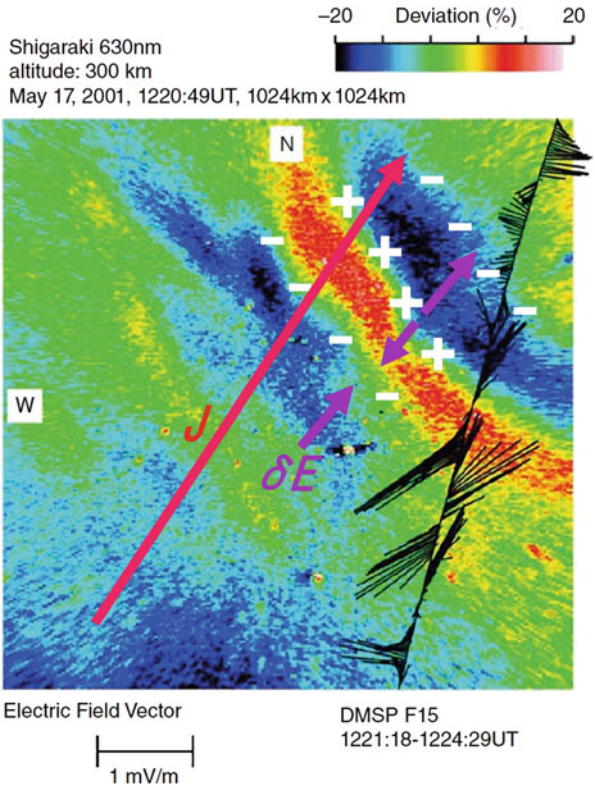

Fig. 7.21 Relationship between 630 nm airglow intensity and polarization electric fields associated with MSTIDs (Shiokawa et al. 2003). Ionospheric current $\boldsymbol{J}$, polarization electric field, and positive and negative charges are added in the figure

in the region where it decreases. These electric field perturbations associated with MSTIDs are explained as follows: A neutral wind, $\boldsymbol{U}$, in the thermosphere is driven by the balance between the pressure gradient force and the ion drag force. Therefore, a neutral wind blows southeastward at mid-latitudes in the Northern Hemisphere during the night. The Pedersen current, $\boldsymbol{J}$, driven by $\boldsymbol{U}$ is in the direction of $\boldsymbol{U} \times \boldsymbol{B}$, where $\boldsymbol{B}$ is the geomagnetic field. Because the 630 nm airglow intensity is proportional to the product of the electron density and the molecule oxygen density, it is proportional to the Pedersen conductivity, Σ_P, which is proportional to the product of the electron and neutral densities. Therefore, the spatial variation in the 630 nm airglow intensity can be regarded as that of Σ_P (Makela and Kelley 2003). To maintain the continuity of $\boldsymbol{J}$, positive or negative charges accumulate as shown in the figure, and a polarization electric field $\delta\boldsymbol{E}$ is generated. Because $\delta\boldsymbol{E}$ has an east (west) component, due to the $\boldsymbol{E} \times \boldsymbol{B}$ drift, the *F*-region plasma moves upward (downward) to higher (lower) altitudes, where the neutral density decreases (increases), resulting in a decrease (increase) in the airglow intensity or Σ_P. The polarization electric field plays an important role in generating nighttime MSTIDs. The magnetic field lines of the Earth through the ionosphere are highly conductive, and the ionospheric electric field is transmitted to the *F* region in both the Northern and Southern Hemispheres almost without attenuation (Farley 1960). This results in the simultaneous generation of MSTIDs with mirror-image structures in both hemispheres (Otsuka et al. 2004).

As mentioned above, the electron density perturbations caused by nighttime MSTIDs are accompanied by polarization electric fields. The Perkins instability has been proposed as a mechanism for the growth of these electron density and polarization electric field perturbations (Perkins 1973). In the mid-latitude *F* region during nighttime, the downward gravitational diffusion velocity is balanced by the upward component of the plasma velocity under the equatorward neutral wind being parallel to the magnetic field lines and the $\boldsymbol{E} \times \boldsymbol{B}$ drift caused by the eastward electric field. However, the Perkins instability is a phenomenon in which such a balance is disturbed. Figure 7.22 shows the field-line-integrated Pedersen current on the plane normal to the geomagnetic field lines at mid-latitudes, $\boldsymbol{J}$, Pedersen conductivity (Σ_P) variation, and polarization electric field. $\boldsymbol{J}$ flows across the conductivity variation. Here, $\boldsymbol{J}$ is divided into the two components, which are perpendicular and parallel to the wave front of the MSTID ($\boldsymbol{J}_\perp$ and $\boldsymbol{J}_\parallel$, respectively). $\boldsymbol{J}_\perp$ flows continuously; therefore, $\boldsymbol{J}_\perp$ is constant in the high- and low-conductivity regions. However, $\boldsymbol{J}_\parallel$ is larger in the high-conductivity region and smaller in the low-conductivity region. Therefore, the $\boldsymbol{J} \times \boldsymbol{B}$ force acting on the plasma is larger (smaller) downward in the high (low)-conductivity region, and the plasma receives a downward (upward) force. Consequently, the plasma in the high (low)-conductivity region moves to lower (higher) altitudes, and the conductivity becomes further higher (lower). Thus, the initial small perturbation grows with time. The Perkins instability grows when the wave number vector of the fluctuations is between $\boldsymbol{J}$ and the east–west direction. Because most of the MSTIDs observed during nighttime have a northwest–southeast (southwest–northeast) wave front in the Northern (Southern) Hemisphere, the Perkins instability is considered to be their generation mechanism.

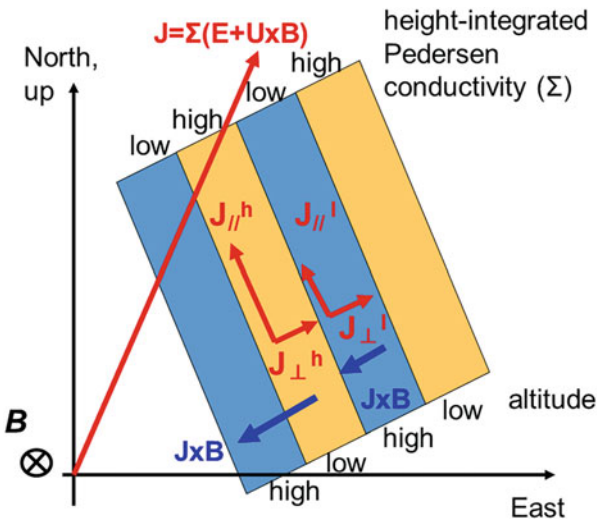

Fig. 7.22 Relationship between current and conductivity variations in plane orthogonal to magnetic field lines. Current and conductivity are integrated along magnetic field lines over entire *F* region. Generation mechanism of Perkins instability is shown

7.6 Plasma Bubbles

7.6.1 Introduction

In the *F* region of the ionosphere, the electron density reaches a peak at approximately 300 km altitude. At the magnetic equator, where the magnetic field lines of the Earth are horizontal, the bottomside of the *F* region can become unstable, and the electron density can significantly decrease locally. This phenomenon is called a plasma bubble, analogous to a rising bubble in water. A plasma bubble occurs at the bottomside of the *F* region after sunset and rises to higher altitudes. The left figure in Fig. 7.23 shows all-sky images of the 630 nm airglow emitted by oxygen atoms, which are simultaneously captured by the all-sky airglow imagers at Sata, Japan, and Darwin, Australia (Otsuka et al. 2002). Because the brightness of the airglow is proportional to the electron density in the bottomside of the *F* region, the plasma bubble corresponds to the dark region of the airglow extending poleward in both all-sky images. The width of the plasma bubble is approximately 100 km in the east–west direction. The structure of the plasma bubble observed simultaneously in the Northern and Southern Hemispheres is symmetrical to the equator. This is because the same magnetic field lines pass through Sata and Darwin, as shown in the right side of Fig. 7.23, and the low-electron density region due to the plasma bubble has a structure elongating along the magnetic field lines of the Earth.

As described in Sect. 7.1, a plasma bubble causes a large error in satellite positioning such as the GPS because the electron density rapidly decreases in a localized area. In addition, because of the existence of electron density irregularities with various spatial scales inside a plasma bubble, the amplitude and phase of the

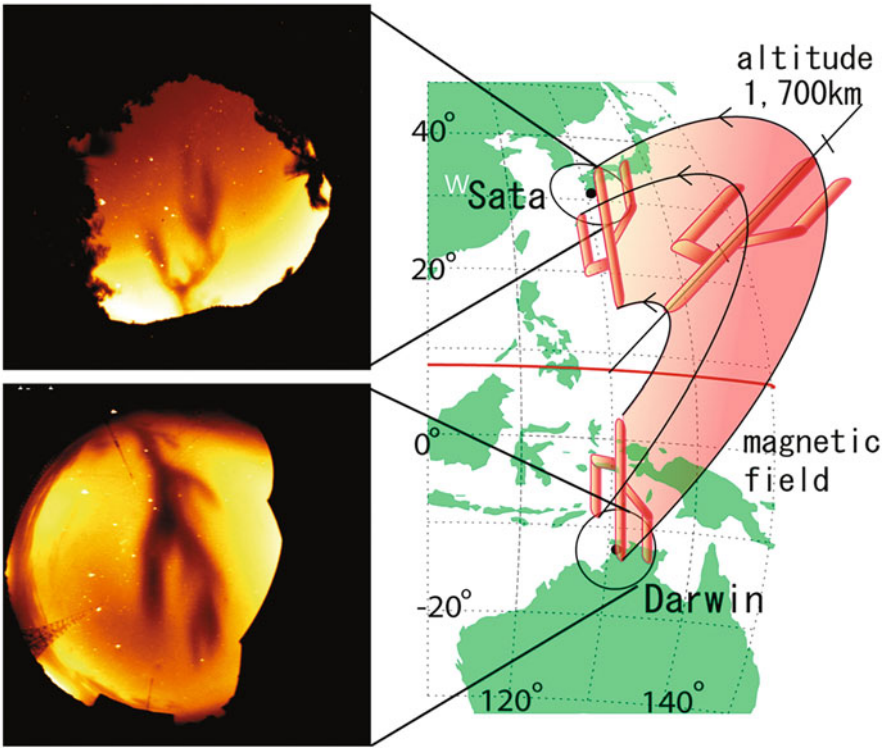

Fig. 7.23 (Left) All-sky image of airglow at wavelength of 630 nm observed simultaneously in Sata, Japan, and Darwin, Australia. Dark airglow region corresponds to plasma bubble (Otsuka et al. 2002). (Right) Schematic of structure of plasma bubble

radio waves passing through it fluctuate (scintillation), which affects satellite broadcasting, communication, and satellite positioning.

7.6.2 Rayleigh–Taylor Instability

In general, the Rayleigh–Taylor instability occurs when a heavy fluid is above a light fluid; it also occurs at the magnetic equator in the bottomside of the *F* region, where the electron density increases with the altitude. Figure 7.24 shows a schematic of the electron density distribution at the bottomside of the *F* region at the magnetic equator. In the *F* region, electrons move gyroscopically around the magnetic field lines, whereas ions with large masses move in $\boldsymbol{g} \times \boldsymbol{B}$, where $\boldsymbol{g}$ is the gravitational acceleration and $\boldsymbol{B}$ is the geomagnetic field. Consequently, the electric current flows eastward at the magnetic equator where the magnetic field is horizontal and northward. At the sunset terminator, a phenomenon called pre-reversal enhancement

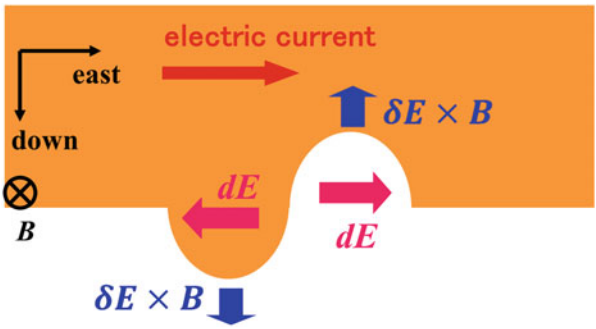

Fig. 7.24 Schematic of Rayleigh–Taylor instability

occurs, in which the eastward electric field increases. This eastward electric field also contributes to the eastward Pedersen current. As shown in the figure, we consider a small perturbation of the iso-electron density surface at the bottomside of the F region. Because this electron density perturbation creates inhomogeneity of the conductivity, to maintain the electric current continuity, positive charges are accumulated where the eastward current flows into the low-electron density region, whereas negative charges are accumulated where it flows out. These charges produce an eastward polarization electric field where the iso-electron density surface is uplifted to higher altitude, and the drift of the polarization electric field, $\boldsymbol{E} \times \boldsymbol{B}$, lifts the plasma to an even higher altitude. In contrast, where the iso-electron density surface is initially pushed down to a lower altitude, a westward electric field is generated, and the downward $\boldsymbol{E} \times \boldsymbol{B}$ drift pushes it further down to further lower altitudes. Thus, the initial small perturbation grows with time. This instability is the Rayleigh–Taylor instability, which causes the formation of plasma bubbles.

The linear growth rate of the Rayleigh–Taylor instability, γ, is expressed as

$$\gamma = \left(\frac{E_0}{B} + \frac{g}{\nu_{in}} \right) \frac{1}{n_e} \frac{\partial n_e}{\partial z} \tag{7.10}$$

where E_0 is the eastward electric field, B is the magnetic field, g is the gravitational acceleration, ν_{in} is the ion-neutral collision frequency, n_e is the electron density, and z is the altitude. Eq. (7.10) suggests that the growth rate increases under the following conditions: (1) the eastward electric field, E_0, is large, (2) the plasma is lifted to higher altitudes, where ν_{in} is smaller, and (3) the altitude gradient of the electron density, $(\frac{1}{n_e} \frac{\partial n_e}{\partial z})$, is large. Under conditions (7.1) and (7.2), the eastward current increases. The eastward electric field not only is related to condition (7.1) but also contributes to condition (7.2) by lifting the plasma to higher altitudes under the $\mathbf{E} \times \mathbf{B}$ drift. Therefore, the eastward electric field is an important factor for plasma bubble generation. In the linear theory of the Rayleigh–Taylor instability, the instability grows only at the bottomside of the F region, where the electron density gradient is upward. However, actually a plasma bubble can exceed the F_2 peak altitude, where the electron density gradient is downward.

7.6.3 Characteristics of Plasma Bubble Generation

Plasma bubble occurrence rate is high under high solar activity conditions and shows pronounced seasonal and longitude variations (Burke et al. 2004; see Fig. 7.25). The solid line in Fig. 7.25 shows the day when the sunset terminator line and the magnetic field lines are parallel. It is found that the plasma bubble occurrence rate is highest when the solar terminator is parallel to the geomagnetic field. This occurs during the equinoxes at the longitude where the declination of the geomagnetic field is almost equal to zero and near the December solstice at the longitude where the declination is tilted westward. Such seasonal and longitudinal variations in the plasma bubble occurrence rate are consistent with those in the eastward electric field at sunset (Tsunoda 1985). However, there are some features that cannot be explained, such as the high occurrence rate in the Atlantic region, low occurrence rate in the Pacific region, and asymmetry in its occurrence rate between spring and autumn. It has been pointed out that these phenomena may be caused by effects in the neutral atmosphere such as atmospheric gravity waves propagating upward from the lower atmosphere to the ionosphere and the thermosphere.

Geomagnetic disturbances are also known to affect plasma bubble generation. It has been reported that during the main phase of geomagnetic storms, when a strong eastward electric field is applied to the ionosphere, plasma bubbles are probably generated, and some of them can reach mid- and high latitudes (Aa et al. 2018). However, during the recovery phase of geomagnetic storms, the circulation of the neutral atmosphere in the thermosphere changes from the geomagnetic quiet conditions, as described in Sect. 7.3, and plasma bubble generation tends to be suppressed (Carter et al. 2014).

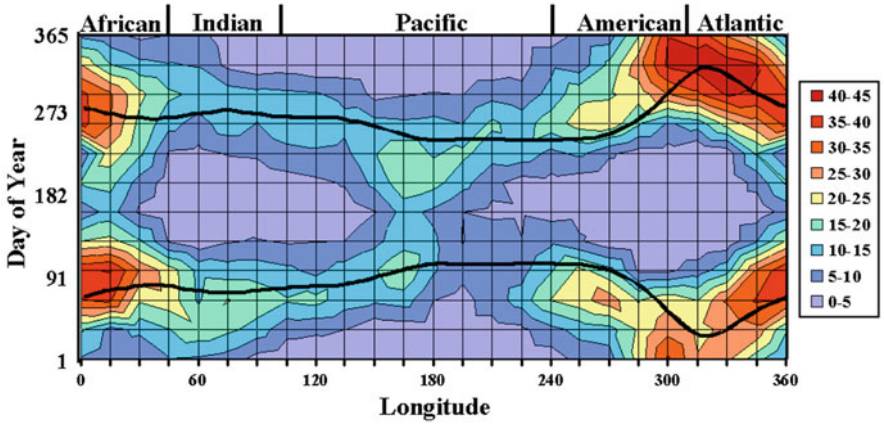

Fig. 7.25 Longitudinal and seasonal variations in plasma bubble occurrence rate (Burke et al. 2004). Black lines represent days when sunset line is parallel to magnetic field lines of Earth

7.6.4 Increase in Eastward Electric Field at Sunset

After sunset, when plasma production ceases, the electron density in the *E* region decreases rapidly owing to the recombination of ions. Consequently, the spatial gradient of the conductivity in the *E* region becomes large between the day and nightsides across the sunset terminator line, creating a strong zonal electric field. The electric field in the ionosphere is eastward during the daytime and reverses its direction to westward during nighttime. Because the eastward electric field increases at sunset before it reverses the direction, this phenomenon is called pre-reversal enhancement. The mechanism was explained by Farley et al. (1986). As shown in Fig. 7.26, (1) in the *F* region, the neutral wind, **U**, blows eastward around the sunset terminator. (2) By **U**, dynamo electric field $\mathbf{E} = -\mathbf{U} \times \mathbf{B}$ is generated (**B** is the magnetic field, which is horizontal and northward at the magnetic equator). (3) **E** is transmitted to the *E* region along the magnetic field lines. (4) **E** drives the westward Hall current in the *E* region on the dayside. (5) Negative charges are accumulated along the sunset terminator in the *E* region. (6) An electric field $\mathbf{E}_\phi$, which is eastward in the dayside and westward in the nightside, is induced. (7) $\mathbf{E}_\phi$ is transmitted to the *F* region. This eastward electric field drives the upward $\mathbf{E} \times \mathbf{B}$ drift, which lifts the plasma to higher altitudes and produces the condition in which plasma bubbles are probably generated.

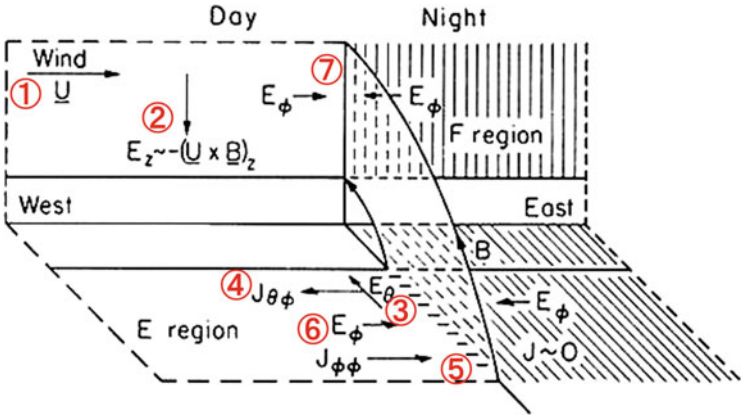

Fig. 7.26 Schematic of generation mechanism of pre-reversal enhancement (numbering added to Fig. 7.7 in Farley et al. (1986)). Horizontal direction of figure represents east–west direction, with left side representing dayside and shaded area on right side representing nightside. Upper part of figure represents *F* region and lower part represents *E* region. These regions are connected by magnetic field lines (curves)

7.6.5 *Prediction of Plasma Bubble Occurrence by Numerical Model*

It can be considered that a plasma bubble is generated when the growth rate of the Rayleigh–Taylor instability is high. Efforts have been made to predict plasma bubble occurrence by calculating the daily growth rate of the Rayleigh–Taylor instability using a numerical model that can simulate the atmosphere of the Earth (Carter et al. 2014; Shinagawa et al. 2018). Shinagawa et al. (2018) used the global atmosphere–ionosphere coupling model, Ground-to-topside model of Atmosphere and Ionosphere for Aeronomy (GAIA), to calculate the linear growth rate and compare it with actual plasma bubble occurrence. Figure 7.27 shows the maximum value of the linear growth rate of the Rayleigh–Taylor instability for each day in 2011. It is compared with the plasma bubble occurrence observed by the Equatorial Atmospheric Radar (EAR) at Kotthabang, Indonesia (100.32°E, 0.2°S), and the GPS at the same site. It is found that the high growth rate tends to coincide with the actual plasma bubble occurrence. This result suggests that the global atmosphere–ionosphere coupling model is useful for predicting plasma bubble generation.

However, plasma bubble generation is determined not only by the growth rate of the Rayleigh–Taylor instability but also by "seeding" of the instability by atmospheric waves (Tsunoda 2010). Yokoyama et al. (2019) developed a local high-precision model that numerically solves the motion of the ionospheric plasma and showed that vertical neutral-atmosphere winds of a few meters per second can act as seeding for plasma bubbles. Therefore, the prediction of atmospheric wave propagating into the ionosphere and thermosphere also needs to be considered to predict plasma bubble generation.

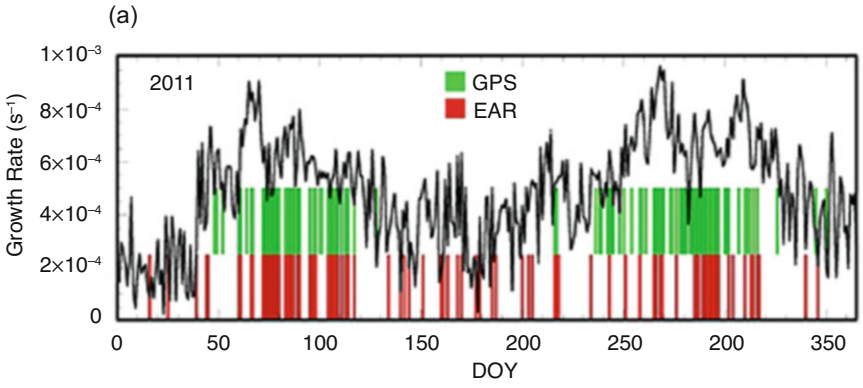

Fig. 7.27 Daily variation in linear growth rate of Rayleigh–Taylor instability in 2011 at Kotthabang, Indonesia, as determined by GAIA (solid black line). Bars indicate days when plasma bubbles are observed by GPS (green) and EAR (red) (Shinagawa et al. 2018)

7.7 Sporadic E Layers

7.7.1 Introduction

A sporadic *E* layer (*Es* layer) is known as a layer of dense electrons and ions that appears transiently in the ionospheric *E* region. In this section, basic information about *Es* layers is reviewed. For more detailed information, it is recommended to refer to several review papers on *Es* layers (Whitehead 1970, 1989; Mathews 1998).

7.7.2 Effects on Radio Wave Propagation

Figure 7.28 shows *Es* layer observations by an ionosonde as the most representative observation example. An ionosonde is an instrument that observes the ionosphere by sweeping the frequency of high-frequency (HF) radio waves. It can obtain information of the electron density in the ionosphere using the fact that the reflection frequency of the transmitted radio waves depends on the electron density at the reflection altitude. The relationship between the maximum reflection frequency, f, of a transmitted radio wave and the electron density, n_e, in the ordinary wave mode

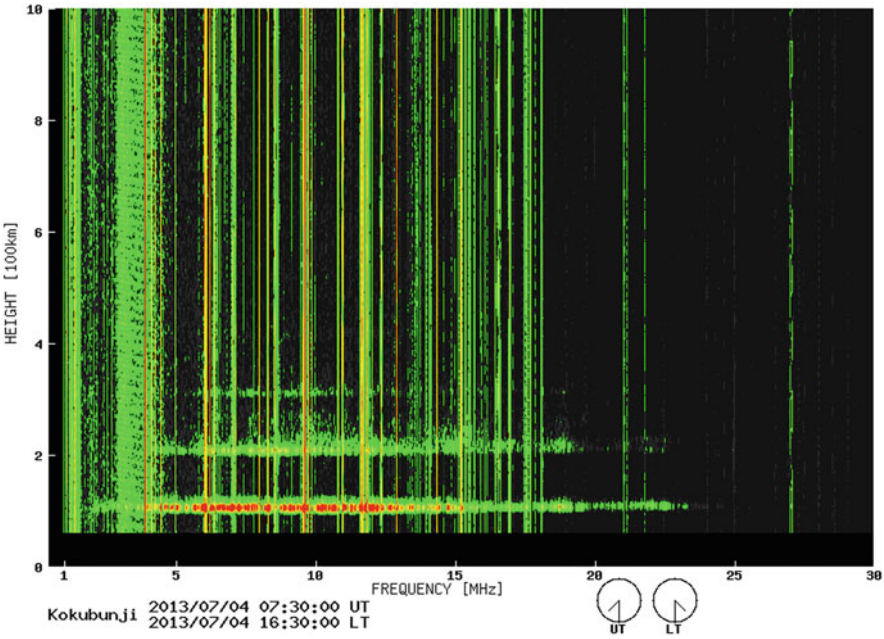

Fig. 7.28 *Es* layer observed by ionosonde at Kokubunji, Tokyo (35.71°N, 139.49°E) at 07:30 universal time (UT) on July 4, 2013 (from WDC for Ionosphere and Space Weather, Tokyo, NICT: http://wdc.nict.go.jp/IONO/HP2009/ISDJ/index-E.html)

(O-mode), as described in Sect. 7.1, is based on the Appleton–Hartree equation as follows:

$$f = \sqrt{e^2 n_e / 4\pi^2 m_e \varepsilon_0} \approx 8.98 \times \sqrt{n_e}\ [\mathrm{Hz}] \tag{7.11}$$

The strong reflection echo at an altitude of ~100 km (see Fig. 7.28) reaches a frequency of >20 MHz and, therefore, is considered to be a reflection echo from an *Es* layer with an electron density of $>5 \times 10^{12}$ m^{-3} (Hunsucker and Hargreaves 2002). In the case of oblique incidence of the transmitted radio wave, the frequency, f, at which the radio wave is reflected, as described in Eq. (7.3) in Sect. 7.1.2, is given by

$$f = \frac{f_p}{\cos i_0} \tag{7.12}$$

Thus, as the incident angle, i_0, increases, the reflection frequency, f, becomes larger than f_p, which is the reflection frequency in the case of vertical incidence (Hunsucker and Hargreaves 2002). For example, when a transmitted radio wave is incident on an *Es* layer with an electron density of $>5 \times 10^{12}$ m^{-3} at an incident angle of 80°, the reflection frequency is >115 MHz, which implies that the Es layer can reflect very high-frequency (VHF) radio waves. Thus, the influence of an *Es* layer on the radio propagation environment, particularly on communication and broadcasting in the shortwave/ultra-shortwave band (HF/VHF band) (e.g., Sakai et al. 2019), has been pointed out as a social impact. Therefore, it is important to monitor and predict the dynamics of *Es* layers.

7.7.3 *Characteristics and Observations of* es *Layers*

An example of *Es* layer observations by an incoherent scattering radar (Mathews et al. 1997) is presented in Fig. 7.29. A gradual decay of the ionosphere by daytime sunlight can be seen after sunset (at 15–17 local time (LT)). The timescale of this decay is considered to be dependent mainly on the lifetime (recombination time constant) of NO^+ and O_2^+, which are the major ions in the ionospheric *E* region. In contrast, a high-electron density *Es* layer with a narrow altitude range is observed to be maintained for a long time after sunset. This is considered to be due to an *Es* layer being mainly composed of metallic ions with relatively long lifetimes. The source of metallic species is abrasion of meteors entering the atmosphere, and various metallic species such as Fe, Mg, Si, Na, Ca, and K are distributed in atomic, ionic, and molecular forms in the lower part of the ionosphere (cf. Plane et al. 2015). For example, the lifetime of Fe^+ ion, the major species of metallic ions (cf. Kopp 1997), is 10^6 s (approximately 11 days) at an altitude of 110 km (cf. Matuura et al. 2013). In the example of Ca^+ ion observations obtained by laser sensing (lidar) shown in

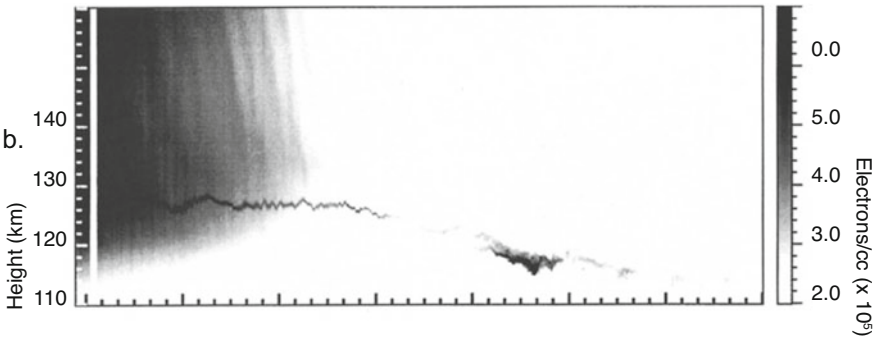

Fig. 7.29 *Es* layer observed by incoherent scattering radar (transmission frequency: 430 MHz) at Arecibo (18.34°N, 66.75°W) on September 2, 1994 (Mathews et al. 1997)

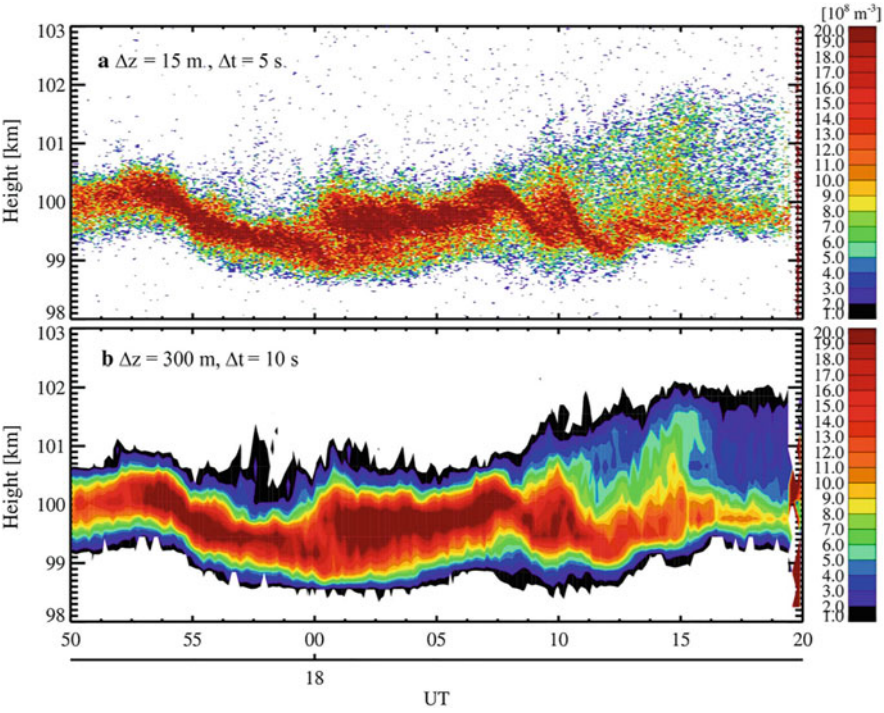

Fig. 7.30 High-density Ca^+ layers observed by Ca^+ lidar (transmission wavelength: 393 nm) at Tachikawa, Tokyo (35.7°N, 139.4°E), on December 24, 2014 (Ejiri et al. 2019). (**a**) Ca^+ data with time resolution of 5 s and altitude resolution of 15 m and (**b**) Ca^+ data with time resolution of 10 s and altitude resolution of 300 m

Fig. 7.30 (Ejiri et al. 2019), a dense layer of Ca^+ ions with a narrow altitude width (>1 km) corresponding to an *Es* layer is also captured. This suggests that metallic ions constitute the main component ions of this *Es* layer.

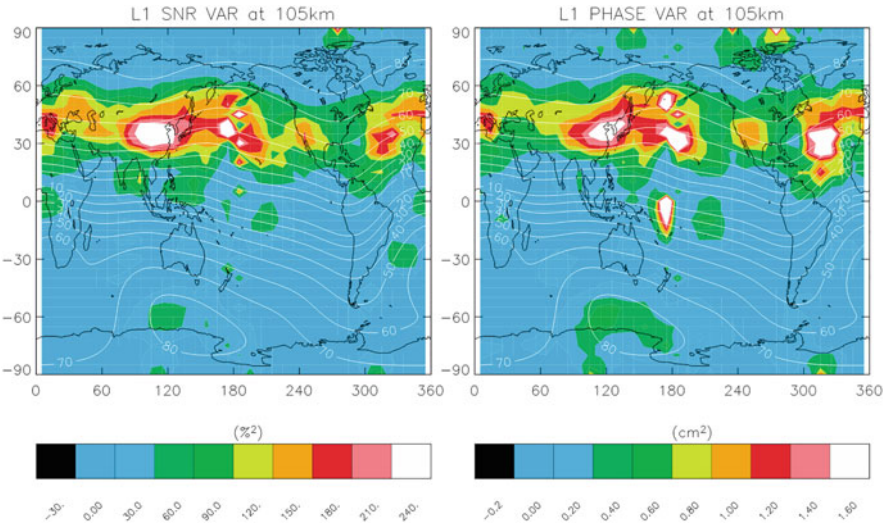

Fig. 7.31 Global distributions of (left) variances of signal noise ratio (SNR) and (right) variances of phase in GPS L1 band (approximately 1.6 GHz) occultation observations from June to August 2002 at altitudes of 100–110 km (Wu et al. 2005)

The high-resolution Ca^+ lidar data (see Fig. 7.30a) shows fine irregular structures in the *Es* layer, which can cause amplitude and phase scintillations of the ultrahigh-frequency (UHF) radio waves used in the Global Navigation Satellite System (GNSS). Therefore, their influence on the GNSS positioning accuracy is of concern. However, information on the global distribution of *Es* layers is obtained by observing its scintillation (see Fig. 7.31). It can be confirmed that the scintillation by an *Es* layer is more significant at mid-latitudes in the summer hemisphere. These characteristics are well-known general seasonal and latitudinal characteristics of *Es* layers, which have been investigated by ionospheric observations using ionosondes and other instruments for many years. *Es* layers are also known to occur at low and high latitudes, although not as pronounced as at mid-latitudes (cf. Whitehead 1970, 1989; Mathews 1998).

7.7.4 Generation Mechanism: Wind Shear Theory

We introduce the wind shear theory, which is a most probable generation mechanism of the most prominent mid-latitude *Es* layer. The motion of ions is affected by a neutral wind (motion of the neutral atmosphere) through collisions, and when the effects of the magnetic field of the Earth are considered, the horizontal neutral wind generates a vertical ion motion (ion velocity). Based on the equation of motion of ions with some assumptions, the vertical ion velocity, w_i, can be expressed as

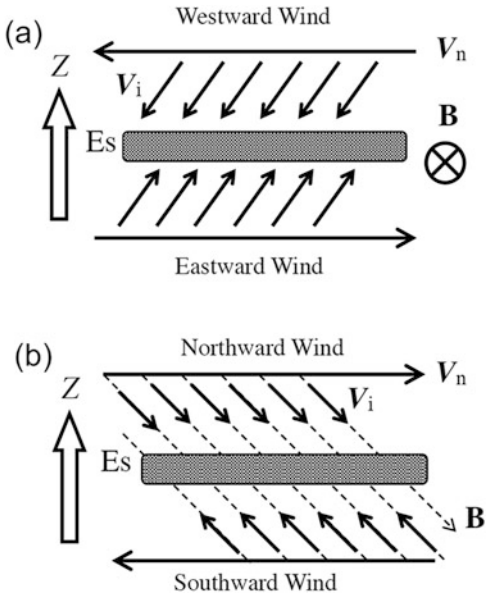

Fig. 7.32 Schematic of *Es* layer formation by wind shear theory (Shinagawa et al. 2017). (**a**) for wind shear in east–west component of neutral wind and (**b**) for wind shear in north–south component of neutral wind

$$w_i = [v \cos I \sin I + (\nu_{in}/\omega_i) u \cos I] / \left[1 + (\nu_{in}/\omega_i)^2\right] \tag{7.12}$$

Here, ν_{in} is the collision frequency between ions and neutrals and ω_i is the gyrofrequency of ions (Shinagawa et al. 2017). The signs of southward wind v and eastward wind u can change the direction of the vertical ion velocity. If there is an altitudinal change in the horizontal wind direction (horizontal wind shear) that produces a downward ion velocity above an altitude and an upward ion velocity below that altitude, ions can accumulate in that altitude region to form a high-density layer of electrons and ions (*Es* layer). Figure 7.32 shows an example of an *Es* layer produced by horizontal wind shear in the Northern Hemisphere. In the case of the Northern Hemisphere, an *Es* layer can be formed by the accumulation of ions in the altitude region between the vertical downward ion velocities produced by westward horizontal winds at higher altitudes and the upward ion velocities produced by eastward horizontal winds at lower altitudes. Similarly, in the conditions of northward winds at higher altitudes and southward winds at lower altitudes, *Es* layer formation can be also induced. Based on this wind shear theory, derivation of the vertical ion accumulation rate using neutral wind data calculated by a global atmospheric model has been attempted (Shinagawa et al. 2017). As shown in Fig. 7.33, the global distribution of the vertical ion accumulation rates derived from a global atmospheric model reproduces the general characteristics of *Es* layers well, which are prominent at mid-latitudes in the summer hemisphere, suggesting that the wind shear theory plays an important role in the formation of *Es* layers.

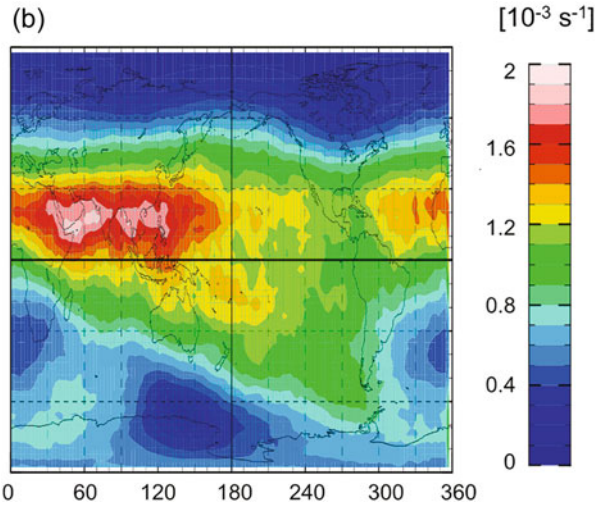

Fig. 7.33 Global distribution of vertical ion accumulation rates for June–August 2002 calculated by GAIA (Shinagawa et al. 2017)

References

Aa, E., Huang, W., Liu, S., Ridley, A., Zou, S., Shi, L., et al.: Midlatitude plasma bubbles over China and adjacent areas during a magnetic storm on 8 September 2017. Space Weather. **16**, 321–331 (2018). https://doi.org/10.1002/2017SW001776

Basu, S., MacKenzie, E., Basu, S.: Ionospheric constraints on VHF/UHF communication links during solar maximum and minimum periods. Radio Sci. **23**, 363 (1988)

Bland, E.C., Heino, E., Kosch, M.J., Partamies, N.: SuperDARN radar-derived HF radio attenuation during the September 2017 solar proton events. Space Weather. **16**, 1455–1469 (2018). https://doi.org/10.1029/2018SW001916

Bowman, B.R., Tobiska, W.K., Marcos, F.A., Huang, C.Y., Lin, C.S., Burke, W.J.: A new empirical thermospheric density model JB2008 using new solar and geomagnetic indices, AIAA/AAS Astrodynamics Specialist Conference. 18–21 August 2008, Honolulu, Hawaii, paper AIAA 2008–6438 (2008)

Bruinsma, S.: The DTM-2013 thermosphere model. J. Space Weather Space Clim. **5**, A1 (2015). https://doi.org/10.1051/swsc/2015001

Burke, W.J., Huang, C.Y., Gentile, L.C., Bauer, L.: Seasonal-longitudinal variability of equatorial plasma bubbles. Ann. Geophys. **22**, 3089–3098 (2004)

Carter, B.A., Retterer, J.M., Yizengaw, E., Groves, K., Caton, R., McNamara, L., et al.: Geomagnetic control of equatorial plasma bubble activity modeled by the TIEGCM with Kp. Geophys. Res. Lett. **41**, 5331–5339 (2014). https://doi.org/10.1002/2014GL060953

Chamberlin, P.C., Woods, T.N., Eparvier, F.G.: Flare irradiance spectral model (FISM): flare component algorithms and results, space. Weather. **6**, S05001 (2008). https://doi.org/10.1029/2007SW000372

Chen, H., Liu, H., Hanada, T.: Storm-time atmospheric density modeling using neural networks and its application in orbit propagation. Adv. Space Res. **53**, 558–567 (2014)

Dellinger, J.H.: Sudden disturbances of the ionosphere. Proc. IRE. **25**, 1253 (1937)

Ejiri, M.K., Nakamura, T., Tsuda, T.T., Nishiyama, T., Abo, M., Takahashi, T., Tsuno, K., Kawahara, T.D., Ogawa, T., Wada, S.: Vertical fine structure and time evolution of plasma irregularities in the es layer observed by a high-resolution Ca+ lidar. Earth Planets Space. **71**, 3 (2019). https://doi.org/10.1186/s40623-019-0984-z

Farley, D.T.: A theory of electrostatic fields in the ionosphere at nonpolar geomagnetic latitudes. J. Geophys. Res. **65**, 869–877 (1960)
Farley, D.T., Bonelli, E., Fejer, B.G., Larsen, M.F.: The prereversal of the zonal electric field in the equatorial ionosphere. J. Geophys. Res. **91**, 13723–13728 (1986)
Fedrizzi, M., Fuller-Rowell, T.J., Codrescu, M.V.: Global joule heating index derived from thermospheric density physics-based modeling and observations. Space Weather. **10**, S03001 (2012). https://doi.org/10.1029/2011SW000724
Foster, J.C.: Storm time plasma transport at middle and high latitudes. J. Geophys. Res. **98**, 1675–1689 (1993)
Foster, J.C., Rideout, W.: Storm enhanced density: magnetic conjugacy effects. Ann. Geophys. **25**, 1791–1799. www.ann-geophys.net/25/1791/2007/ (2007)
Foster, J.C., et al.: Multiradar observations of the polar tongue of ionization. J. Geophys. Res. **110**, A09S31 (2005). https://doi.org/10.1029/2004JA010928
Hedin, A.E.: Extension of the MSIS thermosphere model into the middle and lower atmosphere. J. Geophys. Res. **96**, 1159–1172 (1991)
Hines, C.O.: Internal atmospheric gravity waves at ionospheric heights. Can. J. Phys. **38**, 1441–1481 (1960)
Hocke, K., Schlegel, K.: A review of atmospheric gravity waves and traveling ionospheric disturbances: 1982–1995. Ann. Geophys. **14**, 917–940 (1996)
Hooke, W.H.: Ionospheric irregularities produced by internal atmospheric gravity waves. J. Atmos. Terr. Phys. **38**, 1441–1481 (1968)
Hosokawa, K., Shiokawa, K., Otsuka, Y., Ogawa, T., St-Maurice, J.-P., Sofko, G.J., Andre, D.A.: Relationship between polar cap patches and field-aligned irregularities as observed with an all-sky airglow imager at Resolute Bay and the PolarDARN radar at Rankin inlet. J. Geophys. Res. **114**, A03306 (2009). https://doi.org/10.1029/2008JA013707
Hosokawa, K., Otsuka, Y., Ogawa, Y., et al.: Observations of GPS scintillation during an isolated auroral substorm. Prog Earth Planet Sci. **1**, 16 (2014). https://doi.org/10.1186/2197-4284-1-16
Huang, Y., Richmond, A.D., Deng, Y., Chamberlin, P.C., Qian, L., Solomon, S.C., Roble, R.G., Xiao, Z.: Wavelength dependence of solar irradiance enhancement during X-class flares and its influence on the upper atmosphere. J. Atmos. Sol. Terr. Phys. **115–116**, 87–94 (2014)
Hunsucker, R.D.: Atmospheric gravity waves generated in the high-latitude ionosphere: a review. Rev. Geophys. Space Phys. **20**, 293–315 (1982)
Hunsucker, R.D., Hargreaves, J.K.: The high-latitude ionosphere and its effects on radio propagation Cambridge Atmospheric and Space Science Series. Cambridge University Press, Cambridge (2002). https://doi.org/10.1017/CBO9780511535758
Jacchia, L.G.: New static models of the thermosphere and exosphere with empirical temperature profiles. Smithson. Astrophys. Special Report 313 (1970)
Jin, H., Miyoshi, Y., Fujiwara, H., Shinagawa, H., Terada, K., Terada, N., Ishii, M., Otsuka, Y., Saito, A.: Vertical connection from the tropospheric activities to the ionospheric longitudinal structure simulated by a new Earth's whole atmosphere-ionosphere coupled model. J. Geophys. Res. **116**, A01316 (2011). https://doi.org/10.1029/2010JA015925
Jones, J.B.L., Bentley, R.D., Hunter, R., Iles, R.H.A., Taylor, G.C., Thomas, D.J.: Space weather and commercial airlines. Adv. Space Res. **36**, 2258–2267 (2005). https://doi.org/10.1016/j.asr.2004.04.017
Kikuchi, T., Lühr, H., Kitamura, T., Saka, O., Schlegel, K.: Direct penetration of the polar electric field to the equator during a DP2 event as detected by the auroral and equatorial magnetometer chains and the EISCAT radar. J. Geophys. Res. **101**(A8), 17161–17173 (1996)
Kikuchi, T., Hashimoto, K.K., Nozaki, K.: Penetration of magnetospheric electric fields to the equator during a geomagnetic storm. J. Geophys. Res. **113**, A06214 (2008). https://doi.org/10.1029/2007JA012628
Kopp, E.: On the abundance of metal ions in the lower ionosphere. J. Geophys. Res. **102**, 9667–9674 (1997). https://doi.org/10.1029/97JA00384

Kress, B.T., Mertens, C.J., Wiltberger, M.: Solar energetic particle cutoff variations during the 29–31 October 2003 geomagnetic storm. Space Weather. **8**, S05001 (2010). https://doi.org/10.1029/2009SW000488

Le, H., Liu, L., Wan, W.: An analysis of thermospheric density response to solar flares during 2001–2006. J. Geophys. Res. **117**, A03307 (2012). https://doi.org/10.1029/2011JA017214

Livneh, D.J., Seker, I., Djuth, F.T., Mathews, J.D.: Continuous quasiperiodic thermospheric waves over Arecibo. J. Geophys. Res. **112**, A07313 (2007). https://doi.org/10.1029/2006JA012225

Lu, G., Richmond, A.D., Lühr, H., Paxton, L.: High-latitude energy input and its impact on the thermosphere. J. Geophys. Res. Space Physics. **121**, 7108–7124 (2016). https://doi.org/10.1002/2015JA022294

Makela, J.J., Kelley, M.C.: Using the 630.0-nm nightglow emission as a surrogate for the ionospheric Pedersen conductivity. J. Geophys. Res. **108**(A6), 1253 (2003)

Mathews, J.D.: Sporadic E: Current views and recent progress. J. Atmos. Sol. Terr. Phys. **60**, 413–435 (1998). https://doi.org/10.1016/S1364-6826(97)00043-6

Mathews, J.D., Sulzer, M.P., Perillat, P.: Aspects of layer electrodynamics inferred from high-resolution ISR observations of the 80–270 km ionosphere. Geophys. Res. Lett. **24**, 1411–1414 (1997). https://doi.org/10.1029/97GL01320

Matsumura, M., et al.: GAIA simulations of the ionospheric response to successive X-class solar flares on September 6, 2017. unpublished manuscript, (2018)

Matuura, N., Tsuda, T.T., Nozawa, S.: Field-aligned current loop model on formation of sporadic metal layers. J. Geophys. Res. Space Physics. **118**, 4628–4639 (2013). https://doi.org/10.1002/jgra.50414

Milan, S.E., Lester, M., Yeoman, T.K.: HF radar polar patch formation revisited: summer and winter variations in dayside plasma structuring. Ann. Geophys. **20**, 487–499 (2002). https://doi.org/10.5194/angeo-20-487-2002

Misra, P., Enge, P.: Global Positioning System: Signals, Measurements, and Performance (Re-vised Second Edition). Ganga-Jamuna Press, Lincoln (2006)

Miyoshi, Y., Jin, H., Fujiwara, H., Shinagawa, H.: Numerical study of traveling ionospheric disturbances generated by an upward propagating gravity wave. J. Geophys. Res. Space Phys. **123**, 2141–2155 (2018). https://doi.org/10.1002/2017JA025110

Nishioka, M., Tsugawa, T., Jin, H., Ishii, M.: A new ionospheric storm scale based on TEC and foF2 statistics. Space Weather. **15**, 228–239 (2017). https://doi.org/10.1002/2016SW001536

Ogawa, T., Sinno, K., Fujita, M., Awaka, J.: Severe disturbances of VHF and GHz waves from geostationary satellites during a magnetic storm. J. Atmos. Terr. Phys. **42**(7), 637–644 (1980)

Otsuka, Y., Shiokawa, K., Ogawa, T., Wilkinson, P.: Geomagnetic conjugate observations of equatorial airglow depletions. Geophys. Res. Lett. **29**(15) (2002). https://doi.org/10.1029/2002GL015347

Otsuka, Y., Shiokawa, K., Ogawa, T., Wilkinson, P.: Geomagnetic conjugate observations of medium-scale traveling ionospheric disturbances at midlatitude using all-sky airglow imagers. Geophys. Res. Lett. **31**, L15803 (2004). https://doi.org/10.1029/2004GL020262

Perkins, F.: Spread F and ionospheric currents. J. Geophys. Res. **78**, 218–226 (1973)

Picone, J.M., Hedin, A.E., Drob, D.P., Aikin, A.C.: NRLMSISE-00 empirical model of the atmosphere: statistical comparisons and scientific issues. J. Geophys. Res. **107**(A12), 1468 (2002)

Pitteway, M.L.V., Hines, C.O.: The viscous damping of atmospheric gravity waves. Can. J. Phys. **41**, 1935–1948 (1963)

Plane, J.M.C., Feng, W., Dawkins, E.C.M.: The mesosphere and metals: chemistry and changes. Chem. Rev. **115**, 4497–4541 (2015). https://doi.org/10.1021/cr500501m

Qian, L., Burns, A.G., Chamberlin, P.C., Solomon, S.C.: Flare location on the solar disk: modeling the thermosphere and ionosphere response. J. Geophys. Res. **115**, A09311 (2010). https://doi.org/10.1029/2009JA015225

Roble, R.G., Ridley, E.C., Dickinson, R.E.: On the global mean structure of the thermosphere. J. Geophys. Res. **92**(A8), 8745–8758 (1987). https://doi.org/10.1029/JA092iA08p08745

Sakai, J., Hosokawa, K., Tomizawa, I., Saito, S.: A statistical study of anomalous VHF propagation due to the sporadic-E layer. Radio Sci. **54**, 426–439 (2019). https://doi.org/10.1029/2018RS006781

Secan, J.A., Bussey, R.M., Fremouw, E.J., Basu, S.: An improved model of equatorial scintillation. Radio Sci. **30**(3), 607–617 (1985). https://doi.org/10.1029/94RS03172

Shinagawa, H., Miyoshi, Y., Jin, H., Fujiwara, H.: Global distribution of neutral wind shear associated with sporadic E layers derived from GAIA. J. Geophys. Res. Space Physics. **122**, 4450–4465 (2017). https://doi.org/10.1002/2016JA023778

Shinagawa, H., Jin, H., Miyoshi, Y., Fujiwara, H., Yokoyama, T., Otsuka, Y.: Daily and seasonal variations in the linear growth rate of the Rayleigh-Taylor instability in the ionosphere obtained with GAIA. Prog Earth Planet Sci. **5**, 16 (2018). https://doi.org/10.1186/s40645-018-0175-8

Shiokawa, K., Otsuka, Y., Ihara, C., Ogawa, T., Rich, F.J.: Ground and satellite observations of nighttime medium-scale traveling ionospheric disturbance at midlatitude. J. Geophys. Res. **108**(A4), 1145 (2003). https://doi.org/10.1029/2002JA009639

Smithtro, C., and S., Solomon: An improved parameterization of thermal electron heating by photoelectrons, with application to an X17 flare. J. Geophys. Res. 113, (2008) 10.1029/2008JA013077

Solomon, S.C., Qian, L.: Solar extreme-ultraviolet irradiance for general circulation models. J. Geophys. Res. **110**, A10306 (2005). https://doi.org/10.1029/2005JA011160

Tsugawa, T., Saito, A., Otsuka, Y., Yamamoto, M.: Damping of large-scale traveling ionospheric disturbances detected with GPS networks during the geomagnetic storm. J. Geophys. Res. **108**(A3), 1127 (2003). https://doi.org/10.1029/2002JA009433

Tsunoda, R.T.: Control of the seasonal and longitudinal occurrence of equatorial scintillations by the longitudinal gradient in the integrated E-region Pedersen conductivity. J. Geophys. Res. **90**, 447 (1985)

Tsunoda, R.T.: On equatorial spread F: Establishing a seeding hypothesis. J. Geophys. Res. **115**, A12303 (2010). https://doi.org/10.1029/2010JA015564

Tsurutani, B.T., et al.: Extreme solar EUV flares and ICMEs and resultant extreme ionospheric effects: comparison of the Halloween 2003 and the bastille day events. Radio Sci. **41**, RS5S07 (2006). https://doi.org/10.1029/2005RS003331

Vadas, S.L., Liu, H.: Generation of large-scale gravity waves and neutral winds in the thermosphere from the dissipation of convectively generated gravity waves. J. Geophys. Res. **114**, A10310 (2009). https://doi.org/10.1029/2009JA014108

Whitehead, J.D.: Production and prediction of sporadic E. Rev. Geophys. **8**, 65–144 (1970). https://doi.org/10.1029/RG008i001p00065

Whitehead, J.D.: Recent work on mid-latitude and equatorial sporadic-E. J. Atmos. Sol. Terr. Phys. **51**, 401–424 (1989). https://doi.org/10.1016/0021-9169(89)90122-0

Wu, D.L., Ao, C.O., Hajj, G.A., Juarez, M.T., Mannucci, A.J.: Sporadic E morphology from GPS-CHAMP radio occultation. J. Geophys. Res. **110**, A01306 (2005). https://doi.org/10.1029/2004JA010701

Yokoyama, T., Jin, H., Shinagawa, H., Liu, H.: Seeding of equatorial plasma bubbles by vertical neutral wind. Geophys. Res. Lett. **46**, 7088–7095 (2019). https://doi.org/10.1029/2019GL083629

Zhang, D.H., Xiao, Z.: Study of ionospheric response to the 4B flare on 28 October 2003 using international GPS service network data. J. Geophys. Res. **110**, A03307 (2005). https://doi.org/10.1029/2004JA010738

Part III
Solar Storms and Space Weather Forecast

Chapter 8
Structure of Solar Atmosphere and Magnetic Phenomena

Kiyoshi Ichimoto, Toshifumi Shimizu, Kazumasa Iwai, and Hisayoshi Yurimoto

8.1 Solar Constant and Overall Structure of Solar Atmosphere

The Sun is a standard star (G-type star) in the universe. It has a diameter of approximately 1.4 million km, mass of 2×10^{30} kg, and rotation period of approximately 27 days. It is a gaseous body consisting mainly of hydrogen and helium, and its surface gravity is 28 times greater than that of the Earth. The thermal energy generated in the center of the Sun by the fusion reaction of hydrogen is transported by radiation to approximately 70% of its radius and subsequently carried to the surface by convective motion. The temperature of the photosphere is approximately 5800 K, and the entire sphere emits 4×10^{26} J of energy per second as light. The Earth is approximately 150 million km from the Sun and receives 1.37 kW/m^2 of the energy outside the atmosphere (which is called the "solar constant"). Most of this energy is contained in the near-ultraviolet, visible, and near-infrared regions (Fig. 8.1); however, the spectrum emitted by the Sun covers a wide range, from X-rays to radio waves. Although it accounts for only a small fraction of the total radiant energy, the intensity of the radiation from X-rays to ultraviolet rays

K. Ichimoto (✉)
Astronomical Observatory, Graduate School of Science, Kyoto University, Kyoto, Japan
e-mail: ichimoto@kwasan.kyoto-u.ac.jp

T. Shimizu
Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Sagamihara, Kanagawa, Japan

K. Iwai
Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Aichi, Japan

H. Yurimoto
Department of Earth and Planetary Sciences, Hokkaido University, Sapporo, Hokkaido, Japan

K. Kusano (ed.), *Solar-Terrestrial Environmental Prediction*,
https://doi.org/10.1007/978-981-19-7765-7_8

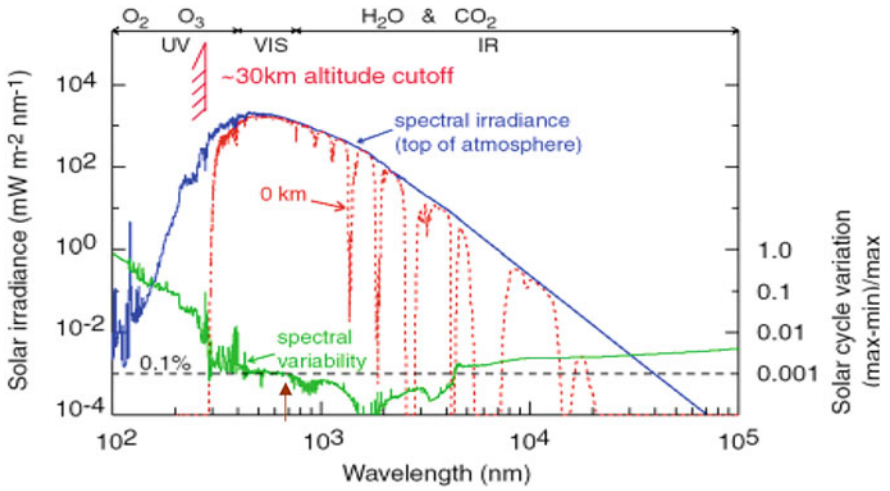

Fig. 8.1 Solar radiation spectrum. Blue line shows spectrum outside the atmosphere of the Earth, and red line shows that at the surface of the Earth. Green line (right-hand side scale) shows fraction of variation associated with solar cycle activity (Lean 2000)

significantly varies with the solar activity and is an important cause of environmental changes on the Earth.

The solar spectrum in the visible to infrared region contains numerous absorption line spectra corresponding to various elements in the solar atmosphere. Absorption lines are formed because the temperature of the photosphere decreases with height. However, at wavelengths shorter than 190 nm, the line spectrum turns into emission lines. This indicates the existence of a higher temperature region (chromosphere or transition region) above the photosphere. In addition, the soft X-ray emission suggests the existence of a temperature region above one million K (corona). The temperature and density structure of the average solar atmosphere constructed to reproduce the observed solar spectrum are shown in Fig. 8.2. Here, the origin of the altitude (the horizontal axis) is conventionally taken to be the point where the optical depth in continuum light at 500 nm is 1. Above the convection layer is the photosphere, where the atmospheric temperature decreases with the altitude and most of the energy is emitted as near-ultraviolet, visible, and infrared light. At an altitude of approximately 500 km, the temperature begins to rise, and the chromosphere remains at approximately 10,000 °C until an altitude of approximately 2000 km. The density in the chromosphere decreases by approximately four orders of magnitude with the altitude; however, because ionization of hydrogen proceeds along the height, the electron density is almost constant with the height in the chromosphere. The chromosphere can be observed by absorption lines with high extinction coefficients, such as the Hα line (656.2 nm), and by emission lines in the ultraviolet region. At an altitude of approximately 2000 km, the temperature abruptly rises to more than one million K, forming a corona that emits soft X-rays and the

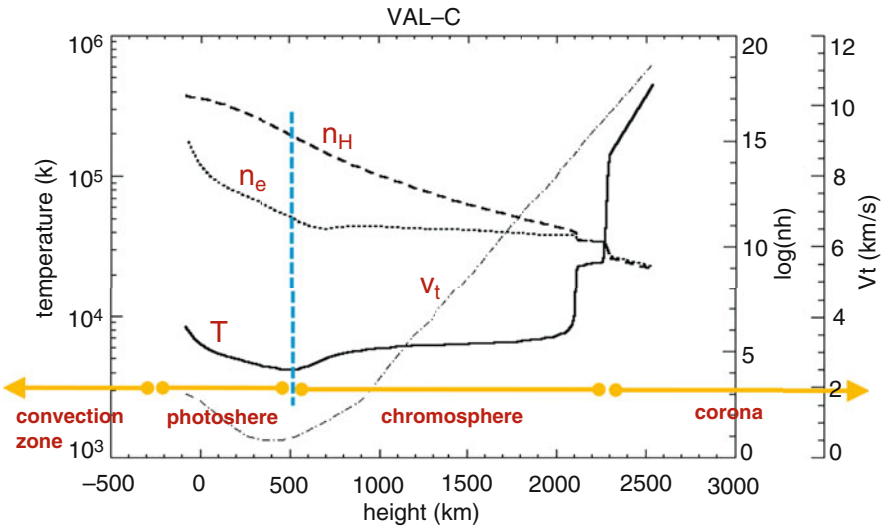

Fig. 8.2 Reference model of solar atmosphere (VAL-C; Vernazza et al. (1981))

extreme ultraviolet (EUV) radiation. The boundary region between the chromosphere and the corona is called the transition layer, from which emission lines of various ions are emitted.

Figure 8.3 shows solar images seen in different lights. It is noticeable that there are sunspots (groups) in the photosphere, which connect to the bright structures of the chromosphere and a corona by a strong magnetic field. A region containing sunspots and a bright corona is called active region, and the other region is called a quiet region. In quiet regions, the photosphere appears to shine almost uniformly, whereas the corona has structures such as coronal bright points (relatively small scale) and dark coronal holes (large scale). Because ultraviolet and X-rays have locally bright regions, the intensity of the radiation reaching the Earth at these wavelengths varies with a period of approximately 27 days as the Sun rotates.

8.2 Photosphere: Interaction of Convection and Magnetic Fields

The first striking feature in a high-resolution image of the Sun in a quiet region is the granular patch that covers the entire surface of the Sun (Fig. 8.4). It contains convection cells (granulations) of approximately 1500 km in size, which form and disappear on a timescale of approximately 10 min and maintain the surface temperature of the Sun by carrying heat energy from inside it. The gaps between the granulations contain rows of bright points with sizes of 100–200 km. This is the area where a magnetic field of approximately 1 kG is concentrated, and time series of

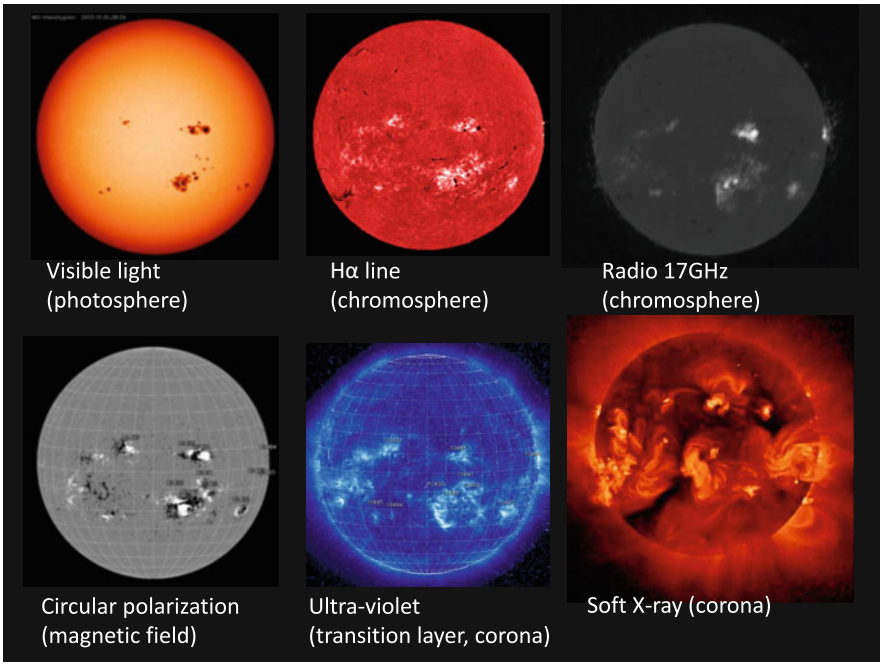

Fig. 8.3 Sun seen at different wavelengths. From top left to bottom right, SOHO/Michelson Doppler Imager (MDI), Hida Observatory/Solar Magnetic Activity Research Telescope (SMART), Nobeyama Radio Heliograph, SOHO/MDI, SOHO/Extreme ultraviolet Imaging Telescope 171A (October 30, 2003); bottom right, "Yohkoh"

high-resolution images show that the convective motion of the granulations is constantly shaking them horizontally. These regions, where bright points are aligned, are formed when the magnetic field in the quiescent region is swept by the horizontal flow, which has a larger spatial scale than that of the granulations. This large-flow structure is called a supergranulation, which typically has a size of approximately 30,000 km and a lifetime of approximately 1 day. Other intermediate convection cells with sizes of 10,000–20,000 km are also known to exist on the solar surface, and thus, convection has a hierarchical structure.

Figure 8.5 shows the distribution of circularly (center) and linearly (right) polarized signals in FeI 656.2-nm absorption line in a quiescent region observed by the Hinode satellite near the center of the Sun. Because the solar surface is being viewed from above, the circularly polarized signal is the magnetic field component perpendicular to the solar surface, whereas the linearly polarized signal is the horizontal magnetic field component along the solar surface. As can be seen, the magnetic field exists in a discrete form everywhere on the solar surface, even in the quiet region. The magnetic fields perpendicular to the solar surface are distributed as a network of kilogauss (kG)-intensity flux tubes, particularly concentrated at the boundaries of supergranular or intermediate-granular cells. The horizontal magnetic field is

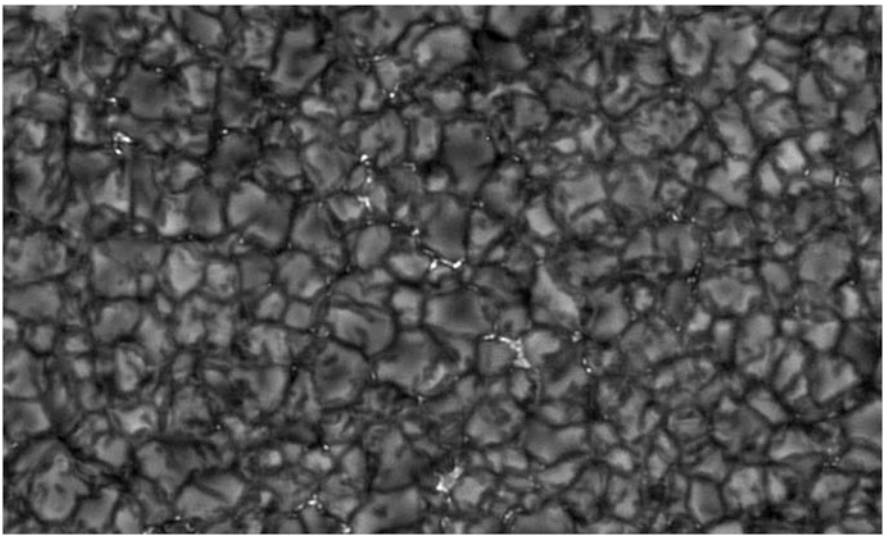

Fig. 8.4 High-resolution image of quiet region at 430 nm (G-band) by Hinode Solar Optical Telescope (SOT) (November 22, 2006)

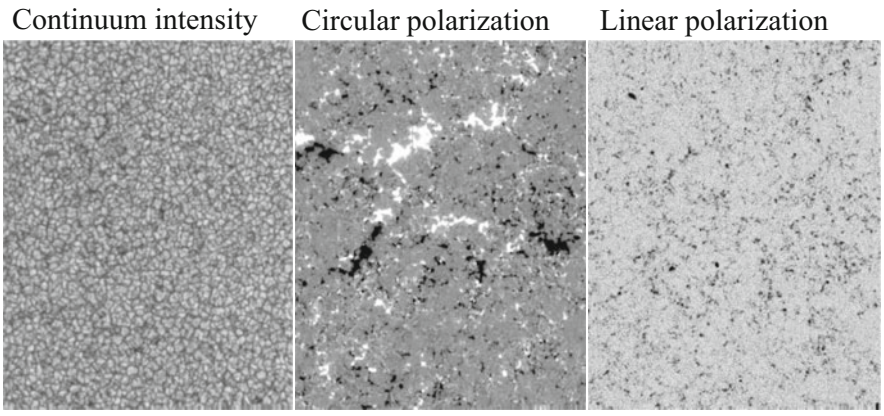

Fig. 8.5 Quiet region at solar disk center. From left to right, continuum intensity, circular polarization, and linear polarization in FeI 630.2 nm. By Hinode SOT (from *The Sun as a Guide to Stellar Physics*, Ichimoto 2019)

estimated to be several hundred gauss and is continuously generated and dissipated on the same timescale as that of the individual granulations. It was pointed out that the energy supply by these magnetic fields may play an important role in the formation of the upper atmosphere (Ishikawa and Tsuneta 2009).

When a strong magnetic field is concentrated on a large scale, the convective motion, which is the source of energy from below, is suppressed, resulting in a reduced surface temperature and the formation of sunspots. Although the typical

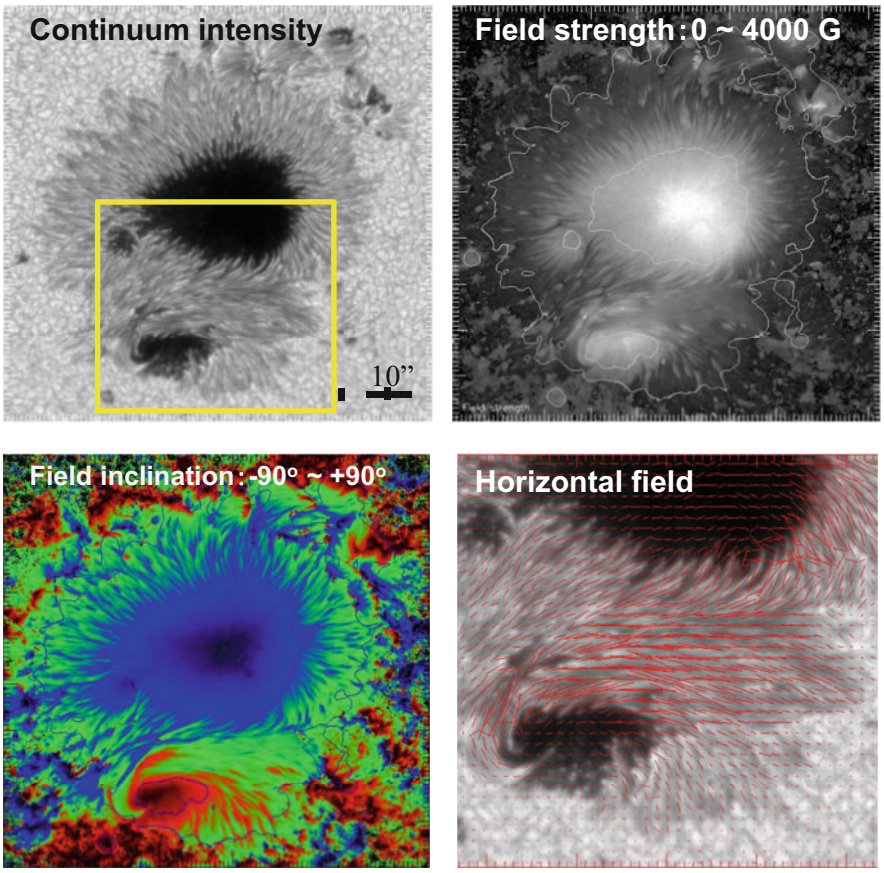

Fig. 8.6 Magnetic field structures of sunspots by Hinode SOT. Continuum image (top left), field strength (top right), inclination of magnetic field (bottom left), and horizontal field direction (bottom right)

magnetic field intensity of a sunspot is 2000–4000 G, intensities as high as 6250 G have been reported (Okamoto and Sakurai 2018). Figure 8.6 shows the distribution of photosphere magnetic fields in an active region observed by the Hinode satellite. The dark area in the center of the sunspot (an umbra) does not show convection-like granules; instead weak magnetized convection phenomena are observed as many bright points (umbral dots), maintaining a temperature of 3000–5000 K inside the sunspot. A penumbra surrounding the umbra is formed by many radially extending striations, and in the photosphere, there is a violent outward flow exceeding several kilometer per second along the striations with a nearly horizontal magnetic field (Evershed 1909). It is understood that these are also thermal convection phenomena occurring under the influence of a strong magnetic field (Ichimoto et al. 2007).

The upper and lower sunspots in Fig. 8.3 have different polarities, S and N, respectively, as can be seen from the elevation angle of the magnetic field (lower left

panel). The map at the lower right shows the direction of the horizontal magnetic field connecting the two sunspots, which is very different from the simple connection between N and S polarities. It also presents a strongly sheared structure almost along the magnetic neutral line. This indicates that the magnetic field structure significantly deviates from the current-free potential field (rot $\mathbf{B} = 0$) and that the magnetic field has a large free energy associated with the strong electric current. This is the origin of a flare explosion, and analyzing the vector magnetic field of the photosphere is important for space weather research, as described in more detail in Chap. 10.

In the photosphere, the pressure of gases dominates the pressure of the magnetic field, except in sunspots, where the magnetic field is particularly strong. Moreover, the energy is transmitted to the upper atmosphere by the motion of the gases caused by convection, which distorts the magnetic field penetrating the corona. Some of the high-frequency components of the oscillations of the photosphere, which have peak power at a period of 5 min, are also considered to transmit energy to the upper atmosphere as waves.

8.3 Chromosphere: From Gas-Dominated to Magnetic Field-Dominated

The chromosphere is a region in which the density decreases with height and the magnetic pressure dominates the gas pressure. Thus, the shapes of the density structures that form the chromosphere are determined by the magnetic field. The most universal chromospheric structures are spicules protruding from the surface. They are frequently observed at the solar limb in the Hα and calcium HK lines (Fig. 8.7), and their lifetimes begin with the ascent of the gases and end with their descent or disappearance in several minutes. Their typical size is approximately 10,000 km in length and 100–300 km in width. The lateral motion of spicules captures the upward passage of Alfvénic waves (Okamoto and de Pontieu 2011). Spicules are densely packed in the boundary region of supergranulation, and the total flux of the gases launched by them is estimated to be 100 times larger than the mass flowing out of the corona as the solar wind. It has been suggested that a compressive wave is generated by the pressure enhancement under the foot, which is nonlinearized into a shock wave in the process of upward propagation. Moreover, a shock wave passes through the transition layer and ejects chromospheric gases (Suematsu et al. 1982; Shibata 1982); however, the origin of the increase in the pressure beneath the feet has not yet been clarified.

Figure 8.8 shows images of an active region in visible continuum light (top) and the Hα line (bottom). This region consists of bipolar sunspots aligned in the east–west direction. The Hα line image shows that the chromosphere is composed of striated density structures (fibrils). Fibrils connecting sunspots of different polarities and radiating outward from the sunspots are characteristic.

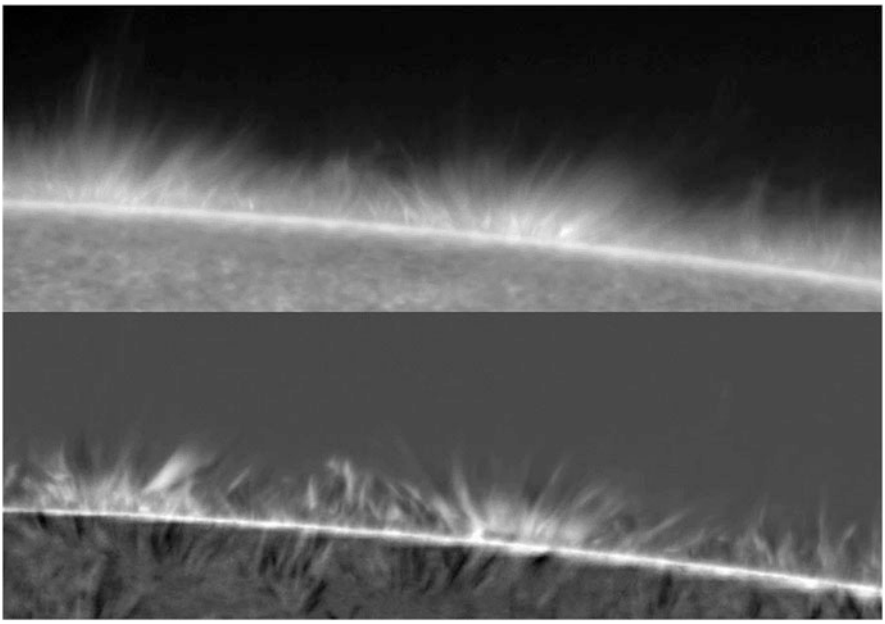

Fig. 8.7 Spicules seen at solar limb. CaH intensity (top) and Hα Doppler map (bottom) by Hinde SOT

In addition, there are small sunspots between the two sunspot groups and several short arch-like streaks. This is the region where a magnetic flux newly emerges from under the photosphere. Magnetic levitation is a common phenomenon during the growth phase of a sunspot group and is an important factor in predicting the onset of flares because it can induce large-scale instability of the magnetic field structure. Innumerable fine striations constituting the chromosphere are accompanied by gas flows.

Prominences (or dark filaments seen on the solar disk) are plasmas of approximately 10,000 K floating in a hot corona of 1 MK. The sizes of prominences extensively vary; however, relatively large ones can reach hundreds of thousands of kilometers in length and tens of thousands of kilometers in height, and they remain stable for more than a week. The pressure balance causes the density of prominences to be two orders of magnitude higher than that of the surrounding corona, and without a supporting force, they can fall to the surface in approximately 10 min under the strong gravity of the Sun. A prominence is supported by the magnetic field, i.e., a Lorentz force generated by the electric current. The concept that prominences are strongly related to the magnetic field is based on the fact that they appear above the global magnetic neutral line of the photosphere (Fig. 8.9).

Prominences can be broadly classified into quiescent and active region prominences (Fig. 8.10). The former type consists of tall, vertically elongated threads in which small-scale plasmas move up and down continuously, similar to convection,

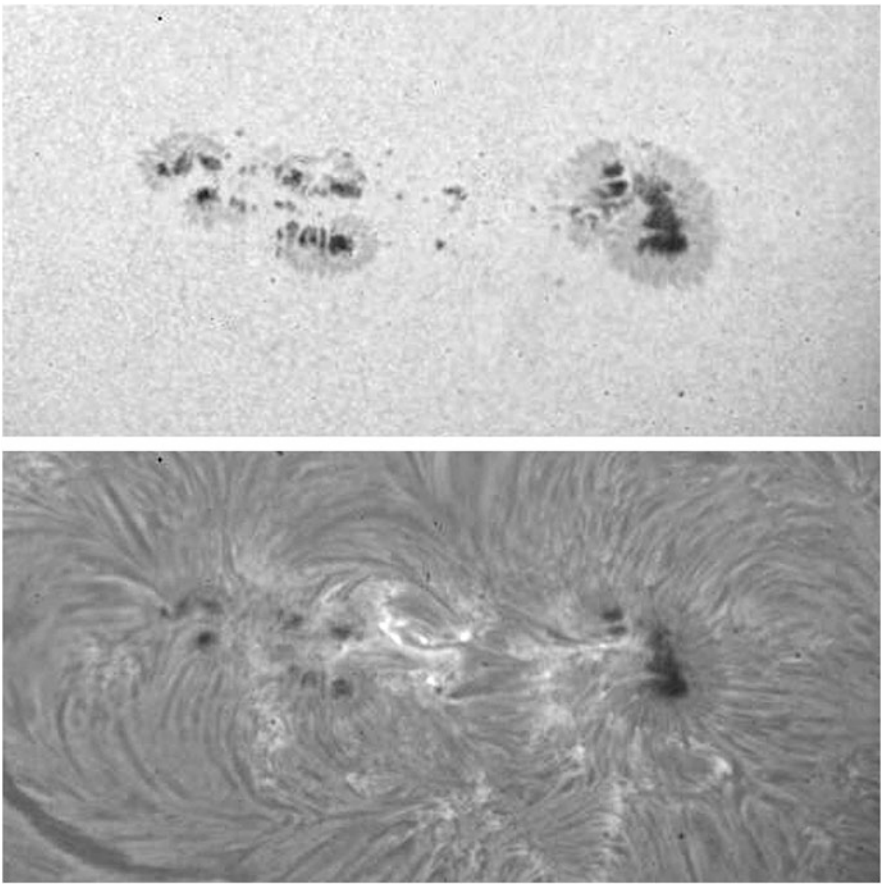

Fig. 8.8 Images of active region in visible continuum (top) and Hα (bottom). Domeless solar telescope of Hida Observatory

despite their global stability (Berger et al. 2008). In contrast, the latter type consists mainly of horizontally extended striations, in which Alfvénic transverse waves are frequently excited (Okamoto et al. 2007).

The structure of the magnetic field supporting a prominence is depicted as a convex magnetic field line in the corona connecting the positive and negative polarities of the photosphere, where the coronal plasma cools and accumulates (Hillier et al. 2012). However, the measurement of the magnetic field is difficult owing to weak polarization signals, and details of the magnetic field structures of prominences are still unidentified. It is known that prominences can abruptly destabilize, rise, and fly away from the gravity of the Sun, frequently accompanied by coronal mass ejections (CMEs). Therefore, one of the important issues in space weather research is to accurately measure the magnetic fields of prominences and analyze their stability.

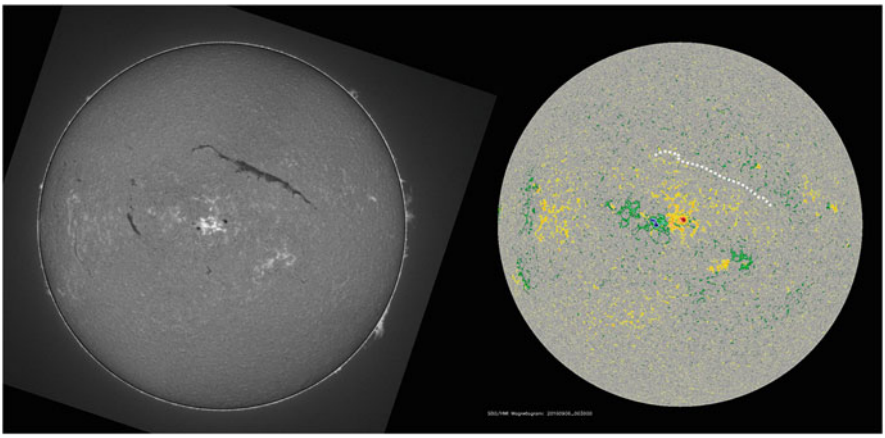

Fig. 8.9 Full-disk Hα image (left, SMART at Hida Observatory) and line of sight magnetic field (right, MDI on SOHO) (Ichimoto 2019)

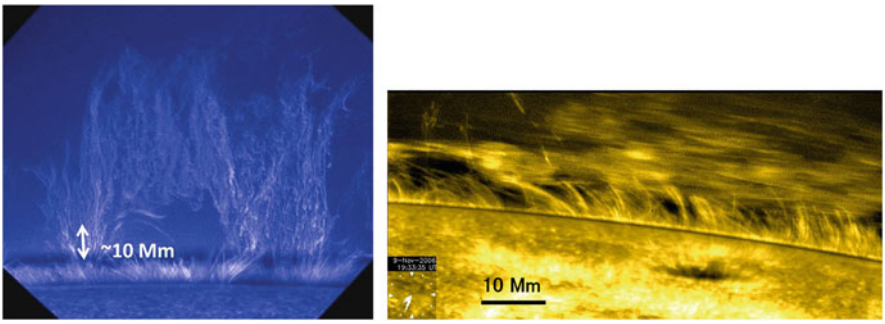

Fig. 8.10 Quiescent (left) and active (right) prominences. Hinode SOT (Ichimoto 2019)

8.4 Corona and Solar Wind: Hot Atmosphere Dominated by Magnetic Fields

The solar corona is the outer atmosphere of the Sun, with a temperature of more than one million K, formed above the solar surface (photosphere) at approximately 6000 K. The corona contains diverse dynamics created by the interplay of the solar magnetic field and plasma. In this section, we focus on the solar corona.

The most massive coronal dynamics are flares, whose source is the magnetic energy stored in the magnetic field above the solar surface, including the corona. When this energy is converted into heat and kinetic energy in a short time, abrupt plasma heating occurs, and a flare is recognized as an abrupt intensification of X-rays and ultraviolet rays. A flare may also be accompanied by a plasma jet (jet or a CME).

The occurrence of a flare can be identified from the increase in the intensity of soft X-rays by the Geostationary Operational Environmental Satellite (GOES). From the

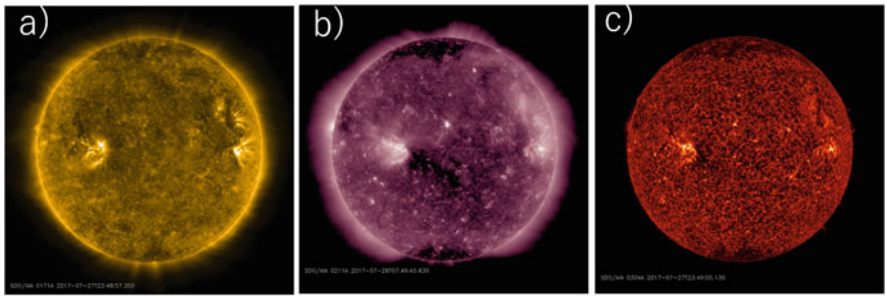

Fig. 8.11 Ultraviolet images acquired by AIA on board SDO. (**a**) 17.1-nm band image containing Fe IX emission line (approximately 800,000 K) and (**b**) 21.1-nm band image containing Fe XIV emission line (approximately two million K) capture temperature of coronal plasma. (**c**) In ultraviolet region, He II 30.4-nm emission line (50,000 K) is observed originating from chromospheric plasma and a 30.4-nm band image including it captured by AIA (credit: NASA/SDO)

images of ultraviolet and soft X-rays captured as two-dimensional spatial information, we can determine the location on the solar disk where a flare occurred and its form (morphology). In addition, dynamic motions such as plasma eruptions can be captured by time-lapse imaging. Figure 8.11 shows examples of EUV images recorded by the Atmospheric Imaging Assembly (AIA) telescope on board the Solar Dynamics Observatory (SDO) of the National Aeronautics and Space Administration (NASA), which has been constantly observing the coronal atmosphere since 2010. The AIA telescope images in various wavelength bands using the property of multilayers to reflect and select only certain wavelength bands. These EUV images are effective in capturing the spatial distribution of the coronal plasma limited to a certain temperature. In contrast, soft X-ray telescopes are good at capturing a wide range of temperatures from one million to over ten million K. Figure 8.12 is an example image captured by the X-ray telescope (XRT) on board the solar physics satellite, Hinode.

EUV and soft X-ray coronal images show that the solar corona is highly structured that it is difficult to imagine it from the flat photosphere visible in the white light. Bright regions are called "active-region" coronas, whereas regions with little or no coronal emission are called "coronal holes." The region in between is called a "quiet-region" corona.

Active regions are regions where many flares occur. Sunspots are frequently present in the photosphere, and the solar magnetic field is particularly concentrated in these regions. Above the photosphere, where the positive- and negative-polarity magnetic patches are distributed, there is a striated coronal structure that connects them, which is called a "coronal loop." As shown in Fig. 8.13, in addition to stable (potential) coronal loops in the shapes of arcs, twisted structures that are presumed to have accumulated energy can be visually observed, which is useful for estimating the magnetic field structure at the source of flare and CME generation.

In Fig. 8.12, we notice that there are many small bright points in the quiet-region corona. These are called X-ray bright points (XBPs), and they have lifetimes ranging

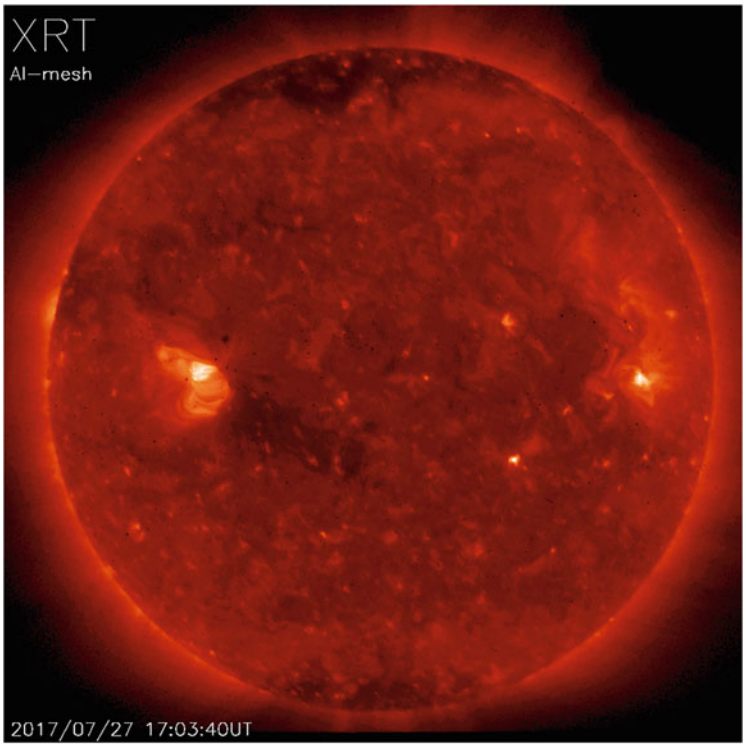

Fig. 8.12 Soft X-ray corona image. Image is taken by XRT on board solar physics satellite, Hinode (credit: JAXA/National Astronomical Observatory of Japan/NASA/SAO)

from a few hours to several tens of hours. In the photosphere, the formation of a short-lived ephemeral region, which is a small magnetic bipolar region, and/or magnetic cancellation, in which positive and negative magnetic patches approach each other and disappear, are occasionally observed. Thus, the evolution of the magnetic field, such as the formation of an ephemeral region and annihilation of the magnetic flux observed as a magnetic cancellation, is considered to be the source of XBPs. In addition, small flares (microflares and nanoflares) are observed in XBPs, suggesting that the magnetic reconnection is driven by the magnetic field change.

Large-scale examples of coronal holes are frequently found in the polar regions of the Sun; however, they are also observed at mid- and low-latitudes. They are considered to be important structures for space weather as sources of the solar wind (details are provided in the next section). A polar region with a coronal hole seen in EUV and soft X-ray coronal images is magnetically unipolar in the photosphere. A unipolar region is a primary origin of the open magnetic field lines connecting to the heliosphere (interplanetary space), which is the basic coronal magnetic structure for the acceleration mechanism of the solar wind at high speeds (more than 700 km/s). The solar wind also has a low-speed component

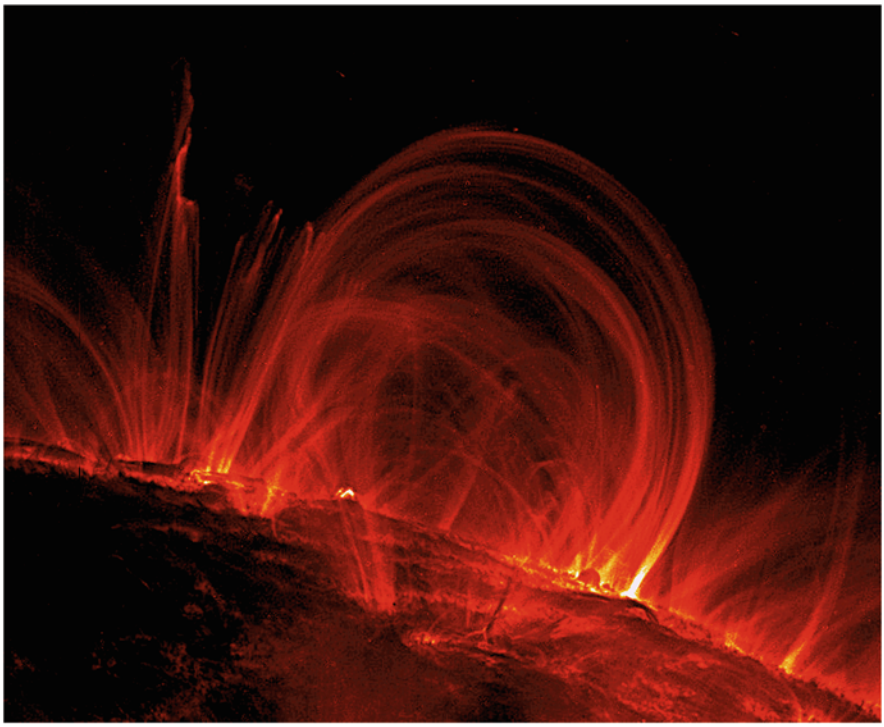

Fig. 8.13 Coronal loops captured by TRACE satellite (NASA) in high-resolution ultraviolet image (credit: NASA/LMSAL)

(300–400 km/s), and there are many theories about its origin and mechanism. Recent Hinode satellite observations suggest that the source is a coronal hole formed immediately next to an active-region corona. A magnetic field line from such a coronal hole may be open to the heliosphere.

The plasma ejected from the solar corona into the heliosphere can be captured by a coronagraph. A coronagraph is a particular telescope that creates an artificial "total eclipse" by removing the very bright solar light from the solar disk and can observe the faint visible light originating from the corona. Figure 8.14 shows an image captured by a Large Angle and Spectrometric Coronagraph (LASCO) on board the Solar and Heliospheric Observatory (SOHO) satellite (the European Space Agency (ESA)/NASA). A coronal plasma erupting from the Sun can be captured at a distance up to tens of times greater than the solar disk.

The solar corona is an atmosphere dominated by the magnetic field, and understanding the magnetic field in the corona is important for physically understanding the coronal dynamics that occur in it. Coronal loops observed in EUV and soft X-ray imaging are used to determine the morphology of the magnetic structures in the corona. Because it is extremely difficult to measure the 3D magnetic field in the corona directly, modeling has been performed to estimate it using maps of the

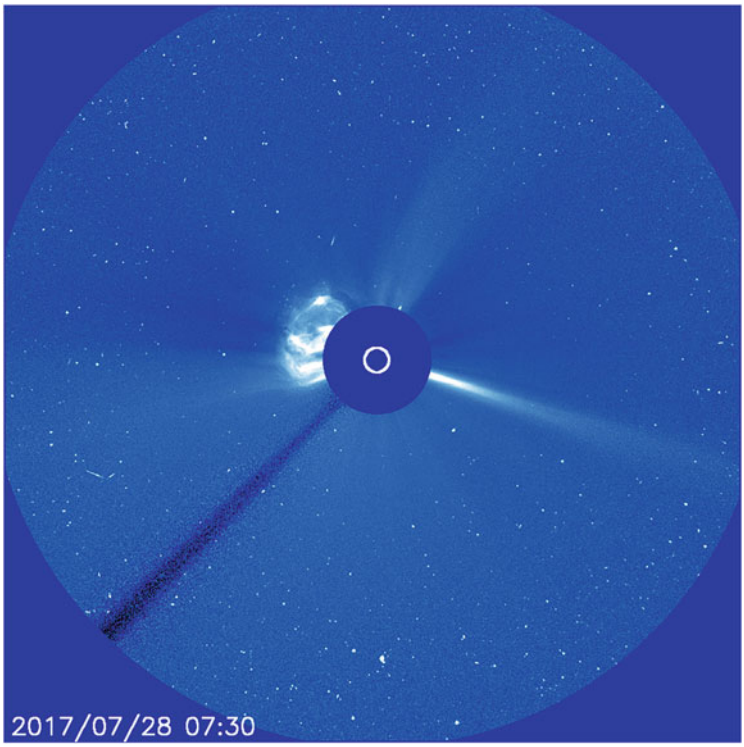

Fig. 8.14 Images from LASCO on board SOHO satellite (credit: ESA/NASA)

magnetic field distribution observed on the photosphere as boundary conditions (details are discussed in Chap. 10).

There is a major problem regarding the corona that has not been solved for several decades, called the "coronal heating problem." As described in Sect. 8.1, as we move upward from the photosphere at approximately 6000 K, the temperature begins to rise, and the temperature of the corona exceeds one million K. The existence of the hot corona cannot be explained by the thermodynamic energy flow. In the corona, a constant loss of energy is due to thermal radiation (mainly at EUV and soft X-ray wavelengths) and heat conduction to the solar surface in closed magnetic field structures. In addition, the mass is removed by the solar wind in coronal holes, where the magnetic field is open to the heliosphere. These losses cool down ultimately. The ubiquitous corona is formed by injection of energy commensurate with these losses by nonthermal processes, and plasma heating occurs. The energy input is 10^4 J/m^2/s in active regions, 10^2 J/m^2/s in quiet regions, and 10^3 J/m^2/s in coronal holes (Withbroe and Noyes 1977), and this physical mechanism is called the "coronal heating problem." The loss mechanisms in different regions are different owing to the variations in the coronal structures. In active and quiet regions, the

mechanism is thermal radiation or heat conduction to the solar surface, whereas in coronal holes, it is outflow by the solar wind.

The sources of the energy for the coronal heating are turbulent convection motions in the photosphere. The energy flux due to turbulent motions is estimated to be 10^5–10^6 J/m/s^2, and it is considered that a part of this kinetic energy is removed by the magnetic field and dissipated in the upper corona. There are two major theories; the first is that waves generated by the turbulent motions propagate through the magnetic field lines and dissipate their energy (the wave heating theory). The other is that the turbulent motions create numerous discontinuities in the coronal magnetic field, in which the energy is dissipated by many small flares (nanoflares) (the nanoflare heating theory). In observational and theoretical investigations, various types of approaches have been adopted to identify the heating mechanism.

It is noted that the definitions of a nanoflare and a microflare are based on the assumption that the energy released by the largest giant flare is approximately 10^{26} J. A nanoflare and a microflare have energy scales of approximately $10^{17–19}$ J and $10^{20–22}$ J, respectively. Nanoflares are frequently referred as sudden energy-release phenomena inside coronal loops.

8.5 Solar Winds and Heliosphere

As mentioned in the previous section, the solar corona is continuously heated and kept hot by some physical mechanism. Therefore, a part of the plasma atmosphere of the corona escapes the gravity of the Sun by its thermal pressure and flows out into the interplanetary space as a supersonic wind, i.e., the solar wind (Parker 1958). The solar wind speeds vary from 300 km/s to 800 km/s. There two types of the solar wind: slow solar wind (300–400 km/s) and fast solar wind (over 700 km/s). Fast solar wind is characterized by lower density and higher temperature than slow solar wind. It is known that fast winds flow out from a coronal hole, which is a region where the magnetic field lines are open toward the interplanetary space. It should be noted that fast solar wind, which has a higher temperature, flows out from those coronal holes that are cooler and less dense than the surrounding regions.

The acceleration mechanism of the solar wind is not well understood and is being actively studied as a major problem in solar and space physics. There are many unanswered questions at various levels, such as "what is the elementary process for accelerating the solar wind?" and the structural question of "from what region the solar wind flows out?", and these questions are related to each other (Viall and Borovsky 2020). The solar wind speed is considered to be related to the magnetic field structure of the source region. The relationship between the solar wind speed and the magnetic field structure on the solar surface has been empirically obtained. The solar wind velocity is correlated with the inverse of the expansion factor (f) *of* the coronal magnetic field (Wang and Sheeley 1990). It is also correlated with the area (A) of the coronal hole (Nolte et al. 1976). Interplanetary scintillation (IPS) observations have suggested that the ratio of the magnetic field strength (B) to the

expansion factor (B/f) is correlated with the solar wind speed (Fujiki et al. 2005), and a theoretical explanation of this correlation was provided by Suzuki (2006). These empirical models are still being improved with increasing observational data.

The solar wind basically travels in the radial direction with the Sun rotating in approximately 25 days. Because the magnetic field is frozen in the plasma in the interplanetary space, the magnetic field in the heliosphere forms a spiral structure, which is a combination of the radial motion of the solar wind and the rotation of the Sun (Parker spiral). The angle between the spiral structure and the radial direction is approximately 45° in the orbit of the Earth, and it becomes larger with increasing radial distance from the Sun. If the source regions of slow solar wind are located on the west side to the source region of fast solar wind, slow solar wind is aligned in front of fast solar wind in the radial direction on the same longitude. In this case, the high-speed wind reaches the low-speed wind and pushes in from behind. In this region, the density and temperature of the solar wind increase. Such a region causes a geomagnetic disturbance when it arrives at the Earth.

The "heliosphere" refers to the region that is controlled by the solar wind. The density of solar wind decreases with the radial distance. Then, the solar wind decelerates as it interacts with the interstellar medium flowing in the interstellar space, and a termination shock is formed in the region where the solar wind becomes subsonic speed. A heliopause is formed at the boundary between the solar wind and the interstellar medium. The outer part of the heliopause is the interstellar space. Voyager 1 and 2, launched in the 1970s, crossed the outer edge of the heliosphere after planetary exploration. Voyager 1 passed through a termination shock at 94 AU in 2004, and Voyager 2 passed through a termination shock at 84 AU in 2007. Both probes then continued to fly further to the outer heliosphere, with Voyager 1 and 2 passing through the heliopause in 2012 and 2018, respectively, at approximately 120 AU from the Sun, becoming the first artifacts to reach the interstellar medium (Burlaga et al. 2013, 2019). Energetic neutral atoms (ENAs) are produced at the outer edge of the heliosphere by interactions between the solar wind and the interstellar medium. The IBEX satellite, which observes ENAs in the orbit of the Earth, has provided many insights into the structure of the heliosphere (McComas et al. 2009).

The global distribution of the solar wind changes in response to the 11-year solar cycle. In general, during the solar minima, coronal holes in the solar pole regions expand, and fast solar wind regions expand to mid- and low latitudes, whereas slow solar wind regions are limited in a narrow band around the equator. During the solar maxima, the polar coronal hole regions shrink, and accordingly, fast wind regions shrink to high latitudes, whereas slow wind regions expand (Fig. 8.15). From the end of solar cycle 23 to the beginning of cycle 24, when a significant decrease in the solar activity occurred, a complex global structure of the solar wind was also observed (Tokumaru et al. 2015), and further detailed observations are required.

The solar wind scatters radio waves. This scattering phenomenon can be detected as an IPS by ground-based radio observations (Hewish et al. 1964). Although in situ observations by a spacecraft can provide detailed plasma parameters of its location, the exploration area is limited. In particular, it is difficult to investigate the

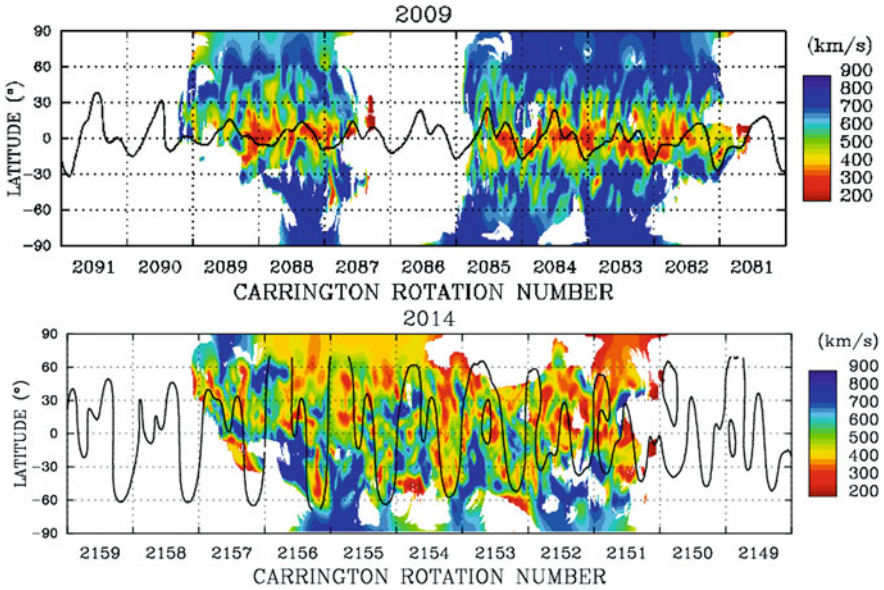

Fig. 8.15 Solar cycle variation in solar wind velocity distribution obtained from IPS observations. Top, solar minima (2009); bottom, solar maxima (2014). Colors from red to blue correspond to slower to higher solar wind speeds

high-latitude region of the heliosphere and the vicinity of the Sun. However, IPS observations can derive the global structure of the heliosphere, including high latitudes, by observing radio sources in various directions from the ground, complementary to in situ observations by a spacecraft. In Japan, Nagoya University has operated radio telescopes dedicated to IPS observations at 327 MHz since the 1980s (Tokumaru et al. 2011, Fig. 8.16). These observations have been used for studies on the solar wind (Tokumaru et al. 2010) and CMEs (Iwai et al. 2019). In the 2020s, many spacecrafts, such as the Parker Solar Probe, which will approach the Sun to approximately 10 solar radii, and the Solar Orbiter, which will explore the high-latitude region, are expected to dramatically increase the observational data of the inner heliosphere. In addition, IMAP, which is planned as a successor of IBEX for the 2020s, is expected to make promising observation of ENAs. Coordinated studies of space- and ground-based observations with solar wind models and simulations will be highly required for the next decades.

Fig. 8.16 Radio telescope for IPS observations operated by ISEE, Nagoya University (Toyokawa, Aichi)

8.6 CMEs Recorded in Recovered Samples from Space

8.6.1 Introduction

The Genesis spacecraft, launched by NASA on August 8, 2001, is the first human solar wind return mission (Burnett 2013). For 2 years and 4 months, from December 3, 2001, to April 1, 2004, the spacecraft deployed four solar wind collection arrays to collect the solar wind at the Lagrangian point (L_1) between the Sun and the Earth (Fig. 8.17). The solar wind is composed of charged particles with an average velocity of approximately 450 km s^{-1}. The solar wind collection method utilized the fact that these high-speed charged particles are implanted into array materials when they collide with their surfaces. During the latter half of the collection period, from the end of October to the beginning of November 2003, the largest solar storm (known as the Halloween solar storm of 2003), ever, observed by the US geostationary meteorological satellite (the GOES), occurred. Although the damage on the ground caused by this solar storm was limited, many satellites and planetary probes in operation at that time were severely affected (NOAA 2004). For example, detectors of the US Advanced Composition Explorer (ACE) spacecraft, which constantly observes the solar wind, were damaged by the solar storm.

The ACE spacecraft was launched on August 25, 1997, and is still observing the Sun at the same Lagrangian point (L_1) as the Genesis solar wind collection array. The accumulated velocity distribution of helium particles observed by the Solar

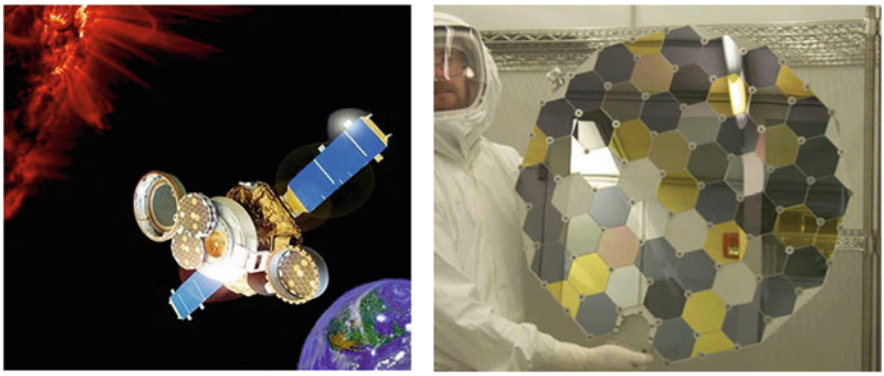

Fig. 8.17 Schematic of Genesis spacecraft of NASA collecting solar wind at Lagrangian point (L_1) (left) and solar wind collection array (right) (credit: Jet Propulsion Laboratory, NASA)

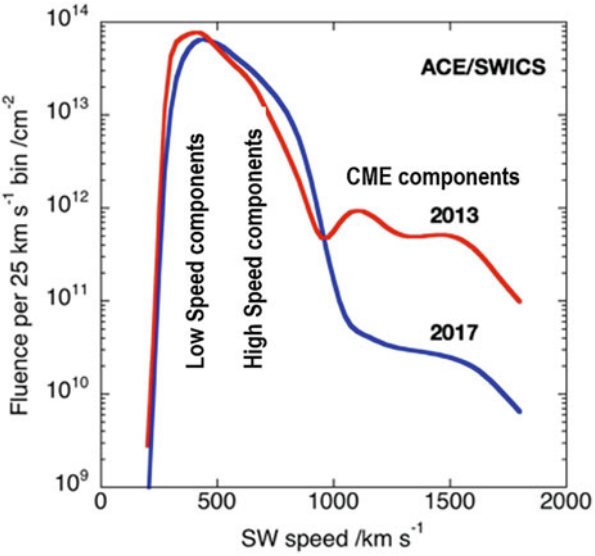

Fig. 8.18 Velocity distribution of solar wind He fluences estimated from ACE/SWICS measurements. Graphs show integrated solar wind He during Genesis operational period (December 2001–March 2004) for 2013 (red line (Reisenfeld et al. 2013) and 2017 (blue line; Reisenfeld, personal communication)

Wind Ion Composition Spectrometer (SWICS) instrument of the ACE spacecraft during the Genesis solar wind collection period is shown in Fig. 8.18. The velocity distributions obtained by the old and new calibrations from the same observation data can be compared. Accordingly, we infer that although the velocity distributions of the low- and high-velocity components of the solar wind are in good agreement, there is a significant discrepancy of more than one order of magnitude for the CME component with a velocity exceeding 1000 km s^{-1}. The time series measurements show that this part of the CME component represents the integrated value of the 2003 Halloween solar wind. This calibration discrepancy may possibly be attributed

to the instrumental failures caused by the high-flux solar wind generated by the Halloween solar storm. Because spacecraft observations are, in principle, one-time measurements, it is difficult to verify the quantitative accuracy when indefiniteness is found in the calibration.

On the other hand, because the solar wind collected by the Genesis collection array is integrated during the collection period, it is impossible to analyze the temporal change over time. However, the solar wind during that period is stored in the material of the solar wind collection array. Therefore, the material of the array can be measured as many times as necessary to investigate reproducibility. However, its measurement methods, including quantitative methods, have been unexplored thus far. In this study, we developed a new secondary neutral mass spectrometry method in addition to a new secondary ion mass spectrometry method and succeeded in measuring the CME solar wind of the Halloween solar storm using the Genesis collection array for the first time. This method can also be applied to determine the history of past solar storms from geological samples.

8.6.2 Analytical Method for Solar Wind Implanted in Solids

The main elements constituting the solar wind are hydrogen (95%) and helium (5%). When charged particles with kinetic energy of several to several hundred keV, such as the solar wind, collide with a solid, they are implanted at a depth of several to several hundred nanometer from the surface of the solid, depending on the energy of each particle. Conventionally, secondary ion mass spectrometry (SIMS) has been used to analyze such particles implanted in solids. Because the detection sensitivity of SIMS for hydrogen is high (Greenwood et al. 2011), it is reasonable to use SIMS for hydrogen analysis. In this study, a micrometer-sized primary ion beam of Cs was irradiated on a solid surface, and the secondary ions generated by sputtering were analyzed by SIMS. The background of hydrogen detection is limited by the amount of residual water vapor adsorbed on the sputtered primary ion-irradiated surface, which was several parts per million (ppm) in this experiment. The spatial resolution in the depth direction was approximately 5 nm.

Concurrently, helium is the element with the highest ionization energy; therefore, its secondary ionization efficiency is very low. Therefore, a new analytical method is needed to replace SIMS. We have developed an isotope nanoscope (laser ionization mass nanoscope (LIMAS), JEOL) to analyze helium implanted in solids (Bajo et al. 2015; Tonotani et al. 2016) (Fig. 8.19). The isotope nanoscope is a new secondary neutral mass spectrometry (SNMS) method in which a nanometer-sized Ga primary ion beam is irradiated on a solid surface. Moreover, the secondary neutral particles generated by sputtering are tunnel-ionized by an intense laser beam and subsequently analyzed by mass spectrometry. Under the conditions of this study, tunnel ionization results in 70% ionization for He and 100% ionization for the other elements. It is important to note that during tunnel ionization, the residual gas in the vacuum in the laser beam is also ionized simultaneously. Because the total

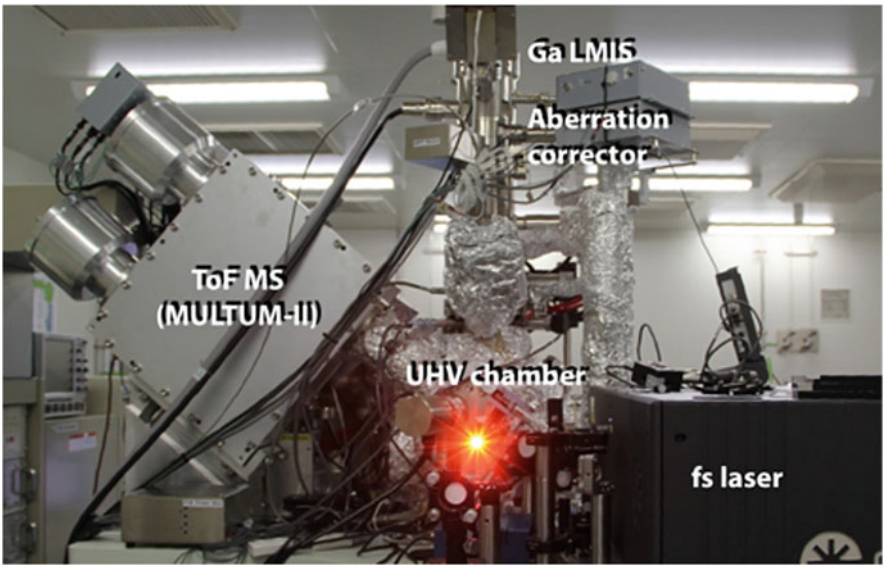

Fig. 8.19 Isotope nanoscope LIMAS. Ga LMIS, Ga liquid metal ion source; UHV chamber, ultrahigh-vacuum sample chamber; fs laser, femtosecond laser; ToF MS, time-of-flight mass spectrometer

pressure in the sample chamber during analysis is 10^{-9} Pa, the partial pressure of He is much smaller; however, He in the ionized residual gas is still detected as a background signal. The background of the He detection corresponded to a few ppm in terms of the concentration of He atoms in the solid. As in SIMS, the spatial resolution in the depth direction is approximately 5 nm.

8.6.3 Measurements of Solar Wind Collected in Space

The solar wind was collected in space and brought to the Earth by the NASA Genesis mission for a period of 2 years and 4 months from December 3, 2001, to April 1, 2004. The solar wind collection array consisted of plates of 15 types of high-purity solid materials (aluminum, sapphire, silicon, germanium, DLC, gold, and others) with mirror surfaces (Fig. 8.20). We analyzed DLC fragments because DLC is implanted in all solar wind hydrogen and solar wind helium irradiated onto the collection array for 2 years and 4 months without escaping the DLC (Heber et al. 2012; Huss et al. 2020). The DLC layer is deposited up to a thickness of 1 μm by plasma CVD on a 500-μm-thick silicon substrate plate (Fig. 8.20).

Depth profiles of the hydrogen and helium concentrations in the DLC are shown in Fig. 8.21. The hydrogen concentration shows a peak at 6.0×10^{21} cm^{-3} (12 atom %) at an implanted depth of 12 nm. Subsequently, the concentration decreases with

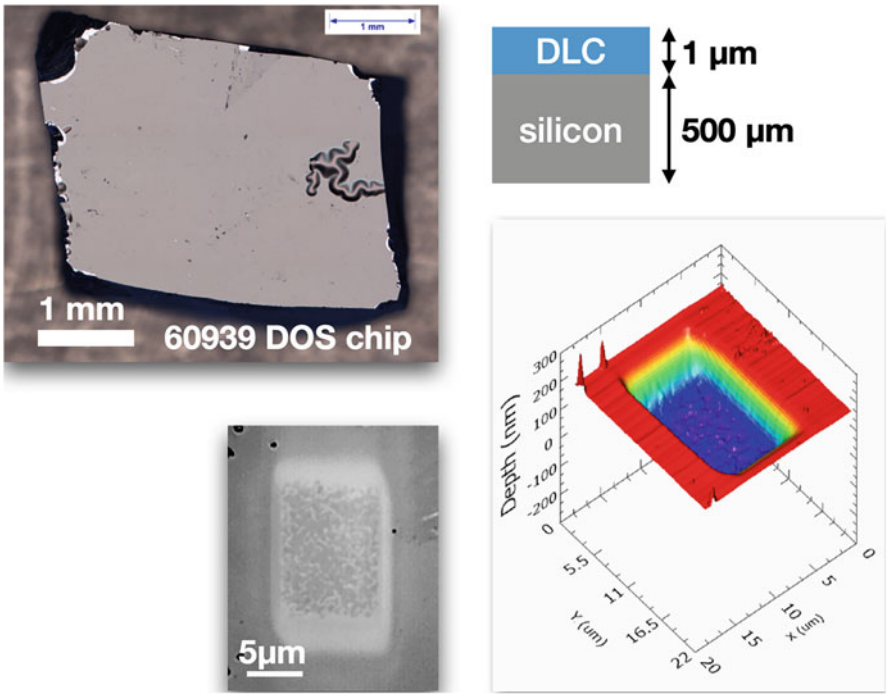

Fig. 8.20 DLC fragment-implanted solar wind (upper left), schematic of cross-sectional structure of DLC (upper right), scanning electron micrograph of sputtered crater after analysis by isotope nanoscope LIMAS (lower left), and geometry of sputter crater (lower right)

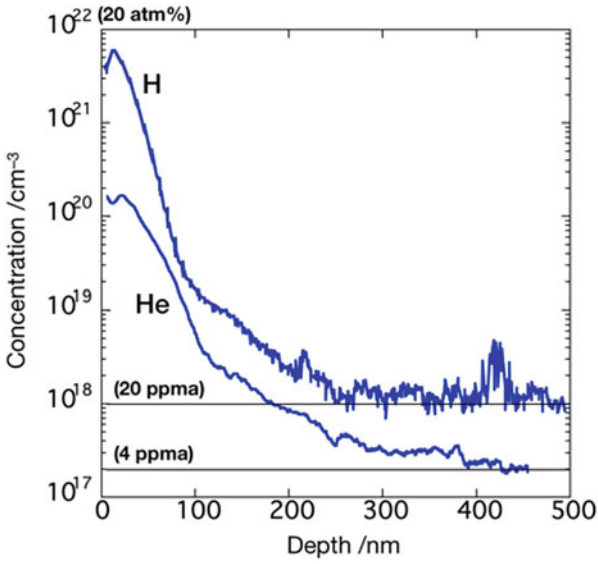

Fig. 8.21 Depth profiles of H and He in DLC. Increments in concentration of hydrogen in spikes at 210 nm and 430 nm are caused by temporal contamination of hydrogen during DLC deposition process

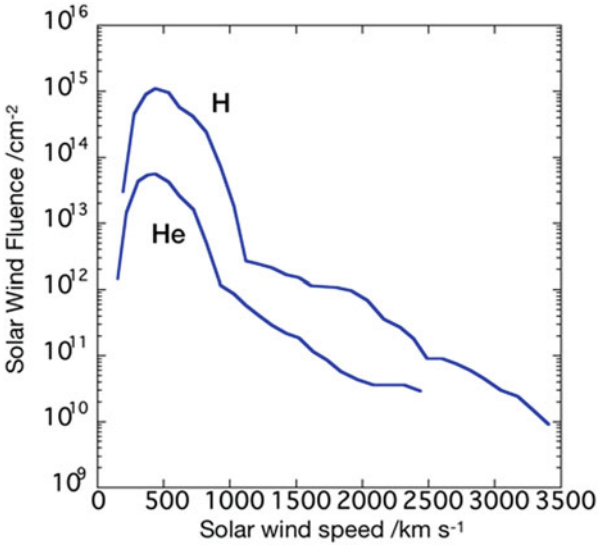

Fig. 8.22 Speed distribution of solar wind reconstructed from depth profile analysis of collected solar wind samples

the depth, the rate of decrease becomes gradual at a depth of 100 nm, and the decrease stops at a depth of 250 nm or more, resulting in a constant concentration of 20 ppm. This constant concentration of hydrogen is the amount originally contained in the DLC film at the time of deposition. In comparison, helium injection shows a peak at 22 nm, slightly deeper than hydrogen injection. The peak concentration is 1.7×10^{20} cm^{-3} (0.34 atom%). Following this, the concentration decreases gradually above a depth of 100 nm and reaches a constant value at a depth of 400 nm. However, note that this constant value of 4 ppm is an apparent value due to the residual He gas in the vacuum of the sample chamber, as described above, and does not indicate the He content of the DLC.

The range of the solar wind in a DLC is determined by the kinetic energy of the colliding charged particles. Therefore, using numerical simulation model of SRIM/TRIM (Ziegler et al. 2012) to calculate the transportation of charged particles in a material from interactions between the colliding charged particles and the material, we can convert the depth profiles shown in Fig. 8.21 to the collision speed distribution of charged particles (Fig. 8.22).

The speed distribution of the solar wind calculated by analyzing the collected solar wind samples has a maximum value of approximately 400 km s^{-1}, which is equal to the mean velocity of the measured solar wind (Fig. 8.22). The speed distribution of He is almost consistent with the velocity distribution observed by the ACE spacecraft (Fig. 8.18). Therefore, the speed distributions of the low-velocity component (interstream component) and the high-velocity component (coronal hole component) of the solar wind during the solar wind collection period of Genesis are reproduced from the results of the depth analysis of the solar wind collection samples.

In contrast, the CME component with velocities >1000 km s^{-1} was indeterminate in the ACE observations (Fig. 8.18). Compared to the results obtained from the solar wind recovery samples, the CME component is close to the ACE 2013 calibration value in terms of the irradiance, whereas the degree of the velocity change is close to the slope of the 2017 calibration value. The maximum velocities are much higher than the 2000 km s^{-1} observed by the ACE spacecraft. Moreover, when the velocity distribution of hydrogen is considered, it is noticeable that the maximum velocity of the solar wind may have exceeded 3000 km s^{-1}. It is now clear for the first time that such an ultrafast CME struck the Earth during the Halloween solar storm of 2003.

8.6.4 Conclusion

We successfully measured the solar wind recovered by the Genesis spacecraft of NASA using the newly developed SIMS and SNMS methods. This enables the analysis of the solar wind in a laboratory on the ground, which is observed only in space. The first merit of analyzing the solar wind in a laboratory is that the past solar wind can be analyzed and verified many times. In satellite observations in the space, the solar wind is observed only once in real time, and it cannot be repeated. The speed distribution of the solar wind for the period of 2 years and 4 months from December 3, 2001, to April 1, 2004, determined in the laboratory agreed with the velocity distribution observed by the ACE spacecraft for interstream and coronal hole components. However, for the CME component, the present measurements showed for the first time that a CME solar wind with a velocity of 3000 km s^{-1} hit the Earth during the Halloween solar storm of 2003. Thus, it is now possible to verify the solar wind events observed by satellites by analyzing the solar wind recovery samples in a laboratory on the ground. Furthermore, by applying this method to lunar samples and asteroidal samples such as Itokawa and Ryugu, it may be possible to reconstruct the history of prehistoric solar activity.

References

Bajo, K.-I., Olinger, C.T., Jurewicz, A.J.G., Burnett, D.S., Sakaguchi, I., Suzuki, T., Itose, S., Ishihara, M., Uchino, K., Wieler, R., Yurimoto, H.: Depth profiling analysis of solar wind helium collected in diamond-like carbon film from genesis. Geochem. J. **49**, 559–566 (2015)

Berger, T.E., Shine, R.A., Slater, G.L., Tarbell, T.D., Title, A.M., Okamoto, T., Ichimoto, K., Katsukawa, Y., Suematsu, Y., Tsuneta, S., Lites, B.W., Shimizu, T.: Hinode SOT observations of solar quiescent prominence dynamics. ApJ. **676L**, 89B (2008)

Burlaga, L.F., Ness, N.F., Stone, E.C.: Magnetic field observations as voyager 1 entered the Heliosheath depletion region. Science. **341**, 147–150 (2013)

Burlaga, L.F., Ness, N.F., Berdichevsky, D.B., Park, J., Jian, L.K., Szabo, A., Stone, E.C., Richardson, J.D.: Magnetic field and particle measurements made by voyager 2 at and near the heliopause. Na. Astron. **3**, 1007–1012 (2019)

Burnett, D.S.: The genesis solar wind sample return mission: past, present, and future. Meteorit. Planet. Sci. **48**, 2351–2370 (2013)

Evershed, J.: Radial movement in sun-spots. Mon. Not. R. Astron. Soc. **69**, 454–457 (1909)

Fujiki, K., Hirano, M., Kojima, M., Tokumaru, M., Baba, D., Yamashita, M., Hakamada, K.: Relation between solar wind velocity and properties of its source region. Adv. Space Res. **35**, 2185–2188 (2005)

Greenwood, J.P., Itoh, S., Sakamoto, N., Warren, P., Taylor, L., Yurimoto, H.: Hydrogen isotope ratios in lunar rocks indicate delivery of nature. Geoscience. **4**, 79–82 (2011)

Heber, V.S., Baur, H., Bochsler, P., McKeegan, K.D., Neugebauer, M., Reisenfeld, D.B., Wieler, R., Wiens, R.C.: Isotopic mass fractionation of solar wind: evidence from fast and slow solar wind collected by the genesis mission. Astrophys. J. **759**, 121 (2012)

Hillier, A., Berger, T., Isobe, H., Shibata, K.: Numerical simulations of the magnetic Rayleigh-Taylor instability in the Kippenhahn-Schluter prominence model. I. Formation of Upflows. ApJ. **746**, 120–132 (2012)

Hewish, A., Scott, P., Wills, D.: Interplanetary scintillation of small diameter radio sources. Nature. **203**, 1214–1217 (1964). https://doi.org/10.1038/2031214a0

Huss, G.R., Koeman-Shields, E., Jurewicz, A.J.G., Burnett, D.S., Nagashima, K., Ogliore, R., Olinger, C.T.: Hydrogen fluence in genesis collectors: implications for acceleration of solar wind and for solar metallicity. Meteorit. Planet. Sci. **55**, 326–351 (2020)

Ichimoto, K.: Spectropolarimetry and magnetic structures. In: Engvold, O., Vial, J.C., Skumanich, A. (eds.) The Sun as a Guide to Stellar Physics, pp. 185–205. Elsevier (2019)

Ichimoto, K., Shine, R.A., Lites, B., et al.: Fine scale structures of the Evershed effect observed by the solar optical telescope aboard Hinode. PASJ. **59**, 593–599 (2007)

Ishikawa, R., Tsuneta, S.: Comparison of transient horizontal magnetic fields in a plage region and in the quiet sun. Astron. Astrophys. **495**, 607–612 (2009)

Iwai, K., Shiota, D., Tokumaru, M., Fujiki, K., Den, M., Kubo, Y.: Development of a coronal mass ejection arrival time forecasting system using interplanetary scintillation observations. Earth Planets Space. **71**, 39 (2019)

Lean, J.L.: Short term, direct indices of solar variability. Space Sci. Rev. **94**, 39–51 (2000)

McComas, D.J., Allegrini, F., Bochsler, P., Bzowski, M., Christian, E.R., Crew, G.B., DeMajistre, R., Fahr, H., Fichtner, H., Frisch, P.C., Funsten, H.O., Fuselier, S.A., Gloeckler, G., Gruntman, M., Heerikhuisen, J., Izmodenov, V., Janzen, P., Knappenberger, P., Krimigis, S., Kucharek, H., Lee, M., Livadiotis, G., Livi, S., MacDowall, R.J., Mitchell, D., Möbius, E., Moore, T., Pogorelov, N.V., Reisenfeld, D., Roelof, E., Saul, L., Schwadron, N.A., Valek, P.W., Vanderspek, R., Wurz, P., Zank, G.P.: Global observations of the interstellar interaction from the interstellar boundary explorer (IBEX). Science. **326**, 959 (2009)

NOAA: Halloween space weather storms of 2003. NOAA Technical Memorandum OAR SEC-88, (2004)

Nolte, J.T., Krieger, A.S., Timothy, A.F., Gold, R.E., Roelof, E.C., Vaiana, G., Lazarus, A.J., Sullivan, J.D., McIntosh, P.S.: Coronal holes as sources of solar wind. Sol. Phys. **46**, 303–322 (1976)

Okamoto, T., de Pontieu, B.: Propagating waves along spicules. ApJL. **736**, L24–L29 (2011)

Okamoto, T., Sakurai, T.: Super-strong magnetic field in sunspots. ApJL. **852**, L16–L21 (2018)

Okamoto, T., Tsuneta, S., Berger, T.E., et al.: Coronal transverse magnetohydrodynamic waves in a solar prominence. Science. **318**, 1577–1580 (2007)

Parker, E.N.: Dynamics of the interplanetary gas and magnetic fields. Astrophys. J. **128**, 664 (1958)

Reisenfeld, D.B., Wiens, R.C., Barraclough, B.L., Steinberg, J.T., Neugebauer, M., Raines, J., Zurbuchen, T.H.: Solar wind conditions and composition during the genesis mission as measured by in situ spacecraft. Space Sci. Rev. **175**, 125–164 (2013)

Shibata, K.: Two types of jets and origin of macrospicules. Sol. Phys. **81**(1), 9–17 (1982). https://doi.org/10.1007/BF00151974

Suematsu, Y., Shibata, K., Nishikawa, T., Kitai, R.: Numerical hydrodynamics of the jet phenomena in the solar atmosphere. Sol. Phys. **75**(1–2), 99–118 (1982). https://doi.org/10.1007/BF00153464

Suzuki, T.K.: Forecasting solar wind speeds. Astrophys. J. **640**, L75–L78 (2006)

Tokumaru, M., Kojima, M., Fujiki, K.: Solar cycle evolution of the solar wind speed distribution from 1985 to 2008. J. Geophys. Res. Space Physics. **115**, A04102 (2010)

Tokumaru, M., Kojima, M., Fujiki, K., Maruyama, K., Maruyama, Y., Ito, H., Iju, T.: A newly developed UHF radiotelescope for interplanetary scintillation observations: solar wind imaging facility. Radio Sci. **46**, RS0F02 (2011)

Tokumaru, M., Fujiki, K., Iju, T.: North-south asymmetry in the global distribution of the solar wind speed during 1985-2013. J. Geophys. Res. **120**, 3283–3296 (2015)

Tonotani, A., Bajo, K.-I., Itose, S., Ishihara, M., Uchino, K., Yurimoto, H.: Evaluation of multi-turn time-of-flight mass spectrum of laser ionization mass nanoscope. Surf. Interface Anal. **48**, 1122–1126 (2016)

Vernazza, J.E., Avrett, E.H., Loeser, R.: Structure of the solar chromosphere. III. Models of the EUV brightness components of the quiet sun. Astrophys. J. Suppl. **45**, 635–725 (1981)

Viall, N.M., Borovsky, J.E.: Nine outstanding questions of solar wind physics. J. Geophys. Res. Space Phys. **125**, e26005 (2020)

Wang, Y.-M., Sheeley, N.R.: Solar wind speed and coronal flux-tube expansion. Astrophys. J. **355**, 726 (1990)

Withbroe, G.L., Noyes, R.W.: Mass and energy flow in the solar chromosphere and corona. Annu. Rev. Astron. Astrophys. **15**, 363 (1977)

Ziegler, J.F., Biersack, J.P., Ziegler, M.D.: SRIM, the stopping and range of ions in matter. Lulu Press Co., Morrisville, NC. http://www.srim.org (2012)

Chapter 9
Origin of Solar Storms

Yoichiro Hanaoka, Kyoko Watanabe, and Seiji Yashiro

9.1 Solar Storm Observation Network

Among the magnetic phenomena on the solar surface, large-scale flares and coronal mass ejections (CMEs), particularly those that have significant impacts on the Earth and the interplanetary space, are called solar storms. Solar storms are accompanied by an order-of-magnitude increase in electromagnetic radiation (particularly X-rays and radio waves), the formation of magnetic plasma clouds that are much denser than the typical solar wind, and the bombardment of large amounts of high-energy particles.

Therefore, monitoring of magnetic phenomena on the solar surface has been continued for a long time. It started with ground-based optical observations and has now evolved to include space-based monitoring of the solar surface at various wavelengths and in situ measurements of the magnetic plasma and energetic particles in space. The originally considered advanced scientific observations have progressed to routine solar storm monitoring, which is currently a contribution of science to social needs.

Y. Hanaoka (✉)
National Astronomical Observatory of Japan, Tokyo, Japan
e-mail: yoichiro.hanaoka@nao.ac.jp

K. Watanabe
National Defense Academy of Japan, Yokosuka, Kanagawa, Japan

S. Yashiro
Catholic University of America, Washington, DC, USA

K. Kusano (ed.), *Solar-Terrestrial Environmental Prediction*,
https://doi.org/10.1007/978-981-19-7765-7_9

9.1.1 Capturing Electromagnetic Radiation from Flare

9.1.1.1 Ground-Based Hα Monitoring Observations

The flare was captured in white light for the first time by Carrington et al. in 1859. However, this was an exceptionally large one, and flares are much more frequently observed as chromospheric phenomena. Subsequently, the accumulation of chromospheric observations, particularly using the Hα line, showed the relationship between solar flares and geomagnetic storms and sudden ionospheric disturbances. Thus, systematic monitoring of flares with the Hα line was conducted early on, and the brightness and area of flare ribbons have been recorded routinely.

Currently, many Hα monitoring observations are being conducted. A network of telescopes worldwide has been organized to perform observations 24 h a day to capture all occurring flares. The United States of America (U.S.) had previously set up its own Hα observation network, and presently, it is operating the Global Oscillation Network Group (GONG). The GONG is a group of instruments primarily for helioseismology, located at six sites globally, and includes Hα instruments for monitoring (Hill 2018). The Global High-Resolution Hα Network (Steinegger et al. 2000), which is a collaborative network of seven observatories in five countries, is also in operation. In Japan, advanced Hα observations are performed by the National Astronomical Observatory of Japan and Kyoto University. They are capturing line-of-sight velocity data based on the Doppler effect in addition to high-resolution Hα images because the Doppler data of the Hα line show the motion of erupting filaments, which are a part of CMEs. Kyoto University is attempting to enhance its observational network and is now deploying two more Hα telescopes outside Japan (called the Continuous H-Alpha Imaging Network, UeNo et al. 2009).

9.1.1.2 Observation of X-Rays from Space

Flares are essentially high-energy phenomena in the corona; brightening events seen in Hα during flares are actually caused by chromospheric heating, which is a consequence of the energy release in the corona. The high-temperature, high-energy component produced by a flare was first directly observed with radio waves emitted by the interaction of high-energy electrons with the magnetic field in the corona. Concurrently, the hot plasma of a flare radiates soft X-rays as thermal emission. Therefore, since X-ray observations from space became possible, soft X-ray measurements have been used as more direct information than radio observations. In radio waves, meter waves are presently used to detect the behavior of high-energy particles in the solar atmosphere far above the solar surface, as described subsequently.

Routine observations of the solar soft X-ray flux have continued since the 1970s using the meteorological satellites of the United States. Initially the Synchronous Meteorological Satellites were used, and subsequently the Space Environment

Monitor (https://www.ngdc.noaa.gov/stp/satellite/goes/) onboard the Geostationary Operational Environmental Satellite (GOES), which is a series of satellites launched consecutively, was employed and is still in use. Since 1975, the times and sizes of the detected flares have been published (https://www.ngdc.noaa.gov/stp/satellite/goes/index.html), which has become the global standard of flare lists, instead of one based on Hα observations. To observe the Sun with a meteorological satellite, which has to face the Earth, solar instruments are mounted on its solar panels, which always face the Sun, and not on the main body of the satellite.

The soft X-ray flux measurements by the GOES detect the integrated soft X-ray flux from the Sun, and it is impossible to determine the locations of the flares. Two-dimensional information has been provided by Hα observations. However, since the 1990s, coronal images in X-rays and extreme ultraviolet (EUV) light, which were formerly captured only occasionally, have been continuously acquired by solar observation satellites launched sequentially for scientific purposes. These satellites were not aimed at space weather monitoring. However, the GOES started monitoring observations with a coronal imager called the Solar X-ray Imager in 2001 and the Solar Ultraviolet Imager in 2016. The acquisition of two-dimensional information about flares in X-rays or EUV light has become a regular part of solar monitoring observations.

9.1.1.3 Observation of the Solar Surface Magnetic Field

Because flares are caused by the evolution of the magnetic field on the solar surface, observation of the magnetic field is important for tracking solar storms from their energy accumulation stage, despite the fact that this is not a direct observation of solar flares. Polarimetric observations for deriving the magnetic field distribution on the solar surface are still under development from both the instrumental aspect and the interpretation of the measurement results. Therefore, in many cases, they have been conducted as ad hoc scientific observations by various observatories and satellites using the latest technologies. Among them, full-disk photospheric magnetic field observations along with the helioseismic observations by the GONG mentioned above, the Michelson Doppler Imager (Scherrer et al. 1995) of the SOlar and Heliospheric Observatory (SOHO) launched by the United States and Europe in 1995 (Domingo et al. 1995, a spacecraft for solar observation located at Lagrangian point L1, 1.5 million km away from the Earth in the direction of the Sun), and the Helioseismic and Magnetic Imager (Scherrer et al. 2012) of the Solar Dynamics Observatory satellite (SDO, Pesnell et al. 2012) launched by the United States in 2010, as well as the high-resolution magnetic field observations by the Solar Optical Telescope (Tsuneta et al. 2008) on the Hinode satellite (Kosugi et al. 2007) have been providing data for years. These observations have also contributed to space weather monitoring.

In addition, it has been recognized that measuring the magnetic field of the chromosphere above the photosphere is very important for estimating the coronal magnetic field that causes flares. Routine observations of the chromospheric

magnetic field will be indispensable in future space weather observations. Such observations have been attempted by the Synoptic Optical Long-term Investigations of the Sun by the United States (Keller et al. 2003) and the Solar Flare Telescope of Japan (Sakurai et al. 2018) using some infrared absorption lines. In particular, polarimetry observations of a helium absorption line with the Solar Flare Telescope have enabled obtaining the magnetic fields of solar filaments regularly. An analysis in the Project for Solar–Terrestrial Environmental Prediction (PSTEP) based on such observations revealed the hemispheric dependence of the field patterns of solar filaments (Hanaoka and Sakurai 2017). The information on the magnetic fields of filaments enables inferring the possibility of a filament causing a magnetic storm when it erupts as a CME (Hanaoka et al. 2020). This is one of the important keys to space weather prediction. Therefore, the next-generation synoptic observation network, ngGONG (Hill et al. 2019), which is an extension of the GONG currently under consideration, also plans to introduce routine measurements of the magnetic field in the chromosphere. Specifically, the helium absorption line will be used to predict the magnetic fields of geomagnetic storms as a new category in space weather forecast.

9.1.2 Capturing Mass Ejections from the Sun

9.1.2.1 Imaging Coronal Mass Ejections

Although it was presumed around 1930 that plasma clouds are ejected from the Sun and could cause geomagnetic storms (see, e.g., Cliver 2006), it was not until the 1960s that the ejection of solar coronal material into the interplanetary space was actually observed by radio wave and ground- and space-based coronagraphic observations. In particular, the coronagraphs on board the Skylab, a space station launched by the United States, revealed that CMEs are frequent and common phenomena (see, e.g., Gopalswamy 2016).

The plasma clouds of CMEs can be tracked with space-based coronagraphs until they separate widely from the solar surface in visible light because the electrons in the plasma clouds scatter solar light via the Thomson scattering. For instance, the Large Angle and Spectrometric COronagraph (LASCO, Brueckner et al. 1995) of the SOHO, which carries three coronagraphs, can capture CMEs up to 30 solar radii. After the LASCO started its observations, most of the CMEs that cause geomagnetic storms have been captured, revolutionizing the studies on CMEs. As of 2021, two of the three coronagraphs of the LASCO are still in operation. In addition, the Solar TErrestrial RElations Observatory (STEREO, spacecrafts of the United States launched in 2006; Kaiser et al. 2008) is equipped with coronagraphs. The two spacecrafts of the STEREO departed from the Earth in opposite directions along the orbit of the Earth. Therefore, they can observe Earth-directing CMEs from the side, whereas the LASCO captures them only from the Earth's direction. Motions along the line-of-sight cannot be determined with the LASCO, whereas the STEREO

enables obtaining a three-dimensional (3D) velocity field, although it is for a limited period.

These satellites were originally designed for research observations of the Sun and not intended to continue uninterrupted observations for monitoring purposes, unlike the GOES solar instruments. Therefore, a plan to place a new spacecraft primarily aiming at space weather monitoring at L5 (the Lagrangian point 60° east of the Sun as seen from the Earth) has been proposed (https://www.esa.int/Safety_Security/Lagrange_mission2).

9.1.2.2 Capturing Motions of Erupted Filaments

Although direct imaging is a powerful method for observing CMEs, it is difficult to obtain information on the line-of-sight velocity regularly, and it is only possible to capture a plasma cloud after it has erupted into the interplanetary space. In contrast, Hα observations of filaments, including the line-of-sight velocity measurements, can provide information on the 3D evolution of a filament plasma from before the start to the early stage of the eruption. As mentioned above, Japan has various observations to capture such 3D velocity fields. In particular, the Solar Dynamics Doppler Imager (SDDI, Ichimoto et al. 2017), which captures the 3D velocity of the filament material in a wide range of ±400 km/s, was newly developed in the PSTEP research. The data from the SDDI has been used to detect precursor phenomena of filament eruptions (Seki et al. 2017) and the features of filament eruptions that develop into CMEs (Seki et al. 2019). The Hα observations, including velocity field measurements, are expected to contribute to the early prediction of space weather phenomena.

9.1.2.3 Type-II Burst in Radio Waves

As mentioned above, CMEs can be detected using meter-wave radio emissions. Of particular use at present is the observation of the dynamic spectrum (time variations in the spectrum) of the radio waves from Type-II bursts, which are emitted from the shock front of the plasma cloud of a CME. The frequency of these radio waves reflects the plasma frequency at the site where they are emitted; moreover, the electron density in the corona, which determines the plasma frequency, depends on the height from the solar surface. Therefore, the frequency and its time variations in a burst indicate the height and motion of the shock front. Although the abovementioned spectrum does not have a two-dimensional spatial resolution, unlike coronagraph observations, it provides height information, regardless of the direction in which the CME is ejected. Such observations have been conducted in various observatories worldwide. In Japan, the Yamagawa Dynamic Spectrometer (70–9000 MHz) of the National Institute of Information and Communications Technology (NICT) is currently in operation (Iwai et al. 2017). This instrument can track Type-II bursts to a height of approximately 2 solar radii. Because ground-based radio observations cannot be conducted at frequencies below approximately

30 MHz because of the interference by the ionosphere, observations at lower frequencies (i.e., higher above the Sun) are made by spacecrafts. The Wind satellite of the United States (launched in 1994 and located at L1) can track Type-II bursts as low as 20 kHz.

9.1.2.4 Interplanetary Scintillation

Plasma clouds moving in the interplanetary space far from the Sun can be captured by interplanetary scintillation (IPS) observations. The interplanetary space has a finite electron density owing to the solar wind, and the electron density fluctuates. Radio waves from stable cosmic sources show variations because of refraction by the fluctuating interplanetary plasma (i.e., IPS). Even larger radio perturbations are observed when plasma clouds with higher electron densities, such as CME plasmas, pass. When scintillation from radio sources is observed at various locations on a celestial sphere, the movement of the plasma cloud can be detected against the background of these radio sources. Presently, several IPS observation instruments are in operation worldwide. In Japan, the Institute for Space–Earth Environmental Research of Nagoya University is operating three 327-MHz antennas (Kojima and Kakinuma 1990).

9.1.2.5 In Situ Observations in Interplanetary Space

In addition to the above observations using electromagnetic waves from a distance, the electron densities and magnetic fields of CME plasma clouds are measured in situ by spacecrafts in the interplanetary space. The above-mentioned GOES and Wind spacecrafts have some instruments for such measurements. Moreover, the Advanced Composition Explorer (ACE, Stone et al. 1998), launched by the United States in 1997 and currently working at the L1 point, has a particularly extensive set of instruments to measure the particles and magnetic fields in the solar wind. In addition, the ACE can detect the arrival of the plasma clouds of CMEs via an abrupt increase in particle density. Because the L1 point is 1.5 million km from the Earth toward the Sun, the detected plasma clouds are expected to reach the Earth in a short time and cause a magnetic storm. The Deep Space Climate Observatory (https://www.nesdis.noaa.gov/content/dscovr-deep-space-climate-observatory) spacecraft, launched in 2015 by the United States, aims to monitor space weather and space climate in addition to conducting Earth observations. It is also located at the L1 point, and one of its objectives is to detect the arrival of CME plasma clouds before they hit the Earth. Thus, the instruments focusing on space weather monitoring are now actively being developed.

9.1.3 Capturing High-Energy Particles from the Sun

Galactic cosmic rays were the first high-energy particles to be discovered to originate from outside the Earth. Subsequently, as a consequence of their ground-based observations, it was shown that high-energy particles from the Sun, i.e., solar cosmic rays, also travel during solar flares. However, the frequency of the detection of solar cosmic rays at the ground (ground level enhancement, GLE) is low, and there have been only 72 cases from 1942 to the present (2021) (https://gle.oulu.fi/#/). Cosmic rays are significantly attenuated by the atmosphere of the Earth. Thus, much more energetic particles originating in solar flares frequently precipitate outside of the atmosphere, where currently many satellites are flying and astronauts are working. The high-energy particles that they are exposed to are harmful. Therefore, observation instruments for high-energy particles are currently installed on some satellites, and constant monitoring is being conducted.

The Space Environment Monitor of the GOES is a representative example of such instruments; it measures high-energy particles such as protons, in addition to soft X-rays. As the GOES X-ray flare detection information is published in a fixed format, the measurement of high-energy particles (proton flux) is also published in three ranks: above 10 MeV, above 50 MeV, and above 100 MeV. Furthermore, an alert is issued when the particle flux exceeds a certain threshold; this operation is quite similar to weather warnings.

9.2 Flare Emission

9.2.1 Effects of Flare Emission on Space Weather

9.2.1.1 Solar Flare Emission

When a solar flare occurs, electromagnetic waves of all wavelengths, from radio to γ-rays, and numerous energetic particles are emitted, and they show rapid time variations. The temporal variations in the flare emissions depend on the type of electromagnetic waves. In the early phase of a flare (impulsive phase), radio waves and hard X-ray emissions show rapid temporal variations, and in the main phase (gradual phase), Hα and soft X-ray emissions generally present gradual temporal variations (Kane 1974).

The scale of solar flares is expressed by the peak flux of the soft X-rays at 1–8 Å observed by the GOES (Bornmann et al. 1996). In this book, the X-ray peak flux at 1–8 Å (flare class) is denoted as $F_{1-8Å}$. The X class is the largest flare, with a flux of $>10^{-4}$ W/m^2, and the peak flux decreases by one order of magnitude with each decreasing scale for the M, C, B, and A classes. Figure 9.1 shows the light curves of the soft X-ray emissions from an X9.3-class flare that occurred on September 6, 2017. It can be seen that for this flare, $F_{1-8Å} = 9.3 \times 10^{-4}$ W/m^2. Thus, the flux

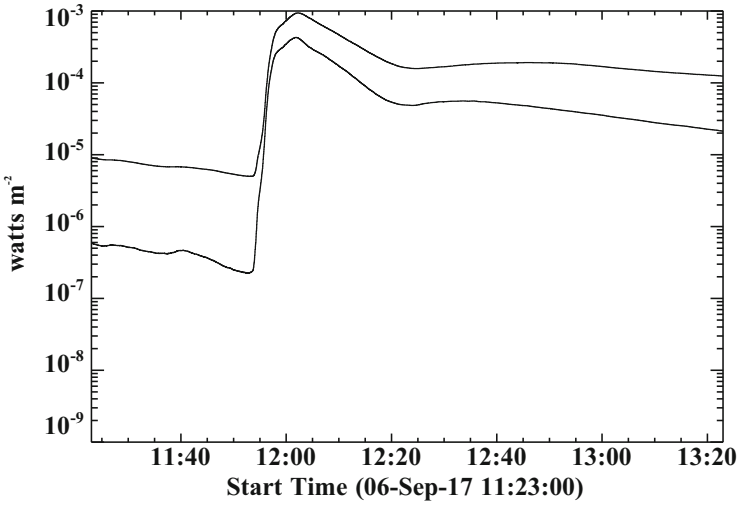

Fig. 9.1 Example of light curves of soft X-ray emissions observed by GOES (X9.3-class flare on September 6, 2017). Upper and lower lines represent time variations in X-rays in 1–8 Å and 0.5–4 Å bands, respectively

of the soft X-ray emissions changes rapidly by several orders of magnitude when a solar flare occurs.

Although not all solar flare emissions have the same time profiles as X-rays, it is known from statistical studies of flare emissions that there is a certain relationship between the GOES flux and the total radiative energy of a flare (Reep and Knizhnik 2019).

9.2.1.2 Flare Emissions Affecting the Ionosphere of Earth

Solar flare emissions such as UV rays and X-rays can cause an anomalous increase in the electron density in the ionosphere of the Earth. Figure 9.2 shows the total ion production rate (= absorption height) of the UV flare emissions in the atmosphere of the Earth. In particular, long-wavelength UV emissions (>100 nm) affect the lower layers of the atmosphere of the Earth, whereas short-wavelength UV emissions and X-rays (<10 nm) are thought to affect the mesosphere (height < 80 km). In fact, X-rays and short-wavelength UV emissions reach the D region of the ionosphere at an altitude of 60–90 km, causing an increase in the electron density in this region. This change in the electron density absorbs high-frequency (HF) radio waves, causing a communication disturbance called the Dellinger phenomenon (Dellinger 1937). Because HF radio waves are routinely used not only for shortwave radio broadcasting but also for offshore aviation and shipboard radio, forecasts of the occurrence of the Dellinger phenomenon and predictions of the duration of the effect are very important for our daily life.

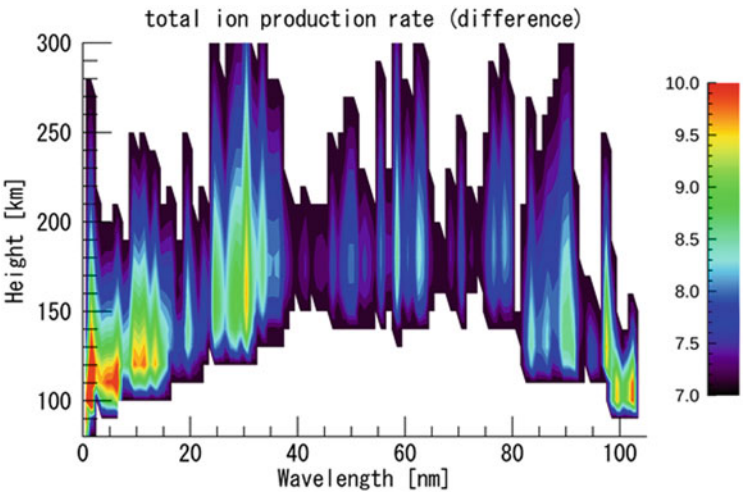

Fig. 9.2 The total ion production rate (= absorption height) of the UV emissions in the atmosphere of the Earth due to the X9.3 flare that occurred on September 6, 2017. (From Watanabe et al. 2021)

The Dellinger phenomenon is generally considered to be caused by the occurrence of solar flares of M class or higher ($F_{1-8Å} > 10^{-5}$ W/m^2), and the scale of a solar flare is currently used for forecasting it. However, the Dellinger phenomenon has been reported to occur even in C-class flares, and to not occur even in X-class flares (Tao et al. 2020). These observations suggest that the flare emissions contributing to the occurrence of the Dellinger phenomenon may be influenced by the emissions that are not proportional to the X-rays.

9.2.1.3 Observation of Flare Emissions

To verify which wavelengths of solar flare emissions contribute to the occurrence of the Dellinger phenomenon, it is necessary to possess observation data of the entire spectrum of the solar flare emissions. However, observation data of UV and X-ray spectra, which are considered to contribute significantly to the Dellinger phenomenon, are limited by the observation wavelengths and the period when the observation satellites were operated.

First, X-ray emissions have been continuously observed by the GOES since 1975. However, only two wavelength bands, 0.5–4 Å and 1–8 Å, have been observed (Bornmann et al. 1996). In contrast, various satellites have been launched to observe EUV emission spectra from the perspective of space weather prediction. The Solar EUV Monitor (SEM) onboard the SOHO satellite (Judge et al. 1998), Solar EUV Experiment (SEE) onboard the Thermosphere Ionosphere Mesosphere Energetics and Dynamics (TIMED) satellite (Woods et al. 2005), and EUV Variability Experiment (EVE) onboard the SDO satellite (Woods et al. 2012) are still observing flare

EUV emission spectra. The SOHO/SEM has been continuously observing EUV emissions with a temporal resolution of 15 s since January 1996. However, it has no wavelength resolution because it observes only 260–340 Å and 1–500 Å. The TIMED/SEE has been observing in the wavelength range of 1–1940 Å with a resolution of 4 Å since January 2002. It has spectral information that the SOHO/SEM does not have; however, its time resolution is approximately few times per day; therefore, it cannot track short time variations such as the time variations in flares. The SDO/EVE has been observing 1–1060 Å emissions with a wavelength resolution of 1 Å and a time resolution of 10 s since May 2010 and has sufficient capability to study the time variability in the solar flare EUV emission spectra. However, the MEGS-A instrument, which observes the 50–370 Å wavelength range in the SDO/EVE, has not been available since May 2014 owing to a failure. As described above, although solar EUV emission data are available, only the SDO/EVE has sufficient wavelength and time resolutions for studying flare emission spectra; however, and its observation time/data are limited.

9.2.1.4 Models of Flare Emission Spectra

Because observations of solar flare emission spectra are limited, models have been constructed to complement and predict them even in the absence of observations. Currently, the most commonly used model is the flare irradiance spectral model (FISM; Chamberlin et al. 2006, 2007, 2008). The FISM can calculate solar flare emission spectra in the wavelength range of 1–1900 Å with a wavelength resolution of 10 Å every minute, and the calculation results are available on the web (https://lasp.colorado.edu/lisird/). Since October 2020, FISM2 data with a wavelength resolution of 1 Å have been made available (Chamberlin et al. 2020). Hereafter, the FISM used before the release of FISM2 will be referred to as FISM1. FISM1 and FISM2 assume that the time evolution of a flare emission is proportional to that of the GOES soft X-rays, and thus, cannot reproduce the differences in the time evolutions of EUV line emissions, which have a different origin from soft X-ray emissions. To understand and reproduce flare emission spectra accurately, it is necessary to consider the physical process of solar flares (Klimchuk et al. 2008; Cargill et al. 2012). However, the FISM is an empirical model and cannot understand this physical process. Therefore, the development of a flare emission model with clear physical processes is required.

9.2.2 Statistical Properties of Flare Emissions

A statistical study on flare emission observations was performed to investigate the spectral and temporal properties of flare emissions and explore the aspects of solar flares that influence these properties (Nishimoto et al. 2020). Events with flare

classes above M3 and observed EUV emissions by the SDO/EVE MEGS-A were used in this analysis. A total of 50 flare events fit this condition.

To investigate how the geometric structures of these flares affect the flare emission spectra, the lengths of the flare emission regions (flare ribbons) and the inter-ribbon distances near the solar surface were determined using SDO/atmospheric imaging assembly (AIA) 1600 Å images, and a statistical study on these data was also performed. A total of 32 events where flare ribbons of sufficient intensity to perform this analysis were observed by the SDO/AIA.

9.2.2.1 Example of Flare Emission Observations

First, the method of deriving the data for statistical analysis is explained using an M6.3-class flare that occurred on November 1, 2013, as an example.

Because the flux of solar soft X-ray emissions is constantly observed by the GOES and the peak flux of a flare (F1–8Å) is an indicator of the flare class, we treated the values obtained from soft X-ray data as fundamental parameters to describe the flare aspect. The fundamental parameters used in this study were the flare class, rise time, and decay time. The rise time is the time from the start of a flare to its peak, and the decay time is the time from its peak to its end. The start and end times of the above flare were derived from the time derivative data of the soft X-ray flux. Figure 9.3 shows the light curve of the soft X-rays (upper panel) and the time variations in the time derivative (lower panel). Because the time derivative has a large time variation in a short period, the flare start and end times were determined using the moving average of the time derivative over 30 s (red line in Fig. 9.3). The flare start time was defined as the time before the flare peak when the 30-s moving average of the time derivative of the soft X-ray emissions (red line) last fell below 5% peak. The flare end time was defined as the time after the flare peak when the red line first exceeded a 5% negative peak.

Subsequently, the EUV line emissions observed by the SDO/EVE were used for comparison with the fundamental flare parameters. As an example, Fig. 9.4 presents the spectrum of the EUV line emissions observed by the SDO/EVE at 19:54 universal time (UT) on November 1, 2013, which shows that the UV emissions consist almost entirely of line emissions. Among them, six EUV lines (Fe VIII-Fe XX 131 Å, Fe XV 284 Å, Fe XVI 335 Å, Fe XVIII 94 Å, Fe XX 133 Å, He II 304 Å), which were considered to have large variations during flares, were selected for the analysis. The EUV lines are emitted from ionized ions and reflect the formation temperature of these ions, e.g., Fe XX represents the emission from the 20th-order ionized iron ions, which are mainly emitted from solar flare plasmas at approximately 9–13 MK. The ionization of iron ions increases with the plasma temperature, with Fe XV originating from the plasma of 2 MK, Fe XVI from 3 MK, and Fe XVIII from 6 MK. The line at 131 Å is affected by both Fe VIII and Fe XX components, with a formation temperature of 0.4–13 MK. The 131 Å line has low-temperature aspects when no flare occurs, whereas it shows high-temperature

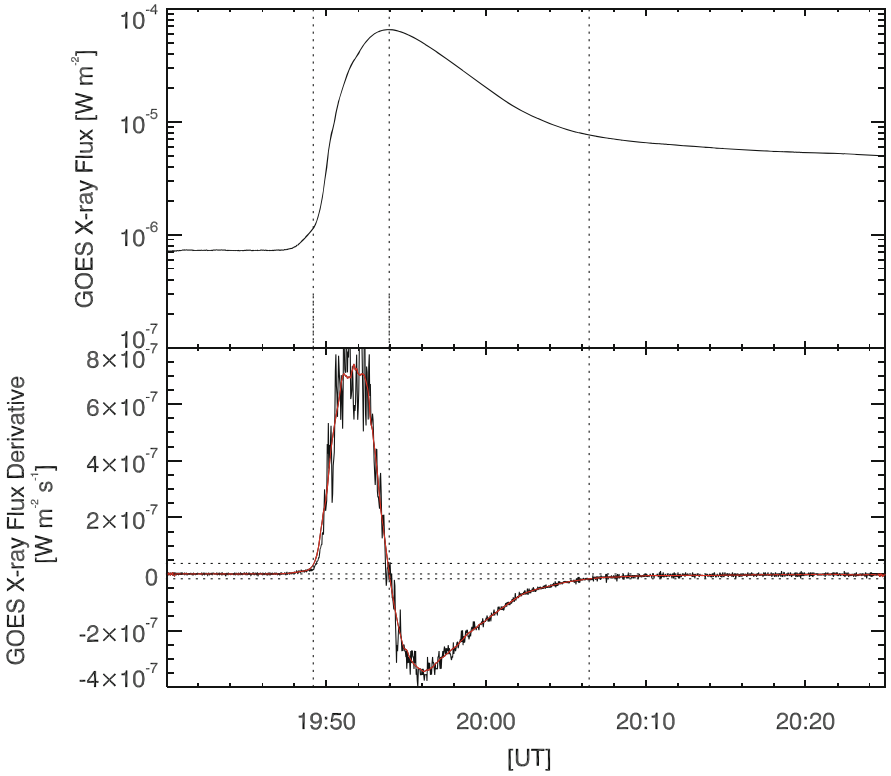

Fig. 9.3 (Upper panel) Light curve of soft X-ray flux from M6.3-class flare observed by GOES on November 1, 2013. (Bottom panel) Time variations in the time derivative (black solid line) and its 30-s moving average (red solid line). Vertical dotted lines represent the start time (19:49 universal time (UT)), peak time (19:54 UT), and end time (20:06 UT) of the flare, respectively. Horizontal dotted lines represent, from the top, 5% peak, 0, and 5% negative peak of the 30-s moving average of the soft X-ray emission time derivative, respectively. (From Nishimoto et al. 2020)

aspects during the flare. He II has a lower formation temperature and is emitted from a 0.05-MK plasma.

From the light curves of the six EUV lines selected above, the peak flux and rise time of each emission were examined. As can be seen in Fig. 9.5, the EUV line light curves (black line) show short time variations; therefore, the 110 s average value (red line) was used for the analysis to obtain an approximate flare variation. The rise time of the EUV line was defined as the time from the start of the soft X-ray emissions until the EUV line flux reached its peak.

The lengths of the flare ribbons and the distances between the ribbons were taken as characteristic values of the geometrical structures of the flare loops. These values were derived from SDO/AIA 1600 Å images. Figure 9.6a shows the lengths of flare ribbons, and Fig. 9.6b shows an example of the derivation of the inter-ribbon distance. The flare ribbon region is defined as the area where the intensity from the

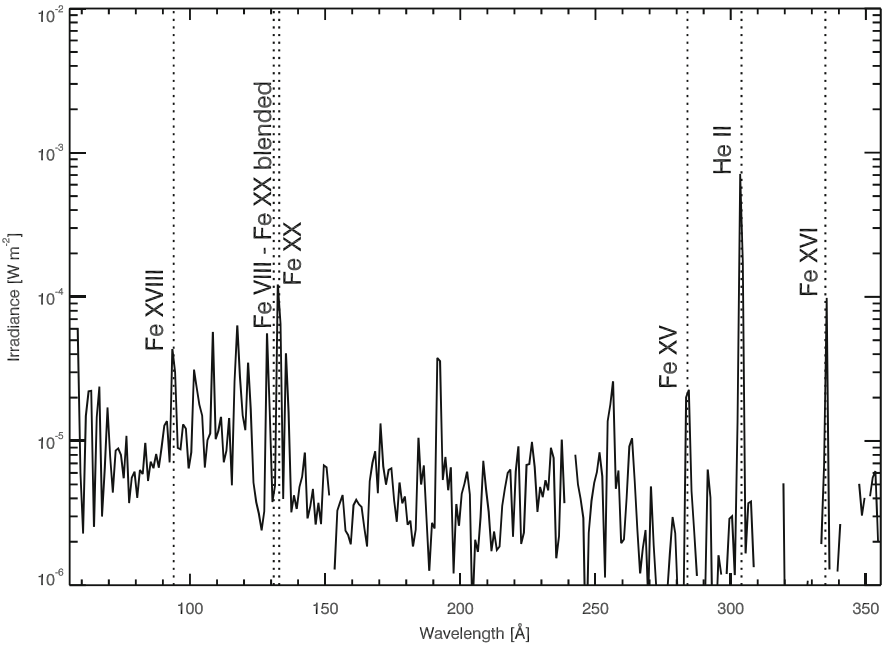

Fig. 9.4 Spectrum of EUV emission observed by SDO/EVE during the peak time (19:54 UT) of the M6.3-class flare on November 1, 2013. Six EUV lines used in this analysis are shown as vertical dotted lines. (From Nishimoto et al. 2020)

background is more than 40σ near the flare peak time, and the length of a flare ribbon is the sum of the longest lengths (two red arrows) of two consecutive regions that are considered to be flare ribbons. The distance between ribbons is defined as the shortest distance between the two points with the highest intensity (red arrows) in the region where the intensity is more than 40σ from the background before and after the flare start time.

9.2.2.2 Statistical Properties of Flare Emissions (Particularly EUV Emissions)

To obtain the statistical properties of the considered flare emission, the EUV data presented in the previous section along with the fundamental parameters of the flare were compared. The relationship between the peak flux of each EUV line and the soft X-ray peak flux is shown in Fig. 9.7. From these results, it is found that the Fe XX, Fe VIII-Fe XX, and Fe XVIII lines have good correlations with soft X-rays, whereas the correlations of Fe XV and Fe XVI are worse than those of other iron lines, and there is no correlation for He II. These EUV lines show a good correlation with the soft X-ray intensity at short wavelengths ($\lambda < 140$ Å) and a poor correlation at long wavelengths ($\lambda > 280$ Å).

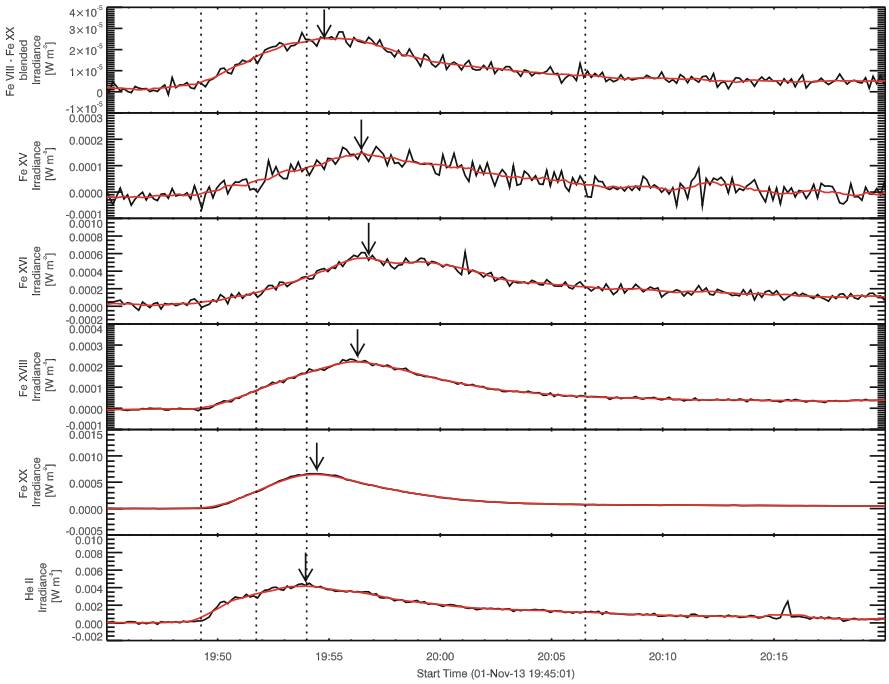

Fig. 9.5 Light curve of each EUV line emission observed by SDO/EVE during the M6.3-class flare on November 1, 2013. Red line is the 110-s moving average. (From Nishimoto et al. 2020)

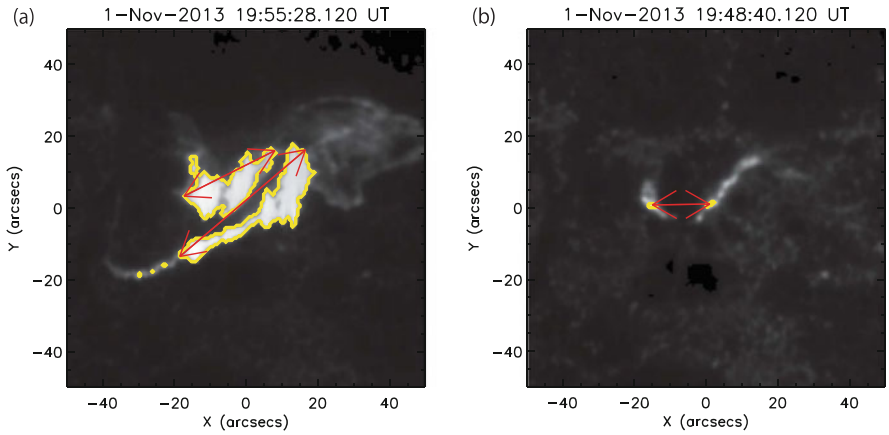

Fig. 9.6 (**a**) Flare ribbon (contour region) and flare ribbon length (red arrow) observed at SDO/AIA 1600 Å. (**b**) Flare ribbon appearance (contour region) and distance between ribbons (red arrows) at the beginning of the flare. (From Nishimoto et al. 2020)

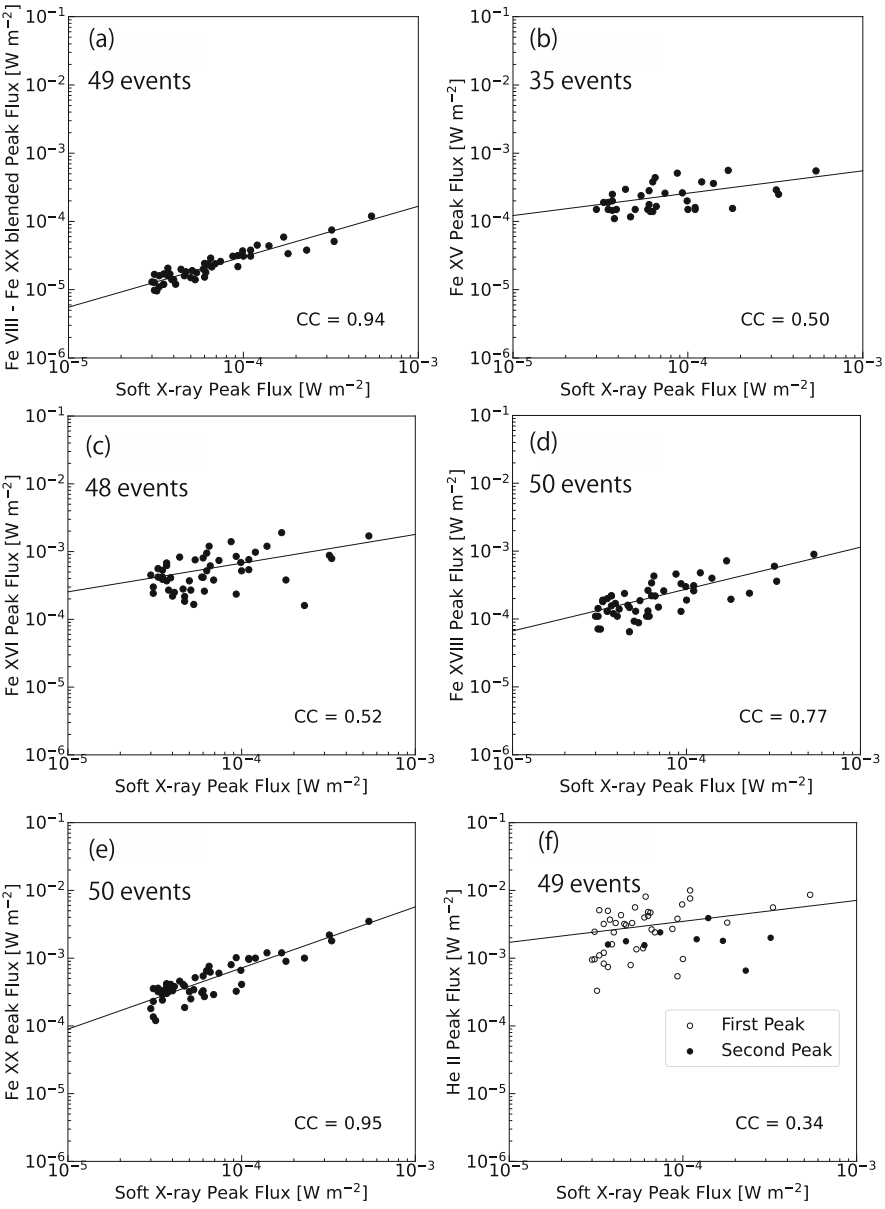

Fig. 9.7 Relationship between soft X-ray peak flux and EUV line peak flux of Fe VIII–Fe XX blended (a), Fe XV (b), Fe XVI (c), Fe XVIII (d), Fe XX (e), and He II (f). Straight lines are approximations. (f) Because He II light curve may have two peaks, the white circle represents the peak before the soft X-ray peak, and the black circle denotes the peak after the soft X-ray peak

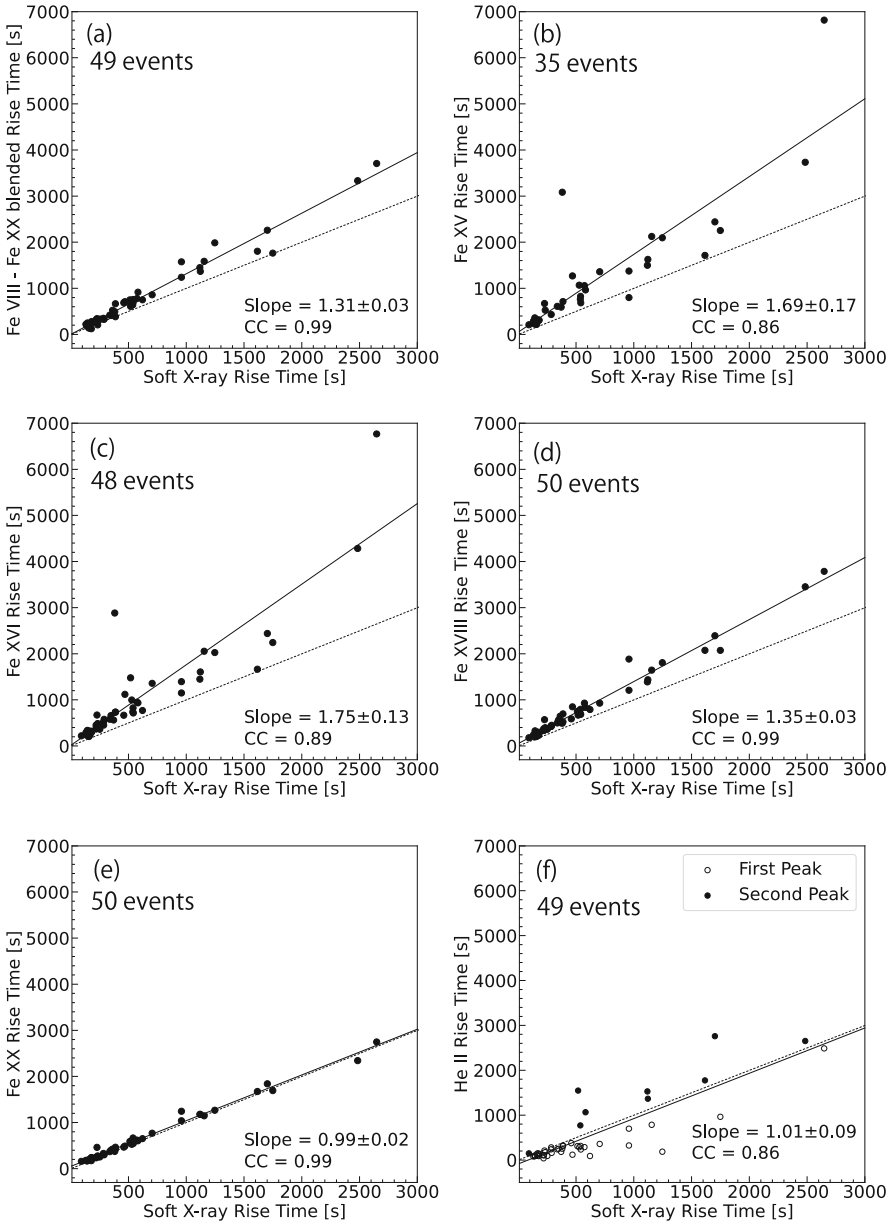

Fig. 9.8 Relationship between the soft X-ray rise time and EUV line rise time of Fe VIII–Fe XX blended (a), Fe XV (b), Fe XVI (c), Fe XVIII (d), Fe XX (e), and He II (f). Solid line is an approximate line, and dotted line is a line with slope 1. (**f**) White and black circles are the same as in Fig. 9.7

Subsequently, we compared the rise times of the EUV lines and the soft X-ray emissions. The results are shown in Fig. 9.8. The rise times of all EUV lines show a good correlation with those of the soft X-ray emissions. If the slope is greater than 1 (dotted line), the rise time of that EUV line is longer than that of the soft X-rays, and if the slope is smaller than 1, the rise time of that EUV line is shorter than that of the soft X-rays. As can be seen from Fig. 9.5, the time in which each EUV line reaches its peak is shorter for higher iron ionization (higher line formation temperature), and in fact, for Fe XVI–Fe XX, the rise times tend to be shorter for a higher line formation temperature. However, the rise time of Fe XVI is longer than that of Fe XV, and such a trend is not generally observed. Although the formation temperature of He II is low, its rise time is similar to that of soft X-rays. A possible reason for this is the blending of EUV lines at other temperatures. According to the CHIANTI database (Dere et al. 1997, 2019), Fe XV is blended with the Ni XIX (~10 MK) line, Fe XVI is blended with Fe XXI (~13 MK), and Fe XVIII could be blended with Fe XX (~13 MK). The He II line is known to be of transition layer origin and presents a nonthermal aspect (Woods et al. 2011); therefore, its rise time is expected to be short even at low formation temperatures. Because of the influence of the various blended lines on the EUV lines, the tendency for the rise time to be shorter at a higher line formation temperature is not statistically clear.

9.2.2.3 Relationship between Geometry of Flare Loop and Flare Emission

The relationships between the geometrical features of the flare loops (inter-ribbon distance and ribbon length) and the fundamental parameters of the flares (soft X-ray rise time and decay time) were statistically investigated, and the results are shown in Fig. 9.9. The distance between the ribbons correlated well with both the soft X-ray rise time (a) and decay time (b). In contrast, the ribbon length did not correlate well with either the soft X-ray rise time (c) or decay time (d). It is considered that the distance between ribbons is proportional to the loop length, whereas the ribbon length is proportional to the depth of the flare loop. These results suggest that the rise time of a flare can be predicted from the distance between flare ribbons (loop length).

9.2.3 Flare Emission Model

9.2.3.1 Construction of Flare Emission Model

To reproduce observed flare emissions, a flare emission model with clear physical processes was developed. In our model, we used the coordinated astronomical numerical software (CANS) one-dimensional (1D) package (http://www-space.eps.s.u-tokyo.ac.jp/~yokoyama/etc/cans/index-e.html) combined with 1D hydrodynamic calculations using the CHIANTI atomic database (Dere et al. 2019). The objective was to reproduce the physical processes of the plasma in a flare loop to

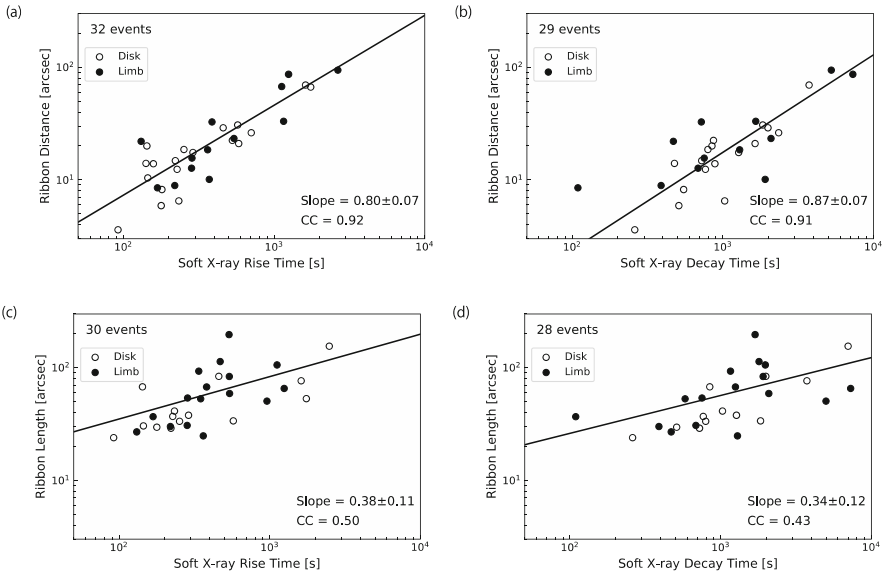

Fig. 9.9 Relationships between distance between flare ribbons (**a**, **b**), ribbon length (**c**, **d**), rise time (**a**), (**c**), and decay time (**b**), (**d**) of soft X-ray emissions observed at 1600 Å by SDO/AIA. (From Nishimoto et al. 2020)

obtain solar flare EUV and X-ray emission spectra (Imada et al. 2015; Kawai et al. 2020). Hereafter, we refer to this model as the "solar flare emission model."

CANS 1D, which is used in the solar flare emission model, has a 1D hydrodynamic model that considers the physical processes of the plasma in a flare loop. This package can calculate the behavior of the plasma in a flare loop after energy injection into the loop and simulates the 1D fluid motion and energy transfer along a magnetic loop. The cross-section of the magnetic loop is assumed to remain unchanged with time, and the fluid is considered to be nonviscous and compressible. Heat transfer and radiative cooling are included, and gravity is also considered. The input values to the model's calculations are the length of the flare loop and the amount of energy injected into the loop.

The CHIANTI database contains a large amount of atomic data for analyzing astrophysical spectra, including atomic energy levels, wavelengths, radiative transition probabilities, ionization rate coefficients, and data for calculating various types of emissions (Dere et al. 1997). The solar flare emission model deals only with the plasma in the solar corona; therefore, the CHIANTI solar coronal composition model is used. By adapting the CHIANTI database to plasma distribution in flare loops, we can obtain flare emission data that can be compared with observations.

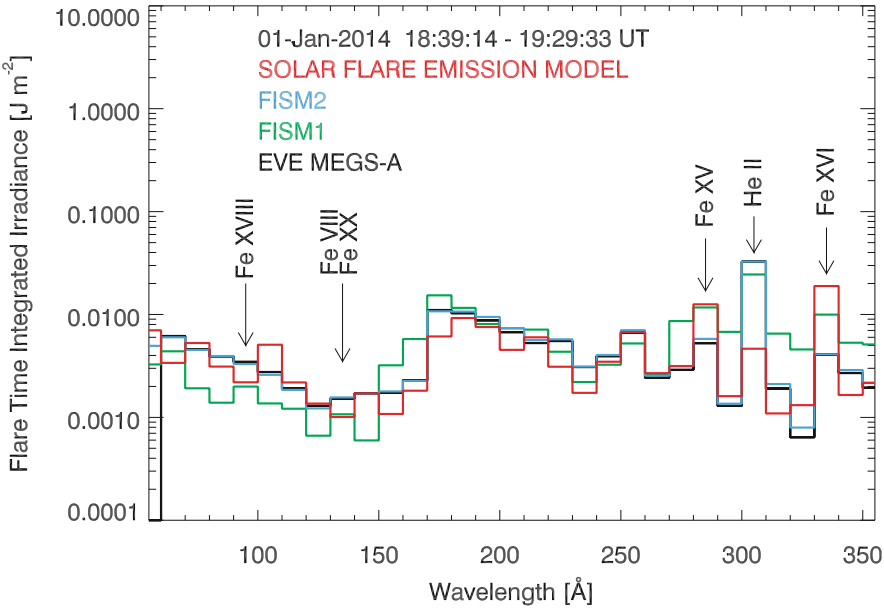

Fig. 9.10 Integrated spectrum of M9.9-class flare that occurred on January 1, 2014. Integration time is from the start to the end of the flare. Solid red, blue, green, and black lines represent calculated spectra by the solar flare emission model, FISM2, FISM1, and observed spectra by SDO/EVE MEGS-A, respectively. Wavelength resolution is set at 10 Å. Arrows represent EUV lines focused on in this study

9.2.3.2 Comparison of Calculation Results of Solar Flare Emission Model with Observations

Nishimoto et al. (2021) performed a comparison of the calculation results of the solar flare emission model with observations. First, the solar flare emission spectra of 21 solar flare events of M3 class or higher observed by the SDO/EVE MEGS-A were obtained using this solar flare emission model. For the loop length, we derived the inter-ribbon distance from the flare ribbon aspect per flare onset time (SDO/AIA 1600 Å) and used a semicircle with the inter-ribbon distance as its diameter. For the heating rate of a flare, we used the heating rate that could reproduce the flow rates of 1–8 Å and 0.5–4 Å observed by the GOES. To treat the flare emissions in the corona, the solar flare emissions in the flare loops calculated accordingly below the altitude where the temperature becomes less than 100,000 K are considered to be emissions from the chromosphere, and not used in this study. This is the same treatment as adopted by Kawai et al. (2020).

Subsequently, we compared the model's calculation results with the observed data. First, we compared the spectra and found that the calculated results of the solar flare emission model reproduced well the time-integrated spectrum of the EUV emissions observed by EVE MEGS-A at the time of the solar flare. An example of

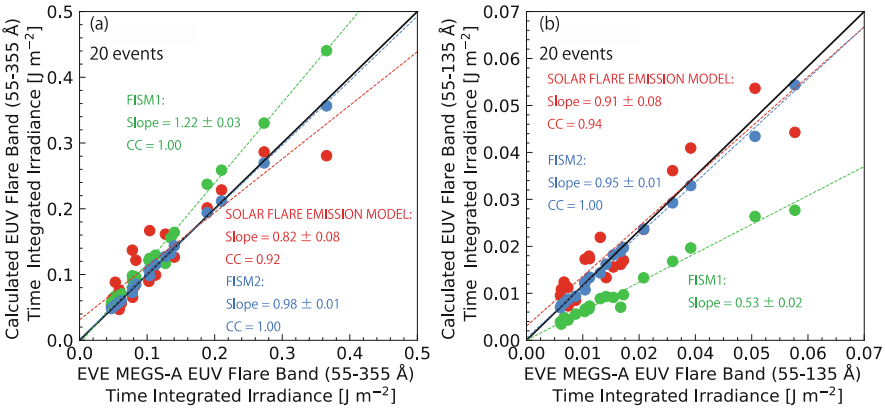

Fig. 9.11 Comparison of model-calculated (solar flare emission model: red, FISM1: green, FISM2: blue) and observed (SDO/EVE) EUV emission energies. (**a**) EUV emission energy at the observed wavelength of EVE MEGS-A and (**b**) EUV emission energy below 140 Å wavelength

these comparisons is shown in Fig. 9.10. In comparison, FISM1 and FISM2 roughly reproduce the observed spectrum.

Subsequently, the EUV emission energies calculated by the solar flare emission model, FISM1, and FISM2 were compared with the observed values of the SDO/EVE (Fig. 9.11a), and both models obtain values close to the observed values. However, when the EUV energy at wavelengths below 140 Å, which is known to be poorly predicted by FISM1, is compared with the observed value (Fig. 9.11b), the solar flare emission model reproduces the observed value well. However, the energy measured by FISM1 is approximately half of the observed value. This underestimation on the short-wavelength side is improved by FISM2, and it yields energy that is almost consistent with the observed value.

The rise times of the six EUV lines obtained by the solar flare emission model are compared with the observed values in Fig. 9.12. Many data points for He II are far from the dotted line with slope 1, whereas the other data points for all the Fe lines present a good correlation.

By comparing the observed and calculated values of the solar flare emission model, we confirmed the agreement of the integrated spectra during the flare period, the agreement of the total EUV emission energy, and the proportional relationship of the rise times of the EUV lines. It is considered that our solar flare emission model can reproduce the time variations in each EUV line; however, it cannot achieve this completely when compared with individual observed values. As an example of the comparison of the calculated and observed values, Fig. 9.13 shows the light curves of each EUV line for an M9.9-class flare that occurred on January 1, 2014. The solar flare emission model reproduces the observed differences in the peak times of the EUV lines. These results show that our solar flare emission model reproduces the observed EUV emission spectrum and its time variation relatively well.

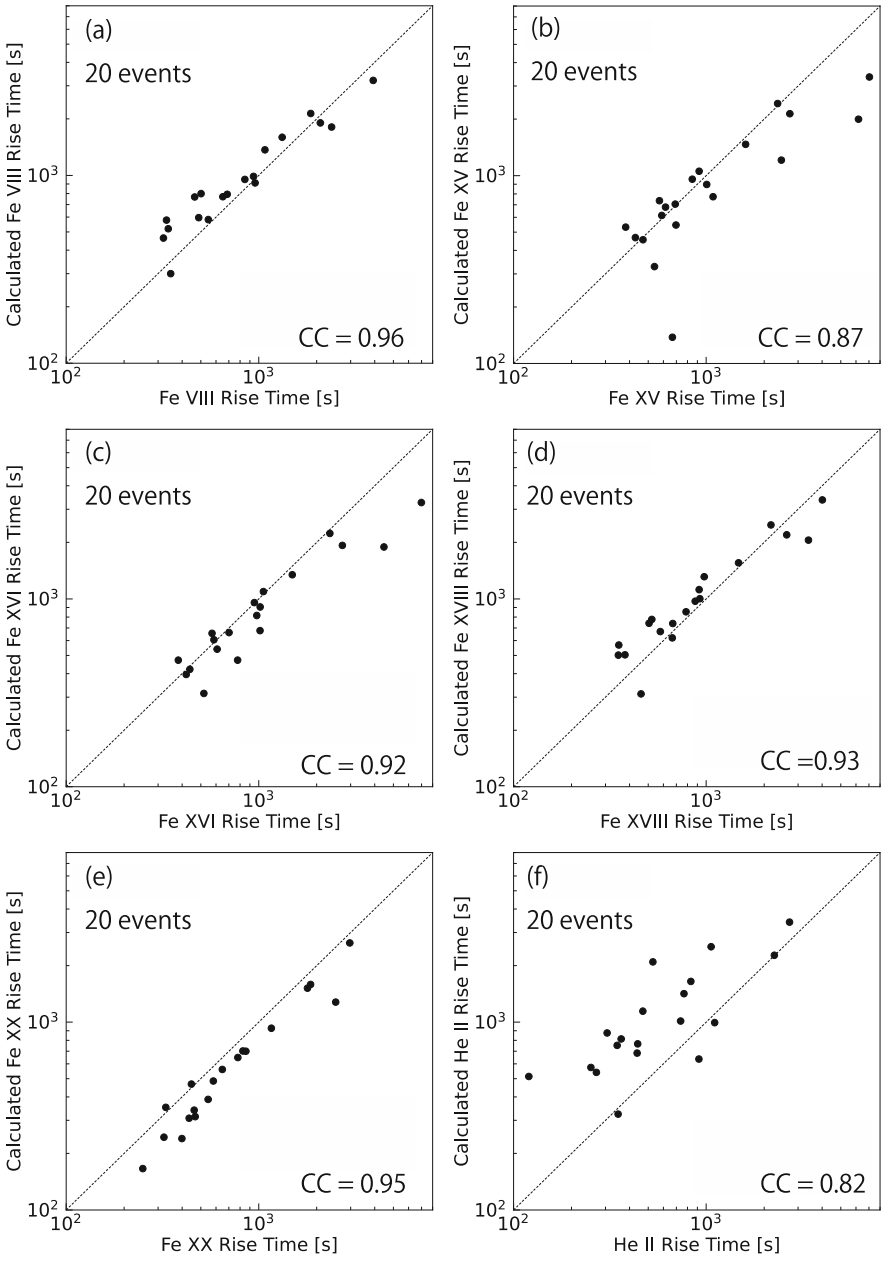

Fig. 9.12 Comparison of model-calculated (solar flare emission model) and observed (SDO/EVE) rise times of EUV line for (a) Fe VIII, (b) Fe XV, (c) Fe XVI, (d) Fe XVIII, (e) Fe XX, and (f) He II. Dotted line represents the line with slope 1. Points above the dotted line indicate that the model-calculated rise time is longer than the observed one, and points below the dotted line indicate that the model-calculated rise time is shorter than the observed one

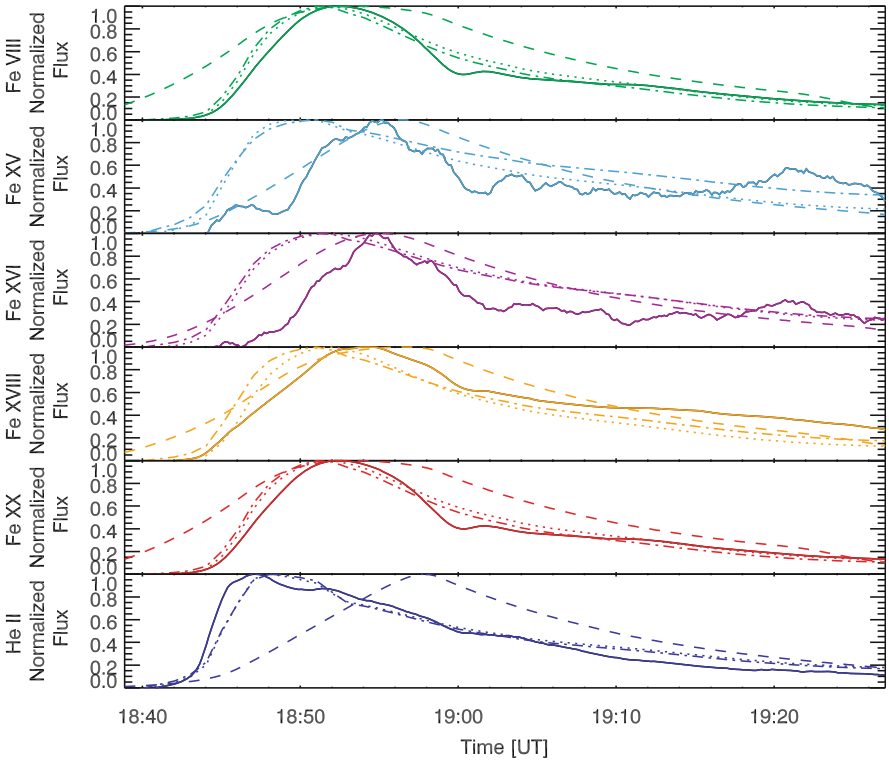

Fig. 9.13 Light curves for each EUV line in M9.9-class flare that occurred on January 1, 2014. Solid lines are SDO/EVE MEGS-A observations, dashed lines are values obtained with the solar flare emission model, and dotted and point-dashed lines are FISM1 and FISM2 values, respectively

9.2.4 Summary and Future Studies

To derive the solar flare spectra that affect the Dellinger phenomenon, we conducted statistical studies on observed flare emission data and constructed a flare emission model. It was found that the main EUV line emissions from a flare correlate well with the GOES X-ray intensity variations; however, there are some discrepancies in the time evolution, such as the peak time, which is related to the temperature profile of the plasma in a flare loop. Subsequently, a solar flare emission model was constructed, and a flare emission spectrum was reproduced. The calculated results were compared with observed soft X-ray and EUV spectra, and it was found that the peak flux and time variations in some EUV line emissions were successfully reproduced. However, the flux of some EUV line emissions, particularly the He II line, cannot be reproduced. This suggests that the actual plasma in the loop is not accurately reproduced, whereas the total energy is almost reproduced, and the spectral distributions with FISM2 are good.

It is possible to improve the accuracy of reproducing solar flare emission spectra using the solar flare emission model. The solar flare spectra obtained from the solar flare emission model can be input to GAIA (Jin et al. 2011), developed by the NICT to simulate the ionospheric effects of solar flare emissions. The responses of the ionosphere of the Earth to solar UV emissions have not yet been established, not only during flares. Therefore, the solar flare emission spectra obtained by the solar flare emission model can be introduced and compared with the observed total electron number distribution and the Dellinger phenomenon on the ground. This will result in improved prediction accuracy of the ionospheric density structure and enable evaluation of the responses to solar flares.

In the future, we will construct a model that can estimate the spectrum of solar flare emissions from observations of X-rays and flare ribbons and will predict the impact on the ionosphere of the Earth by connecting the above two numerical models. Finally, it will be possible to provide space weather forecasts about communication failures caused by solar flare emissions.

9.3 Coronal Mass Ejections

Sudden plasma eruptions from the Sun into the interplanetary space are called CMEs. Their average eruption speed is approximately 450 km/s. If a CME is sufficiently faster than the solar wind, a shock wave is formed in front of the CME, and particles are accelerated by the shock. If this CME originates from the western hemisphere, the accelerated particles reach the Earth along the spiral of the interplanetary magnetic field and are observed as solar energetic particles (SEPs) (Reames 1999). If a huge CME is ejected toward the Earth, it will reach the Earth 18 h to 5 days after the eruption. When the CME has a southward magnetic field, it will cause a magnetic storm (Gosling et al. 1991; Zhang et al. 2007). Thus, CMEs are the main cause of space weather phenomena that affect human life.

Figure 9.14 shows the time evolution of a CME observed by the LASCO coronagraph onboard the SOHO mission. The bright structure in the center is called the core, which corresponds to a prominence (filament) before the eruption (Webb and Hundhausen 1987; Gopalswamy et al. 2003). The outermost semicircular loop is called the leading edge (LE), and the dark region between the core and the LE is called a cavity or void. These LE, cavity, and core are called the three-part structure and are the basic structure of a CME (Illing and Hundhausen 1985; Webb 1988).

9.3.1 History of CME Observations

CMEs are now known to be fundamental in space weather research; however, their discovery was much later than other relevant phenomena (Table 9.1). The Seventh Orbiting Solar Observatory (OSO-7; 1971–1973) observed the solar corona using

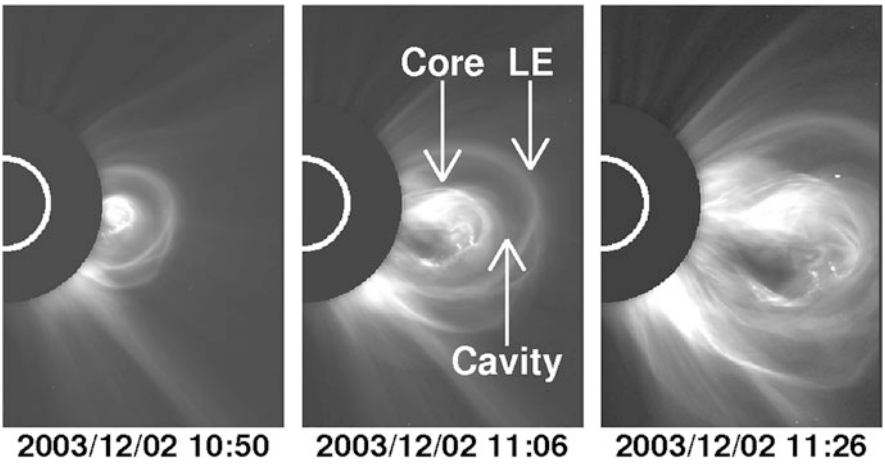

Fig. 9.14 CME observed by LASCO coronagraph onboard the SOHO spacecraft; white semicircle represents the location and size of Sun

Table 9.1 History of the discovery of major space weather phenomena

Discovery of space weather phenomena		
Obs. Year	Phenomena	Discoverer (publication year)
1859	Solar flare	Carrington (1859)
1891	Prominence eruption	Fenyi (1892)
1908	Sunspot magnetic field	Hale (1908)
1942	Solar energetic particle	Forbush (1946)
1947	Type II radio burst	Payne-Scott et al. (1947)
1962	Interplanetary shock	Sonett et al. (1964)
1971	Coronal mass ejection	Tousey (1973)

coronagraphs from space to avoid the atmosphere of the Earth. Thanks to the coronagraphic observation without sky, OSO-7 observed 20 CMEs in 2.5 years. The discovery of CMEs recognized the importance of coronagraphic observations from the space.

When a CME was first discovered, it was called a "coronal transient." At the time of writing, the Japanese version of Wikipedia states that the reason for the name was that "at the time it was thought to be a blast associated with a solar flare." However, this explanation seems to be incorrect. Earlier press releases (NRL SSD Research Achievements: 1960–1970) stated that the new phenomenon was a plasma cloud eruption associated with a prominence eruption. Gosling et al. (1974) used the term "mass ejection coronal transient" as a category of phenomena that show sudden changes in the corona as observed by coronagraphs. The term "mass ejection coronal transient" was subsequently renamed "coronal mass ejection." It was discovered that fast CMEs originate from solar active regions and are well associated with strong flares (MacQueen et al. 1974; Hildner et al. 1976; Gosling et al. 1976). The P78–1

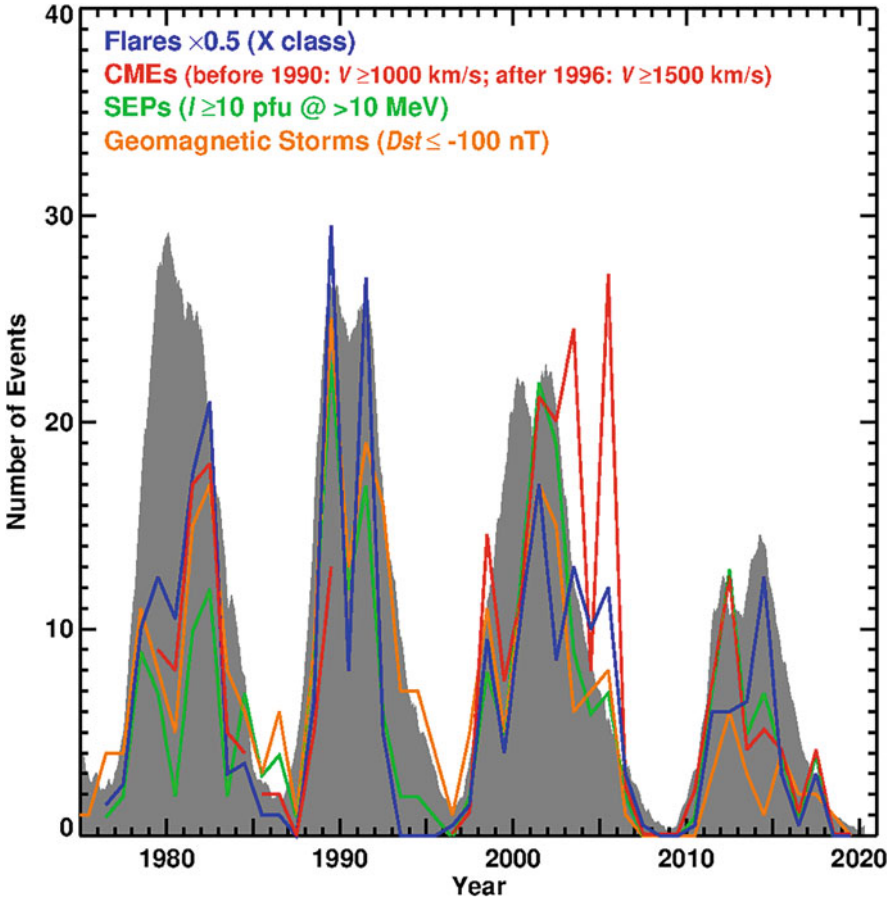

Fig. 9.15 Annual numbers of space weather events (X-class solar flares (blue), fast CMEs (red), SEPs (green), and geomagnetic storms (orange)) superposed on sunspot numbers (gray)

satellite (or Solwind; 1979–1985) observed approximately 1000 CMEs and showed the solar cycle variations of CMEs (Howard et al. 1985). The Solar Maximum Mission (SMM; 1980, 1984–1989) also observed approximately 1000 CMEs and clarified the basic structure of CMEs (Illing and Hundhausen 1985), their characteristics, and solar cycle variations (Hundhausen 1993; Hundhausen et al. 1994).

The LASCO onboard the SOHO (1995–present) began observations in 1996 and has observed more than 30,000 CMEs (http://cdaw.gsfc.nasa.gov/; Yashiro et al. 2004). Long-term observations spanning more than 26 years enabled the study of inter-solar cycle variations. CME interactions between two CMEs and associated particle accelerations were also discovered (Gopalswamy et al. 2001). The launch of the STEREO (2006–present) in 2006 enabled detailed space weather research through multipoint CME observations.

Figure 9.15 shows the solar cycle variation in the number of major space weather events. CMEs (red) are the number of fast CMEs observed by the Solwind in

1979–1985, SMM in 1984–1989, and SOHO since 1996. Because the performance of the coronagraphs differs, the Solwind and the SMM show CMEs with velocities of 1000 km/s or higher, whereas the SOHO shows the number of CMEs observed with velocities of 1500 km/s or higher. Solar flares (blue) are the number of X-class flares classified by the GOES. Half of them are plotted to adjust the scale. SEPs (green) are the number of events with >10 pfu (particle flux unit; proton/(cm^2/s sr)) at >10 MeV flux. Geomagnetic storms (orange) are the number of events with Dst index less than −100 nT (World Data Center for Geomagnetism, Kyoto University). The gray background indicates changes in the sunspot number (obtained from the Royal Observatory of Belgium).

The variations in all space weather phenomena approximately correspond to the variations in the sunspot number. The number of events increases during the solar maximum and decreases during the minimum period. Although Solar Cycle 21 (SC21) does not apply, we can see that the number of flares, CMEs, SEPs, and geomagnetic storms decrease in response to the decrease in the peak sunspot number in SC22, SC23, and SC24. It is known that giant sunspots appear in the decay phase of the solar activity, and a few giant sunspots cause numerous space weather events. Therefore, the number of space weather events in 2005 was large even though the sunspot number was small.

During SC24, the number of the fast CMEs peaked in 2012, and the number of large flares peaked in 2014. The peaks of the SEP and magnetic storm numbers are in 2012, which coincide with the peaks of the fast CMEs. Although flares and CMEs are closely related, approximately 10% of X-class flares are not accompanied by CMEs (Yashiro et al. 2005). Such CME-less X-class flares do not cause SEPs and geomagnetic storms.

9.3.2 CME Rate

Figure 9.16 shows the average occurrence frequencies of CMEs (red; width > 30°) and solar flares (blue; >C class) observed by the LASCO coronagraph for each Carrington rotation (27.3 days). Similar to other active phenomena such as flares, CMEs occur less frequently during the minimum period and more frequently during the maximum period. There are many days without CMEs in the minimum period, and the average number of CMEs is less than 0.5 per day. During the solar maximum, the number of CMEs can exceed 3 per day. Comparing SC23 and SC24, the average number of sunspots decreased by 40%, whereas the number of CMEs has not changed much (Gopalswamy et al. 2020). The CACTus catalogue (http://sidc.be/cactus/), which automatically recognizes CMEs, shows similar results.

Figure 9.16 also shows the frequencies of fast CMEs (red; velocity > 900 km/s) and M-class or larger flares (blue). Comparing SC23 and SC24, we can see that the number of fast CMEs decreased correspondingly to the decrease in the sunspot number. This is because most of the fast CMEs occur from in active regions.

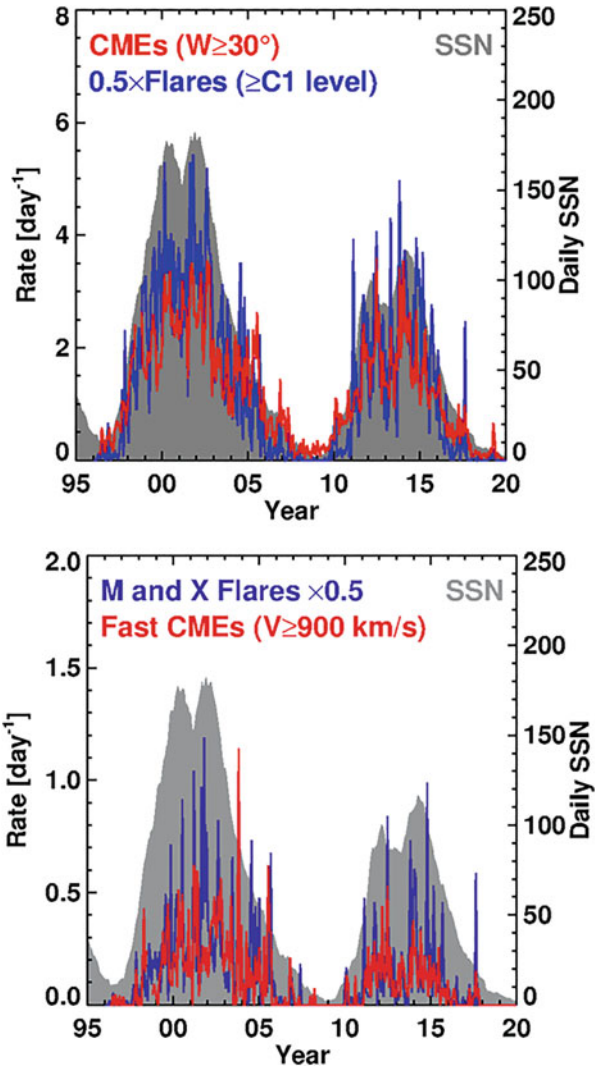

Fig. 9.16 Solar cycle variations in a number of solar flares (blue) and CMEs (red). The gray shadow represents the sunspot numbers (SSNs)

9.3.3 Halo CME

When a CME occurs near the center of the solar disk (or directly behind the Sun) and propagates along the Sun–Earth Line, it is observed as a CME that covers the area around the occulting disk of the coronagraphs (Fig. 9.17). Such CME is called a halo CME (Howard et al. 1982). Frontside halo CMEs are very important in space weather studies because they hit the Earth.

Figure 9.18 shows the speed distributions of halo CMEs (top) and all CMEs (widths >30°; bottom). The projection effects are not corrected. The peak of the

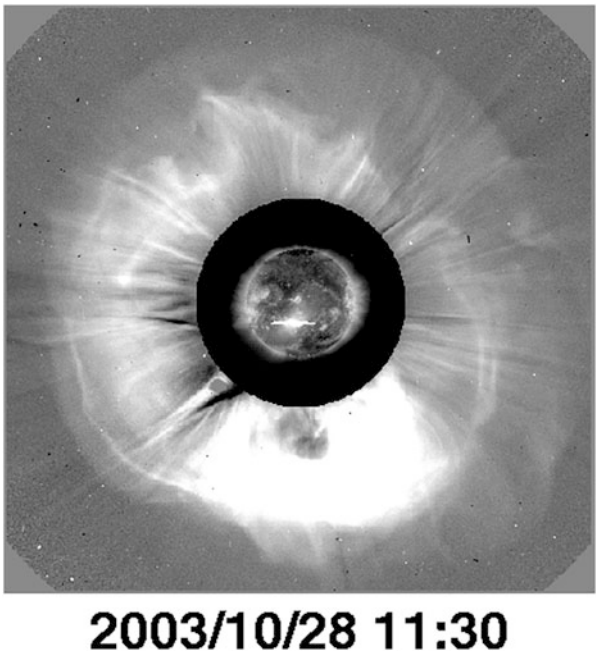

Fig. 9.17 Halo CME observed by LASCO

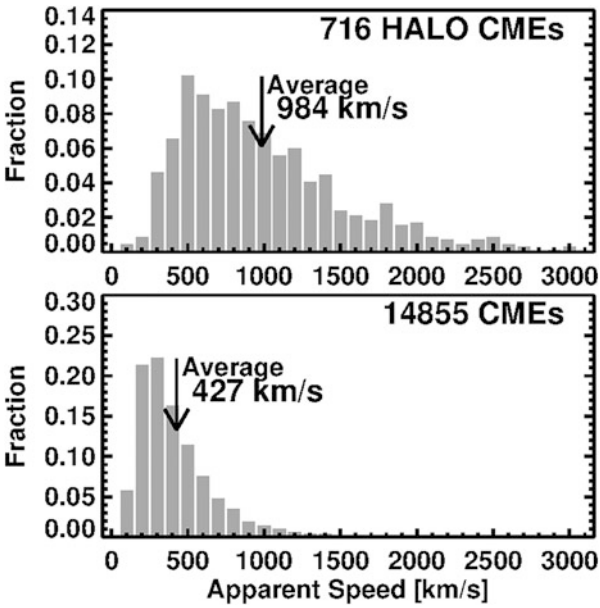

Fig. 9.18 Velocity distributions for halo CMEs (top) and CMEs over 30° width (bottom)

distribution is at 500 km/s for the halo CMEs, whereas it is at 300 km/s for all CMEs. The distribution of the halo CMEs is much wider on the right side (high-speed side), and the average speed of the halo CMEs is more than twice that of all CMEs (984 km/s compared to 427 km/s for all CMEs). One might think that the halo CMEs from the disk center should have a lower apparent speed than the ordinary limb CMEs because the projection effect is greater.

To be observed as a halo CME, a CME must cross the Sun–Earth line. Therefore, inherently wide CMEs have more chance of being observed as halos. The actual width of the halo CME is estimated to be approximately 120°, which is more than double that of ordinary CMEs (Michałek et al. 2003). A wide CME tends to have a high speed (Yashiro et al. 2004). Therefore, CMEs with higher speeds and larger widths have a higher probability of being observed as halo CMEs. Because of this selection bias, a halo CME is a collection of very fast CMEs.

9.4 Solar Energetic Particles

SEPs are one of the major space weather phenomena caused by CMEs. SEPs are charged particles (electrons, protons, and ions) accelerated up to several GeV by the fast CMEs and propagate to the Earth along the interplanetary magnetic field. The speed of a few gigaelectron volts (GeV) protons is very high, approximately 90% speed of light; therefore, the SEPs reach the orbit of the Earth within a few minutes after the onset of a solar storm observation. SEPs cause space weather damage, including exposure of astronauts, passengers, and crew in high-latitude aircrafts, and satellite failure.

9.4.1 SEP Observation

SEPs are monitored by the GOES, a U.S. meteorological satellite in a geostationary orbit. The magnetic field strength in a geostationary orbit is approximately 100 nT; therefore, the Larmor radii of protons (when the pitch angle is 90°) are 0.7 Re (10 MeV), 2.3 Re (100 MeV), and 8.9 Re (1 GeV), respectively. Here, Re is the radius of the Earth. Because the Larmor radius of a proton below 100 MeV is smaller than the size of the magnetosphere of the Earth (10 Re on the Sun side), it might be considered that such an energetic proton is shielded and cannot reach a geostationary orbit (6.6 Re). However, it has been found from a comparison of satellite data in and out of the magnetosphere that protons above 20 MeV actually penetrate freely into geostationary orbits (Paulikas and Blake 1969). Low-energy SEPs enter a geostationary orbit from the tail of the magnetosphere. Although these particles are accelerated in the vicinity of the Sun, they are isotropic, except in the early stages of the event.

There are two types of proton fluxes: differential and integral fluxes, which represent the scales of SEPs. The units are particles/(cm^2 s sr MeV) and particles/(cm^2 s sr), and the latter is also denoted as pfu. The U.S. National Oceanic and Atmospheric Administration uses the integral flux, particularly >10 MeV flux, to measure the scales of SEPs.

9.4.2 *Impulsive SEPs and Gradual SEPs*

SEPs are classified into impulsive and gradual SEPs based on their composition and duration (Reames 1999, 2013). Table 9.2 summarizes their characteristics. The most important difference between them is the abundance of ^{3}He: the ^{3}He/^{4}He ratio in gradual SEPs is the same as that in the solar wind, whereas that in impulsive SEPs is approximately 1, which is very high in ^{3}He. The impulsive SEPs are considered to be accelerated by particles heated by flares due to the high ionization of iron. The difference in the spread of the region where accelerated electrons are observed naturally reflects the difference in the size of the acceleration region: impulsive SEPs have a narrow range, within 30°, whereas gradual SEPs have a spread of more than 120°. Therefore, it is considered that a gradual SEP is accelerated by the shock wave generated in front of a widely expanding CME. Although there is no perfect one-to-one correspondence, the accompanying radio bursts are also different. Impulsive SEPs are accompanied by Type-III radio bursts (radio emissions generated when flare-accelerated electrons escape into the interplanetary space), whereas gradual SEPs are accompanied by Type-II radio bursts (radio emissions caused by shock waves driven by high-speed CMEs). Thus, it is concluded that impulsive SEPs are from solar flares and gradual SEPs are accelerated by shock waves in front of CMEs.

Table 9.2 Classification of SEPs. (From Reames (1999, 2013))

Impulsive SEP and Gradual SEP		
	Impulsive SEP	Gradual SEP
Another name	^{3}He-Rich	Long-duration
Composition	Electron rich	Proton rich
^{3}He/^{4}He ratio	~1	~0.0005
Fe/O Ratio	~1	~0.1
Iron ionization	~20	10–14
Duration	A few hours	A few days
Span of source longitude	<30°	~180°
Radio burst	Type III	Type II
Origin	Solar flare	CME shock

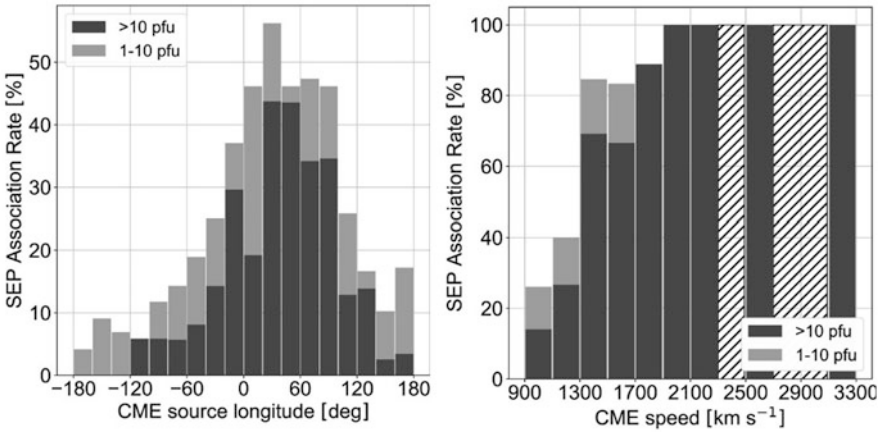

Fig. 9.19 (Left) Longitude dependence of SEP concomitant rate for fast CME (>900 km/s). (Right) Dependence of SEP incidental rate on CME velocity for high-speed CME occurring at E10–W100. No data are available in the shaded bins owing to a lack of samples. (From Kihara et al. 2020)

9.4.3 Gradual SEPs and CMEs

SEPs were first discovered as an increase in cosmic rays observed in conjunction with a solar flare that occurred on February 28, 1942 (Forbush 1946). When SEPs are observed on the ground, they are called GLE events. At the time of discovery, it was thought that gradual SEPs were also accelerated by solar flares; however, 30 years later, CMEs were discovered (Tousey 1973). SEP acceleration by CME shock waves was proposed and extensively accepted (Kahler et al. 1978, 1984). It has been clarified that CMEs are intrinsically important for the acceleration of gradual SEPs because flares without CMEs do not produce gradual SEPs, even if they are X-class flares.

The left panel in Fig. 9.19 shows the SEP association rate of a fast CME (velocity of 900 km/s or higher) as a function of the source longitude. SEPs are more probable to be observed when a fast CME occurs in the range of E10°–W100°. The peak of the SEP association rate is located at west 40°–60°, which roughly corresponds to the root of the interplanetary magnetic field spiral connecting to the Earth.

The CME speed dependence of the SEP association rate at a well-connected longitude (E10°–W100°) is shown in Fig. 9.19. The association rate of the major SEPs is slightly above 10% at 1000 km/s CME, increases with increasing speed, and reaches 100% at >2000 km/s CME. This indicates that the CME speed (i.e., shock wave velocity) is essential for the SEP acceleration.

The relationship between the CME velocity and the SEP flux is shown in Fig. 9.20. The y-axis in the left panel is the differential flux around 2 MeV and 20 MeV, whereas that in the right panel is the integral flux at >10 MeV. In both cases, the peak SEP fluxes increase with increasing CME speed, suggesting that

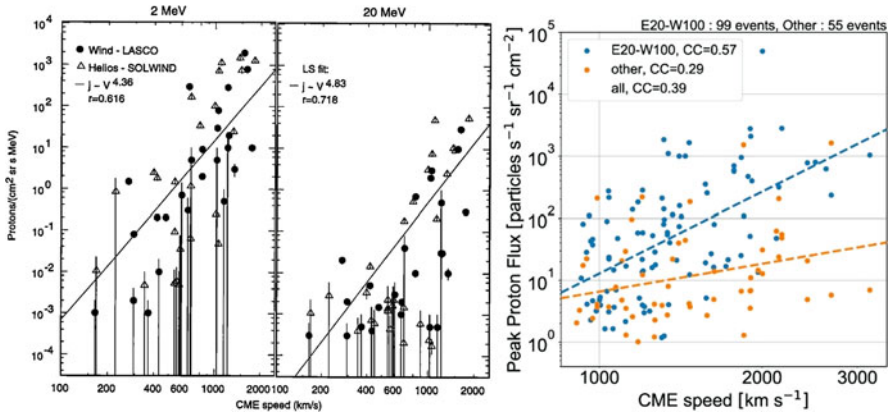

Fig. 9.20 SEP differential flux (left; from Kahler 2001) and SEP integral flux (right; from Kihara et al. 2020) versus CME velocity

faster CMEs are more important for space weather. Kihara et al. (2020) showed that the SEP flux of well-connected longitude events (blue) is stronger on average than the outside events (orange). However, even for events occurring in a well-connected longitude, the SEP flux for a given CME speed spreads by three orders of magnitude. The CME speeds and locations are not good enough to predict the SEP flux. Our current understanding of the phenomenon of SEPs is severely lacking.

References

Bornmann, P.L., Speich, D., Hirman, J., et al.: GOES X-ray sensor and its use in predicting solar-terrestrial disturbances. Proc. SPIE. **2812**, 291–298 (1996). https://doi.org/10.1117/12.254076

Brueckner, G.E., Howard, R.A., Koomen, M.J., Korendyke, C.M., Michels, D.J., Moses, J.D., Socker, D.G., Dere, K.P., Lamy, P.L., Llebaria, A., Bout, M.V., Schwenn, R., Simnett, G.M., Bedford, D.K., Eyles, C.J.: The large angle spectroscopic coronagraph (LASCO). Sol. Phys. **162**, 357–402 (1995). https://doi.org/10.1007/BF00733434

Cargill, P.J., Bradshaw, S.J., Klimchuk, J.A.: Enthalpy-based thermal evolution of loops. III. Comparison of zero-dimensional models. Astrophys. J. **758**(1), 5 (2012). https://doi.org/10.1088/0004-637X/758/1/5

Carrington, R.C.: Description of a singular appearance seen in the sun on September 1, 1859. Mon. Not. R. Astron. Soc. **20**, 13–15 (1859). https://doi.org/10.1093/mnras/20.1.13

Chamberlin, P.C., Woods, T.N., Eparvier, F.G.: Flare irradiance spectral model (FISM) use for space weather applications. Proc. ILWS Workshop, 153 (2006)

Chamberlin, P.C., Woods, T.N., Eparvier, F.G.: flare irradiance spectral model (FISM): daily component algorithms and results. Space Weather. **5**(7), S07005 (2007). https://doi.org/10.1029/2007SW000316

Chamberlin, P.C., Woods, T.N., Eparvier, F.G.: flare irradiance spectral model (FISM): flare component algorithms and results. Space Weather. **6**(5), S05001 (2008). https://doi.org/10.1029/2007SW000372

Chamberlin, P.C., Eparvier, F.G., Knoer, V.: The flare irradiance spectral model-version 2 (FISM2). Space Weather. **18**(12), e2020SW002588 (2020). https://doi.org/10.1029/2020SW002588

Cliver, E.W.: The 1859 space weather event: then and now. Adv. Space Res. **38**, 119–129 (2006). https://doi.org/10.1016/j.asr.2005.07.077

Dellinger, J.H.: Sudden disturbances of the ionosphere. J. Appl. Phys. **8**, 732–751 (1937). https://doi.org/10.1063/1.1710251

Dere, K.P., Landi, E., Mason, H.E., et al.: CHIANTI - an atomic database for emission lines. Astron. Astrophys. Suppl. Ser. **125**, 149–173 (1997). https://doi.org/10.1051/aas:1997368

Dere, K.P., Del Zanna, G., Young, P.R., et al.: CHIANTI—an atomic database for emission lines. XV. Version 9, improvements for the X-ray satellite lines. Astrophys. J. **Suppl. Ser. 241, 2**, 22 (2019). https://doi.org/10.3847/1538-4365/ab05cf

Domingo, V., Fleck, B., Poland, A.I.: The SOHO Mission: an overview. Sol. Phys. **162**, 1–37 (1995). https://doi.org/10.1007/BF00733425

Fenyi, J.: On the enormous velocity of a solar prominence, observed June 17, 1891. Astron. Astrophys. **11**, 63 (1892)

Forbush, S.E.: Three unusual cosmic-ray increases possibly due to charged particles from the sun. Phys. Rev. **70**, 771–772 (1946). https://doi.org/10.1103/PhysRev.70.771

Gopalswamy, N., Lara, A., Yashiro, S., Kaiser, M.L., Howard, R.A.: Predicting the 1-AU arrival times of coronal mass ejections. J. Geophys. Res. **106**, 29207–29218 (2001). https://doi.org/10.1029/2001JA000177

Gopalswamy, N., Shimojo, M., Lu, W., Yashiro, S., Shibasaki, K., Howard, R.A.: Prominence eruptions and coronal mass ejection: a statistical study using microwave observations. Astrophys. J. **586**, 562–578 (2003). https://doi.org/10.1086/367614

Gopalswamy, N.: History and development of coronal mass ejections as a key player in solar terrestrial relationship. Geosci. Lett. **3**, 8 (2016). https://doi.org/10.1186/s40562-016-0039-2

Gopalswamy, N., Akiyama, S., Yashiro, S.: The state of the heliosphere revealed by limb-halo coronal mass ejections in solar cycles 23 and 24. Astrophys. J. **897**, L1 (2020). https://doi.org/10.3847/2041-8213/ab9b7b

Gosling, J.T., Hildner, E., MacQueen, R.M., Munro, R.H., Poland, A.I., Ross, C.L.: Mass ejections from the sun: a view from Skylab. J. Geophys. Res. **79**, 4581–4587 (1974). https://doi.org/10.1029/JA079i031p04581

Gosling, J.T., Hildner, E., MacQueen, R.M., Munro, R.H., Poland, A.I., Ross, C.L.: The speeds of coronal mass ejection events. Sol. Phys. **48**, 389–397 (1976). https://doi.org/10.1007/BF00152004

Gosling, J.T., McComas, D.J., Phillips, J.L., Bame, S.J.: Geomagnetic activity associated with earth passage of interplanetary shock disturbances and coronal mass ejections. J. Geophys. Res. **96**, 7831–7839 (1991). https://doi.org/10.1029/91JA00316

Hale, G.E.: On the probable existence of a magnetic field in sun-spots. Astrophys. J. **28**, 315 (1908). https://doi.org/10.1086/141602

Hanaoka, Y., Sakurai, T.: Statistical study of the magnetic field orientation in solar filaments. Astrophys. J. **851**, 130 (2017). https://doi.org/10.3847/1538-4357/aa9cf1

Hanaoka, Y., Sakurai, T., Otsuji, K., Suzuki, I., Morita, S.: Synoptic solar observations of the solar flare telescope focusing on space weather. J. Space Weather Space Clim. **10**, 41 (2020). https://doi.org/10.1051/swsc/2020044

Hildner, E., Gosling, J.T., MacQueen, R.M., Munro, R.H., Poland, A.I., Ross, C.L.: Frequency of coronal transients and solar activity. Sol. Phys. **48**, 127 (1976). https://doi.org/10.1007/BF00153339

Hill, F.: The global oscillation network group facility - an example of research to operations in space weather. Space Weather. **16**, 1488–1497 (2018). https://doi.org/10.1029/2018SW002001

Hill, F., Hammel, H., Martinez-Pillet, V., de Wijn, A., Gosain, S., Burkepile, J., Henney, C.J., McAteer, J., Bain, H.M., Manchester, W., Lin, H., Roth, M., Ichimoto, K., Suematsu, Y.: ngGONG: The Next Generation GONG - A New Solar Synoptic Observational Network. In

Astro2020: Decadal Survey on Astronomy and Astrophysics, APC white papers, no. 74. Bull. Am. Astron. Soc. **51**, 74 (2019)

Howard, R.A., Michels, D.J., Sheeley Jr., N.R., Koomen, M.J.: The observation of a coronal transient directed at earth. Astrophys. J. **263**, L101–L104 (1982). https://doi.org/10.1086/183932

Howard, R.A., Sheeley Jr., N.R., Michels, D.J., Koomen, M.J.: Coronal mass ejections: 1979-1981. J. Geophys. Res. **90**, 8173–8192 (1985). https://doi.org/10.1029/JA090iA09p08173

Hundhausen, A.J.: Sizes and locations of coronal mass ejections: SMM observations from 1980 and 1984-1989. J. Geophys. Res. **98**, 13117–13200 (1993). https://doi.org/10.1029/93JA00157

Hundhausen, A.J., Burkepile, J.T., St. Cyr, O.C.: Speeds of coronal mass ejections: SMM observations from 1980 and 1984-1989. J. Geophys. Res. **99**, 6543–6552 (1994). https://doi.org/10.1029/93JA03586

Ichimoto, K., Ishii, T.T., Otsuji, K., Kimura, G., Nakatani, Y., Kaneda, N., Nagata, S., UeNo, S., Hirose, K., Cabezas, D., Morita, S.: A new solar imaging system for observing high-speed eruptions: solar dynamics doppler imager (SDDI). Sol. Phys. **292**, 63 (2017). https://doi.org/10.1007/s11207-017-1082-7

Illing, R.M.E., Hundhausen, A.: J.: observation of a coronal transient from 1.2 to 6 solar radii. J. Geophys. Res. **90**, 275–282 (1985). https://doi.org/10.1029/JA090iA01p00275

Imada, S., Murakami, I., Watanabe, T.: Observation and numerical modeling of chromospheric evaporation during the impulsive phase of a solar flare. Phys. Plasmas. **22**, 101206 (2015). https://doi.org/10.1063/1.4932335

Iwai, K., Kubo, Y., Ishibashi, H., Naoi, T., Harada, K., Ema, K., Hayashi, Y., Chikahiro, Y.: OCTAD-S: digital fast Fourier transform spectrometers by FPGA. Earth Planets Space. **69**, 95 (2017). https://doi.org/10.1186/s40623-017-0681-8

Jin, H., Miyoshi, Y., Fujiwara, H., Shinagawa, H., Terada, K., Terada, N., Ishii, M., Otsuka, Y., Saito, A.: Vertical connection from the tropospheric activities to the ionospheric longitudinal structure simulated by a new Earth's whole atmosphere-ionosphere coupled model. J. Geophys. Res. **116**, A01316 (2011). https://doi.org/10.1029/2010JA015925

Judge, D.L., McMullin, D.R., Ogawa, H.S., Hovestadt, D., Klecker, B., Hilchenbach, M., Mobius, E., Canfield, L.R., Vest, R.E., Watts, R., Tarrio, C., Kuehne, M., Wurz, P.: First solar EUV irradiances obtained from SOHO by the CELIAS/SEM. Sol. Phys. **177**, 161–173 (1998). https://doi.org/10.1023/A:1004929011427

Kahler, S.W.: The correlation between solar energetic particle peak intensities and speeds of coronal mass ejections: effects of ambient particle intensities and energy spectra. J. Geophys. Res. **106**, 20947–20956 (2001). https://doi.org/10.1029/2000JA002231

Kahler, S.W., Hildner, E., Van Hollebeke, M.A.I.: Promt solar proton events and coronal mass ejections. Sol. Phys. **57**, 429–443 (1978). https://doi.org/10.1007/BF00160116

Kahler, S.W., Sheeley Jr., N.R., Howard, R.A., Michels, D.J., Koomen, M.J., McGuire, R.E., von Rosenvinge, T.T., Reames, D.V.: Associations between coronal mass ejections and solar energetic proton events. J. Geophys. Res. **89**, 9683–9694 (1984). https://doi.org/10.1029/JA089iA11p09683

Kaiser, M.L., Kucera, T.A., Davila, J.M., St. Cyr, O.C., Guhathakurta, M., Christian, E.: The STEREO mission: An introduction. Space Sci. Rev. **136**, 5–16 (2008). https://doi.org/10.1007/s11214-007-9277-0

Kane, S. R.: Impulsive (flash) Phase of Solar Flares: Hard X-Ray, Microwave, EUV and Optical Observations. Coronal Disturbances: Proc. 57th IAU Symp. 105 (1974)

Kawai, T., Imada, S., Nishimoto, S., et al.: Nowcast of an EUV dynamic spectrum during solar flares. J. Atmos. Solar-Terr. Phys. **205**, 105302 (2020). https://doi.org/10.1016/j.jastp.2020.105302

Keller, C.U., Harvey, J.W., Giampapa, M.S.: SOLIS: an innovative suite of synoptic instruments. Proc. SPIE. **4853**, 194–204 (2003). https://doi.org/10.1117/12.460373

Kihara, K., Huang, Y., Nishimura, N., Nitta, N.V., Yashiro, S., Ichimoto, K., Asai, A.: Statistical analysis of the relation between coronal mass ejections and solar energetic particles. Astrophys. J. **900**, 75 (2020). https://doi.org/10.3847/1538-4357/aba621

Klimchuk, J.A., Patsourakos, S., Cargill, P.J.: Highly efficient modeling of dynamic coronal loops. Astrophys. J. **682**(2), 1351–1362 (2008). https://doi.org/10.1086/589426

Kojima, M., Kakinuma, T.: Solar cycle dependence of global distribution of solar wind speed. Space Sci. Rev. **53**, 173–222 (1990). https://doi.org/10.1007/BF00212754

Kosugi, T., Matsuzaki, K., Sakao, T., Shimizu, T., Sone, Y., Tachikawa, S., Hashimoto, T., Minesugi, K., Ohnishi, A., Yamada, T., Tsuneta, S., Hara, H., Ichimoto, K., Suematsu, Y., Shimojo, M., Watanabe, T., Shimada, S., Davis, J.M., Hill, L.D., Owens, J.K., Title, A.M., Culhane, J.L., Harra, L.K., Doschek, G.A., Golub, L.: The Hinode (solar-B) Mission: an overview. Sol. Phys. **243**, 3–17 (2007). https://doi.org/10.1007/s11207-007-9014-6

MacQueen, R.M., Eddy, J.A., Gosling, J.T., Hildner, E., Munro, R.H., Newkirk Jr., G.A., Poland, A.I., Ross, C.L.: The outer solar Corona as observed from Skylab: preliminary results. Astrophys. J. **187**, L85 (1974). https://doi.org/10.1086/181402

Michałek, G., Gopalswamy, N., Yashiro, S.: A new method for estimating widths, velocities, and source location of halo coronal mass ejections. Astrophys. J. **584**, 472–478 (2003). https://doi.org/10.1086/345526

Nishimoto, S., Watanabe, K., Imada, S., Kawate, T., Lee, K.-S.: Statistical and observational research on solar flare EUV spectra and geometrical features. Astrophys. J. **904**(1), 31 (2020). https://doi.org/10.3847/1538-4357/abbacb

Nishimoto, S., Watanabe, K., Kawai, T., Imada, S., Kawate, T.: Validation of computed extreme ultraviolet emission spectra during solar flares. Earth Planets Space. **73**, 79 (2021). https://doi.org/10.1186/s40623-021-01402-7

Payne-Scott, R., Yabsley, D.E., Bolton, J.G.: Relative times of arrival of bursts of solar noise on different radio frequencies. Nature. **160**, 256–257 (1947). https://doi.org/10.1038/160256b0

Paulikas, G.A., Blake, J.B.: Penetration of solar protons to synchronous altitude. J. Geophys. Res. **74**, 2161 (1969). https://doi.org/10.1029/JA074i009p02161

Pesnell, W.D., Thompson, B.J., Chamberlin, P.C.: The solar dynamics observatory (SDO). Sol. Phys. **275**, 3–15 (2012). https://doi.org/10.1007/s11207-011-9841-3

Reames, D.V.: Particle acceleration at the sun and in the heliosphere. Space Sci. Rev. **90**, 413–491 (1999). https://doi.org/10.1023/A:1005105831781

Reames, D.V.: The two sources of solar energetic particles. Space Sci. Rev. **175**, 53–92 (2013). https://doi.org/10.1007/s11214-013-9958-9

Reep, J.W., Knizhnik, K.J.: What determines the X-ray intensity and duration of a solar flare? Astrophys. J. **874, 2**, 157 (2019). https://doi.org/10.3847/1538-4357/ab0ae7

Sakurai, T., Hanaoka, Y., Arai, T., Hagino, M., Kawate, T., Kitagawa, N., Kobiki, T., Miyashita, M., Morita, S., Otsuji, K., Shinoda, K., Suzuki, I., Yaji, K., Yamasaki, T., Fukuda, T., Noguchi, M., Takeyama, N., Kanai, Y., Yamamuro, T.: Infrared spectro-polarimeter on the solar flare telescope at NAOJ/Mitaka. Publ. Astron. Soc. Jpn. **70**, 58 (2018). https://doi.org/10.1093/pasj/psy050

Scherrer, P.H., Bogart, R.S., Bush, R.I., Hoeksema, J.T., Kosovichev, A.G., Schou, J., Rosenberg, W., Springer, L., Tarbell, T.D., Title, A., Wolfson, C.J., Zayer, I.: MDI engineering team: the solar oscillations investigation - Michelson doppler imager. Sol. Phys. **162**, 129–188 (1995). https://doi.org/10.1007/BF00733429

Scherrer, P.H., Schou, J., Bush, R.I., Kosovichev, A.G., Bogart, R.S., Hoeksema, J.T., Liu, Y., Duvall, T.L., Zhao, J., Title, A.M., Schrijver, C.J., Tarbell, T.D., Tomczyk, S.: The helioseismic and magnetic imager (HMI) investigation for the solar dynamics observatory (SDO). Sol. Phys. **275**, 207–227 (2012). https://doi.org/10.1007/s11207-011-9834-2

Seki, D., Otsuji, K., Isobe, H., Ishii, T.T., Sakaue, T., Hirose, K.: Increase in the amplitude of line-of-sight velocities of the small-scale motions in a solar filament before eruption. Astrophys. J. Lett. **843**, L24 (2017). https://doi.org/10.3847/2041-8213/aa7559

Seki, D., Otsuji, K., Ishii, T., Hirose, K., Iju, T., UeNo, S., Cabezas, D., Asai, A., Isobe, H., Ichimoto, K., Shibata, K.: SMART/SDDI filament disappearance catalogue. Sun Geosphere. **14**, 95–103 (2019). https://doi.org/10.31401/SunGeo.2019.02.01

Sonett, C.P., Colburn, D.S., Davis, L., Smith, E.J., Coleman, P.J.: Evidence for a collision-free magnetohydrodynamic shock in interplanetary space. Phys. Rev. Lett. **13**, 153 (1964). https://doi.org/10.1103/PhysRevLett.13.153

Stone, E.C., Frandsen, A.M., Mewaldt, R.A., Christian, E.R., Margolies, D., Ormes, J.F., Snow, F.: The advanced composition explorer. Space Sci. Rev. **86**, 1–22 (1998). https://doi.org/10.1023/A:1005082526237

Steinegger, M., Denker, C., Goode, P. R., Marquette, W. H., Varsik, J., Wang, H., Otruba, W., Freislich, H., Hanslmeier, A., Luo, G., Chen, D., Zhang, Q.: The New Global High-Resolution Hα Network: First Observations and First Results. In ESA-SP 463, A. Wilson ed., The Solar Cycle and Terrestrial Climate, Solar and Space weather, 617–622 (2000)

Tao, C., Nishioka, M., Saito, S., Shiota, D., Watanabe, K., Nishizuka, N., Tsugawa, T., Ishii, M.: Statistical analysis of short-wave fadeout for extreme space weather event estimation. Earth Planets Space. **72, 1**, 173 (2020). https://doi.org/10.1186/s40623-020-01278-z

Tousey, R.: In: Rycroft, M.J., Runcorn, S.K. (eds.) The solar corona. In: Space Research XIII, p. 713. Akademie-Verlag, Berlin (1973)

Tsuneta, S., Ichimoto, K., Katsukawa, Y., Nagata, S., Otsubo, M., Shimizu, T., Suematsu, Y., Nakagiri, M., Noguchi, M., Tarbell, T., Title, A., Shine, R., Rosenberg, W., Hoffmann, C., Jurcevich, B., Kushner, G., Levay, M., Lites, B., Elmore, D., Matsushita, T., Kawaguchi, N., Saito, H., Mikami, I., Hill, L.D., Owens, J.K.: The solar optical telescope for the Hinode mission: an overview. Sol. Phys. **249**, 167–196 (2008). https://doi.org/10.1007/s11207-008-9174-z

UeNo, S., Shibata, K., Kitai, R., Ichimoto, K., Nagata, S., Isobe, H., Kimura, G., Nakatani, Y., Kadota, M., Ishii, T.T., Morita, S., Otsuji, K.: The CHAIN-project and installation of flare monitoring telescopes in developing countries. Data Sci. J. **8**, S51–S58 (2009)

Watanabe, K., Jin, H., Nishimoto, S., Imada, S., Kawai, T., Kawate, T., Otsuka, Y., Shinbori, A., Tsugawa, T., Nishioka, M.: Model-based reproduction and validation of the total spectrum of solar flare and their impact on the global environment at the X93 event of September 6 2017. Earth Planets Space. **73**, 96 (2021). https://doi.org/10.1186/s40623-021-01376-6

Webb, D.F.: Erupting prominences and the geometry of coronal mass ejections. J. Geophys. Res. **93**, 1749 (1988). https://doi.org/10.1029/JA093iA03p01749

Webb, D.F., Hundhausen, A.J.: Activity associated with the solar origin of coronal mass ejections. Sol. Phys. **108**, 383 (1987). https://doi.org/10.1007/BF00214170

Woods, T.N., Eparvier, F.G., Bailey, S.M., Chamberlin, P.C., Lean, J., Rottman, G.J., Solomon, S. C., Tobiska, W.K., Woodraska, D.L.: Solar EUV experiment (SEE): mission overview and first results. J. Geophys. Res. **110**, A01312 (2005). https://doi.org/10.1029/2004JA010765

Woods, T.N., Eparvier, F.G., Hock, R., Jones, A.R., Woodraska, D., Judge, D., Didkovsky, L., Lean, J., Mariska, J., Warren, H., McMullin, D., Chamberlin, P., Berthiaume, G., Bailey, S., Fuller-Rowell, T., Sojka, J., Tobiska, W.K., Viereck, R.: Extreme ultraviolet variability experiment (EVE) on the solar dynamics observatory (SDO): overview of science objectives, instrument design, data products, and model developments. Sol. Phys. **275**, 115–143 (2012). https://doi.org/10.1007/s11207-009-9487-6

Woods, T.N., Hock, R., Eparvier, F., Jones, A.R., Chamberlin, P.C., Klimchuk, J.A., Didkovsky, L., Judge, D., Mariska, J., Warren, H., Schrijver, C.J., Webb, D.F., Bailey, S., Tobiska, W.K.: New solar extreme-ultraviolet irradiance observations during flares. Astrophys. J. **739**, 59 (2011). https://doi.org/10.1088/0004-637X/739/2/59

Yashiro, S., Gopalswamy, N., Michalek, G., St. Cyr, O.C., Plunkett, S.P., Rich, N.B., Howard, R. A.: A catalog of white light coronal mass ejections observed by the SOHO spacecraft. J. Geophys. Res. **109**, A07105 (2004). https://doi.org/10.1029/2003JA010282

Yashiro, S., Gopalswamy, N., Akiyama, S., Michalek, G., Howard, R.A.: Visibility of coronal mass ejections as a function of flare location and intensity. J. Geophys. Res. **110**, A12S05 (2005). https://doi.org/10.1029/2005JA011151

Zhang, J., Richardson, I.G., Webb, D.F., Gopalswamy, N., Huttunen, E., Kasper, J.C., Nitta, N.V., Poomvises, W., Thompson, B.J., Wu, C.-C., Yashiro, S., Zhukov, A.N.: Solar and interplanetary sources of major geomagnetic storms (Dst $<= -100$ nT) during 1996-2005. J. Geophys. Res. **112**, A10102 (2007). https://doi.org/10.1029/2007JA012321

Chapter 10
Prediction of Solar Storms

Kanya Kusano, Shin Toriumi, Daikou Shiota, and Takashi Minoshima

10.1 Formation of Flare-Productive Active Regions

10.1.1 Sunspot Classification and Flare Activity

The structure of a sunspot is one of the most important factors for determining flare activity. For example, the McIntosh classification (an extended version of the Zurich classification) categorizes sunspot groups according to their spatial extents, symmetry of the penumbrae, and overall spot distributions, and it has been statistically proven that sunspots with more complex shapes show higher flare activity (Bornmann and Shaw 1994). Another commonly used classification method is the Mount Wilson classification, which focuses on the magnetic structures of sunspot groups and is expressed by a combination of the following four identifiers:

- α: a unipolar spot group,
- β: a simple bipolar spot group consisting of both positive and negative polarities,
- γ: a complex spot group that cannot be classified as β type because sunspots of both polarities are irregularly dispersed,

K. Kusano (✉)
Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Aichi, Japan
e-mail: kusano@nagoya-u.jp

S. Toriumi
Institute of Space and Astronautical Science, Japan Aerospace Exploration Agency, Sagamihara, Kanagawa, Japan

D. Shiota
National Institute of Information and Communications Technology, Tokyo, Japan

T. Minoshima
Japan Agency for Marine-Earth Science and Technology, Yokohama, Kanagawa, Japan

K. Kusano (ed.), *Solar-Terrestrial Environmental Prediction*,
https://doi.org/10.1007/978-981-19-7765-7_10

- δ: a spot group in which the umbrae of opposite polarities are close to each other within 2° and share a common penumbra.

Among them, δ-spots are known to show outstanding flare activity. A statistical study on the flare magnitudes of various sunspot groups showed that there is some correlation between the spot group area and the flare magnitude; however, simultaneously, the magnitude increases with the complexity of the spot group (Sammis et al. 2000). In particular, all flares of X4-class or larger originate from βγδ-spots with an area of more than 1000 MSH (MSH is one-millionth of the solar hemisphere, which corresponds to approximately 3×10^6 km^2).

Although it is commonly accepted that sunspots with larger areas and more complex shapes tend to produce stronger flares, it should be noted that these are not sufficient conditions for producing strong flares. Even if an active region has a large and complex spot, it does not necessarily cause large flares. For instance, the largest sunspot group since the late nineteenth century, which appeared in April 1947, did not show significant flare activity (Aulanier et al. 2013). Therefore, the conditions producing large flares are still unclear. There may be multiple factors that contribute to the occurrence of flares, and it is necessary to elucidate their combinations through further observational and theoretical investigations.

10.1.2 Structures of Flare-Productive Active Regions

Based on many years of observations by the Big Bear Solar Observatory, H. Zirin et al. found that the formation processes of δ-spots can be classified into three categories (Zirin and Liggett 1987):

- Type I: in which magnetic fluxes emerge simultaneously.
- Type II: in which a magnetic flux emerges in the vicinity of an existing sunspot.
- Type III: in which two emerging magnetic fields collide to form a quadrupole sunspot group.

These suggest that the formation of δ-spots is a consequence of large-scale flux emergence, and thus, an understanding of the magnetic flux beneath the solar surface is critical. However, optical observations of the solar interior are impossible.

Therefore, there have been efforts to estimate the shape of the magnetic flux in the solar interior from apparent sunspot motions on the solar surface. For instance, it has been pointed out that the creation of δ-spots involves the emergence of a strongly twisted magnetic flux by a kink instability (Tanaka 1991; Linton et al. 1996; Ishii et al. 1998, 2000) or of a magnetic flux with a U-shaped apex (Poisson et al. 2013). However, these were all inferred from apparent spot motions, and magnetohydrodynamic (MHD) simulations were needed to elucidate the causal relationship between the geometry of the emerging flux and the formation of flare-productive active regions.

In contrast, observational characterization of flaring active regions has been significantly improved by the development of magnetic field diagnostics, particularly by photospheric vector magnetograph. Because solar flares are energy-releasing phenomena that occur via magnetic reconnection, the appearance of their characteristic features in the vicinity of a polarity inversion line (PIL, a line across which the polarity of the magnetic field changes its sign), where magnetic energy accumulates, is expected. It is known that PILs associated with large-scale flares have the following characteristics (Hagyard et al. 1984):

- Magnetic field strength: The horizontal field, which is the magnetic field component parallel to the solar surface, is enhanced.
- Spatial gradient of the magnetic field: The vertical field, which is the magnetic field component perpendicular to the solar surface, has a steep spatial gradient in the direction across a PIL.
- Magnetic shear: The horizontal field is aligned almost parallel to a PIL, and its deviation from the potential field is remarkable.

These magnetic field structures can be formed as a result of the rotational or shearing motions of sunspots (Harvey and Harvey 1976).

Figure 10.1 shows an image of an X-class flare and the vector magnetogram of a sunspot observed by the Hinode satellite. The flare occurs around the PIL between positive and negative polarities (δ-spots), where the horizontal field is strongly sheared. A striped pattern of positive and negative polarities called the "magnetic channel," which often appears in strongly flaring PILs, can also be seen.

Characteristic structures observed in the atmospheres above flaring PILs include a "filament" seen in the Hα and other chromospheric lines and an S-shaped "sigmoid" observed in soft X-rays. Both structures reflect twisted magnetic flux bundles called flux ropes, which are ejected into the space and develop into the magnetic framework of CMEs.

δ-spots, sheared PILs, and spot rotations observed in the photosphere and flux ropes in the atmosphere are formed by the large-scale development of active regions. There have been various efforts to apply these properties to flare prediction, as described in the subsequent sections of this chapter.

10.1.3 Formation of Flare-Productive Active Regions

As has been pointed out by observers, to understand the formation process of flare-productive active regions including δ-spots, it is necessary to trace the flux emergence from the solar interior. As a theoretical solution to this problem, MHD simulations have been conducted to model flux emergence—specifically, a numerical experiment in which a computational domain imitating the stratified solar atmosphere is set up and the magnetic flux is raised from the convection zone to create an active region. By varying the calculation conditions of the flux emergence, efforts to reproduce δ-spots have been made.

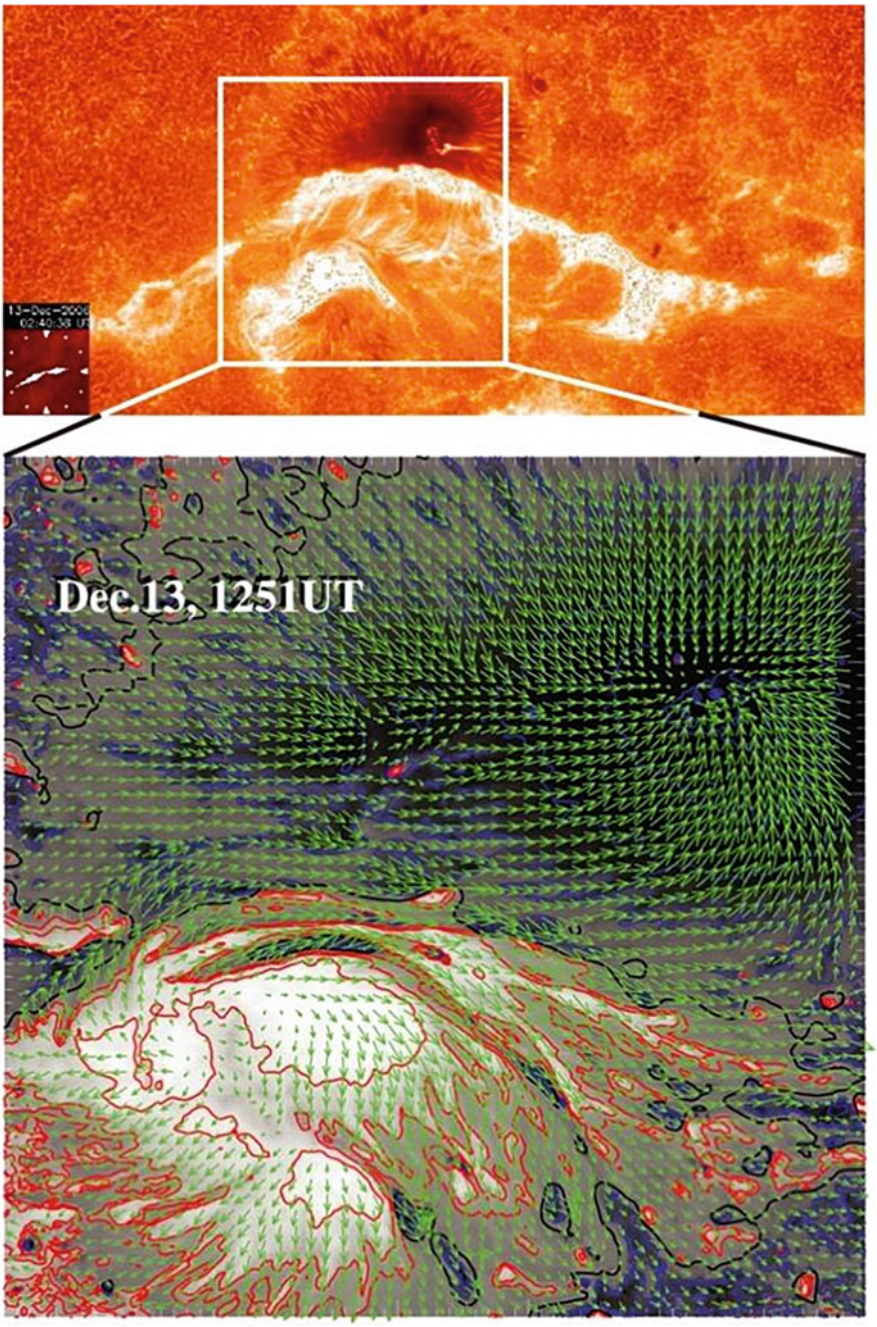

Fig. 10.1 X-class flare observed by Hinode satellite. (Top) Ca II H image. Courtesy of T.J. Okamoto. (Bottom) Photospheric vector magnetogram. Grayscale shows positive and negative components of the line-of-sight magnetic field, and green arrows represent in-plane components. Red, blue, and black lines represent positive and negative spots and PIL, respectively. (Image reproduced with permission from Wang et al. (2008), copyright by AAS)

In the 1990s, the emergence of a strongly twisted flux tube due to kink instability was investigated. The kink instability converts the twist of a flux tube (wrapping of the magnetic field lines around the axis of the tube) into a writhe (rotation of the axis of the flux tube), which causes the flux tube to strongly deform. Consequently, closely packed positive and negative polarities (i.e., δ-spots), complex configurations, and strong sunspot rotations appear on the solar surface (Fan et al. 1998; Takasao et al. 2015). These configurations produced by the kink instability are similar to Type-I δ-spots.

In the 2000s, another model was proposed in which two emerging flux tubes collide with each other to create a δ-structure on the solar surface (Murray and Hood 2007). For instance, if there exists an asymmetry between two flux tubes (e.g., one emerges ahead of the other), this model explains a Type-II δ-spot.

As another alternative, a model with a single magnetic flux tube emerging at two locations was proposed (Toriumi et al. 2014). In this case, a pair of emerging bipoles collide with each other on the solar surface, forming a quadrupole δ-spot with a sheared PIL inside. This scenario corresponds to Type-III δ-spots.

In Fig. 10.2, a comparison of the simulation models of flaring active regions including the three mechanisms discussed above is shown. In this study, various δ-spots were created by modifying the geometry of the initial flux tubes inside the convection zone, such as by increasing the twist of the tubes, introducing an accompanying thin flux tube, or endowing the flux tube with two emerging sections. A careful examination of the formed coronal fields revealed the existence of magnetic flux ropes (i.e., skeletons of CMEs). More recently, by using radiative MHD simulations that reproduce turbulent convection in the solar interior, it has become possible to reproduce even more realistic δ-spots, in which magnetic fields emerge under interaction with convection (Fig. 10.3). Through these numerical experiments, we can explain the observational features of flare-productive active regions self-consistently.

10.1.4 Summary and Future Perspectives

Summarizing, characteristic structures such as δ-spots, sheared PILs, spot rotations, and flux ropes are frequently observed in flaring active regions, and they are formed as a result of a large-scale flux emergence. In particular, the twisting of flux tubes and the interaction with turbulent convection play major roles. This physical picture was obtained by combining observational and theoretical investigations.

In recent years, studies on flaring active regions have not only deepened our physical understanding but also been actively applied to the prediction and forecasting of the occurrence of flares, as discussed in detail in the next section. However, it should be pointed out that there are still some unanswered questions. For example, it is still unclear how an emerging flux, which forms a flare-productive active region, twists in the solar interior, and this remains one of the key issues to be solved in the solar-stellar dynamo theory. In relation to the flare generation mechanism, it has

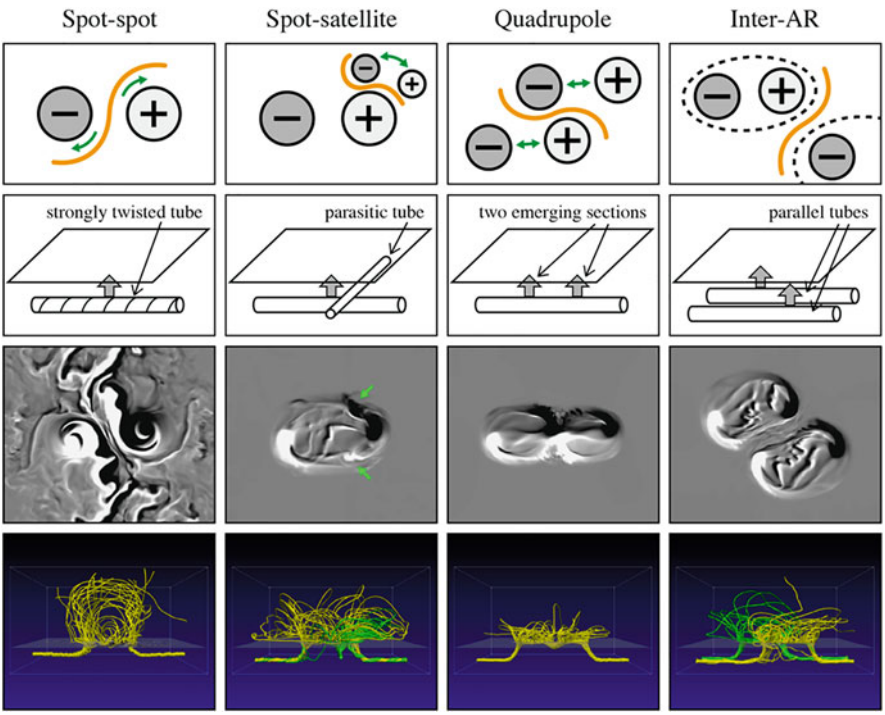

Fig. 10.2 Formation models of flaring active regions by flux emergence simulations. From left to right: Spot–spot (Type I: kinked tube model), Spot–satellite (Type II: interacting tube model), Quadrupole (Type III: multibuoyant segment model), and Inter-AR (case of independently emerging tubes). From the top, observed layouts of magnetic fields, computational conditions, resultant photospheric fields, and field line renderings. (Image reproduced with permission from Toriumi and Takasao (2017), copyright by AAS)

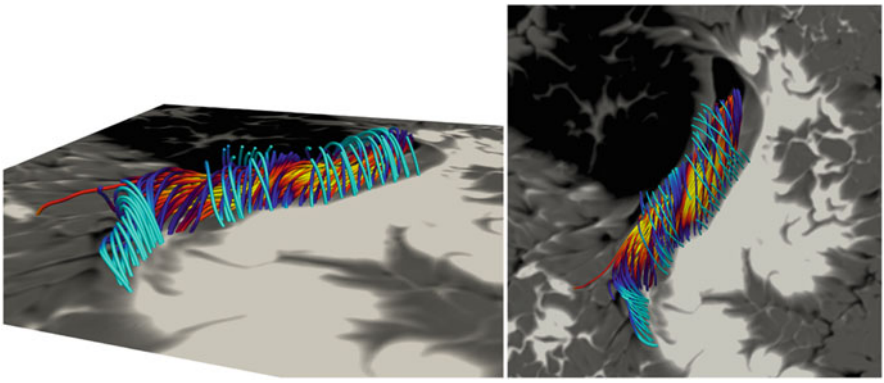

Fig. 10.3 δ-spots reproduced by a radiative MHD simulation. Flux rope is formed above PIL. (Image reproduced with permission from Toriumi and Hotta (2019), copyright by AAS)

been pointed out both observationally and theoretically that there may be a fine-scale magnetic structure that triggers the flare onset. It is necessary to clarify how such structures are formed inside flaring active regions. To clarify these issues, it is essential to improve both observational techniques and theoretical modeling.

10.2 Prediction of Solar Flares

10.2.1 What to Predict

The prediction of solar flares is an important issue in space weather forecasting. Many space weather forecasting agencies, e.g., the National Institute of Information and Communications Technology (NICT) in Japan, provide daily forecasts for the largest Geostationary Operational Environmental Satellite (GOES) flare class, $F_{1-8Å}$, that may occur within 24 h (Kubo et al. 2017). The statistical studies suggested that the GOES flare class is an approximate measure of the total flare energy. However, what to predict (i.e., "predictands") and how long in advance to predict (i.e., the lead time) depend on our concern and interest. Electromagnetic radiations, such as X-rays, (extreme) ultraviolet rays, and radio waves, emitted by a solar flare, propagate at the speed of light and reach the Earth simultaneously as that flare is observed. Then, impulsive SEPs accelerated by a solar flare reach the Earth within tens of minutes to hours after the onset of a flare. Because it is too short to prepare satellite and aircraft operations against the initial impacts of SEPs, we should predict the occurrence of a flare in advance of its onset. Not only the occurrence of a flare but also the intensity and spectrum of the electromagnetic radiation and the energetic particle flux emitted by flares are important predictands. On the other hand, CMEs take approximately a day to a few days to reach the orbit of the Earth and cause magnetic storms after their release from the solar corona. Therefore, even after the onset of flares and CMEs, there is adequate lead time for predicting CME arrivals and geomagnetic storms. In this section, we mainly discuss how we can predict the occurrence and scales of solar flares and how we evaluate flare prediction.

10.2.2 Methods of Prediction

Until now, various methods have been developed to predict solar flares. They can be broadly classified into "empirical prediction" and "physics-based prediction." An empirical prediction is conducted using some observable parameters that have been found to have some correlation with the occurrence of solar flares based on statistical analysis of past observations. As described below, in recent years, several parameters have been proposed as predictors, and various methods using machine learning

have been developed. However, it should be noted that even if a parameter has a certain correlation with flaring, it does not necessarily imply the existence of a causal relationship between that parameter and flaring. In addition, multiple factors may be involved in determining the occurrence and magnitude of a solar flare. Therefore, to achieve good empirical prediction, it is necessary to find a proper predictor through objective and quantitative evaluation of the prediction, as discussed subsequently.

In contrast, physics-based prediction requires predictors derived from the physical mechanisms of solar flares. However, because the onset mechanism of solar flares is still not fully understood, physics-based prediction is under development and has not yet been operationally implemented. To realize physics-based prediction, it is also necessary to deepen the research by objectively verifying the prediction based on physical hypotheses. Such efforts by the authors in recent years are described below.

10.2.3 Empirical Prediction

Various types of empirical predictions of flares have been proposed using various observable parameters. Here, we explain some of the parameters that have been proposed as predictors and introduce the typical methods for empirical prediction.

As explained in the previous section, there is some relationship between the sunspot classification and the flare activity. McIntosh (1990) proposed flare prediction based on morphological properties: revised Zurich sunspot classification (seven types), shape of the semi-dark area of the largest sunspot in each sunspot group (six types), and spread of sunspots (four types). This method has been commonly used for a long time and is still adopted by many space weather forecasting agencies, because sunspot shapes can be easily observed with small telescopes from the ground. However, X-class large flares occur only rarely, and their occurrence probability is only a few percent even though the McIntosh classification type has the highest probability among X-class flares (Gallagher et al. 2002). Therefore, the ability to predict the occurrence of large flares using sunspot morphology classification is limited.

Because a solar flare is an explosive release of magnetic energy, the magnetic field structure of sunspots and active regions should be related to the occurrence of flares. Therefore, there have been several proposals to predict solar flares using the magnetic field structures of active solar regions as predictors. For example, Sammis et al. (2000) found that larger sunspots with more complex magnetic field structures tend to produce larger flares, as previously mentioned. However, the classification of the magnetic field structures and the sunspot areas alone cannot predict more than half of the large flares. Therefore, there are apparently “hidden parameters” that determine the occurrence of flares, and many studies have been conducted to identify them.

A pioneering study was conducted by Hagyard et al. (1984), who investigated the relationship between magnetic shear and solar flares. Magnetic shear is the angle of the horizontal magnetic field vector with respect to the horizontal component of the potential magnetic field on the solar surface.[1] Shear angle is defined as positive when the magnetic field is rotated counterclockwise from the potential field on the PIL. When a shear angle is positive (or negative), the twist of the magnetic field is right-handed (or left-handed) and produces positive (or negative) magnetic helicity, respectively. If the magnetic shear is zero, the magnetic field vector at that point is parallel to the potential field. However, as the shear angle approaches ±90°, the magnetic field has a larger non-potential field component.

The potential field is electric current-free in the corona. Therefore, as long as the normal component of magnetic field on the solar surface remains unchanged, the potential field, $\boldsymbol{B}_{\mathrm{p}}$, corresponds to the minimum energy state. Hagyard et al. (1984) found that an X-class flare occurred in a region where the absolute value of the shear angle was greater than 80–85° along a PIL longer than a certain length. Moreover, they pointed out that the length of a strongly sheared PIL might be an important parameter in determining the onset of large flares.

Many studies have indicated that the flux emergence as well as the flux cancellation in active regions may be related to the occurrence of flares. For instance, Martin et al. (1985) reported that many flares in the active regions of the decay phase are associated with positive and negative flux cancellations. Kusano et al. (2012) showed that flares are more probable to occur when the orientation of the small-scale bipolar field appearing near a PIL is opposite to the large-scale potential field or non-potential field components, based on the comparison of computer simulations and observations. They emphasized the role of such small-scale magnetic fields near a PIL as triggers for large flares.

Schrijver (2007) introduced a parameter R that is defined as the unsigned normal component flux in the vicinity of the point where the gradient of the normal field is strong on a PIL, and showed that the probability of a flare of M1 class or larger within 24 h increases with increasing R. However, even if the R parameter is sufficiently large, the probability of an X-class flare occurring within 24 h is only 10–20%. Furthermore, the probability of the occurrence of a very large flare of X4 class or larger is only 1% even in this case. This suggests that the occurrence of a large flare is not determined by the R parameter alone. Subsequently, Falconer et al. (2008) proposed the length of a PIL with strong magnetic field gradients and a large shear angle as a predictor of solar flare events.

Abramenko (2005) and McAteer et al. (2005) reported the tendency of active regions with complex magnetic field structures to produce active flares, based on multifractal analysis and fractal dimension analysis of the magnetic field. Georgoulis and Rust (2007) proposed a method of first dividing the magnetic field distribution

[1] The potential magnetic field is the electric current-free field given by $\boldsymbol{B}_{\mathrm{p}} = -\nabla\varphi$, where the potential φ, can be obtained by solving the Laplace's equation, $\nabla^2\varphi = 0$, with the boundary condition for the normal component of the magnetic field on the solar surface.

into several elements with a certain criterion and subsequently estimating the susceptibility to flare generation from the structural amount of the magnetic flux connecting them.

In 2010, the Solar Dynamics Observatory (SDO) satellite of the National Aeronautics and Space Administration (NASA) was launched, and its onboard Helioseismic and Magnetic Imager began continuous observations of the vector magnetic fields over the full solar disk (Scherrer et al. 2012). This has enabled extensive studies on the relationship between flare activity and various magnetic field parameters in each active region. In recent years, different parameters of the magnetic field have been calculated for each active region, and there have been many efforts to derive more useful parameters as predictors of the occurrence of flares using machine learning. For example, Bobra and Couvidat (2015) computed 25 different parameters for each active region and used them to predict flares larger than M1 class using support vector machines (Cortes and Vapnik 1995) and showed that they could make relatively good predictions. They also evaluated which parameters contribute more to the flare prediction. They reported that a few parameters such as the total unsigned current helicity and the total magnitude of the Lorentz force were useful in improving the prediction. These parameters can be provided only by vector field observations. Therefore, not only the line-of-sight magnetic field but also the vector magnetic field can help with more accurate flare prediction. However, the results of Bobra and Couvidat (2015) were far from a perfect prediction. Since then, similar studies have been conducted by many researchers, and empirical predictions based on various magnetic field parameters have been tested.

According to Gaizauskas (1989), small-scale emissions can be observed before the onset of a flare by observations in various wavelengths, and these phenomena are called preflares, preflashes, and preheating. For example, Martin (1980) pointed out preflare emissions from observations of the Hα line and the 5303 Å coronal green line. In addition to preflare emissions, various phenomena such as the motion of filament (low-temperature, high-density plasma in the corona supported by the magnetic field) have been reported as candidates for preflare phenomena.

Canfield et al. (1999) pointed out that solar eruptions tend to occur in regions where S (or inverse S)-shaped structures, called sigmoids, are observed by soft X-rays. Since then, many studies have been conducted on the relationship between the sigmoidal structure and explosive phenomena such as flares, filament eruptions, and CMEs (Rust and Kumar 1996; Moore et al. 2001).

Harra et al. (2001) reported based on spectroscopic observations that nonthermal velocities in active regions may increase before the onset of a flare. Zhou et al. (2016), Tan et al. (2016), Chen et al. (2019), and others reported observations of various oscillatory phenomena in active regions before the onset of a flare. Imada et al. (2014) described a large-scale coronal loop expansion before a flare. Wheatland (2004) and others attempted to predict a flare based on the past flare activity in each active region. Recently, Park et al. (2018) presented a probabilistic relationship between the characteristic velocity on the photosphere near a PIL and the occurrence of flares.

Discriminant analysis (DA) is a method for estimating the probability of the occurrence of flares using a certain parameter x as the predictor. Let $N_f(x)$ and $N_q(x)$ be the numbers of cases in which a flare occurred and did not occur in the active regions characterized by x. If there is a parameter space of x in which $N_q < N_f$, x may work as a predictor. Leka and Barnes (2003, 2007) developed a DA to determine the probabilities of the occurrence of flares for various magnetic field parameters of active regions and applied it to flare prediction. In recent years, different methods have been proposed to predict the occurrence of solar flares also by applying various machine learning algorithms (Qahwaji and Colak 2007; Muranushi et al. 2015; Bobra and Couvidat 2015; Nishizuka et al. 2017; Florios et al. 2018).

10.2.4 Physics-Based Prediction

English astronomer Edmund Halley (1656–1742) predicted the return of Halley's comet based on Newtonian mechanics, following which Newtonian mechanics became an indispensable methodology for predicting celestial orbits. Presently, numerical prediction based on the hydrodynamics of the humid atmosphere provides a powerful methodology for predicting meteorological phenomena, which had to rely on empirical methods. This is a result of the progress of computational models and accurate observations. However, even currently, it is still difficult to predict meteorological processes entirely from the first principles of dynamics, and empirical methods called parameterizations are used to calculate complex processes such as cloud dynamics. The history of such predictive research shows that with the development of science and technology, empirical prediction has progressed to predictions based on first principles, with improvement in predictability.

Furthermore, physics-based prediction can provide various opportunities for new discoveries. For example, planet Neptune was predicted before its actual observational discovery based on the irregularity of the orbital motion of Uranus. Therefore, physics-based prediction and the understanding of phenomena stimulate each other and have developed modern science.

This process may also be repeated in the prediction of solar flares and various space weather phenomena. Although physics-based prediction of solar flares has not yet been used for operational forecasting, research for this purpose has been rapidly advancing in recent years. Here, we discuss the current status and future issues of physics-based predictions of solar flares.

To realize physics-based prediction, it is necessary to understand the physical mechanism of the phenomena of solar flares; however, this has not been fully understood yet. Therefore, hypothesis-proven research is necessary. Specifically, we should first formulate a hypothesis and then predict based on it. By verifying the hypothesis by evaluating the prediction, we can mutually develop our understanding and predictive ability.

MHD is a good approximation for large-scale magnetized plasma phenomena such as solar flares. Because solar flares are an explosive release of the energy stored

in the magnetic field, MHD instabilities (Bateman 1978) may be involved in it. Specifically, the magnetic field before a flare is in a stable equilibrium state, which contains free energy, and some instability may be responsible for releasing the free energy during a flare.

The plasma β (the ratio of the thermal pressure to the magnetic pressure) in the solar corona is much smaller than unity; therefore, the stable equilibrium state before the onset of a flare can be approximated by a force-free field (FFF) that satisfies the Lorentz force, $\boldsymbol{J} \times \boldsymbol{B} = 0$, where J is the electric current density and B is the magnetic flux density. Because the displacement current is neglected in MHD, the current density is expressed by $\boldsymbol{J} = \mu_0^{-1} \nabla \times \boldsymbol{B}$, where μ_0 is the permeability of vacuum. Therefore, the FFF is expressed by the force-free equation

$$\nabla \times \boldsymbol{B} = \alpha \boldsymbol{B},$$

where α is a pseud-scalar.[2] From the divergence of the above equation ($\nabla\cdot$), we obtain

$$\boldsymbol{B} \cdot \nabla \alpha = 0.$$

using Maxwell's equations, $\nabla \cdot \boldsymbol{B} = 0$. Therefore, α is constant along the magnetic field lines. If α is spatially uniform in a volume, the force-free equation becomes a linear equation for the magnetic field. Its solution is called a linear FFF. In contrast, when α has different values on different magnetic field lines, the force-free equation becomes nonlinear, and the solution is called a nonlinear FFF (NLFFF).

Several methods have been developed to numerically derive an NLFFF by solving the force-free equation using observed magnetic field vector data on the solar surface as boundary conditions (Schrijver et al. 2006; Inoue 2016). Because an NLFFF may approximate the three-dimensional magnetic field in the solar corona, it can provide information on the stability of active regions. However, different from the corona, the pressure and gravity in the photosphere cannot be neglected. Therefore, the force-free condition is not fully satisfied in the photosphere, and a more realistic magnetic field model is being developed (Miyoshi et al. 2020).

Although there are many different MHD instabilities, kink and torus instabilities have long been considered possible mechanisms of solar flares and solar eruptions. A kink instability (Kruskal & Schwarzschild 1954) is an ideal MHD instability driven by the electric current flowing in a cylindrical and torus plasma. When the electric current flows along the central axis of the cylinder, if the cylindrical plasma undergoes a bending (kink) deformation, the magnetic field created by the current (poloidal field) weakens outside the kink deformation and strengthens inside it. Subsequently, it drives further deformation. This is the basic mechanism for the growth of a kink instability. If there is a magnetic field along the axial current

[2] A scalar is a quantity that is invariant to the rotational transformations of coordinates. However, a pseud-scalar changes its sign with mirror image transformations, unlike a true scalar.

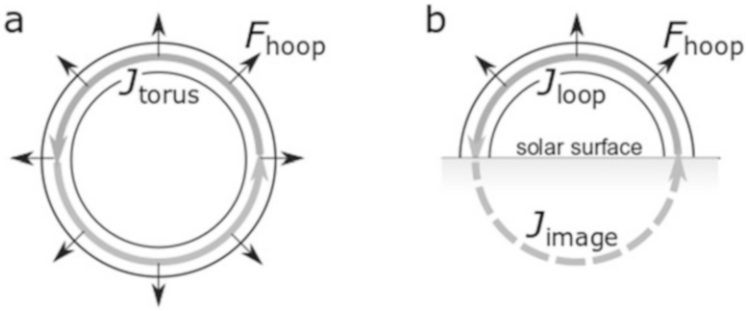

Fig. 10.4 Electric currents (**a**) J_{torus} in the torus and (**b**) J_{loop} in the solar coronal loop. Hoop force F_{hoop} is always acting outward because the boundary condition of the solar surface corresponds to image current J_{image}

(toroidal field), the magnetic tension has the effect of suppressing and stabilizing the deformation. Hence, a kink instability can grow only when the electric current exceeds a certain threshold and the destabilization by the poloidal field dominates the stabilization by the toroidal field. This threshold is expressed by the magnetic twist number, T_w, of the magnetic field lines. The magnetic twist number is the number of rotations of the magnetic field lines around the central axis of the cylinder. If the magnetic field around each magnetic field line can be approximated by a series of small cylindrical magnetic fields, the magnetic twist number can be calculated by a line integral along the magnetic field lines as follows:

$$T_w = \frac{1}{4\pi} \int \alpha dl$$

When the twist number exceeds a certain threshold T_c, the kink instability grows (Berger and Prior 2006).

In the case of a periodic system, i.e., a torus, $T_c = 1$. This corresponds to the Kruskal–Shafranov limit for the electric current that can flow in a torus such as a tokamak. Solar coronal magnetic fields have feet on the solar surface. Because this boundary condition is effective for stabilization, T_c is larger than 1. The threshold, T_c, also depends on the shape of the magnetic field lines, and Hood and Priest (1979) suggested a typical threshold $T_c = 1.65$ in the solar corona.

In contrast to a kink instability, a torus instability is driven by the hoop force, which is always applied outward to the torus current (J_{torus} in Fig. 10.4a). In torus plasma confinement systems such as tokamaks, the hoop force must be canceled by applying an external magnetic field perpendicular to the plane of the torus (toward the plane of the figure), B_{ex}, to maintain the equilibrium state. However, this external

magnetic field rapidly decays with the radius, R, of the torus, and when its decay index

$$n = -\frac{R}{B_{\mathrm{ext}}}\frac{dB_{\mathrm{ext}}(R)}{dR}$$

exceeds the critical value n_c, the torus becomes unstable for outward displacements (Shafranov 1966). This is a torus instability.

Kliem and Török (2006) proposed that this instability causes eruptive dynamics in the solar corona such as eruptive flares and CMEs. Current loops in the solar corona are rooted on the solar surface; however, the normal component of the magnetic field on the solar surface (the photosphere) changes very slightly during the timescale of an eruption. This boundary condition of the magnetic field is equivalent to the image current, J_{image}, flowing below the solar interior (Fig. 10.4b). Hence, the current in the corona is subjected to the same outward hoop force as in a torus current system.

Similar to a torus system, the hoop force is canceled by the external magnetic field orthogonal to the loop current, maintaining the equilibrium state. However, as in a torus system, the current loop is destabilized when the decay index, n, of the external magnetic field exceeds a critical value n_c. Although the value of the critical decay index, n_c, depends on the structure of the current loop and the external magnetic field, it is approximately $n_c \sim 1.5$ in typical cases.

An NLFFF model reconstructed from the observed photospheric magnetic field can help to estimate the T_w and n of active regions. Accordingly, Jing et al. (2018) analyzed the T_w and n of 38 active regions immediately before the onset of M5-class or larger flares. However, these parameters were distributed almost independently of the critical values, and in many regions, they were well below the critical values. A similar study was also conducted by Duan et al. (2019); however, again, no noticeable relationship was found between the critical values of the instabilities and the observed data. However, the above analyses suggest that there is a tendency for eruptive flares to occur when n is large. Therefore, a torus instability may be related to eruptive flares; however, it is difficult to predict a flare onset from these two instability theories.

It is believed that magnetic reconnection plays an important role in solar flares as well as MHD instabilities. In particular, a scenario where an MHD instability and magnetic reconnection synergistically drive each other is promising for explaining the explosive behavior of solar flares (Zhang and Dere 2006; Welsch 2018). In addition, Moore et al. (2001) proposed that a tether-cutting magnetic reconnection between two magnetic loops with large shear at their feet could be responsible for the occurrence of solar flares and solar eruptions, based on a morphological analysis of observational data.

Based on these scenarios, Ishiguro and Kusano (2017) found a new MHD instability that explains the occurrence of flares (Fig. 10.5). They considered that the magnetic reconnection between two magnetic loops with electric currents

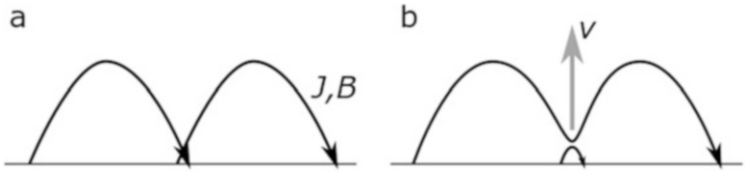

Fig. 10.5 Process of double-arc instability. (**a**) Two magnetic loops with current flowing along magnetic field lines form (**b**) a double arc by magnetic reconnection. If the double arc is destabilized, it starts to move up under the hoop force

flowing along the magnetic field lines forms a double arc (Fig. 10.5b) and numerically analyzed the stability of this double arc. They concluded that when the condition,

$$\kappa = \frac{\int T_w d\phi}{\Phi_{\text{over}}} > \kappa_0,$$

is satisfied, a double arc is destabilized by the hoop force and starts to rise (Fig. 10.5b). The numerator of κ is the magnetic twist flux, which is given by the integration of T_w by the magnetic flux ϕ within the double arc, and the denominator is the magnetic flux overlying the double arc. The threshold, κ_0, depends on the structure of the magnetic field; however, it is generally approximately 0.1. When a double arc becomes unstable and starts to rise, magnetic reconnection is further driven, and the electric current driving the instability increases. With this positive feedback, the instability can explosively grow and may produce a flare.

Based on the theory of double-arc instability, Kusano et al. (2020) developed a prediction scheme (κ-scheme) for large flares. This scheme can evaluate the critical size of the magnetic reconnection, r_c, that can trigger the instability at each point on a PIL and the energy released by the instability, E_r. Because a smaller reconnection is easier to occur, a large flare can be expected to appear at points where r_c is small and E_r is large. Using the κ-scheme, they analyzed all seven active regions (nine events in total) that produced X2-class or larger flares near the center of the solar disk during solar cycle 24 (2008–2019). Consequently, they found that in six of the seven active regions, large flares occurred at the points where E_r exceeded 4×10^{31} erg and r_c was smaller than 1 Mm (see Fig. 10.6). In comparison, all flare-quiet regions did not satisfy this condition even though they had sunspots as large as the flare-active regions.

The above results show that the κ-scheme is a powerful method for not only predicting the occurrence of large flares but also identifying their precise locations. In addition, the results show that the structure of the magnetic twist flux density ($T_w B$) near a PIL is important for determining the onset of large flares. However, the applicability of the κ-scheme for smaller flares remains to be investigated, and efforts are currently underway to implement it for the operational forecasting of solar flares.

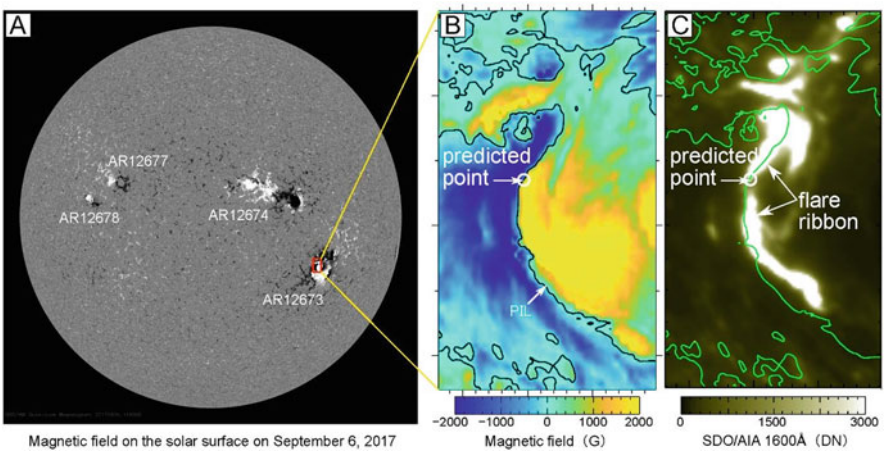

Fig. 10.6 Results of physics-based prediction of X9.3-class flare that occurred in the active region (AR) 12673 on September 6, 2017. (**a**) Distribution of line-of-sight magnetic field components over the full solar disk observed by the SDO satellite. (**b**) Distribution of normal component magnetic fields in AR12673 and predicted point by κ-scheme. (**c**) Predicted point and initial flare ribbon observed by SDO/Atmospheric Imaging Assembly at 1600 Å

Table 10.1 Contingency table

		Observation	
		Positive	Negative
Prediction	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

10.2.5 Evaluation of Prediction

We can classify positive/negative predictions for the occurrence of an event (such as a solar flare) into four groups: successful positive prediction (true positive (TP)), unsuccessful positive prediction (false positive (FP) or false alarm), successful negative prediction (true negative (TN)), and unsuccessful negative prediction (false negative (FN)). Thus, a contingency table (Table 10.1) can be devised, in which all elements—TP, FP, TN, and FN—denote the number of predictions for the corresponding groups.

Various metrics (skill scores) have been proposed to measure predictability (Bobra and Couvidat 2015). For example, the probability of detection (POD) or the hit rate is defined by

$$\mathrm{POD} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}},$$

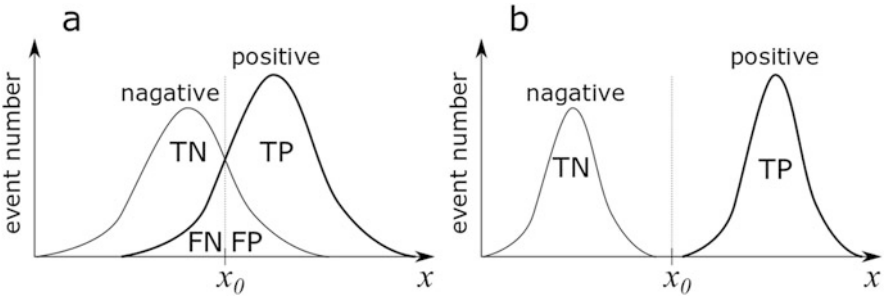

Fig. 10.7 Distribution of event numbers as a function of parameter x. Thick and thin curves represent the number of positive (flaring) events and negative (quiet) events, i.e., $N_f(x)$ and $N_q(x)$, respectively. If we predict an event when $x > x_0$, (**a**) and (**b**) correspond to cases that positive and negative events are partially overlapped and that all events are perfectly discriminated. TP, TN, FP, and FN represent each element in the contingency table

which becomes unity if the predictions are perfect; therefore, FP = FN = 0. The probability of false detection,

$$\mathrm{POFD} = \frac{\mathrm{FP}}{\mathrm{FP} + \mathrm{TN}},$$

and the false alarm ratio,

$$\mathrm{FAR} = \frac{\mathrm{FP}}{\mathrm{TP} + \mathrm{FP}},$$

reflect the ratio of failed predictions to the total negative events and the total positive predictions, respectively, and they become zero in the case of perfect predictions.

Crown (2012) evaluated the flare predictions of the Space Weather Prediction Center of the National Oceanic and Atmospheric Administration from May 1996 to December 2008 for events with an X-class flare within 24 h. The results were TP = 50, FP = 67, FN = 52, and TN = 31,315. The large value of TN relative to the other factors suggests that X-class flares are very rare events. The scores for this prediction were POD = 0.42, POFD = 0.0021, and FAR = 0.57.

Here, let us consider a case in which the prediction is positive when a predictor x is larger than the threshold, x_0 (see Fig. 10.7). In this scenario, the skill scores depend on x_0. Figure 10.8 shows a parameter plane with the POFD on the horizontal axis and the POD on the vertical axis. They range from 0 to 1, and if x_0 is sufficiently small, POFD = 1 and POD = 1, whereas if x_0 is sufficiently large, POFD = 0 and POD = 0. Hence, we can draw a curve connecting these two extreme cases, which is called the receiver operating characteristic curve (ROC). The area of the lower region of the ROC is called the area under the curve (AUC).

In the case in which we can perfectly predict using the predictor, x (Fig. 10.7b), the ROC curve is a rectangular function (dashed line in Fig. 10.8), and the AUC

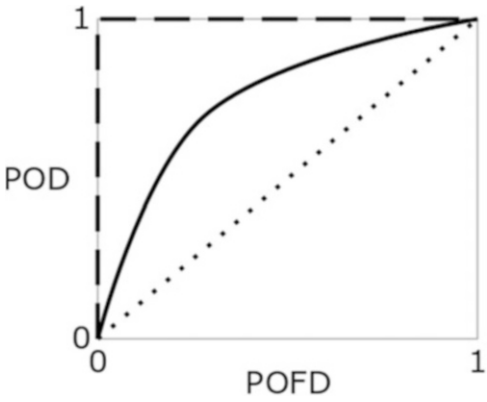

Fig. 10.8 Receiver operating characteristic curve. Dashed line shows a case where ideal prediction is possible, dotted line represents a case where there is no predictive ability, and solid line shows a case where there is some prediction ability. POD and POFD denote the probability of detection and probability of false detection, respectively

becomes unity. When the positive and negative events are partially overlapping as functions of x (Fig. 10.7a), the ROC curve becomes a curve, as shown by the solid line in Fig. 10.8. However, if the distributions of the positive and negative events are completely overlapping, the predictor, x, has no predictive ability. In this case, POD = POFD, and the AUC is 0.5. Therefore, predictability is evaluated by the extent to which the AUC exceeds 0.5 and is close to 1.

Barnes et al. (2016), Leka et al. (2019a), and Leka et al. (2019b) performed cross-evaluations of currently operational flare forecasts. The results showed that there is no outstandingly superior forecasting method. They concluded that the evaluation of the forecasting ability should be multifaceted using multiple scores, because the evaluation results vary depending on the evaluation score. They also showed that the currently operational prediction schemes can rarely accurately predict the first and last large flares in each active region (Park et al. 2020). This suggests that the current flare predictions are biased by the flare activity that had been observed in each active region.

Most of the flare predictions currently conducted by space weather forecasting agencies rely on empirical predictions, and inappropriately, no significant progress has been made in flare prediction capabilities over the past 30 years. However, recently, there have been active efforts to find the best predictor and prediction scheme using the advanced observations of the vector magnetic field and the machine learning technique. Furthermore, studies on prediction based on the theory of MHD instability have been recently initiated.

As in the case of astronomical orbit and meteorological predictions, the author believes that in the future, solar flare prediction will progress from empirical prediction to physical prediction. Simultaneously, physics-based predictions will advance our understanding of the mechanism of solar flares.

10.3 Prediction of Coronal Mass Ejections

10.3.1 Coronal Mass Ejections and Solar Wind

As explained in Chap. 9, a CME is a phenomenon in which the coronal plasma is ejected from the solar corona with magnetic flux. Because the solar wind flows out from the corona in all directions in the interplanetary space, CMEs propagate into the solar wind.

CMEs are phenomena in which the free energy stored in the magnetic loops in the corona is ejected from the Sun by a process of MHD instabilities and magnetic reconnection. In fact, some CMEs observed by coronagraphs have twisted and dense structures in their central regions. Such structures can be understood as structures supported by some of the magnetic field lines of a "twisted flux tube (flux rope)," similar to those seen in prominences. Because twisted magnetic field lines indicate the existence of an internal current, such structures contain magnetic free energy. Hence, the released energy can drive the ejection and propagation of a CME.

Figure 10.9 shows a schematic of the magnetic field lines during the occurrence of a CME. Magnetic reconnection occurs when a magnetic loop in the corona is stretched vertically and the magnetic field lines in opposite directions meet. On the sunward side of the reconnected magnetic field lines, the plasma is heated by the released energy, and a solar flare is generated. On the opposite side, a flux rope is formed and released. By converting the magnetic energy into kinetic energy, the magnetic reconnection replenishes the upward momentum and the magnetic field

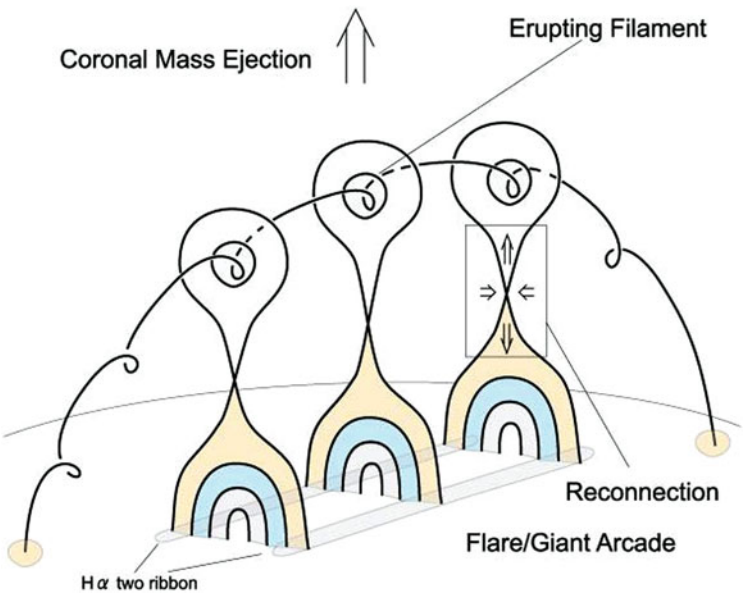

Fig. 10.9 Magnetic field lines during a solar flare associated with a CME (Shiota et al. 2005)

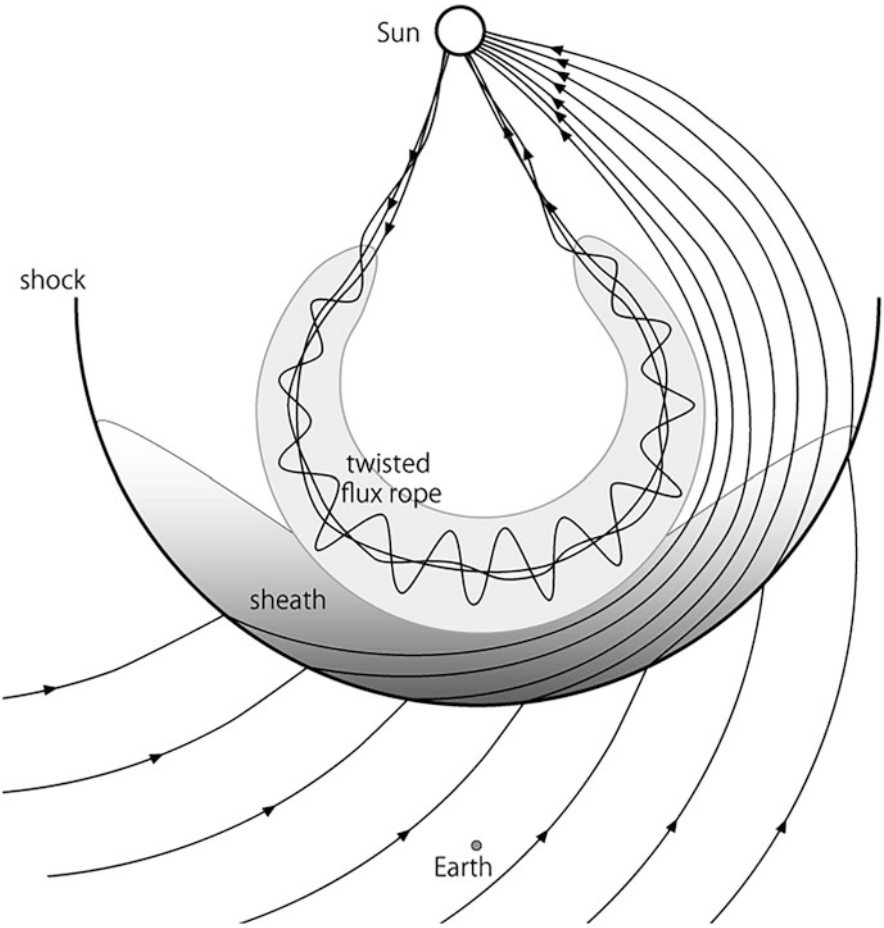

Fig. 10.10 Schematic of a CME propagating in the solar wind. Arrows represent magnetic field lines. Shadows in the background represent relative densities

lines to the flux rope. The increase in the magnetic flux wrapping the flux rope axis and the current in the flux rope as the reconnection proceeds results in the expansion and acceleration of the flux rope. As it moves up, the flux rope expands further because the surrounding magnetic field becomes weaker. If the decay index of the surrounding magnetic field, n, exceeds the threshold for the torus instability, the flux rope expands continuously and is ejected into the interplanetary space while accelerating (Kliem and Török 2006), as described in the previous section.

After a CME is launched into the interplanetary space, it propagates, interacting with the solar wind flowing around it (Fig. 10.10). Such a CME propagating in the interplanetary space is called an interplanetary CME (ICME). It expands and propagates in the solar wind of the interplanetary space while maintaining a certain solid angle in a certain direction. If the speed of a CME is higher than that of the solar

wind flowing in front of it, a CME pushes and compresses the solar wind from behind, forming the sheath region. If the speed difference between a CME and solar wind exceeds the magnetic sound speed, it forms a shock wave. The high velocity of the propagation of the front in the solar wind implies a large Mach number of the shock wave and a large compression ratio of the plasma. The Mach number of a shock wave is determined by the state of the background solar wind and the propagation speed difference. Therefore, if the background solar wind is a low-speed solar wind with a low temperature and a high density, the magnetic sound speed in the fast mode reduces, and thus, a strong shock wave may be formed even if the speed difference is small. On the other hand, if the speed of a CME is lower than that of the surrounding solar wind, the CME is accelerated by the solar wind up to the same speed as the solar wind, and they propagate together.

If the solar wind in the direction of the CME propagation has a nonuniform structure with different velocities and plasma states, each part of the CME propagating in the structure interacts with each background solar wind in different propagation directions. Consequently, as an example, the propagation speed becomes relatively high in the region where a high-speed wind is flowing. In contrast, the propagation velocity becomes relatively low in the region where low-speed solar wind is flowing, and the CME structure is distorted (Odstrcil and Pizzo 1999). When a discontinuity in the solar wind speed exists in the longitudinal direction, far away from the Sun, the interface is angled from the radial direction by the Parker spiral. When a CME encounters this oblique discontinuity, the propagation velocity of the CME may appear to be shifted in the longitudinal direction. This phenomenon is called a CME deflection (Wang et al. 2004).

Solar active regions with high activity frequently produce a series of successive CMEs with a short time interval. In such cases, when the propagation speed of the preceding CME is relatively lower than that of the following CME, the latter meets the preceding one during the propagation, and they interact (Lugaz and Gombosi 2005).

Thus, it is important for a CME prediction to reproduce the structures of both the emitted CME and background solar wind in the direction of the propagating CME accurately to the maximum extent, because their interaction plays an important role in the propagation process of a CME.

10.3.2 ICMEs and Space Weather Disturbances

Here, we describe the features of an ICME observed in situ when it arrives at the Earth. As shown in Fig. 10.10, the spatial scale of an ICME expands to the order of an astronomical unit by the time it reaches the interplanetary space near the Earth. Therefore, the observed features differ depending on which part of the ICME passes through the position of the Earth.

First, consider the case where a high-speed CME is ejected in front of the Earth (near the disk center of the Sun) and the Earth passes through the center of the ICME.

The first structure that arrives at the Earth is the shock wave in front of the ICME. If energetic particles (SEPs) are accelerated by the shock wave, they arrive at the Earth earlier than the shock wave. Because of the involved rapid increase in density, temperature, velocity, and magnetic field strength of the plasma, the shock wave causes a large disturbance when it passes through the magnetosphere of the Earth.

In the sheath region behind the shock wave, the interplanetary magnetic field is compressed and folded with the background solar wind. If the shock wave is strong in the sheath region, the magnetic field vectors tangential to the shock wave surface may randomly switch over a short period, which is called a planar magnetic structure (PMS, Nakagawa et al. 1989). When not only the PMS but also the regional background solar wind is compressed, magnetic fields of the same direction with high intensity may be formed in the sheath region. Magnetospheric disturbances can result when the magnetic field in such sheaths has a southward component.

After the passage of the sheath region, the magnetic field strength increases, whereas the plasma temperature and density decrease, and a structure in which the direction of the magnetic field vector appears to rotate gradually may pass. This structure is called a magnetic cloud (MC, Burlaga et al. 1981) and can be understood as an observation of the passage of a part of the twisted flux rope in the center of the CME. When this part of the magnetic field has a southward component, it becomes a factor in magnetospheric disturbances. Particularly, when the MC (flux rope) has sufficient magnetic flux, the southward magnetic field in it continues for a long time, which increases the possibility of causing a large magnetospheric disturbance.

When the propagation direction of the CME is further away from the Sun–Earth line, the region passing through the position of the Earth as seen from the CME in Fig. 10.10 moves away from the center. When the propagation direction of the CME is sufficiently large with the Sun–Earth line, the flux rope (MC) may not arrive at the Earth; instead only the shock wave or its sheath region arrives. However, when the propagation speed of the CME is lower than that of the solar wind, no shock wave is formed, and only the flux rope (MC) may arrive.

10.3.3 Solar Wind and CME Arrival Prediction Models

To predict the impacts of an approaching CME on the Earth, it is important to accurately reproduce its internal structure (shock wave, sheath, and MC) and subsequently accurately reproduce which region of the CME the Earth will pass through.

One method for the prediction to incorporate these physical factors is a global MHD simulation of the inner heliosphere, including the solar–terrestrial neighborhood. An MHD model commonly used in space weather forecasting is ENLIL (Odstrcil 2003). In this model, the inner solar corona up to 0.1 AU from the solar surface, which is the solar wind acceleration region, and the high-latitude region are excluded from the computational domain. Because these regions require a high computational cost, they are excluded to accelerate the computation and predict

CMEs with a sufficient lead time. The background solar wind is simulated using the observed magnetic field of the solar photosphere, a coronal magnetic field model, and an empirical solar wind velocity model, and the accelerated supersonic solar wind distribution is provided from the inner boundary of the computational domain. Although there are many variants of the magnetic field model and the solar wind velocity model, the most orthodox version is based on the Wang–Sheeley–Arge (WSA) model (Arge et al. 2004) and is called WSA-ENLIL.

The WSA-ENLIL model is typically used for solar wind prediction and does not include CME information. When a CME occurs, simulations with additional CME information are performed. The information on the CME (solid angle extent, direction, and speed of propagation) is inferred from coronagraphs with the cone model (Zhao et al. 2002; Xie et al. 2004). The CME information is input to the inner boundary as a fluid pulse (a mass of plasma with increased velocity and pressure for a certain period) to calculate the propagation process of the CME in the solar wind. As mentioned above, a CME is a clump of plasma with a flux rope in the center; however, the input in the cone model is a clump of plasma without a flux rope inside. Therefore, it is unsuitable for the prediction of the magnetic field observed as an MC; in contrast, it can be used for the prediction of the timing of the arrival of the shock wave and the impact of the sheath magnetic field of a CME, because the shock wave can form a sheath that sweeps the background magnetic field.

In Japan, a space-weather usable system anchoring numerical operations and observations (SUSANOO, Shiota et al. 2014), a numerical simulation developed by Nagoya University and the NICT, which is basically similar to ENLIL, has been in operation. The main difference between the SUSANOO and ENLIL simulations is that the former uses a Yin–Yang grid to calculate the propagation in all directions. Another major difference is that in the SUSANOO framework, CMEs with internal flux ropes can propagate into the solar wind (SUSANOO-CME, Shiota and Kataoka 2016).

A CME input model of the SUSANOO-CME simulations is illustrated in Fig. 10.11, in which a CME is assumed to be expanding and propagating in a self-similar manner while maintaining a solid angle in the propagation direction (Gibson and Low 1998). In this case, the CME is regarded as a fixed state in space when it is on the coordinates of a self-similar system. Figure 10.11 shows a conceptual figure of the CME in the coordinate system of the self-similar system. The state of this fixed CME (e.g., the CME shape, orientation, magnetic field strength) and the relationship between the self-similar system and the real space (e.g., the propagation direction, speed, and onset time) are provided using ten different parameters. From the viewpoint of the coordinates of the self-similar system, the sphere at the inner boundary of the computational domain corresponds to a sphere of radius R'_{cut} that moves inward toward the origin in Fig. 10.11. As shown in Fig. 10.12, using this relation, we can reproduce the propagation process of a CME with a flux rope in the computational domain by providing the cross section of the CME model to the inner boundary of the simulation.

Using this simulation technique for space weather forecast operations is ongoing at the NICT. However, there are still many issues to be solved to employ it for

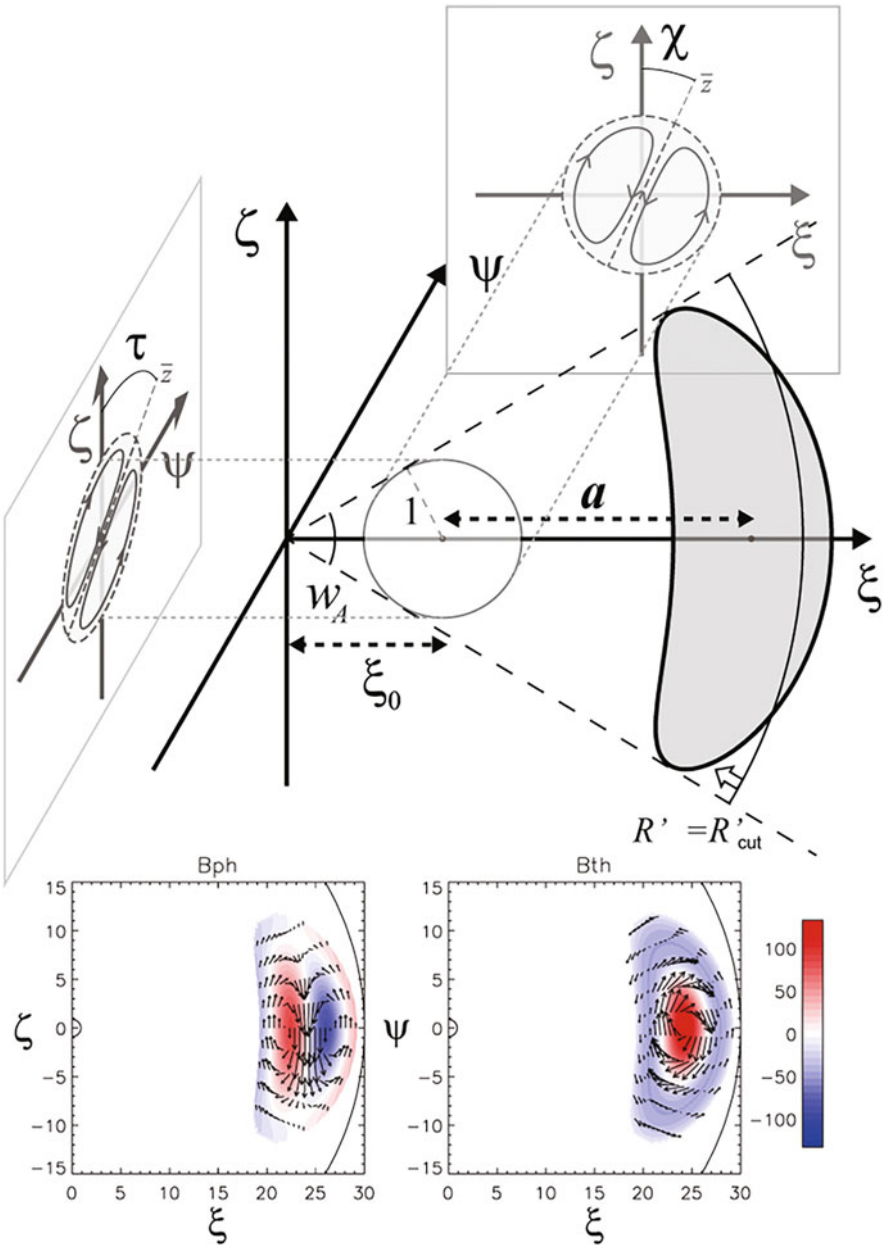

Fig. 10.11 Conceptual diagram of the CME model of SUSANOO-CME (Shiota and Kataoka 2016). CME is assumed to be a spheromak magnetic field contained in a spherical shell, deformed into a dynamically collapsed region (hatch region). Lower panels show the magnetic field when the axis of the spheromak field points south (meridional and equatorial cross sections)

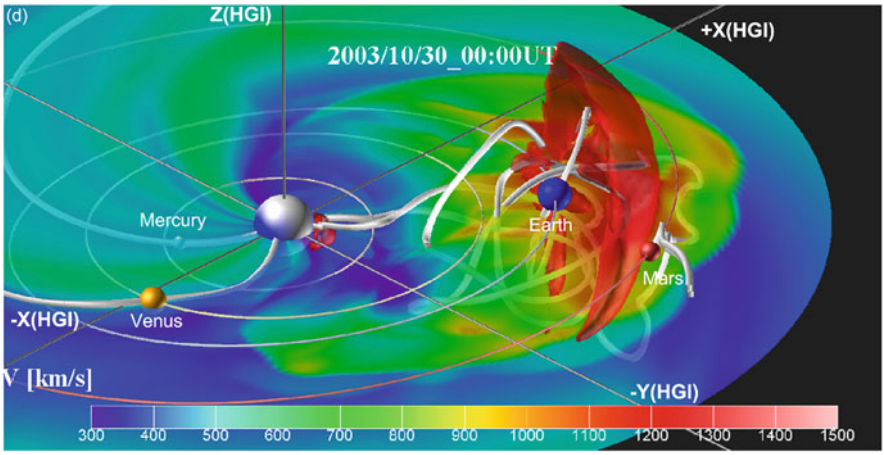

Fig. 10.12 Three-dimensional view of SUSANOO-CME results. Plasma velocities (colors) and magnetic field lines (tubes) during the passage of the 2003 Halloween event CME by SUSANOO-CME through the position of Earth (Shiota and Kataoka 2016)

predicting the impact of actual CME arrivals. First, as mentioned above, there are ten parameters of CMEs, for which a small amount of information can be obtained from observational data (and the parameters may change during propagation). Hence, quantitative prediction of the magnetic field structure of a CME arriving at the Earth is difficult. Another issue that needs to be resolved is the accurate reproduction of the background solar wind, which can significantly affect the propagation process of a CME. To reduce this uncertainty, a system for ensemble prediction has been developed. Furthermore, a method of data assimilation for ensemble prediction simulations with interplanetary scintillation observations of propagating ICMEs at the Institute for Space-Earth Environmental Research, Nagoya University, is also being developed (Iwai et al. 2019, 2021).

10.4 Prediction of Solar Energetic Particles

10.4.1 Introduction

Solar activity produces charged particles (electrons, protons, and heavy ions) with energies ranging from a few kiloelectron volts up to several gigaelectron volts, which are ejected into the interplanetary space, and they are called SEPs. The existence of SEPs was first confirmed using ground-based ion chambers and neutron monitors during the solar activity in February and March 1942. The corresponding event was a ground-level enhancement, in which secondary cosmic rays are generated by the impacts of SEPs on the atmosphere and are observed on the ground. At that time, all SEPs were considered to be produced by flares because flares were the

Table 10.2 Characteristics of two types of SEPs. (Reprinted from Desai and Giacalone (2016) licensed under Creative Commons Attribution 4.0 International License)

	Impulsive SEP	Gradual SEP
Electron/ion	$\sim 10^2 - 10^4$	$\sim 50 - 100$
$^3He/^4He$	~ 1	$\sim 4 \times 10^{-4}$
Fe/O	~ 1	~ 0.1
Ionization state of Fe	~ 20	~ 14
Duration	$< 1 - 20$ h	$< 1 - 3$ days
Radio type	Type III	Type II

most active events on the Sun. However, currently, SEPs are classified into two types: impulsive SEPs produced by flares and gradual SEPs produced by CMEs (in addition, there are cases in which both types of SEPs are produced). The differences between the two types are summarized in Table 10.2 and shown in Fig. 10.13. Impulsive SEPs, which originate from flares, are rich in electrons and heavy ions and have a high ionization level, reflecting the fact that they are generated in the active regions of the lower solar atmosphere. Owing to their short timescales, impulsive SEPs are considered to be produced in flare regions and immediately ejected, ballistically traveling. Because charged particles tend to move parallel to the magnetic field lines, instead of perpendicular to them, the associated active region and the Earth are considered to be connected by the same magnetic field lines when impulsive SEPs are observed. Indeed, the active regions producing impulsive SEPs are distributed on the west side of the Sun (Fig. 10.13, lower right panel).

In contrast, gradual SEPs are considered to be produced by the shock waves that are generated when CMEs propagate from the solar atmosphere to the interplanetary space, and they are characterized by their higher proton abundance and lower ionization level than impulsive SEPs. The main difference from impulsive SEPs is the wide generation region of gradual SEPs. Because a shock can propagate perpendicular to the magnetic field lines, it has a wide generation area in the longitudinal direction, as shown in the upper left panel of Fig. 10.13. Therefore, gradual SEPs originating from CMEs generated at a wide range of longitudes are observed in the orbit of the Earth (Fig. 10.13, lower left panel). In addition, SEPs originating from CMEs on the farside of the Sun have been observed.

The effects of SEPs on the space environment include radio communication problems, malfunction and degradation of equipment onboard aircrafts and satellites, and exposure of astronauts during extravehicular activities. Because these main exposures are protons above approximately 10 MeV, it is necessary in the framework of space weather forecast to identify the generation factors of gradual SEPs, particularly those associated with CME shocks, and to quantitatively predict their flight into the orbit of the Earth. These will become increasingly important in the future as humans expand their activity area beyond the magnetosphere of the Earth. Here, we briefly introduce the progress of the research on the prediction of gradual SEPs based on plasma physics.

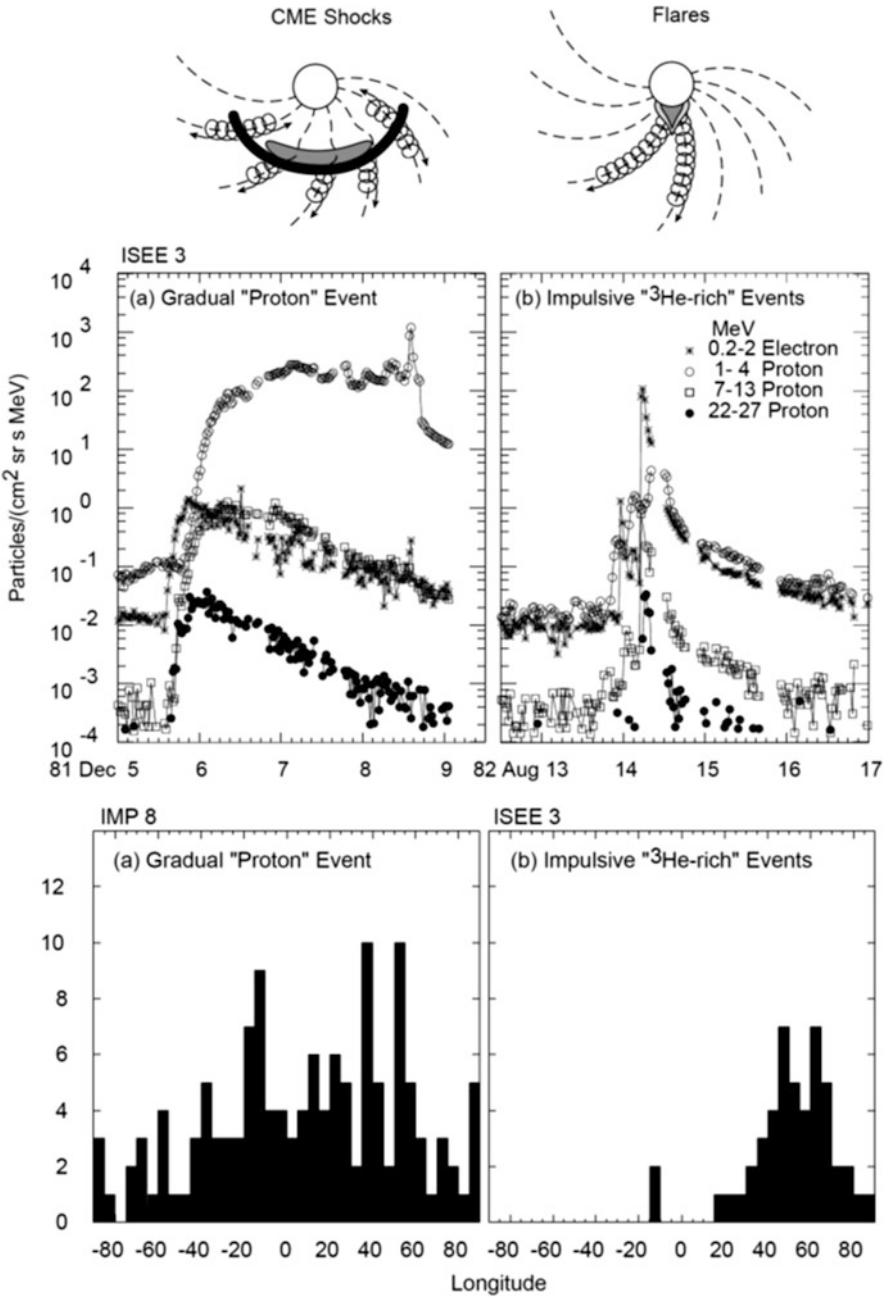

Fig. 10.13 Differences between gradual SEPs (left) and impulsive SEPs (right). (Top) Schematic, (middle) example time profile, and (bottom) distribution of number of SEP events versus solar longitude of the originating flare. Data were acquired by the ISEE-3 satellite between 1979 and 1984. (Image reproduced with permission from Reames (1999), copyright by Springer)

10.4.2 Prediction Based on Transport Model

Generally, SEPs are treated as test particles that do not affect the solar wind structure because their energy density is sufficiently smaller than that of the background solar wind. First, we consider a case in which the solar wind is unmagnetized and flowing radially from the Sun. The Parker equation is adopted to describe a SEP, which is known as a transport equation for galactic cosmic rays in the heliosphere,

$$\frac{\partial f}{\partial t} + V_{sw}\frac{\partial f}{\partial r} - \frac{p}{3}(\nabla \cdot V_{sw})\frac{\partial f}{\partial p} = \frac{\partial}{\partial r}\left(D\frac{\partial f}{\partial r}\right) + S$$

where t, r, and p are the time, distance from the Sun, and momentum of the SEP, respectively; $f(t, r, p)$ is the phase-space distribution function, $V_{sw} = V_{sw}e_r$ is the solar wind velocity, D is the spatial diffusion coefficient, and S is the source term. The Parker equation is an advection–diffusion equation consisting of the transport by the solar wind (second term on the left-hand side), cooling due to the expansion of the solar wind (the third term on the left-hand side), and spatial diffusion (first term on the right-hand side). The origin of the spatial diffusion is the magnetic field disturbance; a high intensity of the disturbance implies a small diffusion coefficient (i.e., it is confined in the disturbed region).

In the following, we consider the effect of the magnetic field. The magnetic field in the steady solar wind has a spiral structure (the Parker spiral), and SEPs have different motions in directions parallel and perpendicular to the magnetic field. Consequently, the velocity distribution becomes anisotropic. In addition, the motion of the particles is affected by the convective electric field in the solar wind, $E = -V_{sw} \times B$. However, by transforming the coordinates to the frame rotating with the Sun (corotating frame), the solar wind velocity in this frame, $V_{sw}^c = V_{sw}e_r - \Omega r e_\phi$ (where Ω is the angular velocity of the solar rotation), becomes parallel to the magnetic field lines, and the electric field disappears (the corotating frame is adopted hereafter). Furthermore, if we assume that SEPs are sufficiently scattered by the disturbance, the velocity distribution becomes isotropic and the Parker equation becomes reasonable. As an example, this can be used, e.g., to track the SEPs downstream of a sufficiently disturbed shock.

However, the SEPs that first arrive in the orbit of the Earth are not necessarily isotropic. Because the SEPs leak from the acceleration region to arrive ballistically along the magnetic field lines, the velocity distribution is enhanced in the direction parallel to the magnetic field lines due to the magnetic mirror effect. The Parker equation is insufficient from the viewpoint of SEP "prediction." Therefore, we adopt a model that considers the magnetic field effect and the associated velocity anisotropy, called the focused transport equation (Skilling 1975; le Roux et al. 2007; le Roux and Webb 2009). Here, we use it along the Parker spiral in the steady solar wind (Ruffolo 1995) as follows:

$$\frac{\partial f}{\partial t}+\frac{\partial}{\partial z}\left(\mu v+V_{sw}^{c}\right)f+\frac{\partial}{\partial \mu}\left[-\frac{v+\mu V_{sw}^{c}}{2}\frac{\partial \ln B}{\partial z}-\mu\frac{\partial V_{sw}^{c}}{\partial z}\right]\left(1-\mu^{2}\right)f+\frac{\partial}{\partial p}$$
$$\times\left[\frac{\left(1-\mu^{2}\right)V_{sw}^{c}}{2}\frac{\partial \ln B}{\partial z}-\mu^{2}\frac{\partial V_{sw}^{c}}{\partial z}\right]pf=\frac{\partial}{\partial \mu}D_{\mu\mu}\frac{\partial f}{\partial \mu}+S$$

where v and μ are the velocity and the cosine of the pitch angle (the angle between the velocity vector and the magnetic field vector) of the SEP measured in a frame moving with the solar wind (with velocity V_{sw}^{c}), respectively. z represents the spatial coordinate along the magnetic field line, B is the magnetic field strength, and $D_{\mu\mu}$ is the pitch angle diffusion coefficient. The second term of the left-hand side includes the free stream of the SEP itself. The pitch angle diffusion coefficient can be given from the following quasi-linear theory (Jokipii 1971) as

$$D_{\mu\mu}=A|\mu|^{q-1}\left(1-\mu^{2}\right),(1<q<2)$$

where coefficient A is given as a function of the mean free path in the direction parallel to the magnetic field lines. Based on the relationship between the pitch angle diffusion coefficient and the spatial diffusion coefficient,

$$D=\frac{\lambda_{\parallel}v}{3}=\frac{v^{2}}{4}\int_{0}^{1}\frac{\left(1-\mu^{2}\right)^{2}}{D_{\mu\mu}}d\mu\rightarrow A=\frac{3v}{2(4-q)(2-q)\lambda_{\parallel}}.$$

This equation tracks the SEP orbit in a given stationary solar wind and does not specify its origin, which is represented by the source term, S. For example, an impulsive SEP can be simulated by giving a finite source term only at (t, z) = (0, 0). Lario et al. used this approach to model a gradual SEP (Lario et al. 1998). They first performed MHD simulations of a CME shock propagating in a steady solar wind. To simulate the SEP generation by the CME shock, a finite source term is provided at the position of the shock, and the focused transport equation is solved from the shock front to the Earth along the magnetic field lines.

The focused transport equation is frequently employed in SEP prediction as a framework for space weather forecast. Lario et al. (1998) assumed the source term to be a function of the physical quantity of the shock, and they obtained the parameter dependence of the source term empirically by comparing with observed data. The results were compiled as a database in the solar particle engineering code (SOLPENCO) (Aran et al. 2006). In the SOLPENCO, time series of SEP fluxes in the orbit of the Earth is prepared by conducting MHD and focused transport simulations with various parameters (e.g., longitude and shock velocity). When users input parameters, they obtain the prediction data of the SEP flux that precedes the shock arrival at the orbit of the Earth by interpolating the simulation data with the nearest parameters. A model similar to the Lario et al. approach is called the SEP model (SEPMOD) (Luhmann et al. 2017). In the SEPMOD, an SEP flux in the orbit of the Earth is estimated by simulating the focused transport equation on an MHD

simulation of a CME shock using the ENLIL code of NASA. Large-scale and long-time calculations are possible, e.g., to deal with SEP fluxes in the orbit of Mars and multiple CME shocks.

The above approaches deal with CME shocks propagating in the interplanetary space as the sources of SEPs; however, there are also efforts to predict the subsequent transport of the SEPs assuming that they are produced in the vicinity of the Sun (flares and coronal shocks) within a short period of time. Predicting SEPs originating near the Sun is of critical importance because they are expected to have higher energy than those from the interplanetary space. The Earth–Moon–Mars radiation environment module provides the energetic particle radiation environment module (EPREM) (Schwadron et al. 2010). The EPREM estimates SEP fluxes in lunar and Martian orbits using the SEPs observed in the orbit of the Earth as boundary conditions for different longitude points at 1 AU. This is in anticipation of the upcoming manned missions to the Moon and Mars. It should be noted that this method is based on the assumption that the SEP production rate in the magnetic field lines leading to the Moon and Mars orbits is equal to that in the magnetic field lines leading to the orbit of the Earth. The solar particle radiation space weather model (Marsh et al. 2015) traces the transport of SEPs produced in the vicinity of the Sun into the interplanetary space, with a particular focus on their motions across the magnetic field lines (drift and cross-field diffusion). Although cross-field motions have been considered in the transport of galactic cosmic rays, they are neglected in the transport of lower energy SEPs (hence, the focused transport equation along the Parker spiral has been adopted). However, simultaneous multisatellite observations have suggested the possibility that SEPs are distributed in a direction perpendicular to the magnetic field lines, unlike what was previously considered.

10.4.3 Prediction Based on Acceleration and Transport Model

An SEP prediction model mainly tracks the transport of energetic particles produced, and does not necessarily clarify their origin. Self-consistent treatment of the particle acceleration mechanism requires first-principles calculations, which are impractical from the viewpoint of the computational cost to apply them to the real scale of the heliosphere. However, the Parker equation and the focused transport equation can also treat particle acceleration within an adiabatic approximation.

The momentum advection term in the Parker equation represents the cooling associated with the expansion of the solar wind; however, conversely, if it is compressed (if $\nabla \cdot V_{sw}$ is negative), the momentum of the particles will increase (the corresponding term in the focused transport equation is $\partial V_{sw}^{c}/\partial z$). This is a condition satisfied by a shock wave, and if a particle continues to be trapped by the shock front, it will be continuously accelerated. This is called the diffusive shock acceleration (Fermi acceleration) theory (Blandford and Ostriker 1978). Here, we

consider a particle that can cross a shock front in a shock rest frame, where the left and right sides are upstream and downstream, respectively. We assume that a magnetic disturbance develops before and after the shock and the particle is reflected. Because reflection occurs elastically in the rest frame of the fluid, the momentum of the particle increases before and after the collision in the case of a head-on collision at the upstream, whereas it decreases in the case of a follow-up collision at the downstream. Consequently, the momentum change of the particle after one round trip of the shock is $4m(V_{\rm up} - V_{\rm down})/3 > 0$ ($V_{\rm up/down}$ are the velocities upstream and downstream of the shock, respectively, and factor 4/3 is the weight of the angular distribution of the particle.). The steady-state solution of the Parker equation for a one-dimensional shock is $p^2 f \propto p^{-\Gamma}$, where exponent Γ is given as a function of the compression ratio of the shock, r, as $\Gamma = (r + 2)/(r - 1)$. The result is independent of the shock velocity and the diffusion coefficient in the limit of a strong shock, i.e., $r \to 4$ and $\Gamma \to 2$.

The diffusive shock acceleration theory is the most promising theory to explain the origin of cosmic rays because it provides the remarkable result that a power spectrum with a universal exponent can be constructed. However, there is an unsolved problem in this theory. In the diffusive shock acceleration theory, it is necessary for particles to travel back and forth through a shock, whereas thermal protons cannot do so because (fluid velocity) $>$ (thermal velocity). To reflect particles, a magnetic disturbance (an Alfvén wave) is required not only downstream but also upstream of the shock, and particles with much higher velocities than the Alfvén velocity are reflected. These are known as injection problems of particle acceleration at shocks, where the particles to be statistically accelerated by the shock must have been pre-accelerated to some extent.

The above theory is for shocks in which the magnetic field and the normal vector of the shock surface are parallel. However, in general, they are oblique, and there is a convective electric field in the oblique shock. Because the magnetic field is amplified through a shock, particles are trapped in the shock front by the magnetic mirror effect and become accelerated by drifting in the direction of the convective electric field. This is called shock drift acceleration, and Fermi and drift acceleration coexist in general oblique shocks. However, the same argument can be developed for oblique shocks by transforming the coordinates into the de Hoffmann–Teller frame, in which a convective electric field disappears. Although the acceleration time in a plane oblique shock is shorter than that in a parallel shock, the steady-state spectrum remains unchanged (Drury 1983).

The original focused transport equation includes Fermi and drift acceleration. Therefore, by solving it from the CME shock front, it would be possible to simultaneously track the acceleration and transport of the SEPs within the adiabatic approximation. The focused transport equation (Ruffolo 1995) can be extended as follows:

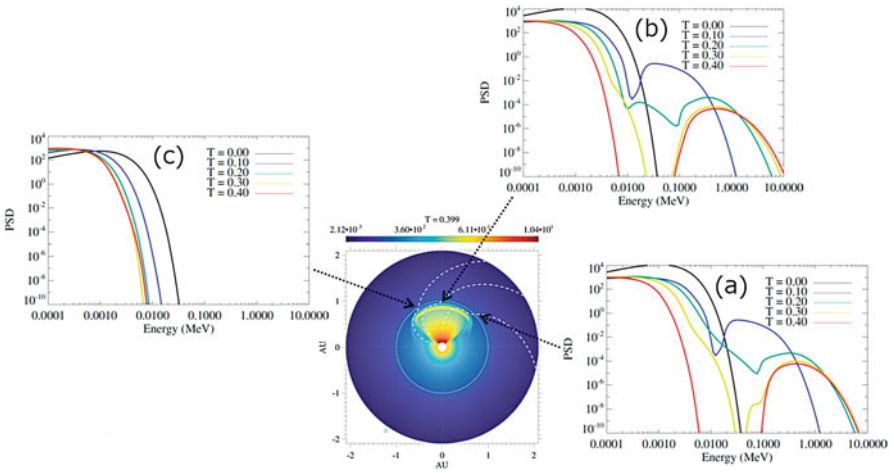

Fig. 10.14 MHD and focused transport simulations. (**a–c**) Time evolution of the energy spectrum at the shock front for different magnetic field lines. Time T is normalized by Alfvén time for 1 AU

$$
\begin{aligned}
&\frac{\partial f}{\partial t} + \frac{\partial}{\partial z}(\mu v + V_{\parallel}) f + \frac{\partial}{\partial \mu} \\
&\quad \times \left[-\frac{\mu}{2}\frac{DlnB}{Dt} + \frac{V_E}{v} \cdot \left(\frac{D}{Dt} + \mu v \frac{\partial}{\partial z} \right) b - \frac{1}{v}\left(\frac{D}{Dt} + \mu v \frac{\partial}{\partial z} \right) V_{\parallel} - \frac{v}{2}\frac{\partial\, ln\, B}{\partial z} \right] \\
&\quad \times (1-\mu^2) f + \frac{\partial}{\partial p} \\
&\quad \times \left[\frac{1-\mu^2}{2}\frac{DlnB}{Dt} + \frac{\mu V_E}{v} \cdot \left(\frac{D}{Dt} + \mu v \frac{\partial}{\partial z} \right) b - \frac{\mu}{v}\left(\frac{D}{Dt} + \mu v \frac{\partial}{\partial z} \right) V_{\parallel} \right] pf \\
&\quad = \frac{\partial}{\partial \mu} D_{\mu\mu} \frac{\partial f}{\partial \mu}
\end{aligned}
$$

where $V = V_{\parallel} b + V_E$ is the fluid velocity, $V_E = E \times B/B^2$ is the electric field drift, $b = B/|B|$ and $D/Dt = \partial/\partial t + V \cdot \nabla$ is the Lagrange derivative. The first three terms in the brackets of the pitch angle advection term (the third term on the left-hand side) and the momentum advection term (the fourth term on the left-hand side) are the gradient-B drift acceleration, curvature drift acceleration, and Fermi acceleration terms, respectively. The terms that were not included in the focused transport equation in (Ruffolo 1995) are zero in the steady solar wind, have finite values only at the shock front, and contribute to particle acceleration.

An example of the numerical simulation of this model is presented. We first conduct an MHD simulation of plasma ejection into a two-dimensional stationary solar wind to generate a shock and then solve the focused transport equation along a specific magnetic field line. Figure 10.14 plots the time evolution of the energy

spectrum at a shock front for three magnetic field lines, which shows particle acceleration in the magnetic field lines where the shock nose passes (panels (a) and (b)), although the source term is not provided in the focused transport equation. Depending on the angle between the shock normal and the Parker spiral, the Fermi acceleration dominates near the Sun, whereas the gradient-B drift acceleration dominates away from the Sun. An attempt to implement this model into the space weather forecast simulator, SUSANOO-CME (Shiota and Kataoka 2016), is in progress.

10.4.4 Summary

A physical model for the prediction of SEPs is presented in which the focused transport equation approach is useful owing to the trade-off between the level of approximation and the computational cost. Although the main purpose of the model is to track the transport from the acceleration region, an attempt to extend it to study SEP acceleration and transport by CME shocks is in progress because the model can describe Fermi and drift acceleration at shocks. However, the present model can only deal with interplanetary shocks. High-energy SEPs with energies exceeding 100 MeV, which may have serious impacts on the space environment, are considered to be generated by flares and coronal shocks in the vicinity of the Sun, and thus, their modeling is also required.

In diffusive shock acceleration, it is necessary that the seed particles are pre-accelerated and that there is a disturbance to reflect particles. For example, in an event where CMEs occur repeatedly, there are observation results that suggest SEPs are probable to be generated by a succeeding CME shock. Moreover, if it is considered that the first CME generates a super-thermal component that becomes a seed particle and disturbs the magnetic field, the results are consistent with the diffusive shock acceleration theory. Focused transport simulations depend strongly on the estimates of the temperature and the mean free path of the seed particles, which may be clarified by comparing the parameter survey of the simulation with observations.

With a few exceptions, until now, SEP observations have been mainly single-point observations in the orbit of the Earth. However, recent simultaneous multisatellite observations have provided new insights. For example, simultaneous observations at different longitudes by the STEREO satellites have shown that SEPs are transported in a direction that crosses the magnetic field lines more than expected (Wiedenbeck et al. 2013). Therefore, it may be necessary to modify the scenario. The future direction is conducting simultaneous observations at different radii. The Parker Solar Probe and the Solar Orbiter, which are scheduled to be fully operational in the 2020s, will observe the solar wind plasma, electromagnetic field, and energetic particles from approximately 10 solar radii to Earth orbit. This will allow separating the acceleration and transport processes of SEPs. It will also enable answering remaining questions such as the nature of seed particles, direct observation of

acceleration sites such as shock waves and magnetic field disturbances, and multidimensional transport processes, which are expected to significantly contribute to heliospheric physics. Thus, it will be necessary to develop a complex SEP simulation model as a complementary tool to interpret unprecedented observational data.

References

Abramenko, V.I.: Multifractal analysis of solar magnetograms. Solar Phys. **228**, 29 (2005)
Aran, A., Sanahuja, B., Lario, D.: SOLPENCO: a solar particle engineering code. Adv. Space Res. **37**, 1240 (2006)
Arge, C.N., Luhmann, J.G., Odstrcil, D., Schrijver, C.J., Li, Y.: Stream structure and coronal sources of the solar wind during the may 12th, 1997 CME. J. Atmos. Solar Terr. Phys. **66**, 1295 (2004)
Aulanier, G., Demoulin, P., Schrijver, C.J., Janvier, M., Pariat, E., Schmieder, B.: The standard flare model in three dimensions. II. Upper limit on solar flare energy. Astron. Astrophys. **549**, A66 (2013)
Barnes, G., et al.: A comparison of flare forecasting methods. I. Results from the 'all-clear' workshop. Astrophys. J. **829**, 89 (2016)
Bateman, G.: MHD instabilities, Cambridge. MIT Press, Mass (1978)
Berger, M.A., Prior, C.: The writhe of open and closed curves. J. Phys. A Math. Gen. **39**, 8321 (2006)
Blandford, R.D., Ostriker, J.P.: Particle acceleration by astrophysical shocks. Astrophys. J. Lett. **221**, 29 (1978)
Bobra, M.G., Couvidat, S.: Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm. Astrophys. J. **798**, 135 (2015)
Bornmann, P.L., Shaw, D.: Flare rates and the Mcintosh active region classifications. Sol. Phys. **150**, 127–146 (1994)
Burlaga, L., Sittler, E., Mariani, F., Schwenn, R.: Magnetic loop behind an interplanetary shock: voyager, Helios, and IMP 8 observations. J. Geophys. Res. **86**, 6673 (1981)
Canfield, R.C., Hudson, H.S., McKenzie, D.E.: Sigmoidal morphology and eruptive solar activity. Geophys. Res. Lett. **26**, 627–630 (1999)
Chen, X., et al.: Quasi-periodic pulsations before and during a solar flare in AR 12242. Astrophys. J. **878**, 78 (2019)
Cortes, C., Vapnik, V.: Support-vector networks. Mach. Learn. **20**, 273–297 (1995)
Crown, M.D.: Validation of the NOAA space weather prediction Center's solar flare forecasting look-up table and forecaster-issued probabilities. Space Weather. **10**, S06006 (2012)
Desai, M., Giacalone, J.: Large gradual solar energetic particle events. Living Rev. Sol. Phys. **13**, 3 (2016)
Drury, L.: An introduction to the theory of diffusive shock acceleration of energetic particles in tenuous plasmas. Rep. Prog. Phys. **46**, 973 (1983)
Duan, A., et al.: A study of pre-flare solar coronal magnetic fields: magnetic flux ropes. Astrophys. J. **884**, 73 (2019)
Falconer, D.A., Moore, R.L., Gary, G.A.: Magnetogram measures of Total nonpotentiality for prediction of solar coronal mass ejections from active regions of any degree of magnetic complexity. Astrophys. J. **689**, 1433 (2008)
Fan, Y., Zweibel, E.G., Linton, M.G., Fisher, G.H.: The rise of kink-unstable magnetic flux tubes in the solar convection zone. Astrophys. J. Lett. **505**(1), 59–63 (1998)
Florios, K., et al.: Forecasting solar flares using magnetogram-based predictors and machine learning. Sol. Phys. **293**, 28 (2018)

Gaizauskas, V.: Preflare activity. Sol. Phys. **121**, 135–152 (1989)
Gallagher, P.T., Moon, Y.-J., Wang, H.: Active-region monitoring and flare forecasting I. data processing and first results. Sol. Phys. **209**, 171–183 (2002)
Georgoulis, M.K., Rust, D.M.: Quantitative forecasting of major solar flares. Astrophys. J. **661**, L109–L112 (2007)
Gibson, S.E., Low, B.C.: A time-dependent three-dimensional magnetohydrodynamic model of the coronal mass ejection. Astrophys. J. **493**, 460 (1998)
Hagyard, M.J., Smith Jr., J.B., Teuber, D., West, E.A.: A quantitative study relating observed shear in Photospheric magnetic fields to repeated flaring. Sol. Phys. **91**, 115–126 (1984)
Harra, L.K., Matthews, S.A., Culhane, J.L.: Nonthermal velocity evolution in the precursor phase of a solar flare. Astrophys. J. **549**, L245 (2001)
Harvey, K.L., Harvey, J.W.: A study of the magnetic and velocity fields in an active region. Sol. Phys. **47**, 233–246 (1976)
Hood, A.W., Priest, E.R.: Kink instability of solar coronal loops as the cause of solar flares. Sol. Phys. **64**, 303–321 (1979)
Imada, S., Bamba, Y., Kusano, K.: Coronal behavior before the large flare onset. Public. Astron. Soc. Jpn. **66**, S17 (2014)
Inoue, S.: Magnetohydrodynamics modeling of coronal magnetic field and solar eruptions based on the photospheric magnetic field. Prog Earth Planet Sci. **3**, 19 (2016)
Ishiguro, N., Kusano, K.: Double arc instability in the solar Corona. Astrophys. J. **843**, 101 (2017)
Ishii, T.T., Kurokawa, H., Takeuchi, T.T.: Emergence of a twisted magnetic flux bundle as a source of strong flare activity. Astrophys. J. **499**, 898–904 (1998)
Ishii, T.T., Kurokawa, H., Takeuchi, T.T.: Emergence of twisted magnetic-flux bundles and flare activity in a large active region, NOAA 4201. Public. Astron. Soc. Jpn. **52**, 337–354 (2000)
Iwai, K., et al.: Development of a coronal mass ejection arrival time forecasting system using interplanetary scintillation observations. Earth Planets Space. **71**, 39 (2019)
Iwai, K., et al.: Validation of coronal mass ejection arrival-time forecasts by magnetohydrodynamic simulations based on interplanetary scintillation observations. Earth Planets Space. **73**, 9 (2021)
Jing, J., et al.: Statistical analysis of torus and kink instabilities in solar eruptions. Astrophys. J. **864**, 138 (2018)
Jokipii, J.R.: Propagation of cosmic rays in the solar wind. Rev. Geophy. **9**, 27 (1971)
Kliem, B., Török, T.: Torus instability. Phys. Rev. Lett. **96**, 255002 (2006)
Kruskal, M., Schwarzschild, M.: Some instabilities of a completely ionized plasma. Proc. R. Soc. A Mathemat. Phys. Eng. Sci. **223**, 348 (1954)
Kubo, Y., Den, M., Ishii, M.: Verification of operational solar flare forecast: case of regional warning center Japan. J. Space Weather Space Clim. **7**, A20 (2017)
Kusano, K., et al.: Magnetic field structures triggering solar flares and coronal mass ejections. Astrophys. J. **760**, 31 (2012)
Kusano, K., Iju, T., Bamba, Y., Inoue, S.: A physics-based method that can predict imminent large solar flares. Science. **369**, 587–591 (2020)
Lario, D., Sanahuja, B., Heras, A.M.: Energetic particle events: efficiency of interplanetary shocks as 50 keV $<$ E $<$ 100 MeV proton accelerators. Astrophys. J. **509**, 415 (1998)
le Roux, J.A., Webb, G.M.: Time-dependent acceleration of interstellar pickup ions at the Heliospheric termination shock using a focused transport approach. Astrophys. J. **693**, 534 (2009)
le Roux, J.A., Webb, G.M., Florinski, V., Zank, G.P.: A focused transport approach to pickup ion shock acceleration: implications for the termination shock. Astrophys. J. **662**, 350 (2007)
Leka, K.D., Barnes, G.: Photospheric magnetic field properties of flaring versus flare-quiet active regions. II. Discriminant Analysis. Astrophys. J. **595**, 1296–1306 (2003)
Leka, K.D., Barnes, G.: Photospheric magnetic field properties of flaring versus flare-quiet active regions. IV. A statistically significant sample. Astrophys. J. **656**, 1173–1186 (2007)
Leka, K.D., et al.: A comparison of flare forecasting methods. II. Benchmarks, metrics, and performance results for operational solar flare forecasting systems. Astrophys. J. Suppl. Ser. **243**, 36 (2019a)

Leka, K.D., et al.: A comparison of flare forecasting methods. III. Systematic behaviors of operational solar flare forecasting systems. Astrophys. J. **881**, 101 (2019b)
Linton, M.G., Longcope, D.W., Fisher, G.H.: The helical kink instability of isolated, twisted magnetic flux tubes. Astrophys. J. **469**, 954–963 (1996)
Lugaz, N., Manchester, I.V.W.B., Gombosi, T.I.: Numerical simulation of the interaction of two coronal mass ejections from sun to earth. Astrophys. J. **634**, 651 (2005)
Luhmann, J.G., et al.: Modeling solar energetic particle events using ENLIL heliosphere simulations. Space Weather. **15**, 934 (2017)
Marsh, M.S., et al.: SPARX: a modeling system for solar energetic particle radiation space weather forecasting. Space Weather. **13**, 386 (2015)
Martin, S.F.: Preflare conditions, changes and events. Sol. Phys. **68**, 217–236 (1980)
Martin, S.F., Livi, S.H.B., Wang, J.: The cancellation of magnetic flux. II - In a decaying active region. Aust. J. Phys. **38**, 929 (1985)
McAteer, R.T.J., Gallagher, P.T., Ireland, J.: Statistics of active region complexity: a large-scale fractal dimension survey. Astrophys. J. **631**, 628 (2005)
McIntosh, P.S.: The classification of sunspot groups. Sol. Phys. **125**, 251–267 (1990)
Miyoshi, T., Kusano, K., Inoue, S.: A magnetohydrodynamic relaxation method for non-force-free magnetic field in magnetohydrostatic equilibrium. Astrophys. J. Suppl. Ser. **247**, 6 (2020)
Moore, R.L., Sterling, A.C., Hudson, H.S., Lemen, J.R.: Onset of the magnetic explosion in solar flares and coronal mass ejections. Astrophys. J. **552**, 833–848 (2001)
Muranushi, T., et al.: UFCORIN: a fully automated predictor of solar flares in GOES X-ray flux. Space Weather. **13**, 778 (2015)
Murray, M.J., Hood, A.W.: Simple emergence structures from complex magnetic fields. Astron. Astrophys. **470**(2), 709–719 (2007)
Nakagawa, T., Nihsida, A., Saito, T.: Planar magnetic structures in the solar wind. J. Geophys. Res. **94**, 11761 (1989)
Nishizuka, N., et al.: Solar flare prediction model with three machine-learning algorithms using ultraviolet brightening and vector magnetograms. Astrophys. J. **835**, 156 (2017)
Odstrcil, D.: Modeling 3-D solar wind structure. Adv. Space Res. **32**, 497 (2003)
Odstrcil, D., Pizzo, V.J.: Three-dimensional propagation of CMEs in a structured solar wind flow: 1. CME launched within the streamer belt. J. Geophys. Res. **104**, 483 (1999)
Park, S.-H., Guerra, J.A., Gallagher, P.T., Georgoulis, M.K., Bloomfield, D.S.: Photospheric shear flows in solar active regions and their relation to flare occurrence. Solar Phys. **293**, 114 (2018)
Park, S.-H., et al.: A comparison of flare forecasting methods. IV. Evaluating consecutive-day forecasting patterns. Astrophys. J. **890**, 124 (2020)
Poisson, M., Lopez Fuentes, M., Mandrini, C.H., Demoulin, P., Pariat, E.: Study of magnetic flux emergence and related activity in active region NOAA 10314. Adv. Space Res. **51**(10), 1834–1841 (2013)
Qahwaji, R., Colak, T.: Automatic short-term solar flare prediction using machine learning and sunspot associations. Sol. Phys. **241**, 195 (2007)
Reames, D.V.: Particle acceleration at the sun and in the heliosphere. Space Sci. Rev. **90**, 413 (1999)
Ruffolo, D.: Effect of adiabatic deceleration on the focused transport of solar cosmic rays. Astrophys. J. **442**, 861 (1995)
Rust, D.M., Kumar, A.: Evidence for helically kinked magnetic flux ropes in solar eruptions. Astrophys. J. **464**, L199 (1996)
Sammis, I., Tang, F., Zirin, H.: The dependence of large flare occurrence on the magnetic structure of sunspots. Astrophys. J. **540**, 583–587 (2000)
Scherrer, P.H., et al.: The Helioseismic and magnetic imager (HMI) investigation for the solar dynamics observatory (SDO). Sol. Phys. **275**, 207–227 (2012)
Schrijver, C.J.: A characteristic magnetic field pattern associated with all major solar flares and its use in flare forecasting. Astrophys. J. **655**, L117–L120 (2007)

Schrijver, C.J., et al.: Nonlinear force-free modeling of coronal magnetic fields part I: A quantitative comparison of methods. Sol. Phys. **235**, 161 (2006)
Schwadron, N.A., et al.: Earth-Moon-Mars radiation environment module framework. Space Weather. **8**, S00E02 (2010)
Shafranov, V.D.: Plasma equilibrium in a magnetic field. Rev. Plasma Phys. **2**, 103 (1966)
Shiota, D., Kataoka, R.: Magnetohydrodynamic simulation of interplanetary propagation of multiple coronal mass ejections with internal magnetic flux rope (SUSANOO-CME). Space Weather. **14**, 56 (2016)
Shiota, D., et al.: Self-consistent MHD modeling of a coronal mass ejection, coronal dimming, and a Giant cusp-shaped arcade formation. Astrophys. J. **634**, 663 (2005)
Shiota, D., et al.: Inner heliosphere MHD modeling system applicable to space weather forecasting for the other planets. Space Weather. **12**, 187 (2014)
Skilling, J.: Cosmic ray streaming - I. Effect of Alfven waves on particles. Mon. Not. R. Astron. Soc. **172**, 557 (1975)
Takasao, S., Fan, Y., Cheung, M.C.M., Shibata, K.: Numerical study on the emergence of kinked flux tube for understanding of possible origin of delta-spot regions. Astrophys. J. **813**(2), 112 (2015)
Tan, B., Yu, Z., Huang, J., Tan, C., Zhang, Y.: Very long-period pulsations before the onset of solar flares. Astrophys. J. **833**, 206 (2016)
Tanaka, K.: Studies on a very flare-active delta group: peculiar delta spot evolution and inferred subsurface magnetic rope structure. Sol. Phys. **136**, 133–149 (1991)
Toriumi, S., Hotta, H.: Spontaneous generation of δ-sunspots in convective magnetohydrodynamic simulation of magnetic flux emergence. Astrophys. J. Lett. **886**(1), L21 (2019)
Toriumi, S., Takasao, S.: Numerical simulations of flare-productive active regions: delta-sunspots, sheared polarity inversion lines, energy storage, and predictions. Astrophys. J. **850**(1), 39 (2017)
Toriumi, S., Iida, Y., Kusano, K., Bamba, Y., Imada, S.: Formation of a flare-productive active region: observation and numerical simulation of NOAA AR 11158. Sol. Phys. **289**, 3351–3369 (2014)
Wang, Y., Shen, C., Wang, S., Ye, P.: Deflection of coronal mass ejection in the interplanetary medium. Sol. Phys. **222**, 329 (2004)
Wang, H., Jing, J., Tan, C., Wiegelmann, T., Kubo, M.: Study of magnetic channel structure in active region 10930. Astrophys. J. **687**(1), 658–667 (2008)
Welsch, B.T.: Flux accretion and coronal mass ejection dynamics. Sol. Phys. **293**, 113 (2018)
Wheatland, M.S.: A Bayesian approach to solar flare prediction. Astrophys. J. **609**, 1134–1139 (2004)
Wiedenbeck, M.E., et al.: Observations of solar energetic particles from 3He-rich events over a wide range of heliographic longitude. Astrophys. J. **764**, 54 (2013)
Xie, H., Ofman, L., Lawrence, G.: Cone model for halo CMEs: application to space weather forecasting. J. Geophys. Res. **109**, A03109 (2004)
Zhang, J., Dere, K.P.: A statistical study of Main and residual accelerations of coronal mass ejections. Astrophys. J. **649**, 1100–1109 (2006)
Zhao, X.P., Plunkett, S.P., Liu, W.: Determination of geometrical and kinematical properties of halo coronal mass ejections using the cone model. J. Geophys. Res. **107**, 1223 (2002)
Zhou, G.P., Zhang, J., Wang, J.X.: Observations of magnetic flux-rope oscillation during the precursor phase of a solar eruption. Astrophys. J. **823**, L19 (2016)
Zirin, H., Liggett, M.A.: Delta spots and great flares. Sol. Phys. **113**, 267–283 (1987)

Chapter 11
Explorations of Extreme Space Weather Events from Stellar Observations and Archival Investigations

Hisashi Hayakawa, Yuta Notsu, and Yusuke Ebihara

11.1 Superflares on Solar-Type Stars

11.1.1 What Is Superflare?

The previous chapters presented how our Earth and society can be affected by solar flares. In contrast, Sect. 11.2 exploits historical documents to show that highly extreme space weather events (comparable to the Carrington event in 1859) may have occurred more frequently than previously considered. However, in the wider universe, there are many more space weather events. In fact, using optical light telescopes and X-ray space telescopes, recent studies have revealed the existence of stars generating gigantic flares (superflares) that are much larger than the largest solar flares since the middle of the twentieth century (Gershberg 2005; Reid and Hawley 2005). These stars are known to possess strong magnetic fields (average magnetic field strength of a few kilogauss (kG)) because they form close binary

H. Hayakawa (✉)
Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Aichi, Japan

Institute for Advanced Research, Nagoya University, Nagoya, Aichi, Japan

RAL Space, Rutherford Appleton Laboratory, Harwell Campus, Science and Technology Facilities Council, Didcot, UK
e-mail: hisashi@nagoya-u.jp

Y. Notsu
Laboratory for Atmospheric and Space Physics, University of Colorado Boulder, Boulder, CO, USA

National Solar Observatory, Boulder, CO, USA

Department of Earth and Planetary Sciences, Tokyo Institute of Technology, Tokyo, Japan

Y. Ebihara
Research Institute for Sustainable Humanosphere, Kyoto University, Kyoto, Japan

K. Kusano (ed.), *Solar-Terrestrial Environmental Prediction*,
https://doi.org/10.1007/978-981-19-7765-7_11

systems or are very young (several hundred million years old), and their rotation speeds are much higher (a rotation period of a few days) than that of the Sun (Reid and Hawley 2005; Güdel 2007; Benz and Guedel 2010; Linsky 2019). In contrast, the Sun is a single, old star with a gradual rotation (4.6 billion years old with a rotation period of 25 days) and a weak magnetic field (average magnetic field strength of a few gauss). Therefore, it has long been considered that superflares (SXR class $>$ X100) do not occur on the Sun.

11.1.2 Discovery of Superflares on Solar-Type Stars Based on Kepler Space Telescope Data

Solar flare research over the past more than 100 years has shown that the frequency of flares follows the rule that "a tenfold increase in the flare energy results in a tenfold decrease in the flare frequency" (Fig. 11.1). If we simply extend this

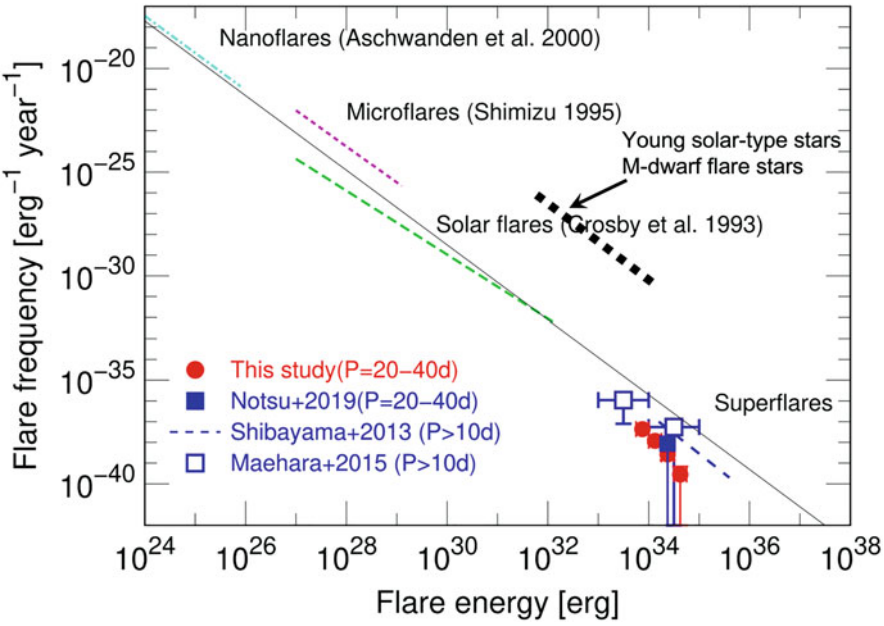

Fig. 11.1 Occurrence frequency distributions of superflares and solar flares (reproduced from © Okamoto et al. 2021). Solar (including nano- and micro-) flares decrease in frequency due to the power law of energy. Flare with energy 100 times larger than that of the largest solar flare occurs approximately once every few thousand years on stars with long rotation periods similar to that of the Sun (Sun-like stars). However, on young solar-type stars (short rotation periods) and M-type flare stars (low-temperature stars with deep convection layers and efficient dynamo mechanisms for magnetic activities), superflares occur approximately 1000 times a year (Güdel 2007; Shibayama et al. 2013; Hawley et al. 2014)

relationship, a superflare that is 100–1000 times larger than the largest solar flare ($\sim 10^{32}$ erg) will occur approximately once every 1000–10,000 years. If superflares were to occur on the Sun, they would be expected to occur at such a low frequency. However, modern observations of the Sun span only ~1.5 centuries with solar flare observations. In the centennial timescale, there were no telescopes or satellites. As such, it is difficult to find them in scientific records. Around 2010, a research group at Kyoto University began to observe multiple stars similar to the Sun (solar-type stars = G-type main sequence stars), instead of analyzing the back history of solar flares. For example, if we observe 10,000 solar-type stars for 1 year, these observations approximate the results obtained from solar observations for 10,000 years. Thus, we might be able to identify indicators to understand if superflares will occur on the Sun.

Around the same time, in the summer of 2010, the Kepler Space Telescope (launched in March 2009), which subsequently made revolutionary achievements in exoplanet exploration, began to present its results. This telescope has been continuously observing the brightness of more than 170,000 stars (at least 50,000 of which are solar-type stars: cf. Okamoto et al. 2021) in the region near Cygnus at intervals of once every 30 min for a long period of 4 years. From these observations, the Kepler Space Telescope could detect even the slightest changes in the brightness of the stars, and the data immediately became available to the space science community. From its data, it became immediately noticeable that the slight brightening of stars caused by superflares was detectable, and Hiroyuki Maehara, who was then a researcher at the Kyoto University Astronomical Observatory, and five undergraduate students from the Faculty of Science (Takuya Shibayama, Yuta Notsu (author of this chapter), Shota Notsu, Takashi Nagao, and Satoshi Kusaba) started examining research on the detection of superflares on solar-type stars.

Figure 11.2 shows examples of superflares detected in a solar-type star. With the quasi-periodic variations of a few days to a few tens of days, spikelike brightness increments occur several times. On closely examining the spikes, it can be seen that rapid brightness increments are followed by gradual decrements. Suspicious events were thoroughly rejected, e.g., when there is another source nearby or only one brightening point with a short duration, by visually examining the individual luminosity curves. Finally, 365 flare events on 148 solar-type stars in the first 90 days of the data were detected (Maehara et al. 2012). The energies of the flares are estimated from the brightness of the flaring star itself, the amplitude of the flare, and its duration, and the values are 10^{33}–10^{36} erg. Because the energy of a Carrington flare, as described in Chap. 9, is estimated to be approximately 10^{32} erg, the flares discovered here are "superflares" that are 10–10,000 times larger than the largest solar flares.

The most important significance of this superflare discovery is the identification of numerous events. Prior to this discovery, only nine "candidate" flare events had been reported for solar-type stars from previous ground-based photographic plate data and visual observations (Schaefer et al. 2000). Given the small number of observations (only nine) and the accuracy of the data used, a statistical study was difficult. However, since the advent of the data from the Kepler Space Telescope, a

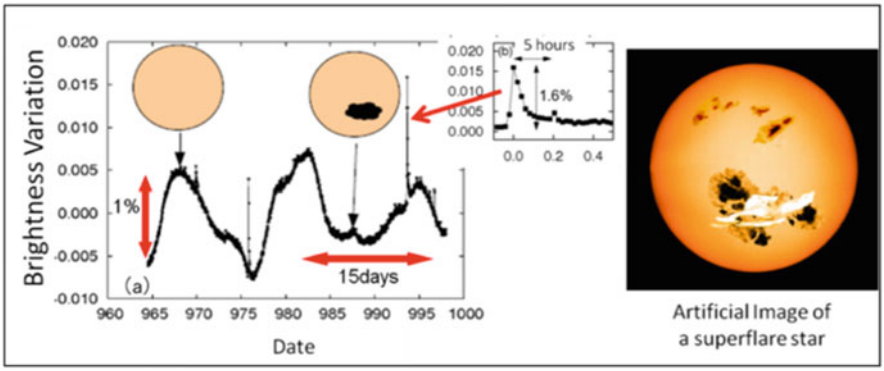

Fig. 11.2 Left: Temporal brightness variations in an example solar-type star showing superflares (from Kepler Space Telescope data: reproduced from © Maehara et al. 2012). In addition to the abrupt brightening of flares, there is gradual change in brightness over a period of approximately 15 days. Star spot size and rotation period can be estimated from brightness variations. Right figure: Artificial image of a solar-type star showing superflares. Superflare (white) is occurring on a large star spot group

longer period of data has been analyzed following the first report in 2012, and more than 1000 superflares have been discovered until now (Shibayama et al. 2013; Maehara et al. 2015). The discovery of such numerous events has allowed for the first time to make a statistical argument for superflares on solar-type stars.

11.1.3 Detailed Investigations by Spectroscopic Observations and Gaia Satellite Observation Data

What are the characteristics of a star that causes a superflare, or what are the conditions in which a superflare occurs? From the observations of solar flares, it is known that a more massive flare is more probable to occur if the size of the sunspot (active region) that causes the flare is large and if the structure of the sunspot (active region) is highly complex (e.g., "delta sunspots") (Sammis et al. 2000; Toriumi and Wang 2019: see also Chap. 10). Is there any similarity in the features of the stars that produce superflares?

If we examine the Kepler Space Telescope data of superflare stars in detail, we can find that many stars show quasi-periodic brightness variations lasting from a few days to several tens of days, besides the flare brightness increments. It is considered that these brightness variations are caused by the rotation of the stars with star spots on their surfaces (Fig. 11.2; Notsu et al. 2013). If this assumption is correct, the period of the brightness variation corresponds to the rotation period of the star, and its amplitude corresponds to the size of the star spot. Because the amplitude of the variability is typically 0.1–10%, most stars with superflares are expected to have star

spots larger than sunspots, as shown in Fig. 11.2. Enormous star spots can store a large amount of magnetic field energy, and this is effective for causing a superflare. The values of the rotation period obtained from the brightness variations range from a few days to 30 days, suggesting that these stars can cause superflares even with a rotation period as long as that of the Sun (approximately 25 days) (Notsu et al. 2013).

However, this interpretation is only an assumption, and it is difficult to answer the above questions using only the photometric observation data from the Kepler Space Telescope. Do giant star spots really exist on the surfaces of stars? Can the brightness variations be explained by the rotations of the stars? Do superflares really occur on a single star with a gradual rotation like the Sun? Furthermore, is there any possibility that stars with different characteristics from the Sun, such as binary stars or subgiant stars, are mixed in with the discovered stars? Even after the discovery with the Kepler Space Telescope data, the above fundamental questions remained unanswered.

In view of the above scenario, a group including the author of this paper fully utilized the advantages of "new" methods and data. These were high-dispersion spectroscopic observations (Subaru Telescope and Apache Point Observatory 3.5 m telescope in the United States (U.S.) and the Gaia satellite (a satellite that observes the positions and distances of approximately one billion celestial objects in the entire sky in visible light)). The objective was to examine in detail the findings of the photometric data of the Kepler Space Telescope (Notsu et al. 2015a, b, 2019; Okamoto et al. 2021).

First, in the spectroscopic observations, using the data of 64 bright stars out of approximately 300 superflare stars discovered from the Kepler Space Telescope data, we conducted a detailed analysis of many absorption lines appearing in the wavelength spectra and obtained the following three results:

1. For more than half of the 64 observed stars (43 stars), there were no signs of binarity, such as overlapping absorption lines between two stars or large variations in the radial velocity. Moreover, we could confirm spectroscopically that the temperature and other parameters are in the range of solar-type stars (G-type main-sequence stars).
2. If the above interpretation of "brightness variation = rotation" is correct, the timescale of the brightness variations corresponds to the velocity of the rotation of the star. Spectroscopic observations allow observers to estimate the rotation velocity from the broadening of absorption lines (Fig. 11.3). We confirmed that such spectroscopic velocities match the brightness variation timescales, supporting the above interpretation.
3. It is known from solar observations that when the average magnetic field on the stellar surface becomes stronger owing to the presence of large star spots, the stellar chromosphere is heated and the emission of chromospheric spectral lines becomes stronger. Moreover, absorption lines such as Ca II lines and hydrogen Hα lines become shallower. Based on the above, we measured the depths of Ca II lines of superflare stars. Consequently, we found a positive correlation between the intensity of the Ca II line and the star spot size estimated from Kepler Space

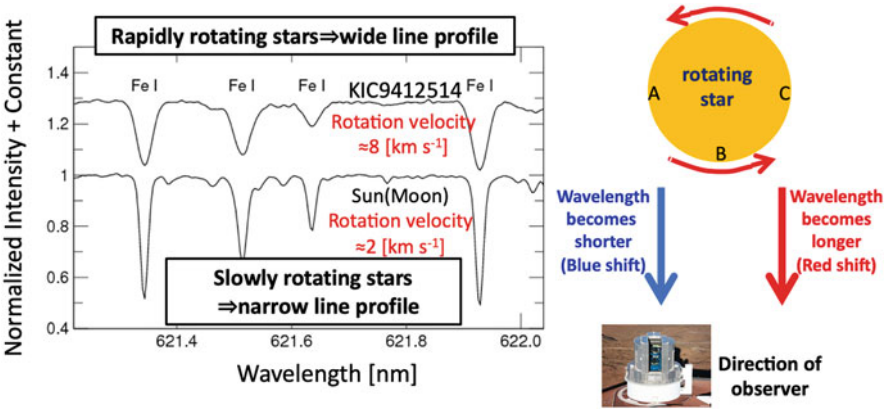

Fig. 11.3 (Right) Light emitted from each location on the surface of a rotating star slightly changes in wavelength under the Doppler effect of light. (Left) Rotation velocities of stars and spreads of spectral lines. Here, four absorption lines of neutral iron (Fe I) are shown. As shown above, absorption lines are broadened by the wavelength shift caused by the Doppler effect of light emitted from the surface of a rotating star. For a gradually rotating star like the Sun, the line is relatively narrower owing to the small wavelength shift caused by the Doppler effect. For a rapidly rotating star, the wavelength shift becomes larger, and the line becomes wider. By measuring the spread of lines, we can determine the rotation velocities of stars

Telescope photometric data (Fig. 11.4). This supported the information obtained from Kepler Space Telescope photometric data and further confirmed that superflare stars have larger star spots than the Sun. This result was confirmed by the high-dispersion spectroscopic observations of Notsu et al. (2015b, 2019) as well as the medium-dispersion spectroscopic observations of Karoff et al. (2016) using the Large-sky Area Multi-Object Fiber Spectroscopic Telescope (LAMOST) in China.

The above results indicated that the quasi-periodic brightness variations of superflare stars can be explained by the rotation of large star spots and that superflare stars have large star spots. Furthermore, these spectroscopic observations were conducted for 64 bright stars out of a total of 300 stars; however, for approximately 90% of the total 300 stars, including the remaining faint stars, the stellar radii were accurately derived from the stellar distance data obtained by the Gaia satellite, which was released in April 2018. This allowed performing a statistical study on superflares using Kepler Space Telescope data by rejecting stars that were not similar to the Sun, such as giants that are also larger in radius and brighter than the Sun. Instead, only stars that can be more accurately described as similar to the Sun (Notsu et al. 2019; Okamoto et al. 2021) were used. Consequently, applying the stellar radius data of the Gaia satellite to all 4 years of Kepler Space Telescope data, Okamoto et al. (2021) reported a total of 2341 superflares on 265 correctly identified solar-type stars. A total of 26 superflares on 15 stars were detected, for gradually rotating Sun-like stars with a rotation period >20 days and a temperature of 5600–6000 K.

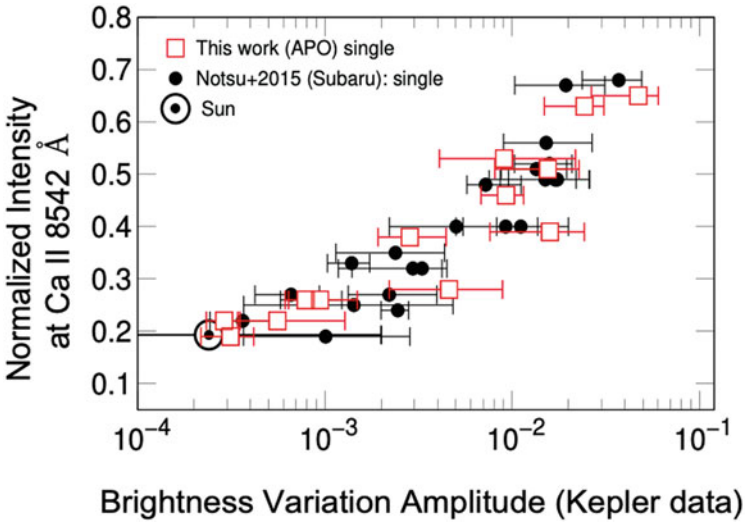

Fig. 11.4 Amplitudes of brightness variations observed with the Kepler Space Telescope (horizontal axis) and normalized intensity of Ca II 8542 Å line center (vertical axis), as reproduced from © Notsu et al. (2019). Difference between points is due to difference in spectroscopic telescopes (Subaru observations (Notsu et al. 2015a, b) and Apache Point Observatory (APO) observations (Notsu et al. 2019)). This figure shows that there is a correlation and good correspondence between the Kepler Space Telescope photometric index (horizontal axis) and spectroscopic index (vertical axis) with respect to star spot size

Based on the above results, assuming that the period and amplitude of the brightness variations shown in Fig. 11.2 represent the rotation period and star spot size of the individual superflare stars, respectively, the latest findings of the statistical properties of superflare stars (Notsu et al. 2019; Okamoto et al. 2021) will be discussed after this subsection.

11.1.4 Star Spot Sizes and Superflare Energies

The above discussion has confirmed that a feature of superflare stars is that they contain large star spots. Flares occur on the Sun when the energy of the magnetic field stored in an active region near the sunspot is abruptly released. If we assume that superflares are generated by the same mechanism as solar flares, then we can explain their energies by the energy of the magnetic field of star spots.

The energy of the magnetic field stored in a star spot, E_{mag}, can be approximated as $E_{\mathrm{mag}} = B^2\, L^3/8\pi$, where B is the magnetic field strength of the star spot and L is the size of the star spot area. Assuming that the area of the star spot region is A_{spot}, energy is released in proportion f of E_{mag} and that ${A_{\mathrm{spot}}}^{3/2} = L^3$, the energy of the flare (E_{flare}) is given by the following relation:

$$E_{\mathrm{Flare}} \approx f E_{\mathrm{mag}} \approx f \frac{B^2 L^3}{8\pi} \approx f \frac{B^2}{8\pi} A_{\mathrm{spot}}^{3/2} \tag{11.1}$$

Figure 11.5 plots the star spot area on the horizontal axis and the flare energy on the vertical axis for superflare data obtained by the Kepler Space Telescope and the solar flare data from 1989 to 2006 by the Geostationary Operational Environmental Satellite (GOES). The data points on the upper right and lower left correspond to superflares and solar flares.

Considering that the spot magnetic field strength of a solar-type star is several kG, the lines based on Eq. (11.1) (for $f = 0.1$) are shown as straight lines in Fig. 11.5 for two cases: $B = 1000$ gauss and $B = 3000$ gauss. For solar flares, in particular, there is a correlation between the sunspot size and the maximum flare energy, and all data points for solar flares and most data points for superflares are located below the line corresponding to Eq. (11.1). These results support the hypothesis that the flare energy can be explained by the spot magnetic energy.

However, some of the superflare data are located above the two lines corresponding to Eq. (11.1). It is difficult to interpret them because a star spot size estimated from the amplitude of the brightness variations includes errors due to the effect of the inclination angle of the rotation axis from the line-of-sight direction and the longitude distribution of the star spots. For example, if the inclination angle of the rotation axis is small, i.e., the star is observed from a direction near the pole, the star spot size is estimated to be smaller than the actual size. Specifically, if we consider the effect of the inclination angle of the rotation axis, we can broadly conclude that the energy of superflares is explained by the energy of star spots.

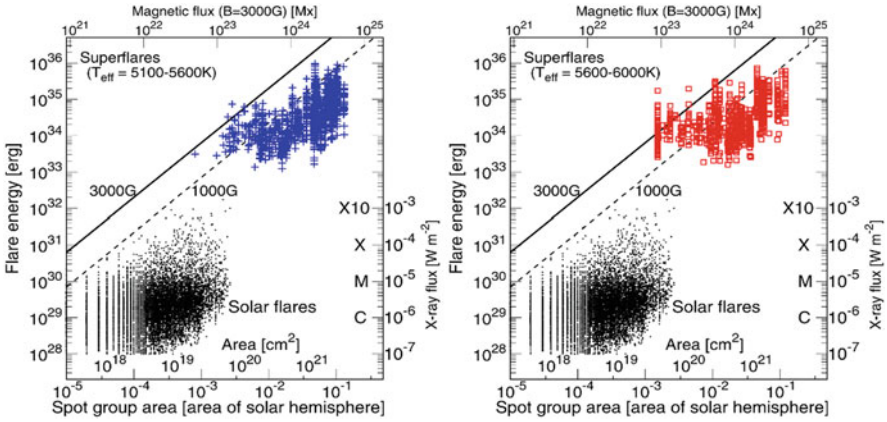

Fig. 11.5 Area of star spots on the horizontal axis and flare energy on the vertical axis for both solar flare data and superflare data found in Kepler Space Telescope data (reproduced from © Okamoto et al. 2021). Vertical axis on the right shows GOES flux intensity, which corresponds to flare energy

11.1.5 Changes in Superflare Activity with Stellar Rotation and Age

The Kepler Space Telescope has also revealed many reliable characteristics of the changes in superflare activity with rotation and age. Figure 11.6 shows the relationship between the frequencies of superflares that release energy above a certain size (10^{34} erg and 5×10^{34} erg) and the rotation period. From this figure, we can see that the frequencies of the superflares decrease with increasing rotation period. The rotation period of a star is known to correlate well with its age (Ayres 1997; Güdel 2007). For example, a young star with a few hundred million years of age rotates approximately an order of magnitude faster than the Sun. Solar-type stars with the same rotation period as the Sun (approximately 25 days, 4.6 billion years old) are found to generate superflares a hundred times less frequently than young solar-type stars with periods of a few days (hundreds of millions of years old). A flare is the release of magnetic energy in the upper atmosphere (corona) of a star, and the source of the magnetic activity is considered to be a dynamo mechanism caused by the rotation and convection of the star. The reason flares occur more frequently in younger stars with shorter rotation periods is that the dynamo mechanism is more active in producing magnetic fields in the latter.

However, we find that the frequencies of the superflares become almost constant when the rotation period is shorter than 2–3 days (Fig. 11.6). A similar relationship is known between the rotation period and the quiescent (steady-state) X-ray luminosity and emission line intensity emitted from the chromosphere, such as Ca II lines, in the case of solar-type stars (Wright 2011).

While the frequency of flares is correlated with the rotation period, what can be inferred about their energy? A star with a shorter rotation period may be able to produce a larger flare. In fact, a scatter plot of the energies of superflares and the rotation period (Fig. 11.7) shows a moderate correlation between the upper limit of the flare energy and the rotation period. Young rapidly rotating solar-type stars with periods of a few days (hundreds of millions of years old) produce flares with energies

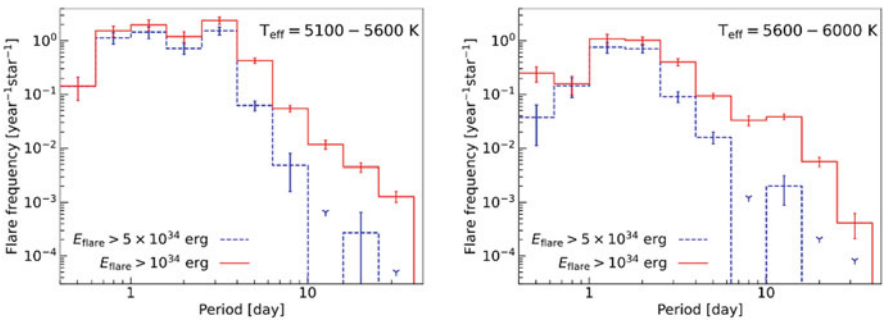

Fig. 11.6 Relationship between frequencies of superflares with energies greater than 10^{34} erg and 5×10^{34} erg and the rotation period of a star. (Reproduced from © Okamoto et al. 2021)

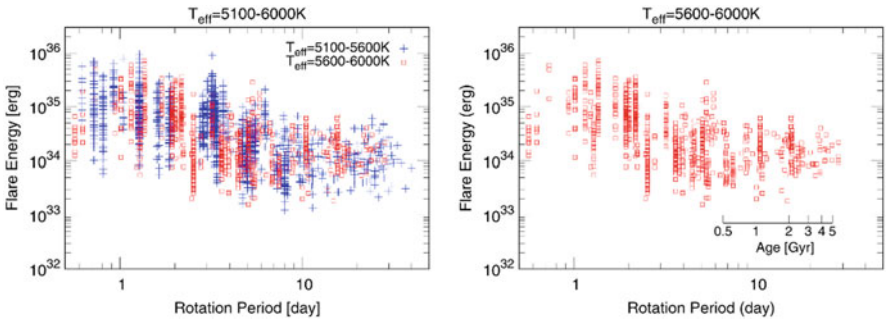

Fig. 11.7 Scatter plots of flare energy versus the rotation period of superflares (reproduced from © Okamoto et al. 2021). They differ based on stellar temperature values: $T_{\rm eff} = 5100$–5600 K and 5600–6000 K. Scale of stellar age is added based on the gyrochronology relation (age versus rotation period relationship) for solar-type stars (Ayres 1997)

up to 10,000 times more (~10^{36} erg) than the largest solar flares. In contrast, solar-type stars with the same rotation period as the Sun (approximately 25 days, 4.6 billion years old) produce only superflares with energies several hundred times larger ($<4 \times 10^{34}$ erg) than those of the largest solar flares.

Summarizing the above results, it can be concluded that superflares several hundred times larger ($\sim 4 \times 10^{34}$ erg) than the largest solar flare ever observed can occur even on the gradually rotating and old Sun, although the frequency is low. However, immediately after the birth of the Sun (several hundred million years old), its rotation speed was high and the mechanism to generate a large amount of magnetic energy necessary for superflares was active. Thus, at that time, superflares up to 10^{36} erg, which are 10,000 times larger (~10^{36} erg) than the largest solar flares, could have occurred, with a flare frequency at least 100 times higher than that of the older slowly rotating Sun.

11.1.6 Do Superflares Occur on the Sun?

Therefore, if superflares occur on the Sun, what is their frequency? Based on the energy frequency distribution of superflares, we can find that a large superflare is associated with a low frequency (a power-law distribution) (Fig. 11.8). As shown in Fig. 11.1, the frequency distribution of superflares in solar-type stars with rotation periods of more than 20 days (the Sun has a rotation period of approximately 25 days) is in good agreement with the extended frequency distribution of solar flares. From Figs. 11.1 and 11.8, we can conclude that superflares exceeding 70 times (7×10^{33} erg) and 100 times (1×10^{34} erg) the largest solar flares occur once every ~3000 and ~6000 years, respectively (Notsu et al. 2019; Okamoto et al. 2021).

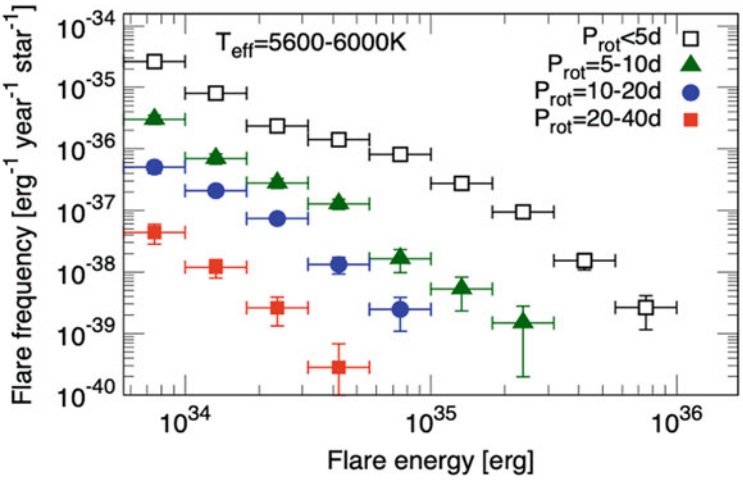

Fig. 11.8 Occurrence frequency distributions of superflares on solar-type stars with temperatures between 5600 and 6000 K (reproduced from © Okamoto et al. 2021). Data are divided by rotation period

Here, it is concluded that a superflare of $\sim 10^{34}$ erg occurs once every several thousand years on a solar-like star. This also corresponds well with the results of statistical studies on the occurrence of large star spots using Kepler Space Telescope data (Figs. 11.9 and 11.10: Maehara et al. 2017, Notsu et al. 2019, Okamoto et al. 2021). Based on Fig. 11.9, the distributions of the rotation periods and star spot areas of solar-type stars show that even Sun-like stars with rotation periods of 20–40 days have star spots with a size of 1% or larger than the solar hemisphere. Using the relation shown in Fig. 11.5, it can be inferred that a star spot of this size has sufficient magnetic energy to produce a 10^{34} erg superflare.

Figure 11.10 shows the frequency of large star spots obtained using these star spot area data of Sun-like stars. It has been pointed out that the frequency of sunspots is log-normal, i.e., the frequency of large sunspots decreases rapidly (Bogdan et al. 1988). It has also been argued that because there have been no giant sunspots exceeding 6500 msh in the past 150 years of modern observations, superflares that reach 10^{34} erg may not occur (Aulanier et al. 2013). In fact, if we examine only the distribution of sunspots in Fig. 11.10, such a trend is apparent at first. However, when the frequencies of large star spots and sunspots are combined, they are broadly represented by a single power-law distribution over a wide range of spot areas from approximately 10^{-3} to 10^{-1} of the solar hemisphere area. In this figure, the values of sunspots and star spots differ by approximately 10^{-2}, which may be caused by only having data for the past 150 years available for sunspots, which is an insufficient observation period. This feature of a common distribution suggests that sunspots and large star spots of Sun-like stars can be explained by a typical physical process. It also shows that sunspots larger than a few percent of the solar hemisphere occur once in a few thousand years, which is similar to the frequency of superflares.

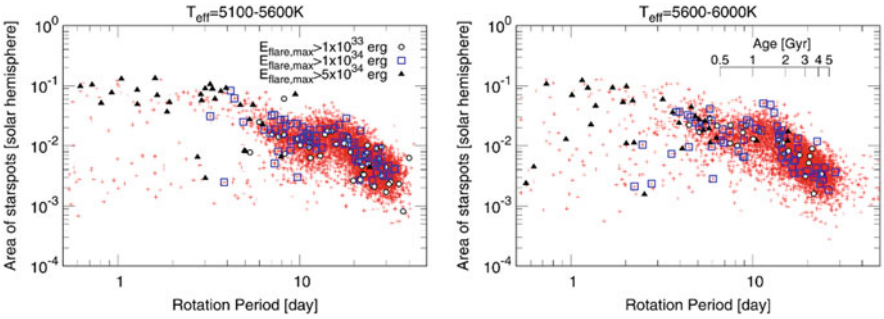

Fig. 11.9 Relationships between rotation periods and star spot areas of solar-type stars with temperatures $T_{\mathrm{eff}} = 5600$–6000 K (including nonsuperflare stars), observed by the Kepler Space Telescope (reproduced from © Okamoto et al. 2021). Black triangles, blue squares, and white circles represent stars that show superflares of $>5 \times 10^{34}$ erg, $>1 \times 10^{34}$ erg, and $> 1 \times 10^{33}$ erg, respectively

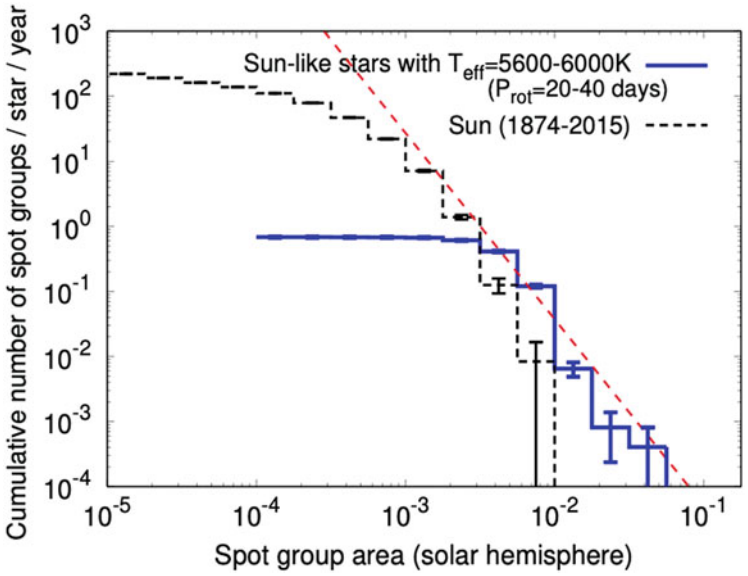

Fig. 11.10 Comparison of appearance frequency versus spot area distributions of star spots on gradually rotating Sun-like stars (temperature 5600–6000 K, rotation period 20–40 days) and of sunspot groups. (Reproduced from © Notsu et al. 2019)

Rubenstein & Schaefer et al. (2000) argued that stars having a hot Jupiter-like companion are good candidates for superflare stars because a hot Jupiter may play the role of a companion star in binary stars, such as RS CVn stars. RS CVn stars are magnetically very active and produce many superflares. Because the Sun does not have any hot Jupiter, Schaefer et al. (2000) concluded that it has never generated superflares. To examine their argument, Okamoto et al. (2021) analyzed whether the

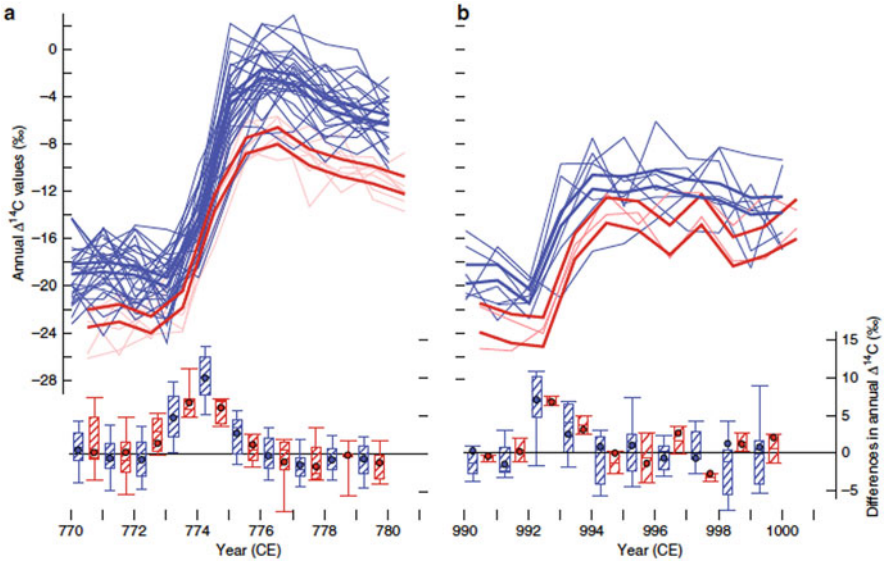

Fig. 11.11 Temporal variations in the amount of ^{14}C around 774–775 A.D. (Panel a) and 993–994 A.D. (Panel b) (Reproduced from © Büngten et al. 2018 under CC-BY)

superflare solar-type stars found in this study had any exoplanets, using the data retrieved from the National Aeronautics and Space Administration (NASA) Exoplanet Archive on May 28, 2020. The results showed that the solar-type superflare stars found in this study had no confirmed exoplanets; however, three candidate exoplanets and three false-positive exoplanets were identified. This suggests that a hot Jupiter is not a necessary condition for superflares on solar-type stars. In addition, Shibata et al. (2013) found from theoretical considerations that a hot Jupiter does not play an essential role in the generation of a magnetic flux in a star if only the magnetic interaction between the star and the hot Jupiter is considered. Although the tidal interaction could be a possible cause of the enhanced dynamo activity, more detailed studies are necessary. This seems to be consistent with the above observational finding that a hot Jupiter is not a necessary condition for superflares on solar-type stars.

Therefore, whether there is evidence of superflares on the Sun needs further examination. Two weeks after Maehara et al.'s (2012) paper on the discovery of superflares was published in Nature, an important study also appeared in the same journal. Miyake et al. (2012) investigated the temporal ratio variations in carbon-14 (^{14}C) isotope amount with annual resolution using the annual rings of Yakusugi cedar and found a sharp increase at 774/775 A.D. (Fig. 11.11). ^{14}C is produced when neutrons produced by cosmic rays colliding with atoms in the upper atmosphere of the Earth collide with nitrogen 14 (^{14}N), following which ^{14}C decays. The fact that the amount of ^{14}C showed a rapid increase suggests that the amount of cosmic ray radiation from the space that reached the Earth increased. This event has also been

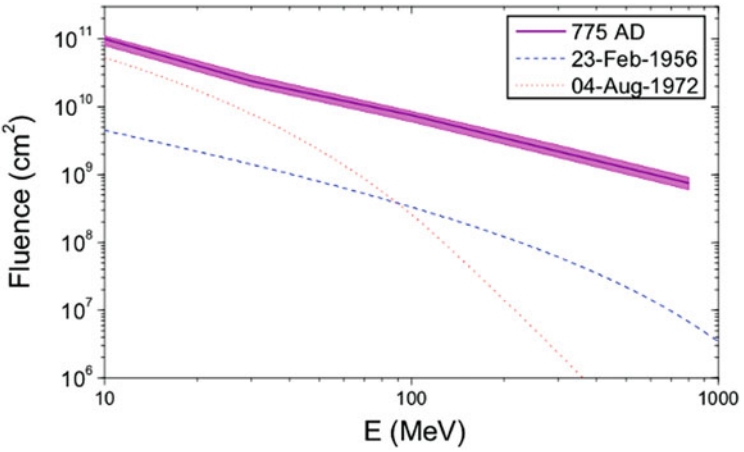

Fig. 11.12 Energy spectra of SEP events (Usoskin 2017), showing an event in 775 A.D. (reproduced from ^{14}C, ^{10}Be, and ^{36}Cl observed values) and two notable examples of observed SEP events in modern times, as reproduced from © Usoskin et al. (2017) under CC-BY. Event on February 23, 1956, is an example of a SEP event with the hardest slope in the energy spectrum, and that on August 4, 1972, is an example of a SEP event with the largest and softest slope in the energy spectrum

found to be global, with detection in trees worldwide other than Yakusugi (Usoskin et al. 2013; Usoskin 2017; Uusitalo et al. 2018; Büntgen et al. 2018a, b).

After the first finding, temporal changes in ^{14}C amount have been explored for longer timescales, and similar events in 993/994 A.D. and 660 B.C. have also been found (Miyake et al. 2013; Park et al. 2017; O'Hare et al. 2019; Usoskin and Kovaltsov 2021). Several possible causes of these cosmic ray increments have been suggested such as nearby supernova explosions, γ-ray bursts, and giant solar flares (Usoskin 2017; Miyake et al. 2019). However, since its discovery in 2012, the temporal changes in ^{14}C amount have been measured in trees worldwide other than Yakusugi cedar. Beryllium-10 (^{10}Be) and chlorine-36 (^{36}Cl) in ice cores (^{10}Be and ^{36}Cl are radioisotopes produced by cosmic rays as well as ^{14}C) have been also measured. These investigations suggested that the ^{14}C concentration of each annual ring is strongly related to its latitude (Uusitalo et al. 2018), and even ^{10}Be, whose concentration does not increase easily in the γ-ray bursts, showed a peak along with ^{36}Cl (Mekhaldi et al. 2015; Büntgen et al. 2018a, b). Presently, solar explosion phenomena are considered the most probable (Mekhaldi et al. 2015; Usoskin 2017). A comparison of the changes in different nuclides—^{14}C, ^{10}Be, and ^{36}Cl—has led to examination of the size of the event compared to the large solar energetic particle (SEP) events captured in modern observations (Fig. 11.12).

In addition to the search for cosmic ray events using tree rings and ice cores, efforts are ongoing to determine whether there are any reports of low-latitude auroras or giant sunspots in the past literature (see Sect. 11.2). If these events are indeed traces of superflares on the Sun, they may suggest that superflares also occur on the

Sun on centennial timescales. However, there is currently considerable uncertainty about the sizes of SEP events and coronal mass ejections (CMEs) associated with superflares. Therefore, it is important to compare the frequency of superflare events with that of terrestrial traces (e.g., large-scale cosmic ray events), such as those mentioned by Schrijver and Beer (2014), Battersby et al. (2019), and Usoskin and Kovaltsov (2021), and those obtained from solar-type star observations. Further studies are needed to determine the relevance of the comparison results for the frequency of superflare occurrence.

11.1.7 Effects of Superflares on Planets (1)

Thus far, we have discussed the observation results of superflares on solar-type stars and the possibility of superflares occurring on the current Sun. What would be the impacts on the Earth if a superflare were to occur on the Sun? As has been discussed in detail in previous chapters, modern society has many applications that could be damaged by solar flares, such as numerous electronic devices and satellite communication. If superflares, which are 10–100 times larger than the largest solar flares, which are extremely rare, occurring only once every several hundred to several thousand years, do occur, they could have enormous impacts on our daily social infrastructure. In particular, the scale of the magnetic storms associated with superflares has been estimated by Takahashi and Shibata (2017) to have a Dst index of 2000 nT, showing it is very important to study their specific impacts (including social and economic effects) in the future (Eastwood et al. 2017; Lingam and Loeb 2017; Riley et al. 2018; Battersby et al. 2019: Kataoka 2020).

The scales of CMEs and SEPs associated with superflares are difficult to predict accurately because stellar CME observations are still in their beginning phase (cf. Moschou 2019; Airapetian et al. 2020). However, extrapolation of the solar upper limits suggests that a superflare 100 times larger than the largest solar flare (10^{34} erg and X1000 flare) could produce a CME with a velocity of 9000 km/s and a SEP flux more than 100 times larger than the largest solar flare (Fig. 11.13 and Takahashi et al. 2016). The results of the radiation transport calculations, in which such high SEP flux values penetrate the planetary atmosphere, suggest that the radiation exposure may increase to a level sufficiently dangerous for life at aircraft altitudes and at the Martian surface. At aircraft altitudes, the atmospheric pressure is lower than that at the surface, and at the Martian surface, there is no magnetic field or sufficient atmosphere like at the surface of the Earth (Fig. 11.14 and Yamashiki et al. 2019). However, as can be seen from the two examples of SEP spectra in Fig. 11.12 and the variations in the data points on the Sun in Fig. 11.13, there is a large scatter between the flares and the magnitudes of the CMEs and the SEPs, even in the range of solar flares. Thus, quantitative estimates of the CMEs and the SEPs associated with superflares require further observation and theoretical studies (cf. Schrijver et al. 2012; Airapetian et al. 2020; Herbst et al. 2019a; Moschou 2019; Lynch et al. 2019; Hu et al. 2022).

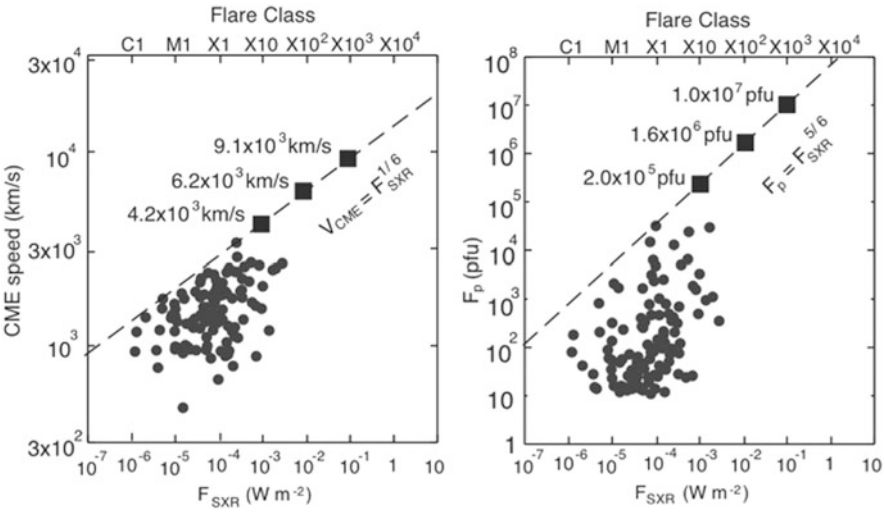

Fig. 11.13 Flare size (GOES X-ray intensity) versus CME velocity and proton flux (from Takahashi et al. 2016; © AAS. Reproduced with permission). Black circle data points are for the solar case. Black square points represent the upper limit in case of superflares

11.1.8 Effects of Superflares on Planets (2)

The discussion in Sect. 11.1.7 was mainly concerned with the case of a superflare occurring in the present solar system. Here, we expand the topic, and first briefly discuss the effects of superflares on past Earth and exoplanet systems, introducing an interesting new theory that has been recently published (Airapetian et al. 2016). For more details in this area, see also Airapetian et al. (2020) and Linsky (2019) for reviews.

11.1.8.1 Faint Young Sun Paradox

The Earth that we live on currently is a very suitable environment for the existence of life, with warm, stable liquid water on the surface and abundant organic matter that forms our bodies. Geological evidences such as rocks suggest that oceans have existed on the Earth since the beginning of life approximately 4 billion years ago and that the present warm environment was already existing. However, according to the standard model of solar evolution, the brightness of the Sun 4 billion years ago was approximately 70% of the current brightness. Moreover, if we assume that the atmosphere of the Earth had the same composition as it does today, the temperature of the surface of the Earth would have been below 0 °C, i.e., liquid oceans would not have existed. This contradiction is called the "faint young Sun paradox," which was proposed by Carl Sagan et al. in 1972 (Segan and Mullen 1972).

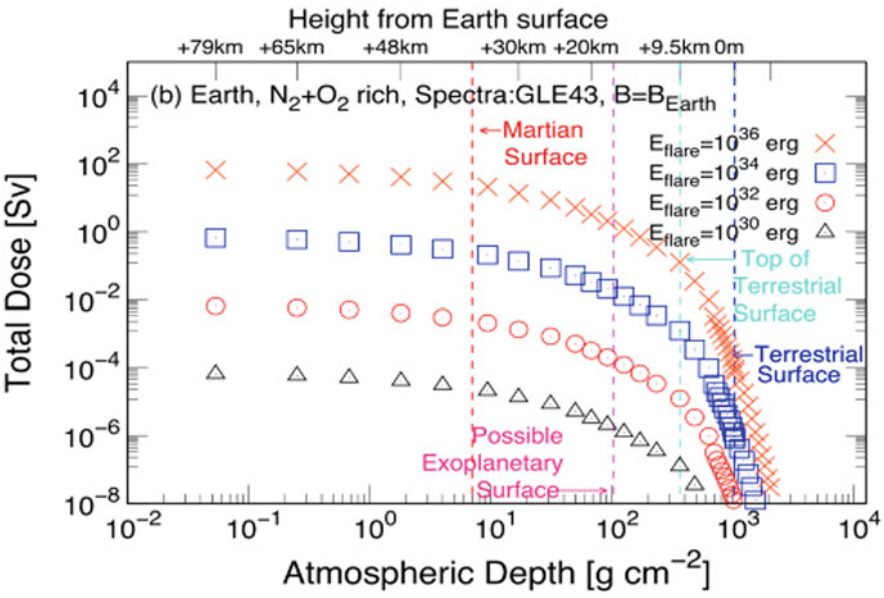

Fig. 11.14 Calculated radiation doses for Earth-like atmosphere injected with energetic particles associated with superflare (assuming 1989 GLE43 event for the shape of the SEP spectrum), from © Yamashiki et al. 2019. Four flare size cases from 10^{30}erg to 10^{36} erg are shown

Various theories have been proposed to resolve the above contradiction (cf. Kasting 2010). They assume that the amount of greenhouse gases in the atmosphere of the Earth was much higher in the past than it is currently. For example, there are theories that the amounts of ammonia (NH_3) and methane (CH_4) were extremely high that the greenhouse effect was effective and that the amount of carbon dioxide (CO_2) was 1000–10,000 times higher than the current concentration. However, none of these theories has been proven to be quantitatively valid, e.g., they do not match the results of geological studies on CO_2 concentrations in the past, and there is currently no definitive solution to this paradox.

11.1.8.2 Effects of Young Sun Superflares on Earth

Approximately 4 billion years ago, the young Sun was 30% darker than it is currently. Therefore, what type of star was it from the perspective of magnetic activities such as flares?

According to the above statistical studies using the Kepler Space Telescope data, the frequency of superflares is much higher on young stars, which rotate approximately 10 times faster than the Sun, than on old ones, exceeding one per day (Fig. 11.1). This result suggests that the young Sun 4 billion years ago was in a much more active state than presently, with frequent superflares, because it was

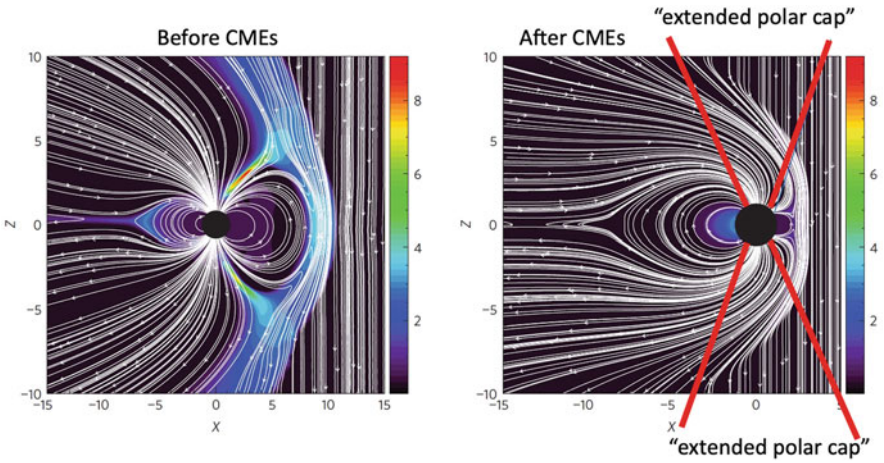

Fig. 11.15 Imaginary view of CMEs ejected by superflare impacting (from the right side of the figure) the magnetosphere of Earth approximately 4 billion years ago (modified from Airapetian et al. 2016 with permission of © Springer-Nature Publishing Group). White lines represent magnetic field lines, and colored bars represent plasma pressure. Magnetosphere of Earth is compressed, resulting in reconnection of magnetic field coordination and the formation of a pathway (region indicated by red lines in the figure) for energetic particles to enter the atmosphere from polar regions. Figure on the right is the final state after reconnection; therefore, pressure is reduced

spinning faster and could easily store the large amount of magnetic field energy needed for superflares. As seen in Fig. 11.13, although there is a large degree of uncertainty in the scale, it is possible that the frequent superflares were accompanied by massive CMEs, which flew toward the Earth. Airapetian et al. (2016) conducted numerical simulations of the interactions of such CMEs with the magnetic field of the Earth when they impacted the Earth. The results showed that the magnetosphere of the Earth was compressed, creating a pathway for energetic particles to enter the atmosphere from the polar regions (Fig. 11.15).

What are the effects of high-energy particles entering the atmosphere? Airapetian et al. (2016) performed chemical reaction calculations assuming the atmospheric composition of the early Earth (nitrogen-rich: 80% N_2, 20% CO_2). The results strongly suggested that high-energy particles chemically reacted with the abundant nitrogen molecules (N_2) in the atmosphere of the Earth to produce large amounts of nitrous oxide (N_2O) and hydrogen cyanide (HCN), when passing through various intermediate products.

11.1.8.3 Formation of Environment Suitable for Birth of Life

N_2O and HCN, which are produced in large quantities, may have played important roles in determining the suitability of the primitive Earth for the birth of life.

N_2O is a powerful greenhouse gas, and considering that a large amount of it is produced and diffused into the atmosphere of the Earth by frequent superflares, it is suggested that the "young Sun paradox" described earlier may be sufficiently resolved. A warm Earth environment with liquid water, which is necessary for the birth of life, may have been formed and maintained on the early Earth owing to superflares.

HCN played another important role. It may be the starting point for the formation of highly complex compounds and the source of amino acids and other biological molecules. Specifically, it may have been the origin of the substances that form our bodies.

11.1.8.4 Effects of Superflares on Exoplanetary Systems

A new theory (Airapetian et al. 2016) has been proposed that active superflares on the young Sun warmed the atmosphere of the primordial Earth and created the environment and materials necessary for life to begin. Although more quantitative studies are needed in the future, it is a very unexpected finding that although solar flares cause damage to the Earth, they may also have been involved in the development of life on the early Earth, which is the origin of our existence.

Many planetary systems outside our solar system have now been discovered, numbering more than 4900 as of January 2022 (according to the NASA Exoplanet Archive). Moreover, efforts to find a "second Earth" are being actively conducted using numerous space and ground-based telescopes. M-type stars, which are cooler than Sun-like stars, are considered to be the most promising targets because they have more inner habitable planets with the right temperature and are easier to observe in detail. However, the magnetic activity level of M-type stars is generally much higher than that of solar-type stars owing to their lower temperatures and more developed convection zones, as shown by the frequency of flares observed by Kepler and TESS (Hawley et al. 2014; Candelaresi et al. 2014; Davenport 2016; Günther et al. 2020; Maehara et al. 2021).

Thus, studies on the effects of stellar magnetic activities on planets, such as superflares, are also being actively conducted for exoplanet systems (see also Airapetian et al. 2020 and Linsky 2019 for reviews). How molecules important to a life-sustaining environment (e.g., O_3, CH_4, H_2O, nitrogen oxides) are affected (produced or destroyed) by energetic particles associated with superflares (Fig. 11.16), and how the results of these studies will be observed by large telescopes (James Webb Space Telescope and ground-based 30-m telescopes) in the near future (Segura et al. 2010; Scheucher et al. 2018; Tilley et al. 2019; Herbst et al. 2019b; Chen 2021). With the progress in the study of stellar superflares, various interdisciplinary investigations are expected to continue.

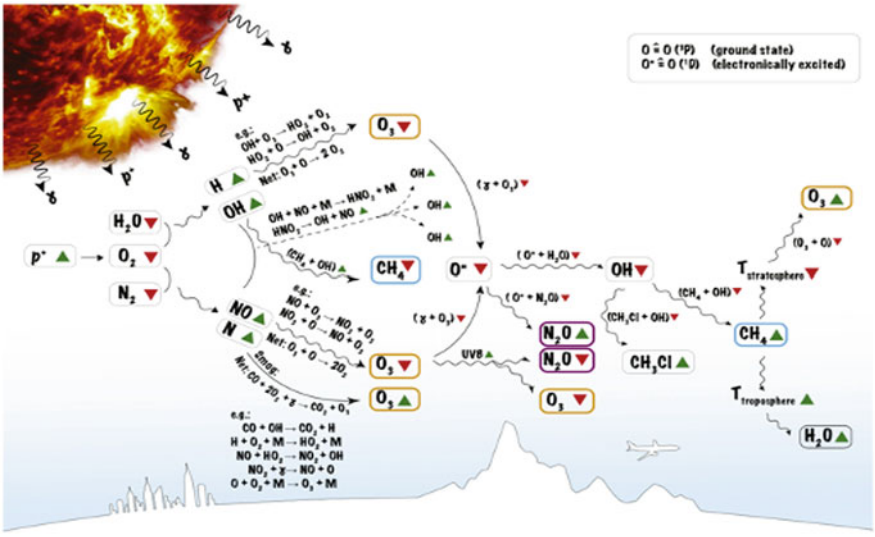

Fig. 11.16 Schematic of chemical reactions in life-related materials caused by energetic particles and ultraviolet light associated with superflare (from Scheucher et al. 2018; © AAS. (Reproduced with permission)

11.1.9 Future Superflare Research

In the first half of this section, we discussed the characteristics of stars that cause superflares based on the latest stellar observation data from the Kepler Space Telescope and other telescopes. It is now shown that even a star like the Sun, which has a long rotation period, could produce large superflares if a huge star spot is generated, although at a low frequency of only once every few thousand years. However, there are still many points to be clarified and investigated before we can answer the important questions such as "Do superflares occur on the Sun?" and "What effects do superflares have on planets?"

First, as shown in Figs. 11.7 and 11.8, the number of slow-rotating superflare stars is small, based on the presently analyzed Kepler Space Telescope data. Moreover, still, no superflare stars have been confirmed by spectroscopic observations to be Sun-like stars with rotation periods of approximately 25 days (see Appendix B of Okamoto et al. 2021). Therefore, further observation and research are important. In this respect, the successors of the Kepler Space Telescope, such as the TESS satellite (launched in July 2018) and the PLATO satellite (to be launched in 2026), are expected to acquire data on many superflare stars in the vicinity that are suitable for spectroscopic observations. This will further increase the accuracy of statistical studies.

Further research on the properties of a large star spot, which is a necessary condition for superflares, is also important. How can such large spots be formed on the slow-rotating Sun? How long are the lifetimes of these giant star spots?

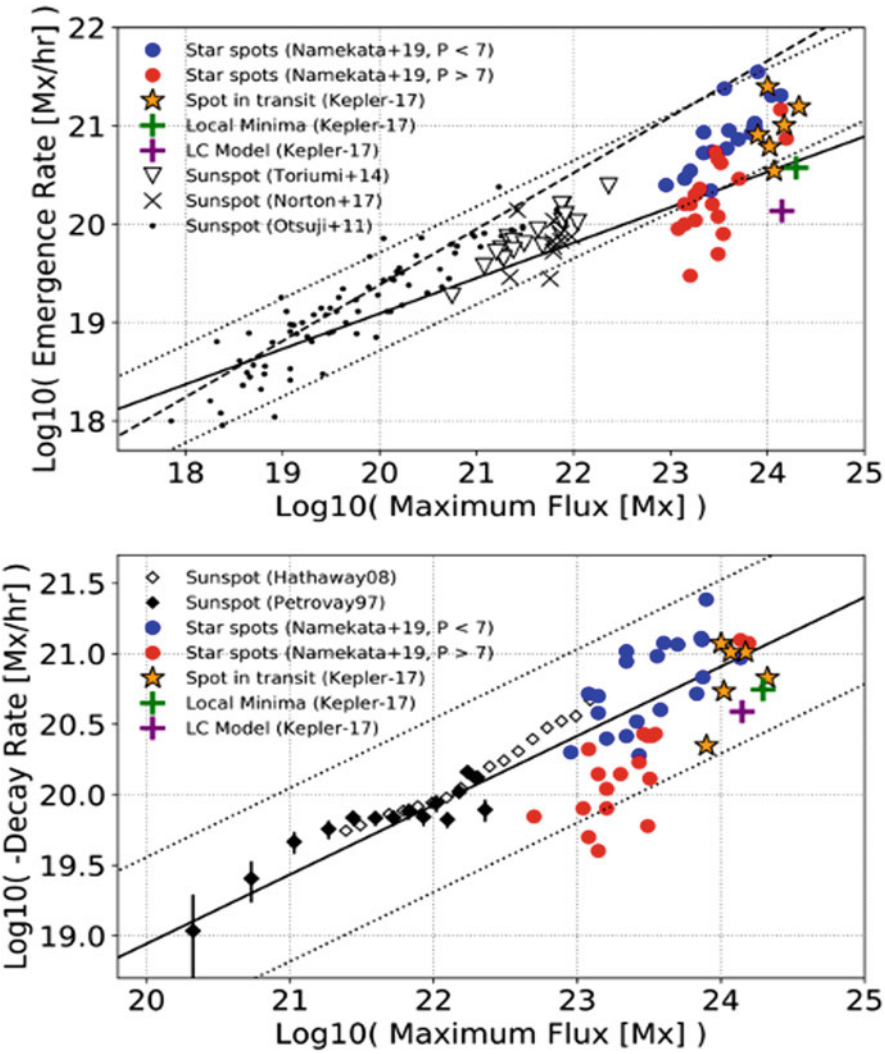

Fig. 11.17 (Upper panel) Maximum value of total magnetic flux of spots versus generation rate of magnetic flux. (Lower panel) Maximum value of the total magnetic flux of sunspots versus the decay rate of magnetic flux. (Both from Namekata et al. 2020a, b) Star spots of solar-type stars observed by the Kepler Space Telescope are located in the general extension of sunspot data. Solar-type stars are color-coded according to their rotation periods (P)

Shibata et al. (2013) proposed the possibility that large sunspots can form in the interior of the Sun based on the dynamo mechanism. Namekata et al. (2019, 2020a, b) using photometric data from the Kepler Space Telescope showed that the formation and extinction rates of large star spots on solar-type stars are in good agreement with those extrapolated from the sunspot formation and extinction rates

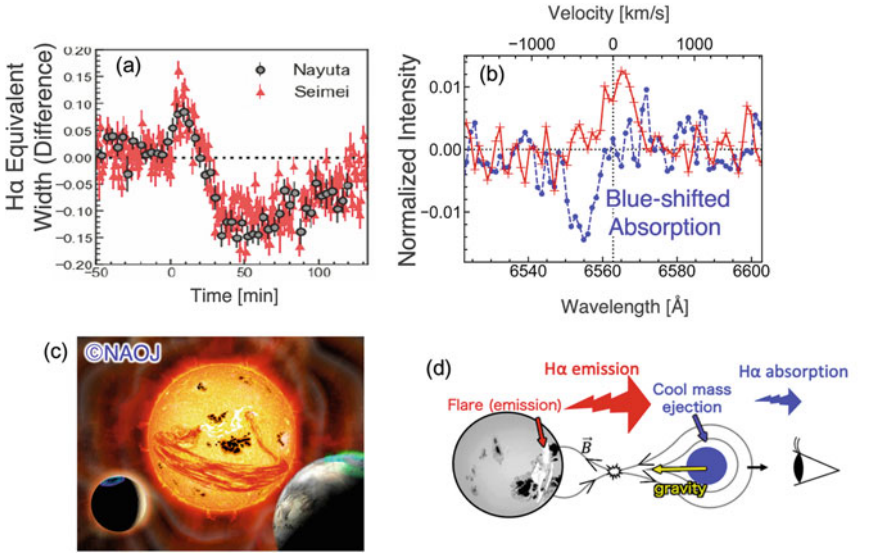

Fig. 11.18 Probable detection of eruptive filament from superflare on young solar-type star EK Dra (from © Namekata et al. 2022a). (**a**) Light curve of equivalent width of hydrogen H-alpha (preflare level is subtracted) from data of two ground telescopes (Seimei and Nayuta). Positive and negative values represent emission and absorption, respectively, compared to the preflare level. At approximately 30 min after the flare peak, H-alpha dimming is caused by possible filament eruptions. (**b**) Preflare-subtracted Hα spectra from Seimei Telescope data at flare peak and in dimming phase. During the dimming phase, blue-shifted absorption caused by possible filament eruptions is observed. (**c**) Artist's depiction of star EK Draconis ejecting gigantic filament (credit: National Astronomical Observatory of Japan). (**d**) Schematic of how an event is observed in Hα (created by Kosuke Namekata for this textbook)

(Fig. 11.17). Moreover, some large star spots have lifetimes of up to 1 year. These results provide indicators for understanding giant, large spots that cause superflares based on the knowledge of solar physics.

However, there are still many ambiguities about giant star spots of superflare stars. For example, in the case of the Sun, in addition to the sunspot size, the "sunspot complexity" is an important condition for generating big flares (Sammis et al. 2000; see also Toriumi and Wang 2019: Chap. 10). In the case of the star spots of superflare stars, what are the "complexities" of the star spot distribution and shape? There have been studies focusing on the phase of the Kepler Space Telescope brightness variations and flare occurrence (Doyle et al. 2018; Roettenbacher and Vida 2018) and studies using exoplanet transits to determine the number of star spots (Morris et al. 2017). Based on these investigations, it is suggested that stellar large star spots consist of numerous star spots with sizes widely distributed on the stellar surface. Further observation studies are needed to confirm whether these are common properties of superflare stars, and theoretical studies on how such star spots are formed are also important. Furthermore, in the case of the Sun, the magnetic activity varies over a period of approximately 11 years; however, what is its relevance in the

case of superflare stars? Observation research from a longer-term perspective is also required.

In addition to large star spots, our understanding of the superflare phenomenon itself is still insufficient. For example, the Kepler Space Telescope observes superflares in "optical continuum light," or i.e., it observes "white light flares." However, we do not know the mechanism of white light emission in the case of solar-type superflares or how much of the total energy released during superflares on solar-type stars is in visible continuum light (cf. the case of solar flares: Emslie et al. 2012). This is partly because the spectroscopic observations of superflares on solar-type stars are still in a very early phase (cf. Namekata et al. 2022a, b). There have been observations of solar white-light flares (e.g., Hudson 2016; Watanabe et al. 2017; Namekata et al. 2017; Jejcic et al. 2018) and spectroscopic observations of visible continuum light in low-temperature M-type stars (e.g., Hawley and Pettersen 1991; Kowalski et al. 2010, 2013; Namekata et al. 2020a, b). However, theoretical considerations are only beginning (e.g., Kowalski and Allred 2018; Heinzel and Shibata 2018; Nizamov 2019; Namekata et al. 2020a, b). In the future, observational considerations will also be important.

Not only continuous light but also the temporal variations in the chromospheric lines during flares are important to understand superflare phenomena based on the knowledge of solar flares. During normal solar flares, the chromospheric lines (e.g., Hα lines) are known to show "red asymmetries," mainly due to the emission from the downward flows of the chromosphere (Ichimoto and Kurokawa 1984). However, many spectroscopic observations of M-type stellar superflares have also presented "blue asymmetries" in the chromosphere (Fuhrmeister et al. 2018; Honda et al. 2018; Vida et al. 2019; Maehara et al. 2021). Is this "blue asymmetry" a common phenomenon of superflares? Moreover, how common or different are the dynamics of the chromosphere in solar flares and superflares? Further observation exploration is needed to understand these points.

Moreover, as described in Sects. 11.1.7 and 11.1.8, the possibility that CMEs associated with superflares have significant impacts on planetary atmospheres is generating interest. If we can determine the physical quantities (mass and velocity) of CMEs associated with superflares, we can assign limits to the intensity and spectra of solar/stellar energetic particles (SEPs), as shown in Figs. 11.12 and 11.13. These are important factors in studying the chemical reactions and radiation intensity of planetary atmospheres, as presented in Figs. 11.14 and 11.16. The limits are also important in examining the relationship (e.g., comparison of frequencies) between superflares and traces of cosmic ray events (e.g., ^{14}C events) that remain on the Earth, as described in Chap. 13. In view of this scenario, some studies and observations of CMEs from superflare stars (particularly M-type stars with frequent flares) are being conducted. Some examples are methods focusing on the dimming in extreme ultraviolet (Harra et al. 2016; Veronig et al. 2021), searches for Type-II radio bursts associated with CMEs (Crosley and Osten 2018), and blue shift of the chromospheric lines in flares (Vida et al. 2019; Maehara et al. 2021; Namekata et al. 2022a). Other instances are studies focusing on the blue shift of the X-ray emission lines in flares (Argiroffi et al. 2019) and on X-ray absorption (Moschou 2019). The

blue shift method has been successfully used to detect the blue-shifted absorption components associated with possible filament eruptions (early phase of CMEs) from young solar-type stars (Namekata et al. 2022a; Fig. 11.18) as well as M-type stars. Further research is required to synthesize these results and compare them with theoretical considerations (Alvarado-Gómez et al. 2018; Lynch et al. 2019; Sun et al. 2021) and others.

As summarized, in future superflare research, dedicated observations that intensively invest a large amount of observation time, such as observations of the time variations in large star spots and spectroscopic observations of superflares themselves, will be important. For example, in Japan, the 3.8-m Okayama Observatory of Kyoto University, which recently started its operation in 2019, is strongly expected to play a major role in the future (cf. Namekata et al. 2022a, b). In addition, there are several space telescopes that are anticipated to collaborate with ground-based telescopes such as the Seimei Telescope. These will include photometric observations by the TESS (started regular observations in July 2018) and PLATO (scheduled launch in 2026) satellites as well as X-ray observations by the NICER mission (started observations in July 2017) and the XRISM (a Japanese X-ray satellite scheduled for launch in 2023). With the progress of these new observations, further collaboration with the solar physics field and cross-disciplinary research with the planetary field (cf. Airapetian et al. 2020; Herbst et al. 2019b) are expected to develop in the direction of superflare research. Moreover, in estimating the impacts on the Earth of a superflare on the present Sun (cf. Eastwood et al. 2017; Lingam and Loeb 2017; Riley et al. 2018), as mentioned in Sect. 11.1.7, it is of critical importance that the entire field of solar–terrestrial research broadly collaborates.

11.2 Extreme Space Weather Events in History

11.2.1 Introduction

As discussed in Chaps. 1 and 2, intense solar eruptions frequently cause severe space weather disasters that significantly affect various modern technological infrastructures such as satellites, power grids, and communication networks. In the last two centuries, the Carrington event in September 1859 has been considered the most extreme space weather event from multiple aspects (Tsurutani et al. 2003; Cliver and Dietrich 2013). If it were to occur in modern times, the equivalent solar superstorm would severely impact the various technological infrastructures on which modern civilization heavily depends (Baker et al. 2008). Hapgood and Thomson (2010, p. 15) warned of the threat of such a superstorm: “Sustained loss of power could mean that society reverts to nineteenth-century practices.”

However, such extreme space weather events occur only infrequently. Statistical analyses have generally shown their occurrence frequency as at best once every century. This low occurrence frequency has also been a major barrier to quantitatively study extreme space weather events (Riley et al. 2018). This is partially

because the operation of systematic quantitative databases on solar eruptions and the resulting magnetic storms mostly started only since the International Geophysical Year (IGY: 1957–1958). Therefore, to understand such low-frequency high-impact space weather events, it is necessary to exploit alternative datasets for extreme space weather events on a long-term basis.

Here, we review three indicators of past space weather events on the basis of archival investigations. First, sunspot records show source active regions of solar eruptions. Second, geomagnetic measurements show disturbances from the resulting magnetic storms. Finally, auroral reports in mid- to low-magnetic latitudes (MLATs) represent an equatorward extension of the auroral ovals. The magnitudes of magnetic storms have frequently been directly assessed using the most negative excursion of the Dst index, as an average of the geomagnetic disturbance variations measured at four mid-latitude stations, since the IGY (e.g., Sugiura 1964; Sugiura and Kamei 1991). Prior to this index, analyses of multiple analog magnetograms allow deriving the Dst estimate. Additionally, it has been empirically shown that the equatorward boundary of an auroral oval correlates well with the Dst index (Yokoyama et al. 1998; Blake et al. 2021), which, in turn, has allowed estimation of storm magnitudes. In this section, we present recent developments on past space weather event reconstruction from these historical records.

11.2.2 Carrington Event

The Carrington event of September 1859 has been generally associated with one of the earliest reported white-light flares (1 September 1859) in observation history (Carrington 1859; Hodgson 1859). This solar flare most probably released an extremely fast and massive interplanetary coronal mass ejection (ICME), which caused one of the largest geomagnetic storms on record. This is evidenced by the geomagnetic disturbances of -1600 nT at Bombay and the equatorial boundary of the auroral visibility down to Honolulu (Chapman 1957; Kimball 1960; Tsurutani et al. 2003; Cliver and Dietrich 2013).

However, the Carrington event (1859) was chronologically located before the onset of most modern instrumental observations and accommodated only fragmentary details, which has inevitably resulted in debates over the magnitude estimates for multiple aspects. The soft X-ray (SXR) fluence for the source flare has been estimated mainly from the amplitudes of the reported solar flare effect (SFE), which is caused by the change in the electrical conductivity of the ionosphere on the arrival of X-rays from the solar eruptions. At that time, the SFE amplitudes were measured as ≈ -110 nT in the horizontal (H) components of the SFE at Greenwich and Kew observatories in the United Kingdom (UK) (Fig. 11.19). Based on these extreme SFE amplitudes, the SXR fluence was initially estimated as $>$ X10 in comparison with a giant flare in November 2003 (Cliver and Svalgaard 2004). Subsequently, it was estimated as $\approx$X45 $\pm$ 5 according to statistical analyses of the SFE amplitudes

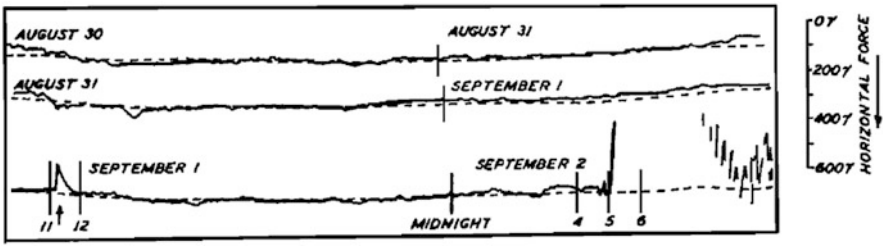

Fig. 11.19 The Kew magnetogram showing SFE and magnetic storm during the Carrington event (Bartels 1937)

and their source SXR flare fluence (e.g., Boteler 2006; Cliver and Dietrich 2013; Curto et al. 2016).

Magnetic storms have been measured with negative excursion in the geomagnetic H component at mid to low latitudes over a period of several days. For the Carrington event, there were only few magnetograms which managed to measure the contemporary geomagnetic disturbances in mid- to low-latitude regions without significant data gaps. The magnetogram at the Colaba Observatory (Bombay in India) was one of such few cases. These geomagnetic disturbances mainly originated from a ring current development. Different from its name, the equatorial ring current is asymmetric and shows a local time dependency, being stronger on the nightside and weaker on the dayside (Ebihara et al. 2002; Le et al. 2004). The field-aligned currents have also been suggested to explain the negative excursion that took place at the Colaba Observatory on the dayside (Ohtani 2022). During the Carrington event, the Colaba magnetogram showed a sharp and large negative excursion of −1600 nT in the geomagnetic H component (Tsurutani et al. 2003). If this geomagnetic disturbance originated from a ring current, the minimum Dst value could be −1760 nT, considering the averaged travel time from Sun to Earth and empirical models describing the relationship between solar wind speed and interplanetary magnetic field (IMF) as well as evolution of Dst (Tsurutani et al. 2003). These estimates were immediately subjected to controversial interpretations, as it is extremely rare to observe such a large negative H excursion at mid- and low latitudes (Akasofu and Kamide 2005).

However, by definition, the Dst index uses an average of four magnetograms at mid- to low-MLATs on the basis of their hourly values, removing the background diurnal variations (Sugiura and Kamei 1991). Siscoe et al. (2006) initially clarified this problem, approximated an hourly average of the Colaba ΔH as ~ −850 nT, and found Colaba at the peak LT as a good proxy for the Dst by coincidence. Later on, Gonzalez et al. (2011) computed the Colaba ΔH as −1050 nT. Cliver and Dietrich (2013) found the range of the min Dst estimate of the Carrington storm as −900 (+50, −150) nT on their basis. Recent archival investigations have allowed us to access copies of Colaba yearbooks and reconstruct time series of the three geomagnetic components. On their basis, Colaba witnessed the disturbance variation up to as ~ −949 ± 31 nT (Fig. 11.20), owing to conflicting peak values in the hourly spot

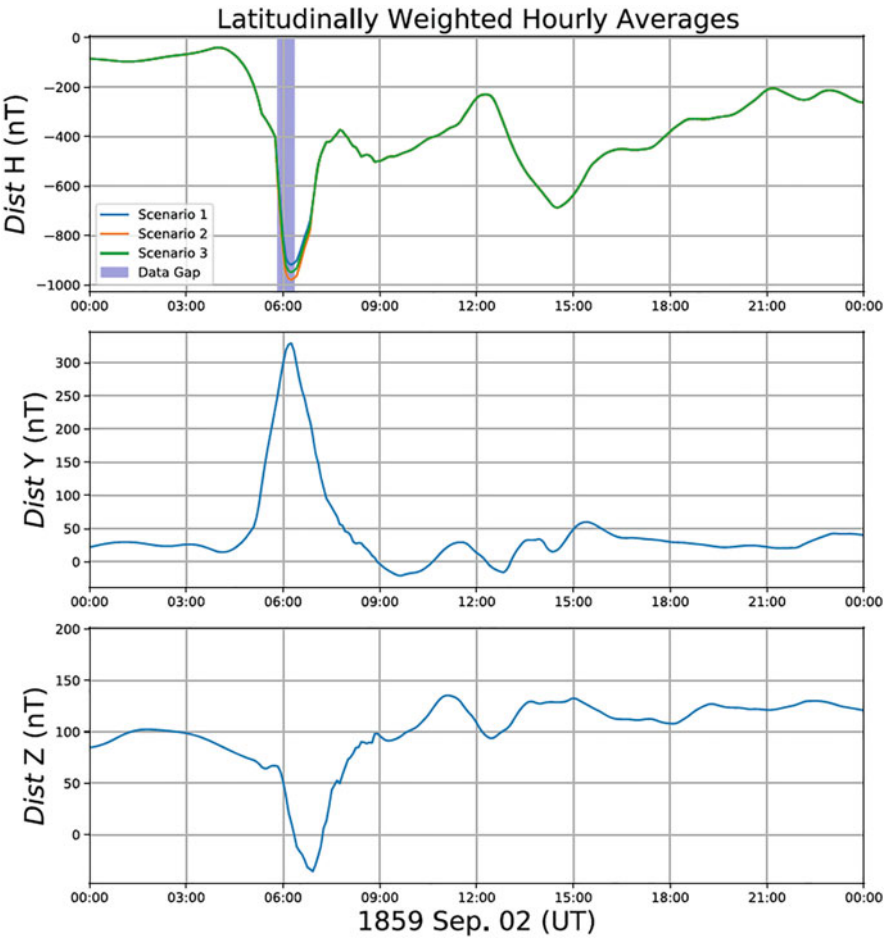

Fig. 11.20 Disturbance variations of three geomagnetic components (H, Y, and Z) at Colaba Observatory during the Carrington magnetic storm, as reconstructed from Fergusson (1860) in Hayakawa et al. (2022a), as reproduced from the said article. (From Hayakawa et al. 2022a, under CC-BY)

measurement and disturbance spot measurement in the Colaba yearbook (e.g., Hayakawa et al. 2022b; c.f., Cliver and Dietrich 2013). This approximates the minimum Dst estimate (Dst*) for the Carrington magnetic storm, whereas caveats must be noted on the uncertainty of their MLT variations. Fortunately, Siscoe et al. (2006) have shown that the LT in Colaba is by coincidence a good proxy for the Dst estimate (see Fig. 11.2 of Siscoe et al. 2006). Although several geomagnetic measurements are known for the same period, their majority had significant data gaps, which made it difficult to further improve our understanding of the current system at that time. It is possible that the effects of local geomagnetic enhancements

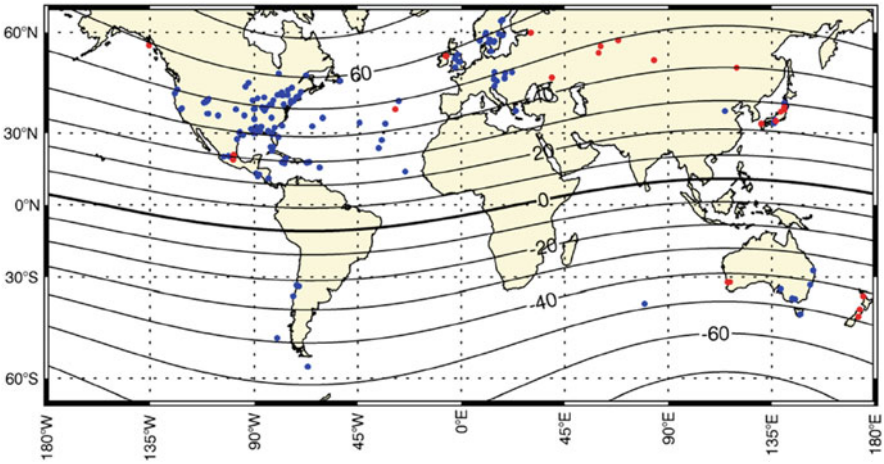

Fig. 11.21 Visible auroral coverage around the time of the Carrington event (1/2–2/3 September 1859) (from © Hayakawa et al. 2019e). Blue dots indicate known observation sites, and red dots represent sites added by Hayakawa et al. (2019e). Black contours represent MLAT

were mixed in with the geomagnetic variations at Bombay (e.g., Love et al. 2019a). Further investigations are needed for contemporary observational records.

The auroral ovals are known to expand equatorward during magnetic storms (Akasofu 1964). There is a good empirical correlation between the equatorward boundary of the auroral oval and the storm magnitude in the Dst index (Yokoyama et al. 1998; Blake et al. 2021). The equatorward boundary of the auroral oval during the Carrington event has been discussed extensively since contemporaneous studies such as Loomis (1860). The auroral visibility has been confirmed down to |23|° MLAT based on the records from Honolulu and Santiago (Tsurutani et al. 2003) and then down to |17|° MLAT based on the naval report from the Pacific Ocean (Hayakawa et al. 2020b). Further investigations have presented auroral records in East Asia, Russia, New Zealand, Mexico, the Iberian Peninsula, and Chile (González-Esparza and Cuevas-Cardona 2018; Hayakawa et al. 2019e, 2020b), as shown in Fig. 11.21 (see also Hayakawa et al. 2020b). To reconstruct the equatorward boundary of the corresponding auroral oval from these records, we need to consider not only the MLAT of these observational sites but also the local elevation angles at the said sites (Fig. 11.22). In this context, great auroral displays were reported almost overhead at Santiago. On this basis, it is possible to reconstruct the equatorward boundary of the auroral oval as $\approx 25.1 \pm 0.5°$ invariant latitude (ILAT) by assuming that the magnetic field of the Earth is a dipole field and the witnessed red aurora is 400 km of altitude (Roach et al. 1960; Ebihara et al. 2017). ILATs are constant along the magnetic field lines and are consistent with the MLAT at the footprint of the magnetic field lines which intersect the terrestrial surface for the dipole magnetic field. As charged particles tend to move along the magnetic field lines, it is useful to employ ILAT for the purpose of describing the distribution of the

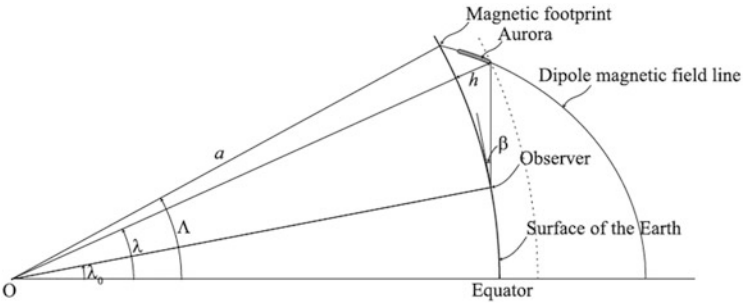

Fig. 11.22 Low-latitude boundary between auroral visible range and associated auroral oval, as reproduced from © Hayakawa et al. (2018e), where λ_0, λ, Λ, a, h, and β are MLAT of the observer, MLAT of the auroral upper edge, ILAT, radius of the Earth, height of the auroral upper edge, and elevation angle of the auroral upper edge, respectively, from the observer

charged particles trapped in the magnetic field of the Earth and the resulting aurora (Hayakawa et al. 2020b).

Thus, the Carrington event has been quantitatively reconstructed as having at least the X-ray intensity of a flare ($\approx$ X45 $\pm$ 5), the size of a magnetic storm (minimum Dst* $\approx -949 \pm 31$ nT), and the equatorward boundary of the auroral oval ($\approx 25.1°$ ILAT). Accordingly, the Carrington event is positioned as one of the most extreme space weather events in history, whereas it is likely not unique anymore, as shown in the next section (Tsurutani et al. 2003; Cliver and Dietrich 2013; Hayakawa et al. 2022a).

It is a close-up sunspot drawing of Carrington (1859) that was used to visualize the solar surface around the Carrington flare. However, recent archival investigations have revealed a number of contemporary sunspot drawings, including those of Carrington himself, as shown in Fig. 11.23 (Hayakawa et al. 2019d). Analyses of these sunspot drawings have shown that the source sunspot active region of the Carrington event was in the eastern hemisphere of the Sun on 28 August, around the solar central meridian from 31 August to 1 September, and over the western rim of the Sun around 7 September. The source sunspot position for the August storm was not favorable located to direct geoeffective ICMEs toward the Earth. In contrast, this giant sunspot was near the central meridian around the white-light flare on 1 September and favorably located for a geo-effective ICME (Hayakawa et al. 2019e).

11.2.3 Extreme Space Weather Events Captured in the Early Modern Observations

It remains debated how unique the Carrington event is, as several "failed" Carrington events have been detected, analyzed, and highlighted in recent studies. For example, a huge ICME was released in the opposite direction to the Earth in July 2012, which

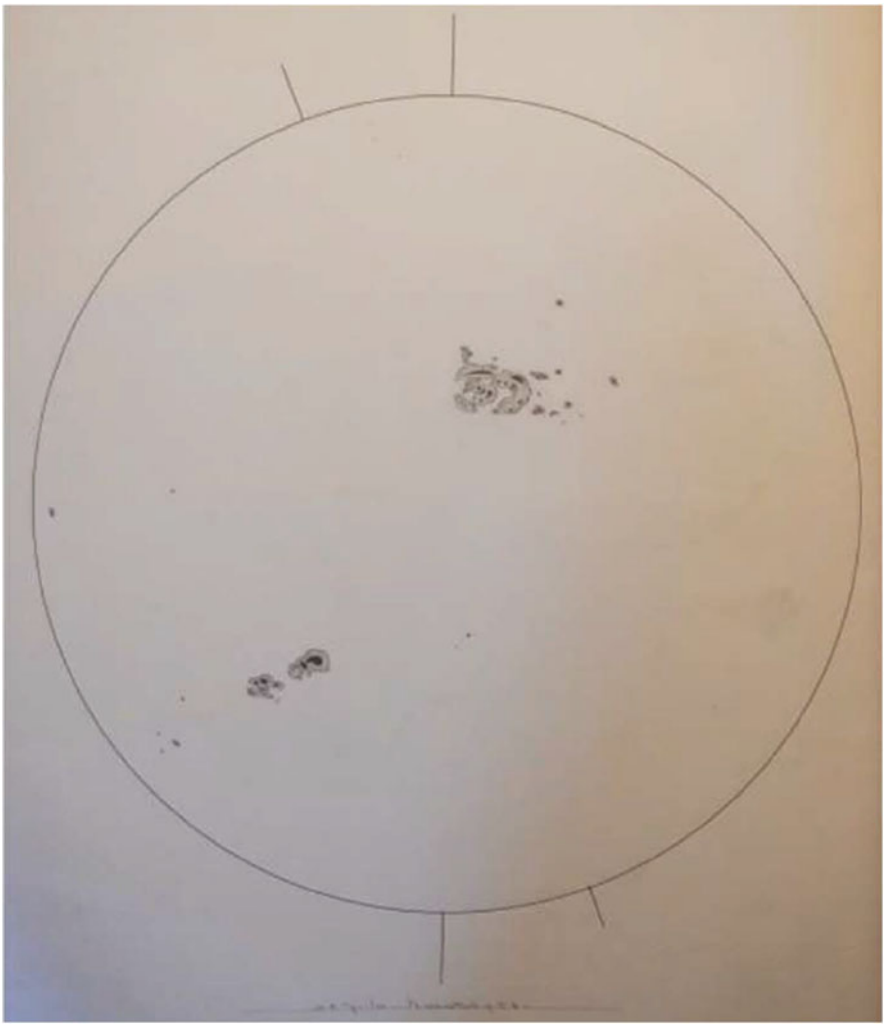

Fig. 11.23 Sunspot sketch handwriting in Carrington at time of the Carrington event (© the Royal Astronomical Society, MS Carrington 3.2, f. 313a). Limb and contrast are enhanced, and the direction of the projection method is corrected (Toriumi and Wang 2019; Hayakawa et al. 2019e)

could have generated an extreme magnetic storm with an estimated minimum Dst ≈ −1182 nT, if it were to hit the Earth with geoeffective IMF orientation (Baker et al. 2013; Ngwira et al. 2013). The ICME in August 1972 was the fastest in observational history and reached the magnetosphere of the Earth in 14.6 h (Russell et al. 2013). Nevertheless, the magnitude of the resultant magnetic storm was modest (minimum Dst ≈ −125 nT), arguably owing to the northerly orientation of the IMF and/or asymmetric geomagnetic activity (Tsurutani et al. 2003; Knipp et al. 2018).

It is also worth considering how unique the Carrington event was based on the early modern observations as well. It was in the 1830s when systematic modern observations of geomagnetic variations and auroras began (e.g., Chapman and Bartels 1940). These records have allowed us to consider a space-weather chronology of less than two centuries. Interestingly, Chapman (1957) contrasted the "outstanding auroras" in the Carrington event of September 1859 with those in February 1872 (Bombay), in September 1909 (Singapore), and in May 1921 (Samoa). These unique cases have been subjected to analyses on their storm magnitude, their equatorward auroral visibility, and their equatorward boundary of the auroral oval (e.g., Silverman and Cliver 2001; Hayakawa et al. 2019d).

The magnetic storm of 4 February 1872 occurred immediately after the maximum of solar cycle 11. Auroras were allegedly observed down to Bombay and robustly confirmed down to MLAT $\approx 19°$, including northern India, western Europe, the Indian Ocean, the Middle East, North and South Africa, and East Asia (Silverman 2008; Hayakawa et al. 2018c). Reliability of the Bombay report was debated owing to the contrast between the extremely low MLAT of Bombay ($\approx 10°$) and the precise and reasonable description (Silverman 2008). However, analysis of auroral records from East Asia has shown that the aurora probably extended approximately 10–15° elevation from the local horizon at Bombay, on the basis of an auroral extension to the local zenith at Shanghai (MLAT $\approx 20°$) and the assumption of the similar spatial extension of the auroral oval in the Indian sector and an auroral elevation of $\approx$ 400 km (Hayakawa et al. 2018c). The auroral observations chronologically coincided with the peak of the magnetic disturbance in Bombay. The Bombay magnetogram suggests the magnetic disturbance has a minimum Dst* < -830 nT.

The space weather event of 25 September 1909 occurred during the declining phase of solar cycle 14. The flare associated with this space weather event occurred at 11–12.5 UT as recorded around the United Kingdom in the optical observations and magnetic measurements (Lockyer 1909; Hayakawa et al. 2019a; Love et al. 2019a). This flare released an ICME and caused a geomagnetic disturbance after 24.75 h. The size of this flare has been estimated as $\geq$ X10 based on the CME propagation time and the synchronized SFE amplitude (Hayakawa et al. 2019a). The magnitude of this magnetic storm has also been reconstructed as minimum Dst* ≈ -595 nT based on magnetic observations from Mauritius, San Fernando, Vieques, and Apia (Love et al. 2019a). The equatorward boundary of the auroral oval has been reconstructed to be $\approx 32°$ ILAT, because auroras were visible at their zenith in Hokkaido and up to 30° elevation in Matsuyama (Hayakawa et al. 2019a). Their temporal evolutions agree significantly well with one another, as shown in Fig. 11.24. Auroras were extensively observed in Japan and Australia, probably because these countries were located in the midnight sector during the main and the recovery phases of this magnetic storm.

The space weather event of 14–15 May 1921 originated from a giant sunspot that appeared near the solar equator in the late declining phase of solar cycle 15 (Newton 1948). The exact eruption chronology is still debated for this magnetic storm (Lundstedt et al. 2015; Lefèvre et al. 2016; Love et al. 2019b). The Sun had been particularly eruptive during 12–20 May and caused at least six sudden storm

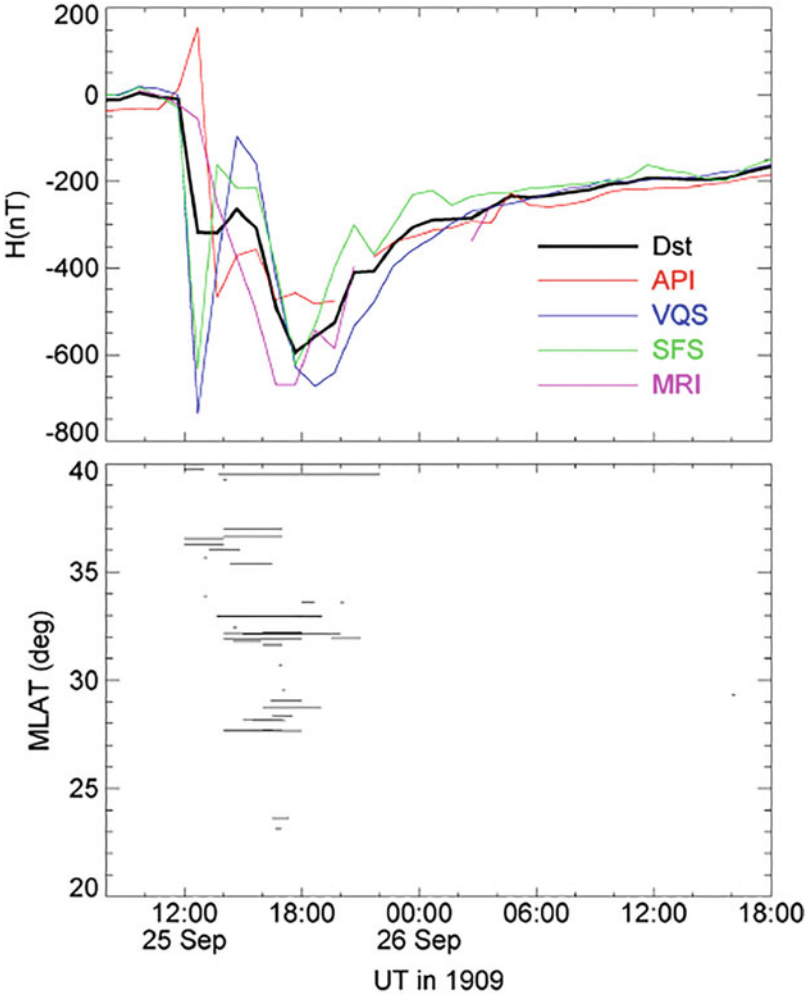

Fig. 11.24 Temporal evolutions of the geomagnetic disturbances (Dst estimate, API (Apia), VQS (Vieques), SFS (San Fernando), and MRI (Mauritius) and low-latitude auroral displays ($< |40|°$ MLAT), as reproduced from © Hayakawa et al. (2019a)

commencements (Newton 1948). Therefore, it is considered that a series of CMEs released during the multiple flares at this time swept the interplanetary space and created preconditioned extreme magnetic storms (Love et al. 2019b). Of these, the magnetic storm on 14–15 May was particularly large to extend auroral visibility down to Samoa Island in the South Pacific Ocean (Silverman and Cliver 2001) and caused telephone line failures in New York and elsewhere, as well as fires at railway stations (Hapgood 2019; Love et al. 2019b). Magnitude of this geomagnetic superstorm was estimated as minimum Dst* $\approx -907 \pm 132$ nT (Love et al.

Table 11.1 Equatorward boundaries of auroral visibility and auroral ovals (absolute values) and magnitudes of magnetic storm (Dst index) for space weather phenomena during modern observations, updated from Hayakawa et al. (2019e), on the basis of subsequent studies (Hayakawa et al. 2020a, b, 2022a). Magnetic storm scale estimates marked with * are based on estimates from magnetic field observations at Bombay, and not the average of four low-latitude stations

Date			EB of auroral visibility (MLAT)	EB of auroral oval (ILAT)	Storm magnitude (Dst: nT)
Year	Mon	Date			
1859	8	28/29	20.2	36.5	≥ -484*
1859	9	1/2	−17.3	25.1±0.5	≈ -918, −949, or −979
1872	2	4	10.0/18.7	24.2	<-830*
1903	10	30	40	44.1	−531
1909	9	25	10.0/23.1	31.6	−595
1921	5	14/15	16.2	27.1	-907 ± 132
1946	3	28	<41.8	<41.8	≤ -512
1989	3	13/14	29	35/40.1	−589

2019b) based on magnetic observations from Apia, Watheroo, San Fernando, and Vassouras. Moreover, the equatorward boundary of the auroral oval was estimated at 27.1° ILAT based on the Apia report (Hayakawa et al. 2019e).

The magnetic storm magnitudes and equatorward auroral boundaries for these extreme space weather events and the March 1989 extreme space weather event are summarized in Table 11.1. This table involves two more geomagnetic superstorms (minimum Dst* < -500 nT) in October 1903 (Fig. 11.25) and March 1946, for which time series and intensity were reconstructed in recent studies (Hayakawa et al. 2020a, b). The October 1903 storm is especially notable because it occurred immediately after a minimum of solar cycle 14, which was one of the smallest solar cycles since the Dalton Minimum (Clette et al. 2014). As confirmed in this table, the Carrington event is hardly the largest and most unique in the observation history but is at best comparable to the extreme space weather events of February 1872 and May 1921, at least in terms of the equatorward auroral boundaries and the minimum Dst estimates.

11.2.4 Extreme Space Weather Events Recorded in Early Modern Observations and Historical Records

Before the modern era of observations (1830s-), extreme space weather events were occasionally recorded in the variations in the concentrations of the cosmogenic isotopes (^{14}C in tree rings and ^{10}Be and ^{36}Cl in ice cores) and reports of sunspots

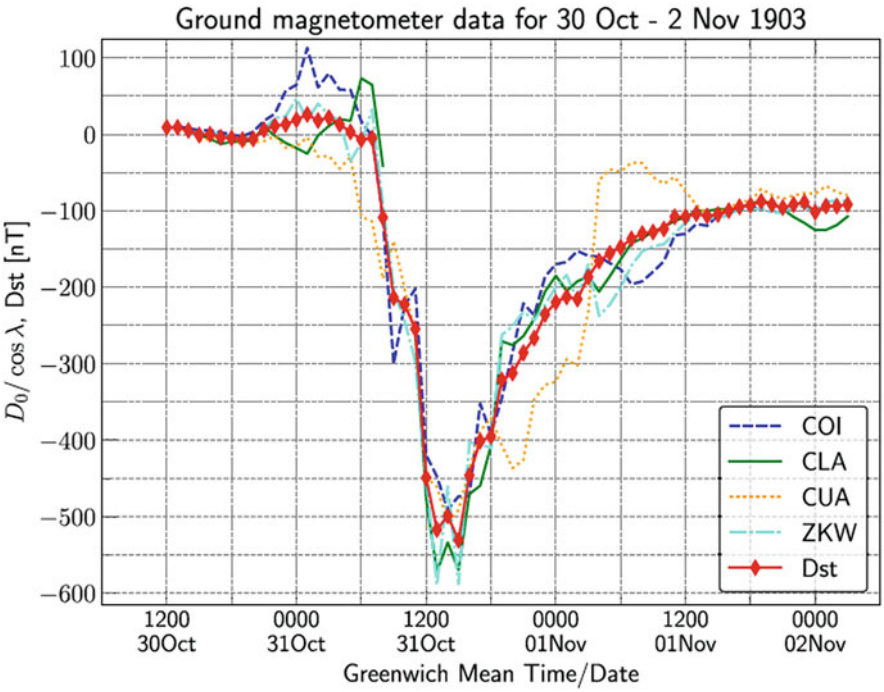

Fig. 11.25 Dst estimate for the geomagnetic superstorm in October–November 1903, on the basis of COI (Coimbra), CLA (Colaba), CUA (Cuajimalpa), and ZKW (Zi-ka-wei). (Reproduced from © Hayakawa et al. (2020b))

and candidate auroras in the historical documents (e.g., Vaquero and Vázquez 2009; Usoskin 2017). Of these, the former is used to identify the footprints of solar energetic particles (e.g., Beer et al. 2012; Usoskin and Kovaltsov 2012; Barnard et al. 2018). In contrast, historical documents have been used to identify footprints of ICMEs from auroral records (e.g., Willis and Stephenson 2000; Hayakawa et al. 2017a, b). Naked-eye sunspot records allow us to identify naked-eye sunspots (giant sunspots) as solar sources of such space weather events (e.g., Willis and Stephenson 2001; Willis et al. 2005; Hayakawa et al. 2017e).

Among such space weather events, those from 1610 can be particularly used for comparisons between historical documents and modern observations, as the sunspot observations have already started (Clette et al. 2014). For example, low-latitude auroras were also observed in East Asia during the space weather events in September 1859 and February 1872 and were recorded as various historical terms such as “red vapor,” “white vapor,” and “red light” (Willis et al. 2007; Hayakawa et al. 2016b, 2018d). Giant sunspots in November 1769 and February 1917 have also been recorded by several historical terms such as “black spots” and “black vapors” (Willis et al. 1996a, 2018; Hayakawa et al. 2019c).

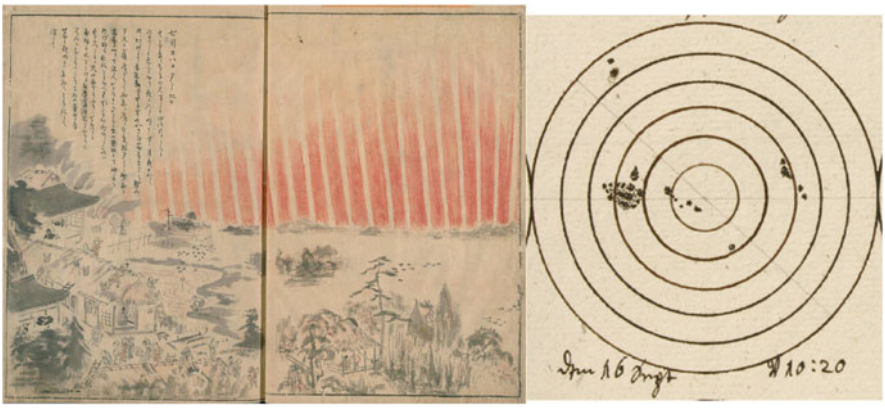

Fig. 11.26 Aurora observed in Nagoya on 17 September 1770 (© National Diet Library) and sunspot sketch on 16 September 1770. (© Potsdam AIP: Arlt 2008)

These terminologies allow us to collect footprints of space weather events in historical documents, especially before the chronological coverage of the systematic geomagnetic measurements. In September 1770, auroral displays were recorded for approximately nine nights in succession across East Asia, down to Lake Dòngtíng (18.9° MLAT) in the Northern Hemisphere and down to Timor (−20.5° MLAT) in the Southern Hemisphere (Fig. 11.26; Willis et al. 1996b; Hayakawa et al. 2017e). The aurora on 17 September was also described as "bright as a moonlit night," which was considered to be a result of a large precipitation of high-intensity low-energy electrons (Ebihara et al. 2017). The reconstruction of the equatorward auroral boundary on 17 September remains somewhat controversial (Ebihara et al. 2017; Kataoka and Iwahashi 2017). Further careful discussions are needed based on the position of the constellation at the exact time and contemporaneous data. Around these extreme space weather events, a giant sunspot group was found on the solar surface, according to Johann C. Staudach's sunspot drawings. This giant sunspot group probably triggered a series of ICMEs which resulted in a long series of low-latitude auroras in East Asia (Hayakawa et al. 2017d; Fig. 11.26).

In February 1730, auroras were also reported in the East Asia sector as far away as Kyoto (25.4° MLAT). A review of sunspot drawings from the same period confirms a significant increase in the sunspot group number in the first half of 1730, indicating enhanced solar activity at that time (Hayakawa et al. 2018c, 2022b).

11.2.5 Space Weather Events Forgotten in Historical Documents

As such, past space weather events are occasionally recorded as candidate auroras in historical documents. These records allow us to trace back the space weather

Fig. 11.27 Trace copies of candidate auroral records from Babylon and Assyria; (**a**) a candidate auroral record in 567 BCE in the Astronomical Diaries from Babylon; and (**b**) another candidate auroral record in 679–655 BCE from the Assyrian Astrological Report. (Reproduced from © Hayakawa et al. (2019e))

chronology and scale magnitudes of the associated magnetic storms. Such historical documents are particularly important for discussions on low-frequency high-impact space weather events. For example, recent analyses have revealed several largest solar particle storms in 7176 BCE, 5259 BCE, 660 BCE, 774/775 CE, and 992/993–993/994 CE, which are unprecedented in observational history, on the basis of cosmogenic isotope data from tree rings and ice cores (e.g., Miyake et al. 2012, 2013; Usoskin et al. 2013; Mekhaldi et al. 2015; Büntgen et al. 2018a, b; Uusitalo et al. 2018; O'Hare et al. 2019; Paleari et al. 2022; Brehm et al. 2022). However, all these extreme space weather events occurred long before the beginning of systematic geomagnetic measurements and telescopic sunspot observations in the last two and four centuries, respectively (Chapman and Bartels 1940; Owens 2013; Clette et al. 2014). Therefore, it is important to survey historical records for space weather events around these anomalous events.

As far as currently documented, candidate auroras in the historical records can be traced back to the tenth century B.C. on a dateable basis (Van der Sluijs and Hayakawa 2022). For a while, it had been considered that the earliest datable auroral record in existence is considered to be that of 12/13 March 567 BCE, recorded in the Astronomical Diaries from Babylon (Stephenson et al. 2004; Hayakawa et al. 2016c; Fig. 11.27a). In contrast, some consider the Biblical vision of Ezekiel to be the

earliest auroral record (e.g., Siscoe et al. 2002; Silverman 2006), whereas further careful philological discussions are needed here. Following the original descriptions, this vision has been dated on 12/13 July 594 BCE or 30/31 July 593 BCE, in case the description is truly reliable (Hayakawa et al. 2019e).

However, recent analyses have further improved space-weather chronology using ancient historical records over Eurasia. Astrological observations had been already operated in Assyria from the seventh century B.C. (Hunger 1992). Their astrological reports have involved at least three references to candidate auroras (Fig. 11.27b). Although these astrological observations are not strictly datable, it is possible to confine the date ranges of these reports to 679–655 BCE, 677–666 BCE, and 679–670 BCE, respectively, following the chronological ranges of the astrologers who signed these reports (Hunger 1992; Hayakawa et al. 2019e).

Chinese records allow us to trace the space-weather chronology three more centuries further into the past, as shown in Van der Sluijs and Hayakawa (2022). Philological analyses have identified a celestial event in the northern night sky in the last year of King Zhāo of Zhōu Dynasty (Fig. 11.28) as "five-colored light" and dated this event to 977 ± 1 BCE or 957 ± 1 BCE. Its event description supports auroral interpretation on the basis of its similarity with the early modern accounts for visual auroras. This report forms a unique space-weather reference before the onset of the Neo-Assyrian Grand Minimum (810–720 BCE), as visualized in Fig. 11.29.

The earliest datable drawings of candidate auroras are found in the Vatican autograph manuscript of the Zūqnīn Chronicle of 771/772 (MS Vat. Sir. 162). This report and its parallel description indicate extreme geomagnetic storms in 771/772 and 773 which extended auroras with distinct ray structures and multiple colorations. In general, it is more difficult to copy graphical records than to copy written records, as can be seen in the example of the sunspot drawings (Hayakawa et al. 2018a; Fujiyama et al. 2019). Therefore, it is important to assess if these graphical records are autographs, copies, or hearsays with careful philological analyses (Uchikawa et al. 2020).

As with candidate auroras, records of giant sunspots observed with the naked eye have been found in historical documents over millennia. Currently, records of naked-eye sunspot observations are considered to date back to at least 165 BCE in ancient China (Yau and Stephenson 1988; Vaquero and Vázquez 2009; c.f., Bicknell 1967). The majority of premodern naked-eye sunspot observations have been recorded in East Asia, although some have been recorded elsewhere (Yau and Stephenson 1988; Vaquero and Vázquez 2009; Hayakawa et al. 2017c). In fact, the earliest known sunspot drawing is dated on 8 December, 1128, in John of Worcester's Chronicle in Fig. 11.30 (MS Oxford CCC 157; Stephenson and Willis 1999). This naked-eye sunspot is considered to have caused an extreme space weather event, based on comparisons with a red vapor record on 13 December, 1128, in the *Goryeosa* (Willis and Stephenson 2001).

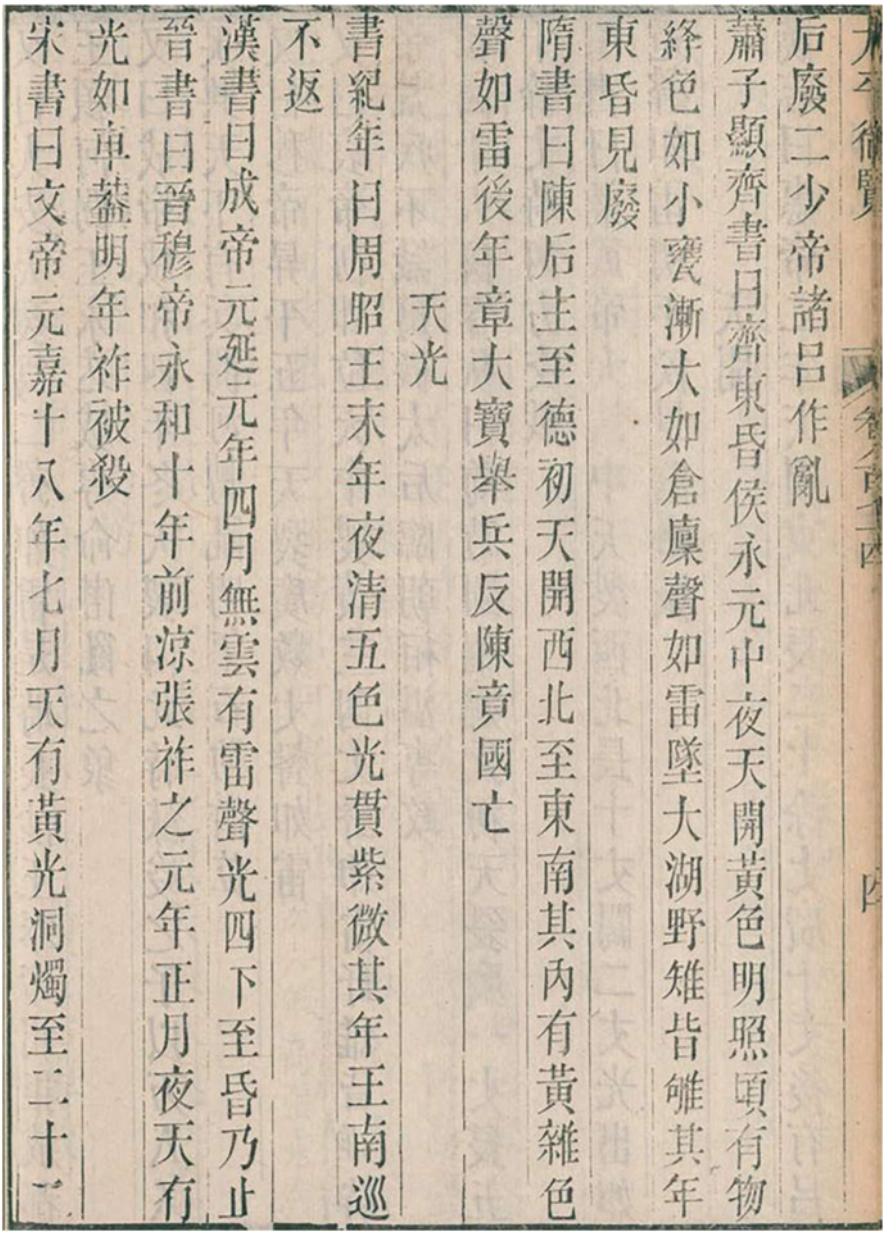
后廢二少帝諸呂作亂
蕭子顯齊書曰齊東昏侯永元中夜天開黃色明照頃有物
終色如小甕漸大如倉廩聲如雷墜大湖野雉皆雊其年
東昏見廢
隋書曰陳后主至德初天開西北至東南其內有黃雜色
聲如雷後年章大寶舉兵反陳竟國亡
天光
書紀年曰周昭王末年夜清五色光貫紫微其年王南巡
不返
漢書曰成帝元延元年四月無雲有雷聲光四下至昏乃止
晉書曰晉穆帝永和十年前涼張祚之元年正月夜天有
光如車蓋明年祚被殺
宋書曰文帝元嘉十八年七月天有黃光洞燭至二十一

Fig. 11.28 One of the variant excerpts from the Bamboo Annals (Ancient Text) in Tàipíng Yùlǎn (v. 874, f. 4b of MS Nu-3 in © the National Diet Library of Japan). See Van der Sluijs and Hayakawa (2022) for detailed discussions on their philological provenances

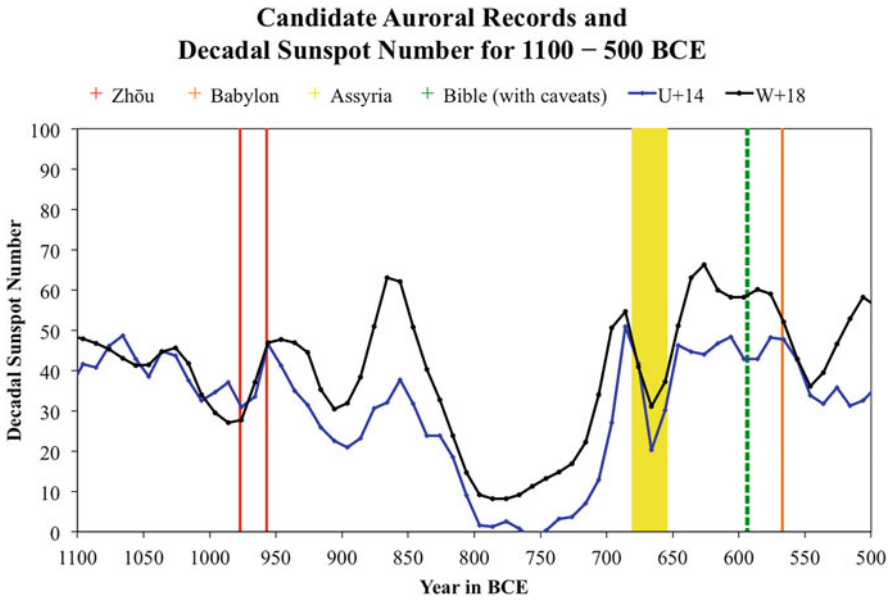

Fig. 11.29 Chronology of the earliest candidate auroras in the Bamboo Annals in Zhōu (Van der Sluijs and Hayakawa 2022), Assyrian Astrological Reports (Hayakawa et al. 2019e), Bible with caveats (Silverman 2006; Hayakawa et al. 2019e), Astronomical Diaries from Babylon (Stephenson et al. 2004), in comparison with the decadal sunspot number reconstructed from cosmogenic isotope data in tree rings and ice cores in U + 14 (Usoskin et al. 2014) and W + 14 (Wu et al. 2018). (Modified from Van der Sluijs and Hayakawa (2022))

Fig. 11.30 Oldest sunspot image in existence (MS 157, f. 192v in © Corpus Christi College of the Oxford University), reproduced from Fig. 11.1 of Willis and Stephenson (2001) under CC-BY

11.2.6 Reconstructions of Ancient and Medieval Space Weather Events

By scrutinizing historical reports of sunspots and candidate auroras, it is possible to reconstruct space weather phenomena over the past three millennia. Before and after the extreme solar particle storm in 774/775 (Miyake et al. 2012), auroras were seen in Amida in northern Mesopotamia in 771/772 and June 773, as recorded in the aforementioned Zūqnīn Chronicle (Hayakawa et al. 2017b). In 776, candidate auroras were reported in England and China (Usoskin et al. 2013). Caveats must be noted here; such historical records of candidate auroras are frequently contaminated with atmospheric optical phenomena (Usoskin et al. 2017) and require careful comparison with modern observations (Hayakawa et al. 2019b; Stephenson et al. 2019). These records are not necessarily consistent with the expected timing of this extreme solar particle storm in the 774 boreal spring and summer (Büntgen et al. 2018a, b), while these records still indicate an enhanced solar activity around the time at the very least (e.g., Usoskin et al. 2013).

In the vicinity of another extreme solar particle storm in 993/994 (Miyake et al. 2012), several candidate auroral records were found in late 992 and early 993. Candidate auroras were recorded in Saxony on 21 October 992, in Saxony and Ulster on 26 December 992, and in Goryeo in December–January 992. It is still unclear how these records relate to the extreme solar particle storms detected in tree rings and ice cores; however, similar to the auroral records around 774/775, they seemingly still confirm enhanced solar activity around these records.

Under enhanced solar activity, ICMEs frequently disturb the magnetosphere of the Earth in succession, as in the Halloween event of October 2003 and the magnetic storm of May 1921 (Newton 1948; Gopalswamy et al. 2005; Love et al. 2019b). Occasionally, auroras can be observed for several nights in succession. In Japan as well, it has frequently been discussed that a “red vapour” was observed around Heian-kyo for three nights, particularly on 21–23 February 1204, as recorded in Omuro Soshoiki and Meigetsuki (Kanda 1933; Saito 1999). On 21 February 1204, a naked-eye sunspot was recorded on the Chinese records. Willis et al. (2005) pointed out that multiple CMEs were emitted from the giant sunspot at that time and that the intense magnetic storm lasted for three nights. This phenomenon was highlighted again in recent studies (Kataoka et al. 2017).

A similar space weather event occurred in March 1582, when auroras were observed across Western Europe on 6–8 March and across lower MLATs in East Asia on 8 March (Hattori et al. 2019) according to Cals3k4b model (Korte and Constable 2011). The minimum Dst* has been estimated at ≈ -600 nT. This is because this space weather event extended the equatorward auroral boundary to a point somewhere comparable to those of the extreme magnetic storms of September 1909 and March 1989.

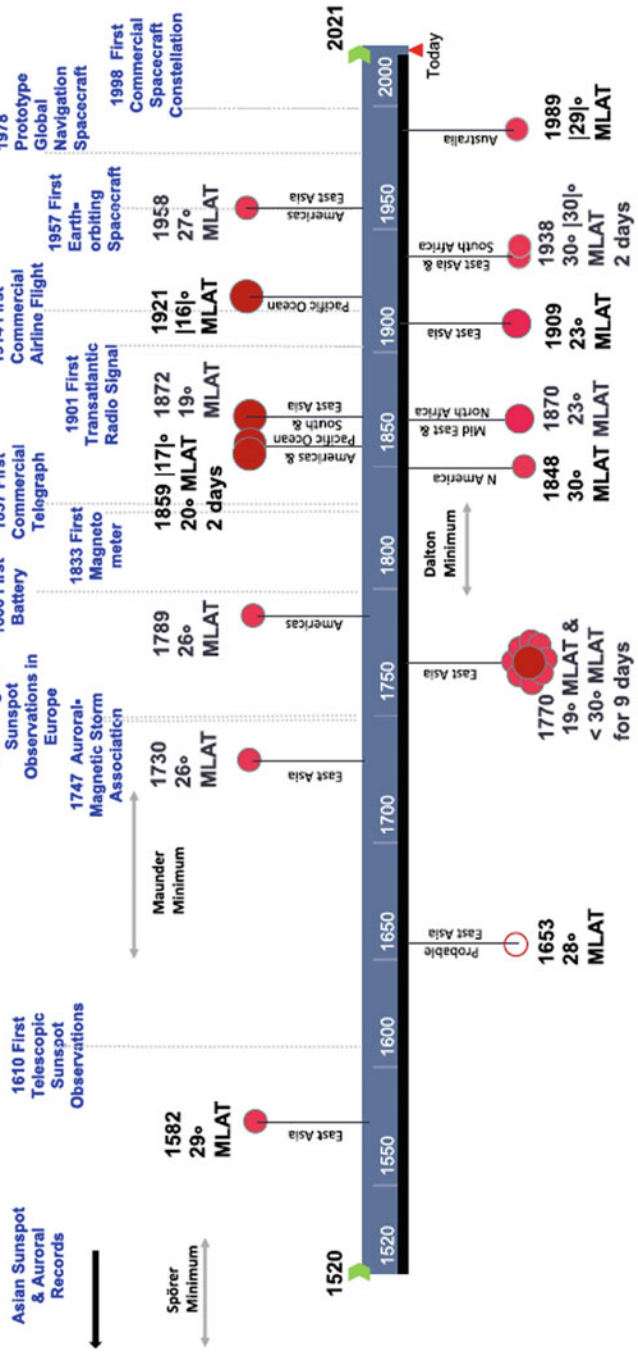

Fig. 11.31 Space weather timeline for the last five centuries, as reproduced from © Knipp et al. (2021) under CC-BY

11.2.7 Conclusion

The Carrington event of September 1859 was regarded as the largest space weather event observed in terms of the flare fluence, the geomagnetic disturbance, and the equatorward extension of the auroral oval. However, early modern observations have indicated that the extreme space weather events in February 1872 and May 1921 were at least comparable to the Carrington event in terms of the magnitude of the geomagnetic storm and their equatorward auroral boundary. These extreme space weather events for the last five centuries have particularly attracted scientific interests, as visualized in the space weather timeline (Fig. 11.31), as visualized in Knipp et al. (2021). Further reconstructions of such space weather phenomena before the IGY (1957–1958) will be necessary, because quantitative evaluations have not been fully conducted yet.

In addition to cosmogenic isotopes of annual rings and ice cores, historical records for sunspots and candidate auroras may visualize footprints of past CMEs, when reconstructing space weather phenomena beyond the time span of modern observations. Because the current auroral and sunspot records date back to 9–12 centuries in graphical sources and 22–30 centuries in textual records, extreme space weather events over a longer time span can be traced by connecting modern observations with historical documents. It is important to continue such quantitative reconstructions of past space weather events and current theoretical studies to improve our understanding of the long-term changes in the solar–terrestrial system. Further investigations are expected in this field.

Acknowledgments This study was supported by JSPS Grants-in-Aid for Scientific Research JP15H05816, JP15H05812, JP15H03732, JP16H03955, JP15H05815, JP17J06954, JP16J00320, and JP21J00106. H. H. has been partly funded by the ISEE director's leadership fund for FYs 2021 and 2022, the Young Leader Cultivation (YLC) programme of Nagoya University, Tokai Pathways to Global Excellence (Nagoya University) of the Strategic Professional Development Program for Young Researchers (MEXT), and the young researcher units for the advancement of new and undeveloped fields, Institute for Advanced Research at Nagoya University of the Program for Promoting the Enhancement of Research Universities. HH has benefited from discussions within the ISSI International Team #510 (SEESUP Solar Extreme Events: Setting Up a Paradigm) and ISWAT-COSPAR S1-01 and S1-02 teams. H. H. and Y. N. were supported by the JSPS Overseas Research Fellowship Program. Y.N. and H.H. acknowledge the International Space Science Institute and the supported International Teams #464: The Role of Solar and Stellar Energetic Particles on (Exo) Planetary Habitability (ETERNAL), #510 (SEESUP Solar Extreme Events: Setting Up a Paradigm), #475 (Modeling Space Weather And Total Solar Irradiance Over The Past Century), and #417 (Recalibration of the Sunspot Number Series). We would like to thank Dr. Fusa Miyake and Dr. Yasuyuki Mitsuma for their tremendous help and advice in writing this paper. H. H. would like to thank Akito D. Kawamura, Alessandra S. Giunta, Ana I. Ribeiro, Ana P. Correia, Bhaskara Veenadhari, David M. Willis, David P. Hand, Delores J. Knipp, Denny M. Oliveira, Edward W. Cliver, Harufumi Tamazawa, Hiroaki Isobe, Hiroko Miyahara, Ichiro Nakanishi, Ilaria Ermolli, Jeffrey J. Love, Frédéric Clette, José M. Vaquero, Marinus van der Sluijs, Mat Owens, Mike Lockwood, Bruno Besser, Jose R. Ribeiro, Kentaro Hattori, Kiyomi Iwahashi, Kosuke Namekata, Kumiko Mase, Lee T. Macdonald, Maricruz Gallego, Matthew N. Wild, Matthew N. Wild, Norio Umemura, Rainer Arlt, Sandeep Kumar, Satoshi Hayakawa, Shin Toriumi, Shyamoli Mukherjee, Takahito Sakaue, Takuya Takahashi, Tomoya Iju, Victor M. S. Carrasco,

Yasuyuki Mitsuma, Yoshikazu Watanabe, and Sam M. Silvermanfor helping me writing up these chapter articles. Y. N. would like to acknowledge that this paper is especially benefited from the collaborative work and discussions with Hiroyuki Maehara, Takuya Shibayama, Satoshi Honda, Kosuke Namekata, Daisaku Nogami, Kazunari Shibata, Soshi Okamoto, Kai Ikuta, Shota Notsu, Yosuke Yamashiki, Adam Kowalski, Isaiah Tristan, James Davenport, Suzanne Hawley, Petr Heinzel, and Vladimir Airapetian. The illustrations used in this paper are based on photographs taken at or provided by the Royal Astronomical Society, the British Museum, the Vatican Library, Corpus Christi College, Oxford, and the National Diet Library. We would like to express our gratitude to them.

References

Airapetian, V.S., et al.: Prebiotic chemistry and atmospheric warming of early earth by an active young Sun. Nat. Geosci. **9**, 452–455 (2016)

Airapetian, V.S., et al.: Impact of space weather on climate and habitability of terrestrial-type exoplanets. Int. J. Astrobiol. **19**(2), 136–194 (2020)

Akasofu, S.: The latitudinal shift of the auroral belt. J. Atmos. Terr. Phys. **26**, 1167–1174 (1964)

Akasofu, S.-I., Kamide, Y.: Comment on "The extreme magnetic storm of 1–2 September 1859" by B. T. Tsurutani, W. D. Gonzalez, G. S. Lakhina, and S. Alex. J. Geophys. Res. Space Physics. **110**, A09226 (2005)

Alvarado-Gómez, J.D., et al.: Suppression of coronal mass ejections in active stars by an overlying large-scale magnetic field: a numerical study. Astrophys. J. **862**(2), 93 (2018)

Argiroffi, C., et al.: A stellar flare-coronal mass ejection event revealed by X-ray plasma motions. Nature Astron. **3**, 742–748 (2019)

Arlt, R.: Digitization of sunspot drawings by Staudacher in 1749 – 1796. Sol. Phys. **247**, 399–410 (2008). https://doi.org/10.1007/s11207-007-9113-4

Aulanier, G., et al.: The standard flare model in three dimensions. II. Upper limit on solar flare energy. Astron. Astrophys. **549**, A66 (2013)

Ayres, T.R.: Evolution of the solar ionizing flux. J. Geophys. Res. **102**(E1), 1641–1652 (1997)

Baker, D.N., et al.: Severe Space Weather Events— Understanding Societal and Economic Impacts. National Academies Press, Washington DC (2008)

Baker, D.N., et al.: A major solar eruptive event in July 2012: defining extreme space weather scenarios. Space Weather. **11**, 585–591 (2013)

Barnard, L., McCracken, K.G., Owens, M.J., Lockwood, M.: What can the annual 10Be solar activity reconstructions tell us about historic space weather? J. Space Weather Space Clim. **8**, A23 (2018)

Bartels, J.: Solar eruptions and their ionospheric effects—a classical observation and its new interpretation. Terr. Mag. Atmos. Elect. **42**, 235–239 (1937)

Battersby, S., et al.: Core concept: what are the chances of a hazardous solar superflare? PNAS. **116**(47), 23368–23370 (2019)

Beer, J., McCracken, K., von Steiger, R.: Cosmogenic Radionuclides. Springer, Berlin (2012)

Benz, A.O., Guedel, M.: Physical processes in magnetically driven flares on the sun, stars, and young stellar objects. Annu. Rev. Astron. Astrophys. **48**, 241–287 (2010)

Bicknell, P.J.: Did Anaxagoras observe a sunspot in 467 B.C.? Isis. **59**, 87–90 (1967)

Blake, S.P., Pulkkinen, A., Schuck, P.W., Glocer, A., Tóth, G.: Estimating maximum extent of auroral equatorward boundary using historical and simulated surface magnetic field data. J. Geophys. Res. Space Physics. **126**, e28284 (2021)

Bogdan, T.J., et al.: Distribution of sunspot umbral areas: 1917—1982. Astrophys. J. **327**, 451 (1988)

Boteler, D.H.: The super storms of august/September 1859 and their effects on the telegraph system. Adv. Space Res. **38**, 159–172 (2006)
Brehm, N., Christl, M., Jull, T., et al.: Tree-rings reveal two strong solar proton events in 7176 and 5259 BCE. Nat. Commun. **13**, 1196 (2022)
Büntgen, U., et al.: Tree rings reveal globally coherent signature of cosmogenic radiocarbon events in 774 and 993 CE, nature. Communications. **9**, id. 3605 (2018a)
Büntgen, U., et al.: Tree rings reveal globally coherent signature of cosmogenic radiocarbon events in 774 and 993 CE. Nat. Commun. **9**, 3605 (2018b)
Candelaresi, S., et al.: Superflare occurrence and energies on G-, K-, and M-type Dwarfs. Astrophys. J. **792**(1), 67 (2014)
Carrington, R.C.: Description of a singular appearance seen in the Sun on September 1, 1859. Mon. Not. R. Astron. Soc. **20**, 13–15 (1859)
Chapman, S.: The Aurora in middle and low latitudes. Nature. **179**, 7–11 (1957)
Chapman, S., Bartels, J.: Geomagnetism, Vol. I: Geomagnetic and Related Phenomena. Oxford University Press, London (1940)
Chen, H.: Persistence of flare-driven atmospheric chemistry on rocky habitable zone worlds. Nature Astron. **5**, 298–310 (2021)
Clette, F., Svalgaard, L., Vaquero, J.M., Cliver, E.W.: Revisiting the sunspot number. A 400-year perspective on the solar cycle. Space Sci. Rev. **186**, 35–103 (2014)
Cliver, E.W., Dietrich, W.F.: The 1859 space weather event revisited: limits of extreme activity. J. Space Weather Space Clim. **3**, A31 (2013)
Cliver, E.W., Svalgaard, L.: The 1859 solar-terrestrial disturbance and the current limits of extreme space weather activity. Sol. Phys. **224**, 407–422 (2004)
Crosley, M.K., Osten, R.A.: Constraining stellar coronal mass ejections through multi-wavelength analysis of the active M Dwarf EQ Peg. Astrophys. J. **856**(1), 39 (2018)
Curto, J.J., Castell, J., Del Moral, F.: Sfe: waiting for the big one. J. Space Weather Space Clim. **6**, A23 (2016)
Davenport, J.R.A.: The Kepler Catalog of Stellar Flares. Astrophys. J. **829**(1), 23 (2016)
Doyle, L., et al.: Investigating the rotational phase of stellar flares on M dwarfs using K2 short cadence data. Mon. Not. R. Astron. Soc. **480**(2), 2153–2164 (2018)
Eastwood, J.P., et al.: The economic impact of space weather: where do we stand? Risk Anal. **37**(2), 206–218 (2017)
Ebihara, Y., et al.: Statistical distribution of the storm-time proton ring current: POLAR measurements. Geophys. Res. Lett. **29**, 1969 (2002)
Ebihara, Y., et al.: Possible cause of extremely bright Aurora witnessed in East Asia on 17 September 1770. Space Weather. **15**, 1373–1382 (2017)
Emslie, A.G., et al.: Global energetics of thirty-eight large solar eruptive events. Astrophys. J. **759**(1), 71 (2012)
Fuhrmeister, B., et al.: The CARMENES search for exoplanets around M dwarfs. Wing asymmetries of Hα, Na I D, and he I lines. Astron. Astrophys. **615**, A14 (2018)
Fujiyama, M., et al.: Revisiting Kunitomo's sunspot drawings during 1835 - 1836 in Japan. Sol. Phys. **294**, 43 (2019)
Gershberg, R.E.: Solar-Type Activity in Main-Sequence Stars, Astronomy and Astrophysics Library. ISBN 978-3-540-21244-7. Springer, Berlin Heidelberg (2005)
Gonzalez, W.D., Echer, E., Tsurutani, B.T., Clúa de Gonzalez, A.L., Dal Lago, A.: Interplanetary origin of intense, superintense and extreme geomagnetic storms. Space Sci. Rev. **158**, 69–89 (2011). https://doi.org/10.1007/s11214-010-9715-2
González-Esparza, J.A., Cuevas-Cardona, M.C.: Observations of Low-Latitude Red Aurora in Mexico during the 1859 Carrington Geomagnetic Storm. Space Weather. **16**(6), 593–600 (2018)
Gopalswamy, N., et al.: Coronal mass ejections and other extreme characteristics of the 2003 October-November solar eruptions. J. Geophys. Res. Space Phys. **110**(A9), A09S15 (2005)

Green, J.L., Boardsen, S.: Duration and extent of the great auroral storm of 1859. Adv. Space Res. **38**, 130–135 (2006)
Güdel, M.: The sun in time: activity and environment. Living Rev. Solar Phys. **4**(1), 3 (2007)
Günther, M.N., et al.: Stellar flares from the first TESS data release: exploring a new sample of M dwarfs. Astrophys. J. **159**(60), 60 (2020)
Hapgood, M.A.: Space Weather: its Impact on Earth and Implications for Business. Lloyd's, London (2010)
Hapgood, M.A.: The great storm of may 1921: an exemplar of a dangerous space weather event. Space Weather. **17**, 950–975 (2019)
Harra, L.K., et al.: The characteristics of solar X-class flares and CMEs: a paradigm for stellar Superflares and eruptions? Sol. Phys. **291**(6), 1761–1782 (2016)
Hattori, K., Hayakawa, H., Ebihara, Y.: Occurrence of great magnetic storms on 6-8 March 1582. Mon. Not. R. Astron. Soc. **487**(3), 3550–3559 (2019)
Hawley, S.L., et al.: Kepler Flares. I. Active and inactive M Dwarfs. Astrophys. J. **797**(2), 121 (2014)
Hawley, S.L., Pettersen, B.R.: The great flare of 1985 April 12 on AD Leonis. Astrophys. J. **378**, 725 (1991)
Hayakawa, H., et al.: East Asian observations of low latitude aurora during the Carrington magnetic storm. Public. Astron. Soc. Jpn. **68**, 99 (2016b)
Hayakawa, H., et al.: Earliest datable records of aurora-like phenomena in the astronomical diaries from Babylonia. Earth Planets Space. **68**, 195 (2016c)
Hayakawa, H., et al.: Historical aurora evidences for great magnetic storms in 990s. Sol. Phys. **292**, 12 (2017a)
Hayakawa, H., et al.: The earliest drawings of datable auroras and a two-tail comet from the Syriac Chronicle of Zūqnīn. Public. Astronomical Soc. Jpn. **69**, 17 (2017b)
Hayakawa, H., et al.: Records of sunspots and aurora candidates in the Chinese official histories of the Yuán and Míng dynasties during 1261-1644. Public. Astron. Soc. Jpn. **69**, 65 (2017c)
Hayakawa, H., et al.: Records of auroral candidates and sunspots in Rikkokushi, chronicles of ancient Japan from early 7th century to 887. Public. Astron. Soc. Jpn. **69**, 86 (2017d)
Hayakawa, H., et al.: Long-lasting extreme magnetic storm activities in 1770 found in historical documents. Astrophys. J. Lett. **850**, L31 (2017e)
Hayakawa, H., et al.: Iwahashi Zenbei's sunspot drawings in 1793 in Japan. Sol. Phys. **293**, 8 (2018a)
Hayakawa, H., et al.: A great space weather event in February 1730. Astron. Astrophys. **616**, A177 (2018c)
Hayakawa, H., et al.: The great space weather event during 1872 February recorded in East Asia. Astrophys. J. **862**, 15 (2018d)
Hayakawa, H., et al.: Low-latitude aurorae during the extreme space weather events in 1859. Astrophys. J. **869**, 57 (2018e)
Hayakawa, H., et al.: The extreme space weather event in September 1909. Mon. Not. R. Astron. Soc. **484**(3), 4083–4099 (2019a)
Hayakawa, H., et al.: The celestial sign in the Anglo-Saxon chronicle in the 770s: insights on contemporary solar activity. Sol. Phys. **294**, 42 (2019b)
Hayakawa, H., et al.: A comparison of graphical Records in the East and the west. Sol. Phys. **294**, 95 (2019c)
Hayakawa, H., et al.: Temporal and spatial evolutions of a large Sunspot group and great auroral storms around the Carrington event in 1859. Space Weather. **17**(11), 1553–1569 (2019d). https://doi.org/10.1029/2019SW002269
Hayakawa, H., Mitsuma, Y., Ebihara, Y., Miyake, F.: The earliest candidates of auroral observations in Assyrian astrological reports: insights on solar activity around 660 BCE. Astrophys. J. Lett. **884**, L18 (2019e)
Hayakawa, H., Ribeiro, P., Vaquero, J.M., et al.: The extreme space weather event in 1903 October/November: an outburst from the Quiet Sun. Astrophys. J. Lett. **897**(1), L10 (2020a)

Hayakawa, H., Ribeiro, J., Ebihara, Y., Correia, A.P., Sôma, M.: South American auroral reports during the Carrington storm. Earth Planets Space. **72**, 122 (2020b)
Hayakawa, H., Nevanlinna, H., Blake, S.P., Ebihara, Y., Bhaskar, A.T., Miyoshi, Y.: Temporal variations of the three geomagnetic field components at Colaba observatory around the Carrington storm in 1859. Astrophys. J. **928**, 32 (2022a)
Hayakawa, H., et al.: An overview of Sunspot observations in 1727–1748. Astrophys. J. (2022b). https://doi.org/10.3847/1538-4357/apjac6671
Heinzel, P., Shibata, K.: Can flare loops contribute to the white-light emission of Stellar Superflares? Astrophys. J. **859**(2), 143 (2018)
Herbst, K., et al.: From solar to stellar flare characteristics. On a new peak size distribution for G-, K-, and M-dwarf star flares. Astron. Astrophys. **621**, A67 (2019a)
Herbst, K., et al.: A new model suite to determine the influence of cosmic rays on (exo)planetary atmospheric biosignatures. Astron. Astrophys. **631**, A101 (2019b)
Hodgson, R.: On a curious appearance seen in the Sun. Mon. Not. R. Astron. Soc. **20**, 15–16 (1859)
Honda, S., et al.: Time-resolved spectroscopic observations of an M-dwarf flare star EV Lacertae during a flare. Public. Astron. Soc. Jpn. **70**(4), 62 (2018)
Hu, J., et al.: Extreme energetic particle events by superflare-associated CMEs from solar-like stars, science. Advances. **8**(12), eabi9743 (2022)
Hudson, H.S.: Chasing white-light flares. Sol. Phys. **291**(5), 1273–1322 (2016)
Hunger, H.: Astrological Reports to Assyrian Kings. Helsinki University Press, Helsinki (1992)
Ichimoto, K., Kurokawa, H.: Hα red asymmetry of solar flares. Sol. Phys. **93**(1), 105–121 (1984)
Jejcic, S., et al.: High-density off-limb flare loops observed by SDO. Astrophys. J. **867**(2), 134 (2018)
Kanda, S.: Records of polar light in our country. Astron. Her. **26**(11), 204–210 (1933)
Karoff, C., et al.: Observational evidence for enhanced magnetic activity of superflare stars, nature. Communications. **7**, id. 11058 (2016)
Kasting, J.F.: Faint young Sun redux. Nature. **464**, 687–689 (2010)
Kataoka, R.: Extreme geomagnetic activities: a statistical study. Earth Planets Space. **72**(1), 124 (2020)
Kataoka, R., Iwahashi, K.: Inclined zenith Aurora over Kyoto on 17 September 1770: graphical evidence of extreme magnetic storm. Space Weather. **15**, 1314–1320 (2017)
Kataoka, R., et al.: Historical space weather monitoring of prolonged aurora activities in Japan and in China. Space Weather. **15**, 392–402 (2017)
Kimball, D.S.: A Study of the Aurora of 1859. Geophysical institute at the university of Alaska, Fairbanks (1960)
Knipp, D.J., Fraser, B.J., Shea, M.A., Smart, D.F.: On the little-known consequences of the 4 august 1972 ultra-fast coronal mass ejecta: facts, commentary, and call to action. Space Weather. **16**, 1635–1643 (2018)
Knipp, D.J., Bernstein, V., Wahl, K., Hayakawa, H.: Timelines as a tool for learning about space weather storms. J. Space Weather Space Clim. **11**, 29 (2021)
Korte, M., Constable, C.: Improving geomagnetic field reconstructions for 0-3 ka. Phys. Earth Planet. Inter. **188**, 247–259 (2011). https://doi.org/10.1016/j.pepi.2011.06.017
Kowalski, A.F., Allred, J.C.: Parameterizations of chromospheric condensations in dG and dMe Model Flare atmospheres. Astrophys. J. **852**(1), 61 (2018)
Kowalski, A.F., et al.: A white light megaflare on the dM4.5e star YZ CMi. Astrophys. J. Lett. **714**(1), L98–L102 (2010)
Kowalski, A.F., et al.: Time-resolved properties and global trends in dMe flares from simultaneous photometry and spectra. Astrophys. J. Suppl. **207**(1), 15 (2013)
Le, G., et al.: Morphology of the ring current derived from magnetic field observations. Ann. Geophys. **22**, 1267–1295 (2004)
Lefèvre, L., et al.: Detailed analysis of solar data related to historical extreme geomagnetic storms: 1868–2010. Sol. Phys. **291**, 1483–1531 (2016)

Lingam, M., Loeb, A.: Risks for life on habitable planets from superflares of their host stars. Astrophys. J. **848**(1), 41 (2017)
Linsky, J.: Host stars and their effects on exoplanet atmospheres: an introductory overview, Lecture Notes in Physics, Volume 955. ISBN 978-3-030-11451-0. Springer Nature Switzerland AG (2019)
Lockyer, W.J.S.: Magnetic storm, 1909 Sept. 25, and associated solar disturbance. Mon. Not. R. Astron. Soc. **70**, 12 (1909)
Love, J.J., Hayakawa, H., Cliver, E.W.: On the intensity of the magnetic superstorm of September 1909. Space Weather. **17**, 37–45 (2019a)
Love, J.J., Hayakawa, H., Cliver, E.W.: Intensity and impact of the New York railroad superstorm of May 1921. Space Weather. **17**, 1281–1292 (2019b)
Lundstedt, H., Persson, T., Andersson, V.: The extreme solar storm of May 1921: observations and a complex topological model. Ann. Geophys. **33**, 109–116 (2015)
Lynch, B.J., et al.: Modeling a Carrington-scale Stellar superflare and coronal mass ejection from {\kappa }^{1}{Cet}. Astrophys. J. **880**(2), 97 (2019)
Maehara, H., et al.: Superflares on solar-type stars. Nature. **485**(7399), 478–481 (2012)
Maehara, H., et al.: Statistical properties of superflares on solar-type stars based on 1-min cadence data. Earth Planets Space. **67**, 59 (2015)
Maehara, H., et al.: Starspot activity and superflares on solar-type stars. Public. Astron. Soc. Jpn. **69**(3), 41 (2017)
Maehara, H., et al.: Time-resolved spectroscopy and photometry of M dwarf flare star YZ Canis Minoris with OISTER and TESS: blue asymmetry in the Hα line during the non-white light flare. Public. Astron. Soc. Jpn. **73**(1), 44–65 (2021)
Mekhaldi, F., et al.: Multiradionuclide evidence for the solar origin of the cosmic-ray events of AD 774/5 and 993/4. Nat. Commun. **6**, 8611 (2015)
Miyake, F., Nagaya, K., Masuda, K., Nakamura, T.: A signature of cosmic-ray increase in AD 774-775 from tree rings in Japan. Nature. **486**, 240–242 (2012)
Miyake, F., Masuda, K., Nakamura, T.: Another rapid event in the carbon-14 content of tree rings 2013. Nat. Commun. **4**, 1748 (2013)
Miyake, F., et al.: Extreme Solar Particle Storms; the Hostile Sun., ISBN: 978-0-7503-2230-0. IOP Ebooks. IOP Publishing, Bristol (2019)
Morris, B.M., et al.: The starspots of HAT-P-11: evidence for a Solar-like Dynamo. Astrophys. J. **846**(2), 99 (2017)
Moschou, S.-P.: The stellar CME─Flare relation: what do historic observations reveal? Astrophys. J. **877**(2), 105 (2019)
Namekata, K., et al.: Statistical studies of solar white-light flares and comparisons with superflares on solar-type stars. Astrophys. J. **851**(2), 91 (2017)
Namekata, K., et al.: Lifetimes and emergence/decay rates of star spots on solar-type stars estimated by Kepler data in comparison with those of sunspots. Astrophys. J. **871**(2), 187 (2019)
Namekata, K., et al.: Optical and X-ray observations of stellar flares on an active M dwarf AD Leonis with the Seimei telescope, SCAT, NICER, and OISTER. Public. Astron. Soc. Jpn. **72**(4), 68 (2020a)
Namekata, K., et al.: Temporal evolution of spatially resolved individual star spots on a planet-hosting solar-type star: Kepler-17. Astrophys. J. **891**(2), 103 (2020b)
Namekata, K., et al.: Probable detection of an eruptive filament from a superflare on a solar-type star. Nature Astron. **6**, 241–248 (2022a)
Namekata, K., et al.: Discovery of a long-duration Superflare on a young solar-type star EK Draconis with nearly similar time evolution for Hα and white-light emissions. Astrophys. J. Lett. **926**(1), L5 (2022b)
Newton, H.W.: "Sudden commencements" in the Greenwich magnetic records (1879–1944) and related sunspot data. Geophys. Suppl. Monthly Notices R. Astron. Soc. **5**, 159–185 (1948)
Ngwira, C.M., et al.: Simulation of the 23 July 2012 extreme space weather event: what if this extremely rare CME was earth directed? Space Weather. **11**, 671–679 (2013)

Nizamov, B.A.: Soft X-ray heating as a mechanism of optical continuum generation in solar-type star superflares. Mon. Not. R. Astron. Soc. **489**(3), 4338–4345 (2019)
Notsu, Y., et al.: Superflares on solar-type stars observed with Kepler II. Photometric variability of superflare-generating stars: a signature of stellar rotation and starspots. Astrophys. J. **771**(2), 127 (2013)
Notsu, Y., et al.: High dispersion spectroscopy of solar-type superflare stars. I. Temperature, surface gravity, metallicity, and vsin i. Public. Astron. Soc. Jpn. **67**(3), 32 (2015a)
Notsu, Y., et al.: High dispersion spectroscopy of solar-type superflare stars. II. Stellar rotation, starspots, and chromospheric activities. Public. Astron. Soc. Jpn. **67**(3), 33 (2015b)
Notsu, Y., et al.: Do Kepler superflare stars really include slowly rotating sun-like stars?—Results using APO 3.5 m telescope spectroscopic observations and Gaia-DR2 data. Astrophys. J. **76**(1), 58 (2019)
O'Hare, P., et al.: Multiradionuclide evidence for an extreme solar proton event around 2,610 B.P. (660 BC). Proc. Natl. Acad. Sci. **116**(13), 5961–5966 (2019)
Ohtani, S.: New insights from the 2003 Halloween storm into the Colaba 1,600 nT magnetic depression during the 1859 Carrington storm. J. Geophys. Res. Space Phys. **127**, e2022JA030596 (2022)
Okamoto, S., et al.: Statistical properties of superflares on solar-type stars: results using all of the Kepler primary mission data. Astrophys. J. **906**(2), 72 (2021)
Owens, B.: Long-term research: slow science. Nature. **495**, 300–303 (2013)
Paleari, C.I., Mekhaldi, F., Adolphi, F., et al.: Cosmogenic radionuclides reveal an extreme solar particle storm near a solar minimum 9125 years BP. Nat. Commun. **13**, 214 (2022)
Park, J., et al.: Relationship between solar activity and Δ14C peaks in AD 775, AD 994, and 660 BC. Radiocarbon. **59**(4), 1147–1156 (2017)
Reid, I.N., Hawley, S.L.: New light on dark stars red dwarfs, low-mass stars. Brown Stars Astron. Astrophys. (2005)
Riley, P., et al.: Extreme space weather events: from cradle to grave. Space Sci. Rev. **214**, 21 (2018)
Roach, F.E., Moore, J.G., Bruner Jr., E.C., Cronin, H., Silverman, S.M.: The height of maximum luminosity in an auroral arc. J. Geophys. Res. **65**, 3575 (1960)
Roettenbacher, R.M., Vida, K.: The connection between starspots and flares on main-sequence Kepler Stars. Astrophys. J. **868**(1), 3 (2018)
Rubenstein, E.P., Schaefer, B.E.: Are superflares on solar analogues caused by extrasolar planets? Astrophys. J. **529**, 1031–1033 (2000). https://doi.org/10.1086/308326
Russell, C.T., Mewaldt, R.A., Luhmann, J.G., Mason, G.M., von Rosenvinge, T.T., Cohen, C.M.S., Leske, R.A., Gomez-Herrero, R., Klassen, A., Galvin, A.B., Simunac, K.D.C.: The very unusual interplanetary coronal mass ejection of 2012 July 23: a blast wave mediated by solar energetic particles. Astrophys. J. **770**, 38 (2013)
Saito, K.: Astronomical Records in Teika's Meigetsuki. ISBN 978-4-87449-029-7. Keiyusha, Tokyo (1999)
Sammis, I., et al.: The dependence of large flare occurrence on the magnetic structure of sunspots. Astrophys. J. **540**(1), 583–587 (2000)
Schaefer, B.E., King, J.R., Deliyannis, C.P.: Superflares on ordinary solar-type stars. Astrophys. J. **529**(2), 1026–1030 (2000)
Scheucher, M., et al.: New insights into Cosmic-ray-induced biosignature chemistry in earth-like atmospheres. Astrophys. J. **863**(1), 6 (2018)
Schrijver, C.J., Beer, J.: Space weather from explosions on the Sun: how bad could it be? EOS Trans. Am. Geophys. Union. **95**(24), 201–202 (2014)
Schrijver, C.J., et al.: Estimating the frequency of extremely energetic solar events, based on solar, stellar, lunar, and terrestrial records. J. Geophys. Res. Space Physics. **117**(A8), A08103 (2012)
Segan, C., Mullen, G.: Earth and Mars: evolution of atmospheres and surface temperatures. Science. **177**(4043), 52–56 (1972)
Segura, A., et al.: The effect of a strong stellar flare on the atmospheric chemistry of an earth-like planet orbiting an M dwarf. Astrobiology. **10**(7), 751–771 (2010)

Shibata, K., et al.: Can superflares occur on our sun? Public. Astron. Soc. Jpn. **65**(3), 49 (2013)
Shibayama, T., et al.: Superflares on solar-type stars observed with Kepler. I. Statistical properties of superflares. Astrophys. J. Suppl. **209**(1), 5 (2013)
Silverman, S.M.: Low latitude auroras prior to 1200 C.E. and Ezekiel's vision. Adv. Space Res. **38**, 200–208 (2006)
Silverman, S.M.: Low-latitude auroras: the great aurora of 4 February 1872. J. Atmos. Sol. Terr. Phys. **70**, 1301–1308 (2008)
Silverman, S.M., Cliver, E.W.: Low-latitude auroras: the magnetic storm of 14-15 May 1921. J. Atmos. Sol. Terr. Phys. **63**, 523–535 (2001)
Siscoe, G.L., Silverman, S.M., Siebert, K.D.: Ezekiel and the northern lights: biblical aurora seems plausible. EOS Trans. Am. Geophys. Union. **83**, 173 (2002)
Siscoe, G., Crooker, N.U., Clauer, C.R.: Dst of the Carrington storm of 1859. Adv. Space Res. **38** (2), 173–179. S0273117705012081 (2006)
Stephenson, F.R., Willis, D.M.: The earliest drawing of sunspots. Astron. Geophys. **40**(6), 21–22 (1999)
Stephenson, F.R., Willis, D.M., Hallinan, T.J.: Aurorae: the earliest datable observation of the aurora borealis. Astron. Geophys. **45**, 6.15–6.17 (2004)
Stephenson, F.R., et al.: Do the Chinese astronomical records dated a.D. 776 January 12/13 describe an auroral display or a lunar halo? A critical re-examination. Sol. Phys. **294**, 36 (2019)
Sugiura, M.: Hourly values of equatorial Dst for the IGY, Ann. Int. Geophys. Year, 35, 9, Pergamon Press, Oxford (1964)
Sugiura, M., Kamei, T.: Equatorial Dst index 1957-1986, *IAGA Bull.*, **40**, ISGI Publication Office, Saint Maur des Fossess (1991)
Sun, X., et al.: Torus-stable zone above starspots. Monthly Notices R. Astron. Soc. **509**(4), 5075–5085 (2021)
Takahashi, T., Shibata, K.: Sheath-accumulating Propagation of Interplanetary Coronal Mass Ejection. Astrophys. J. Lett. **837**(2), L17 (2017)
Takahashi, T., et al.: Scaling relations in coronal mass ejections and energetic proton events associated with solar superflares. Astrophys. J. Lett. **833**(1), L8 (2016)
Tilley, M.A., et al.: Modeling repeated M dwarf flaring at an earth-like planet in the habitable zone: atmospheric effects for an unmagnetized planet. Astrobiology. **19**(1), 64–86 (2019)
Toriumi, S., Wang, H.: Flare-productive active regions. Living Rev. Solar Phys. **16**(1), 3 (2019)
Tsurutani, B.T., Gonzalez, W.D., Lakhina, G.S., Alex, S.: The extreme magnetic storm of 1-2 September 1859. J. Geophys. Res. Space Phys. **108**(A7), 1268 (2003)
Uchikawa, Y., Cowley, L., Hayakawa, H., Willis, D.M., Stephenson, F.R.: Provenance of the cross sign of 806 in the Anglo-Saxon Chronicle: a possible lunar halo over continental Europe? Hist. Geo Space Sci. **11**, 81–92 (2020)
Usoskin, I.G.: A history of solar activity over millennia. Living Rev. Sol. Phys. **14**, 3 (2017)
Usoskin, I.G., Kovaltsov, G.A.: Occurrence of extreme solar particle events: assessment from historical proxy data. Astrophys. J. **757**, 92 (2012)
Usoskin, I.G., Kovaltsov, G.: A.: mind the gap: new precise 14C data indicate the nature of extreme solar particle events. Geophys. Res. Lett. **48**, e2021GL094848 (2021)
Usoskin, I.G., Hulot, G., Gallet, Y., et al.: Evidence for distinct modes of solar activity. Astron. Astrophys. **562**, L10 (2014)
Usoskin, I.G., et al.: The AD775 cosmic event revisited: the Sun is to blame. Astron. Astrophys. **552**, L3 (2013)
Usoskin, I.G., Kovaltsov, G.A., Mishina, L.N., Sokoloff, D.D., Vaquero, J.: An optical atmospheric phenomenon observed in 1670 over the City of astrakhan was not a mid-latitude Aurora. Sol. Phys. **292**, 15 (2017)
Uusitalo, J., et al.: Solar superstorm of AD 774 recorded subannually by Arctic tree rings. Nat. Commun. **9**, 3495 (2018)

Van der Sluijs, M.A., Hayakawa, H.: A candidate auroral report in the Bamboo Annals, indicating a possible extreme space weather event in the early 10th century BCE. Adv. Space Res. (2022). https://doi.org/10.1016/j.asr.2022.01.010

Vaquero, J.M., Vázquez, M.: The Sun Recorded through History: Scientific Data Extracted from Historical Documents. Springer, Berlin (2009)

Veronig, A.M., et al.: Indications of stellar coronal mass ejections through coronal dimmings. Nature Astron. **5**, 697–706 (2021)

Vida, K., et al.: The quest for stellar coronal mass ejections in late-type stars. I. Investigating Balmer-line asymmetries of single stars in virtual observatory data. Astron. Astrophys. **623**, A49 (2019)

Watanabe, K., et al.: Characteristics that produce white-light enhancements in solar flares observed by Hinode/SOT. Astrophys. J. **850**(2), 204 (2017)

Willis, D.M., Stephenson, F.R.: Solar and auroral evidence for an intense recurrent geomagnetic storm during December in AD 1128. Ann. Geophys. **19**, 289–302 (2001)

Willis, D.M., Davda, V.N., Stephenson, F.R.: Comparison between oriental and occidental sunspot observations. Q. J. R. Astron. Soc. **37**, 189 (1996a)

Willis, D.M., Stephenson, F.R., Singh, J.R.: Auroral observations on AD 1770 September 16: the earliest known conjugate sightings. Q. J. R. Astron. Soc. **37**, 733 (1996b)

Willis, D.M., Armstrong, G.M., Ault, C.E., Stephenson, F.R.: Identification of possible intense historical geomagnetic storms using combined sunspot and auroral observations from East Asia. Ann. Geophys. **23**, 945–971 (2005)

Willis, D.M., Stephenson, F.R.: Simultaneous auroral observations described in the historical records of China, Japan and Korea from ancient times to AD 1700. Ann. Geophys. **18**, 1–10 (2000)

Willis, D.M., Stephenson, F.R., Fang, H.: Sporadic aurorae observed in East Asia. Ann. Geophys. **25**, 417–436 (2007)

Willis, D.M., et al.: Sunspot observations on 10 and 11 February 1917: a case study in collating known and previously undocumented records. Space Weather. **16**, 1740–1752 (2018)

Wright, N.J.: The stellar-activity-rotation relationship and the evolution of stellar dynamos. Astrophys. J. **743**(1), 48 (2011)

Wu, C.-J., Krivova, N.A., Solanki, S.K., Usoskin, I.G.: Solar total and spectral irradiance reconstruction over the last 9000 years. Astron. Astrophys. **620**, A120 (2018)

Yamashiki, Y., et al.: Impact of stellar superflares on planetary habitability. Astrophys. J. **881**(2), 114 (2019)

Yau, K.K.C., Stephenson, F.R.: A revised catalogue of far eastern observations of sunspots (165 BC to AD 1918). R. Astron. Soc. Quart. J. **29**, 175–197 (1988)

Yokoyama, N., Kamide, Y., Miyaoka, H.: The size of the auroral belt during magnetic storms. Ann. Geophys. **16**, 566–573 (1998)

Part IV
Prediction of Solar Cycle Activity and Its Impact on Climate

Chapter 12
Eleven-Year Cycle of Solar Magnetic Activity: Observations, Theories, and Numerical Model Predictions

Takashi Sakurai, Hideyuki Hotta, and Shinsuke Imada

12.1 Sunspots and Their Eleven-Year Cycle

12.1.1 Discovery of Sunspots and Their Eleven-Year Cycle

Although large sunspots visible to the naked eye at sunset and other times seem to have been known in East Asia for many years, modern observations with telescopes began only in 1610–1611 (Goldberg 1953; Bray and Loughhead 1964). From early observations by Galileo and Scheiner to approximately 1630, it became known that sunspots have dark and semi-dark areas and the Sun rotates with a period of approximately 1 month (using sunspots as markers). It was also found that the axis of rotation is tilted approximately 7° to the ecliptic plane, sunspots do not appear at high latitudes, and the rotation of the Sun is faster closer to the equator (differential rotation). The next important discovery was of the 11-year periodic variations in sunspots by Schwabe (1844), who observed sunspots over the period of 1825–1865 with 8500 drawings. Motivated by this, Wolf (1851) defined the so-called "sunspot relative number" and made his own observations from 1847 to 1893.

The sunspot relative number, R, as introduced by Wolf, is expressed in terms of the numbers of sunspot groups g and individual sunspots f by.

T. Sakurai (✉)
Solar Science Observatory, National Astronomical Observatory of Japan, Tokyo, Japan
e-mail: sakurai.takashi@nao.ac.jp

H. Hotta
Department of Physics, Graduate School of Science, Chiba University, Chiba, Japan

S. Imada
Department of Earth and Planetary Science, Graduate School of Science, The University of Tokyo, Tokyo, Japan

K. Kusano (ed.), *Solar-Terrestrial Environmental Prediction*,
https://doi.org/10.1007/978-981-19-7765-7_12

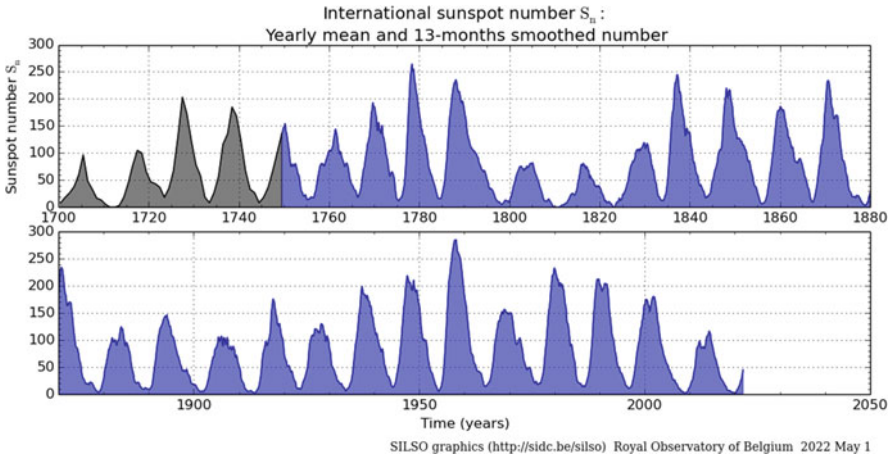

Fig. 12.1 Sunspot relative numbers (http://sidc.be/silso/home). Data shown in blue are monthly averages, whereas only annual averages are provided for the period shown in gray (1700–1749) owing to a lack of detailed data. (SILSO, Royal Observatory of Belgium, Brussels, https://wwwbis.sidc.be/silso/yearlyssnplot)

$$R = k(10g + f).$$

Observation conditions may affect the value of f, and in comparison, the value of g would be more reliable; therefore, the weight was increased for g. Because the average number of sunspots in a group is approximately 10, it can be stated that 10 g and f were so defined to possess similar magnitudes. Wolf (1881) also examined past sunspot observations and determined the monthly mean of the relative number of sunspots. He studied past sunspot observations, derived monthly averages of sunspot relative numbers since 1750, and numbered the cycles with the maximum in 1760 as Cycle No. 1. Wolf derived only the annual averages of the sunspot relative numbers for the years 1700–1749, and prior to that, up to 1610, he only provided the years of the maximum and minimum of the relative numbers.

The constant, k, in the equation defining R was intended to adjust for the deviations from the standard observers and the standard telescope (the 8-cm refractor of Wolf in Zurich). However, Wolf actually chose $k = 1$ for the small telescopes he used during his business trips and for observations in the past (when the telescopes were still primitive) and $k = 0.6$ for the data from his 8-cm telescope in Zurich. This caused confusion subsequently, and in 2015, a major revision of the sunspot relative number was performed, including the definition of k (Clette and Lefèvre 2016). Currently, k is defined as 1 for the standard Zurich observations; therefore, the previous sunspot relative numbers were multiplied by $1/0.6 = 1.67$. Figure 12.1 shows the sunspot relative numbers as compiled by the Sunspot Data Center at Brussels (http://sidc.be/silso/home).

The behavior of sunspots that appear at mid-latitudes during the cycle minimum and move toward the equator near the end of the cycle was discovered by Carrington

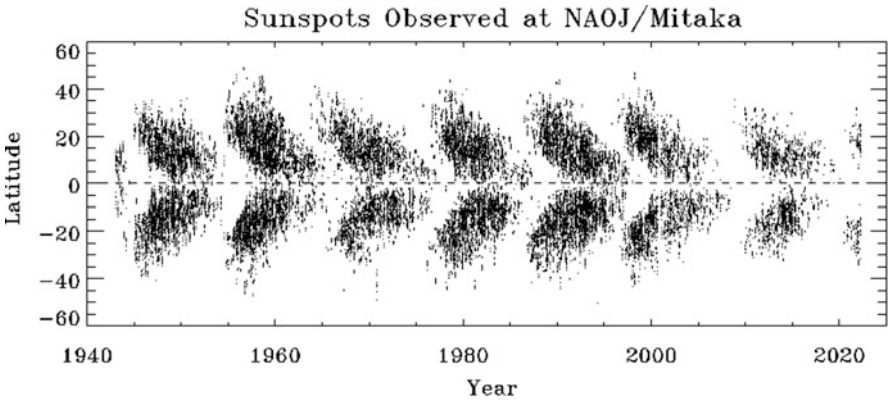

Fig. 12.2 Time variations in the latitudinal distribution of sunspots. (Butterfly diagram; National Astronomical Observatory of Japan)

(1858) and is called Spörer's law after Spörer (1874), who continued detailed observations of sunspots. A diagram showing the latitude distribution of sunspots, with time (year) on the horizontal axis and latitude on the vertical axis, is called the butterfly diagram (Maunder 1904). Figure 12.2 shows the butterfly diagram based on data from the National Astronomical Observatory of Japan.

12.1.2 Beginning of Astrophysical Observations

Hale (1908) discovered in the spectra of sunspots that the absorption lines show the Zeeman effect (the spectral lines split into several components in proportion to the magnetic field strength, as discovered by Zeeman in 1896). From the splitting width, he showed that sunspots have a magnetic field of 1–3 kilo-gauss; this was the first detection of a magnetic field outside the Earth. Apparently, he attempted such observations because he anticipated that sunspots must have a magnetic field, by noticing the spiral structure in the chromosphere in Hα line images captured by a spectroheliograph (which also he invented).

It has been known for a long time that sunspots form groups aligning east–west. Hale and his coworkers found that these groups comprise a pair of magnetic poles (a dipole configuration like a bar magnet). Their alignments (i.e., whether the sunspot located in the direction of rotation (western sunspot, leading sunspot) has positive or negative magnetic polarity) are almost identical in all sunspot groups in each hemisphere and are opposite in the northern and southern hemispheres. Further, they found that this polarity alignment was reversed in the next activity cycle (Hale et al. 1919; Hale and Nicholson 1938). This property is called the Hale–Nicholson law (Fig. 12.3).

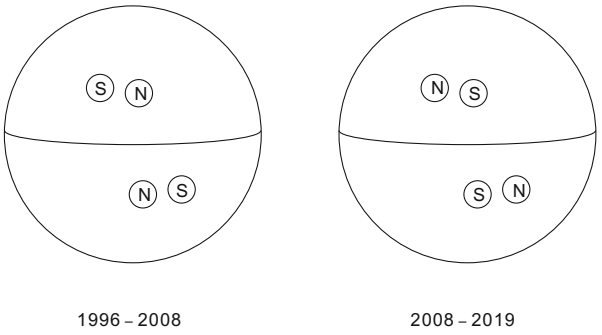

Fig. 12.3 Hale–Nicholson law (north–south antisymmetry in magnetic polarities and 22-year magnetic cycle)

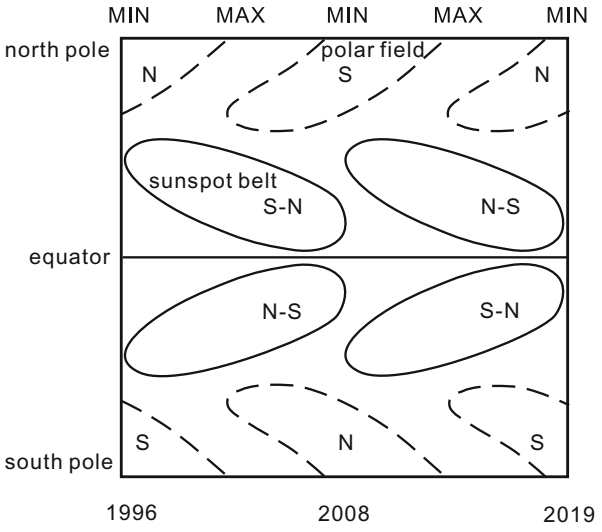

Fig. 12.4 Relationship between the evolution of polar and sunspot fields. This is the schematic of Fig. 12.11

In 1953, Babcock developed an instrument (magnetograph) that could measure the Zeeman effect with high sensitivity by polarization observations (Babcock 1953), and found that the Sun has a magnetic field of several gauss at its poles (Babcock and Babcock 1955) and that the polar field reverses during the active maximum (Babcock 1959). The relationship between the magnetic fields of sunspots and polar regions appears to be such that the trailing sunspots in each hemisphere (sunspots in the eastern part of the group) move toward the poles. Subsequently, they reverse the polar magnetic field near the maximum, which becomes the polar magnetic field for the next cycle (Fig. 12.4).

The next achievement in solar magnetic field observations was the construction of vector magnetographs, which started in the 1980s and was closely related to magnetic energy accumulation in flares. A magnetograph called the Michelson Doppler Imager (MDI) was installed on the Solar and Heliospheric Observatory (SOHO) satellite launched in 1995 and was succeeded by the Helioseismic and Magnetic Imager (HMI) on the Solar Dynamics Observatory (SDO) satellite in 2010.

By observing the magnetic field from space, homogeneous data without gaps due to night or bad weather are currently obtained.

12.2 Solar Dynamo

12.2.1 Development of Plasma Physics (Magnetohydrodynamics) and Classical Dynamo Theory

Chapman, Cowling, Alfvén, Spitzer, and others established the basis of plasma physics and magnetohydrodynamics in 1930–1950. Cowling (1933) showed based on the fundamental equation of magnetohydrodynamics (the induction equation) that an axisymmetric motion cannot maintain a magnetic field (the anti-dynamo theorem). Specifically, the motion that can maintain the magnetic field must be non-axisymmetric, which shows the complexity and difficulty of the problem.

Assuming that the deviation from axisymmetry is small, the velocity field, $\boldsymbol{V}$, and the magnetic field, $\boldsymbol{B}$, are expressed as

$$\boldsymbol{V} = \overline{\boldsymbol{V}} + V', \quad \boldsymbol{B} = \overline{\boldsymbol{B}} + \boldsymbol{B}' \tag{12.1}$$

where $\bar{\boldsymbol{V}}$, $\bar{\boldsymbol{B}}$ are axisymmetric components and $\boldsymbol{V}'$, $\boldsymbol{B}'$ are small perturbations. If we assume that higher-order terms (products of $\boldsymbol{V}'$ and $\boldsymbol{B}'$) in the equation describing the time evolution of $\boldsymbol{B}'$ can be neglected, the mean-field dynamo equation,

$$\frac{\partial \overline{\boldsymbol{B}}}{\partial t} = \nabla \times \left(\overline{\boldsymbol{V}} \times \overline{\boldsymbol{B}} + \alpha \overline{\boldsymbol{B}} \right) + \eta_{\mathrm{t}} \nabla^2 \overline{\boldsymbol{B}} \tag{12.2}$$

$$\alpha \simeq -\frac{\tau}{3} \overline{\boldsymbol{V}' \cdot \nabla \times \boldsymbol{V}'} \tag{12.3}$$

$$\eta_{\mathrm{t}} \simeq \frac{\tau}{3} \overline{V'^2} \tag{12.4}$$

is derived, where τ is the lifetime of the convection driving $\boldsymbol{V}'$, η_{t} is the turbulent diffusion coefficient, and α is a quantity proportional to the kinetic helicity of the convection. Alternatively, in cylindrical coordinates (r, ϕ, z),

$$\bar{B}_{\phi} \equiv B, \qquad \bar{B}_{\mathrm{p}} = \nabla \times (A e_{\varnothing}) \tag{12.5}$$

and introducing the angular velocity of rotation Ω, we obtain

$$\begin{cases} \dfrac{\partial B}{\partial t} = \underset{\Omega\text{-effect}}{r(\nabla\Omega\times\nabla A)_\phi} + \underset{\text{diffusion}}{\eta_t\left(\nabla^2-\dfrac{1}{r^2}\right)B} + \underset{\alpha^2\text{ dynamo}}{\left[\nabla\times\left(\alpha B_p\right)\right]_\phi} \\ \dfrac{\partial A}{\partial t} = \underset{\alpha\text{-effect}}{\alpha B} + \underset{\text{diffusion}}{\eta_t\left(\nabla^2-\dfrac{1}{r^2}\right)A} \end{cases} \tag{12.6}$$

If there is a gradient $\nabla\Omega$ in the rotation velocity, an east–west-oriented magnetic field (toroidal field) B is created from a magnetic field $\boldsymbol{B}$p (poloidal field) in the meridional plane, which is called the Ω effect. If α is nonzero, A (i.e., $\boldsymbol{B}$p) is created from B, which is the α effect. If these overcome the diffusion term, the magnetic field can grow, which is called the $\alpha\Omega$ dynamo.

The cyclonic convection dynamo of Parker (1955) is the earliest theory of this type; however, the so-called Potsdam School (Steenbeck, Krause, and Raedler) provided a more systematic formulation of the $\alpha\Omega$ dynamo after 1966. We can also prove that the generated magnetic field propagates along the isocontours of Ω, equatorward if $\alpha d\Omega/dr < 0$ in the northern hemisphere and poleward if $\alpha d\Omega/dr > 0$, called the Parker–Yoshimura rule (Parker 1955; Yoshimura 1975).

According to the concepts of Parker and the Potsdam School, the α effect is due to convective cells in the convection zone under the influence of the Coriolis force. However, in another concept, one puts more weight on the fact that the leading sunspot appears slightly equatorward and the trailing sunspot slightly poleward. This tilt of the dipole axis from the east–west line (Hale et al. 1919) is called Joy's law (Zirin 1988). In the phenomenological model of Babcock (Babcock 1961; Fig. 12.5), the magnetic fields of trailing sunspots at higher latitudes are diffused to the poles to create the seed fields for the next cycle, and in the kinematic model of Leighton (1969), this process is quantified. Both models can be considered dynamos in which the α effect acts on the solar surface. Hence the process in these models is referred to as the Babcock–Leighton α effect.

12.2.2 Age of Chaos

Studies on solar dynamos by MHD numerical simulations started in the 1980s. In parallel, the distribution of the rotational velocities inside the Sun was determined by helioseismology, a method to infer the solar internal structure from observations of surface oscillations (Fig. 12.6). The results showed that there is a layer of large angular velocity gradient, called a tachocline, around the bottom of the convection zone (dotted line in Fig. 12.6). It is considered a suitable location for the generation of magnetic fields by the Ω effect, because the gradient of Ω is large and the magnetic field can be maintained for a long time owing to low buoyancy. The angular velocity distribution in the convection zone is almost constant in the radial direction, which is different from the Taylor–Proudman theorem for rotating fluids, according to which the angular velocity should be constant along the rotation axis.

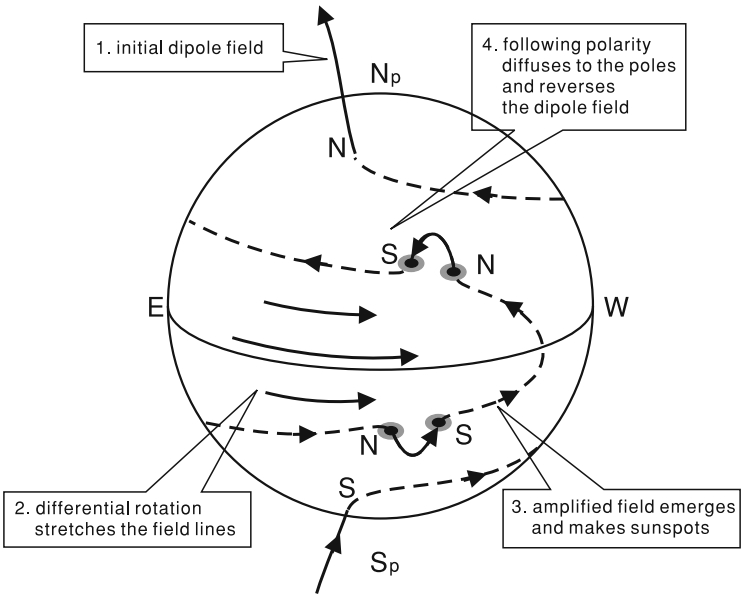

Fig. 12.5 Phenomenological dynamo model of Babcock. (Reproduced from Sakurai (2012), ©Astronomical Society of the Pacific)

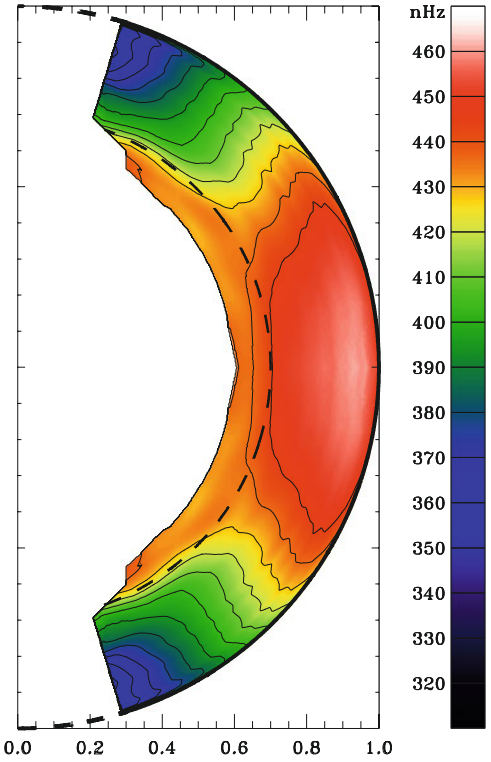

Fig. 12.6 Distribution of rotation velocities in solar interior obtained by helioseismology (Schou et al. 1998; Sekii 2009)

Under such conditions, the classical dynamo theory predicts that the generated magnetic field will migrate toward higher latitudes, which is opposite to the actual case. In a phenomenological model, such as the Wang–Sheeley model (Sheeley et al. 1987), the magnetic flux emergence is taken from observed data, and the magnetic field is driven by a differential rotation and meridional flow and diffuses by random walk. Such a model can reproduce the observed behavior of surface magnetic fields only if a poleward flow is assumed at the surface. This poleward flow (meridional circulation), although slow ($\sim$20 m s^{-1}), has been detected by both direct observations and helioseismology.

12.2.3 Magnetic Flux Transport Dynamo

Because an active latitude (a sunspot latitude) would migrate from low to high latitudes in the classical $\alpha\Omega$ dynamo, the observed (and unexpected) differential rotation led to a new model, the magnetic flux transport dynamo (Choudhuri et al. 1995; Dikpati and Charbonneau 1999). A meridional flow from the equator to high latitudes is observed; thus, a flow from high latitudes to the equator, somewhere in the solar interior, considering the mass conservation law, is expected. Therefore, a flow from the poles to the equator near the bottom of the convection layer was assumed as a possible shape of the meridional flow. We also considered that the bottom of the convective layer where a tachocline is located is a suitable location for the generation of magnetic fields and considered a model as shown in Fig. 12.7. A toroidal magnetic field is generated from the original poloidal field by the Ω effect in the tachocline (panel a). The generated toroidal field rises to the surface owing to the magnetic buoyancy and expands, and simultaneously, it is influenced by the Coriolis force, generating a poloidal field, which is called the Babcock–Leighton α effect (panel b). This poloidal field generated on the solar surface is transported to the bottom of the convection layer. The transport process depends on the parameters of the model, and there are two possible mechanisms: meridional flow and turbulent diffusion (panel c). The transported poloidal magnetic field is again stretched by the shearing flow in the tachocline. The generated field is in the opposite direction to the original toroidal field (panel d). The toroidal field generated here is under the influence of the meridional flow from the poles to the equator. The toroidal magnetic field migrates equatorward. This explains the butterfly diagram. The model has numerous parameters, and well-tuned models can reproduce the butterfly diagram as observed for the real Sun (Fig. 12.8).

Although the flux transport dynamo itself is very successful, there is a major assumption about the shape of the meridional flow. Moreover, owing to its low velocity, it is very difficult to measure the meridional flow inside the solar convection layer. A recent observation by helioseismology showed that the meridional flow does not have a one-cell shape as predicted by the flux transport dynamo; instead it folds back in the middle of the convection layer (Zhao et al. 2013). In contrast, another result suggests a one-cell meridional flow structure (Gizon et al. 2020). This

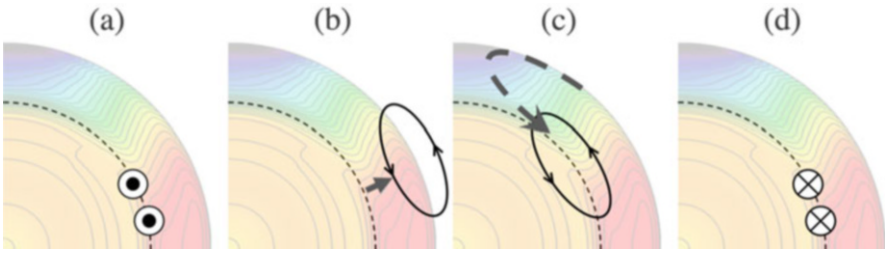

Fig. 12.7 Schematic of flux transport dynamo; (a) Ω effect, (b) α effect, (c) transport process, and (d) field reversal. See the main text for details

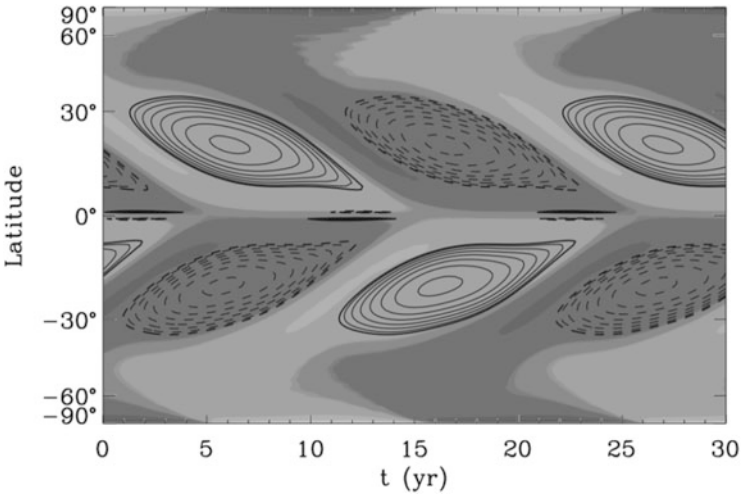

Fig. 12.8 Numerical results from magnetic flux transport dynamo. Gray scale represents the radial magnetic field at the solar surface, and contours represent the toroidal magnetic field at the bottom of the convection zone (Dikpati et al. 2004). ©AAS Reproduced with permission

situation means that the observation studies have not yet converged and that the validity of the flux transport dynamo is also incomplete.

12.2.4 Models for Thermal Convection

In kinematic dynamos, such as the classical αΩ dynamo and the magnetic flux transport dynamo, one assumes a velocity and examines how the magnetic field is generated. Consequently, such a model may break down when new observations, such as a differential rotation or a meridional flow, emerge. In most of these models, axisymmetric two-dimensional calculations are performed and an approximation called the mean field is used, and for nonaxisymmetric components such as

turbulence, the "turbulence model" is used. The turbulence models used here are empirical, instead of those derived from the first principles, and their validity is not always clear.

A completely different approach from these kinematic and mean-field approaches is a three-dimensional (3D) thermal convection model. Using the Boussinesq approximation, which neglects the solar density stratification, in the 1970s, Gilman (1977) already performed 3D thermal convection calculations. They showed how thermal convection carries angular momentum and accelerates the equator. In the 2000s, calculations using an anelastic approximation that considers solar stratification (assuming $\nabla \cdot (\rho_0 \boldsymbol{V}) = 0$, where mean density ρ_0 depends on the radial position) were initiated by a group in Colorado. Consequently, numerous results were obtained for differential rotation and meridional flow generation (e.g., Miesch et al. 2000). In the 2010s, it became possible to reproduce the magnetic cycle using parameters similar to those used for the Sun (e.g., Ghizaru et al. 2010). A detailed analysis of the calculations reproducing the magnetic field periodicity and a discussion of the physical mechanisms driving the large-scale magnetic field and its cyclic behavior showed that most results of 3D thermal convection models can be explained by the $\alpha\Omega$ dynamo (Warnecke et al. 2018). However, there is still no 3D thermal convection model that supports the flux transport dynamo.

There are few undefined parameters in the thermal convection model, which uses the solar standard model (Christensen-Dalsgaard et al. 1996) for the initial stratification. However, we cannot resolve the smallest-scale eddies in the solar interior even with current supercomputers. Concurrently, we have found that simulations that resolve a small-scale turbulence can produce a solar-like periodicity (Hotta et al. 2016). However, currently, there is still a large gap between the thermal convection distributions observed in the real Sun and numerical simulations (e.g., Hanasoge et al. 2012; Lord et al. 2014). In particular, the largest thermal convection observed on the surface of the Sun is supergranulation, which has a scale of approximately 30,000 km. In contrast, when numerical simulations including the entire convection layer are performed, the giant cell, which typically has a scale of approximately 200,000 km, dominates (Lord et al. 2014). Although the cause of this gap is not yet known, it is considered a phenomenon that is critically missing in numerical simulations, such as those of resolution or magnetic field. Featherstone and Hindman (2016) suggested that strong rotation effects in the solar interior could produce supergranular cells. However, the mechanism producing the strong effects of rotation is unknown. Most recently, super high-resolution simulations showed the suppression of large-scale thermal convection (Hotta and Kusano 2021). Although the differential rotation is reproduced well, there still are gaps between observations and simulations in convection velocities. Thus, continued research is required on this topic.

12.3 Prediction of Solar-Cycle Amplitude by Numerical Modeling

12.3.1 Solar Cycle Activity Prediction and its Social Demand

As seen in the construction of the International Space Station and the Lunar Gateway Project, space has become a location for human activities, and there is increasing social interest in and demand for predicting changes in the solar-terrestrial environment, which is the most familiar space. The solar-terrestrial environment in which we live is significantly affected by solar activity, and it is necessary to understand solar variability in order to understand the variability in the solar-terrestrial environment. It is known that the sunspot number varies over a period of approximately 11 years (see Sect. 12.1). In addition, as explained in Part 3, the magnetic activity related to sunspots is a major factor in solar-terrestrial environment variations, and the entire heliosphere also varies over a period of approximately 11 years.

There are three types of solar variabilities: short-timescale variability of approximately 1 h caused by sudden events such as solar flares, decadal variability caused by the solar activity cycle, and long-term variability of several decades or longer. Thus far, space weather research on short timescales, such as flare prediction as described in Part 3, has been vigorously pursued. In contrast, the mechanism of solar cycle activity, which is discussed in Sect. 12.2, is not as well understood as that of solar flares, and there are a limited number of space weather studies on long timescales.

It is known that the degree of solar activity (the number of sunspots) differs in each cycle, with some cycles having numerous sunspots and others having almost no sunspots, such as the Maunder Minimum (see Fig. 12.1). In fact, the current cycle, solar cycle 24, is the most unique cycle with the lowest number of sunspots in the past 100 years. However, the cause of this phenomenon is still not understood well, and it remains to be seen whether the next solar cycle, solar cycle 25, will be a very weak one like the Maunder Minimum or a relatively active one like solar cycle 23 or earlier.

Although various studies worldwide have attempted to predict the solar cycle amplitude using different methods, the prediction results are still debatable (see Fig. 12.9). In recent years, a method of predicting the next cycle activity by estimating the strength of the magnetic field in the polar regions, which is considered the seed of the next solar cycle, has been generating interest. In fact, the correlation between the strength of the magnetic field in the polar regions during the solar minimum and the activity of the next solar cycle has been verified observationally at least in recent cycles.

In this section, we briefly introduce various solar cycle prediction methods and their results. In particular, we explain the methods and results for predicting the next solar cycle activity by predicting the polar magnetic field during the solar minimum period by numerical calculations, which has gained significant attention recently.

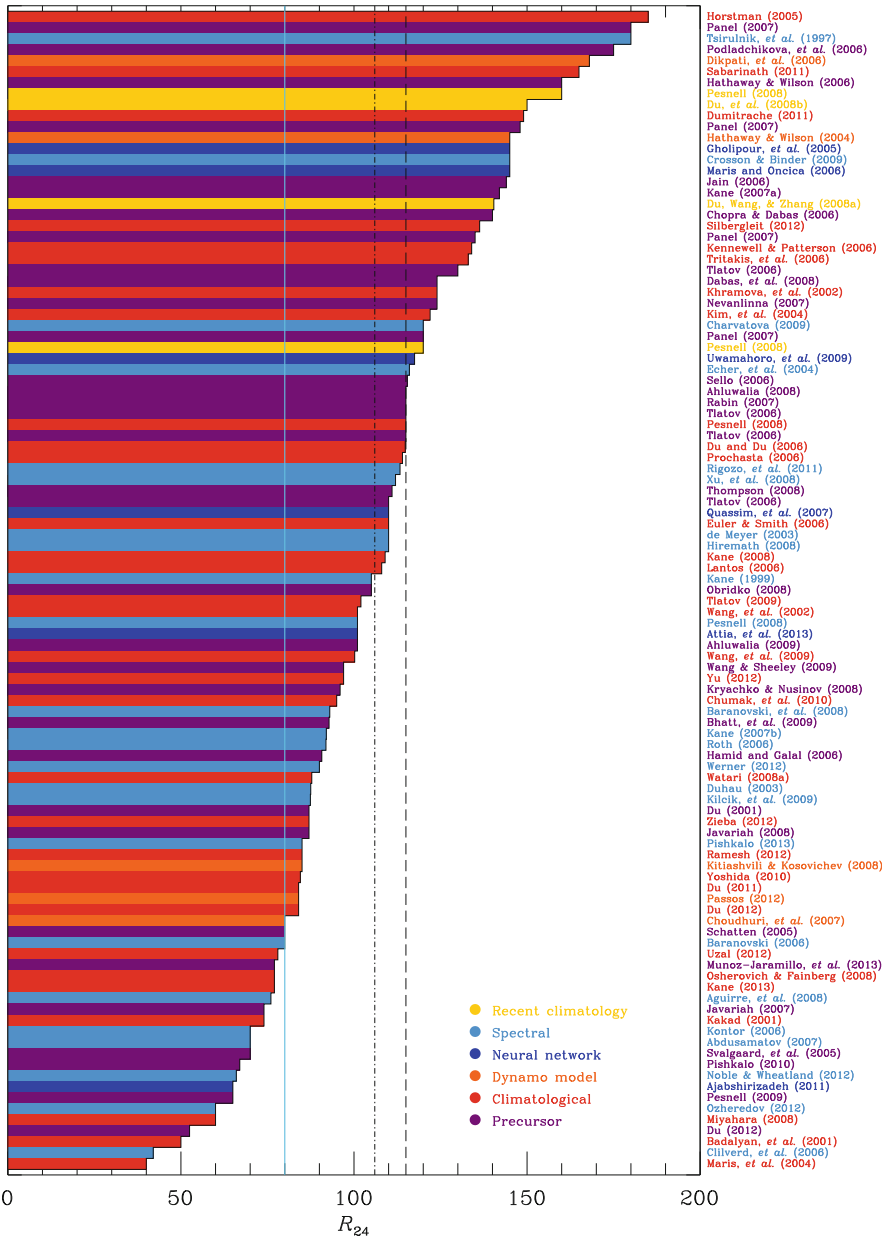

Fig. 12.9 Prediction results for the activity level of solar cycle 24 (Pesnell 2016). Results are shown in descending order from top to bottom. Correct solution for solar cycle 24, sunspot relative number in 2014 (maximum period of solar cycle 24), is represented by a blue line. Dashed and dotted lines are drawn for the entire forecast and average of climatology, respectively. Note: Sunspot relative numbers are before the revision of Clette and Lefèvre (2016). Post-revision values are approximately 1.67 times pre-revision values

12.3.2 Previous Studies on Periodic Activity Prediction

Various methods have been used to study the prediction of solar cyclic activity. According to Pesnell (2016), they are classified into six main categories: climatology (Fig. 12.9: red), recent climatology (yellow), precursor (purple), dynamo model (orange), spectral (light blue), and neural network (blue). These six methods are briefly explained below.

Climatology prediction (e.g., Du and Du 2006) assumes that the future state of a system can be determined only from the statistical properties of its past state. Specifically, this method uses the sunspot number information of all previous cycles as the input to derive the statistical properties and the sunspot number for the next cycle.

Recent climatology prediction determines the future state of a system only from the statistical characteristics of a relatively recent past state of the system. For example, in the prediction of solar cycle 24, the data used are limited to the relatively recent past, such as the data after solar cycle 17.

Precursor prediction (e.g., Svalgaard et al. 2005) determines the future state of a system using features of that system. There are two main types of predictions of solar cycle activity. The first is solar polar field prediction, which uses the estimated polar fields during the solar minimum to predict the activity level for the next cycle. The second is a prediction using the geomagnetic activity level during the solar minimum as an indicator of the solar activity for the next cycle.

Dynamo model prediction (e.g., Choudhuri et al. 2007) uses numerical solutions of physics-based equations, such as the flux transport dynamo described in Sect. 12.2.3, to predict the activity of the next cycle. In these studies, the key is the method to incorporate the observed information of the solar surface magnetic field into the model, and methods such as data assimilation are being intensively investigated.

Spectral forecasting (e.g., Duhau 2003) predicts a future system by performing a frequency analysis of the time series data of a system. For the next solar cycle activity prediction, Fourier analysis (autoregressive prediction, wavelet-based prediction) of sunspot time series data is typically conducted to predict.

The last prediction type uses neural networks (e.g., Quassim et al. 2007), which have rapidly developed in recent years. In this method, the complex relationship between the state of a system (output) and the state of a previous system (input) is modeled by a neural network to predict the future system.

As described above, various methods have been used to predict the next solar cycle activity, and Fig. 12.10 summarizes the variance of the prediction results and skill scores (see Chap. 10 for an explanation of skill scores) for each method. The variance is relatively large for all categories of forecasts, indicating that it is difficult to establish a forecasting method. In addition, the skill scores for forecasting are negative or close to zero in most categories, suggesting that forecasting is difficult. The only prediction with a high skill score of 0.73 is the prediction using the solar polar magnetic fields, indicating that it yields the best result, at least for the prediction of the activity of solar cycle 24.

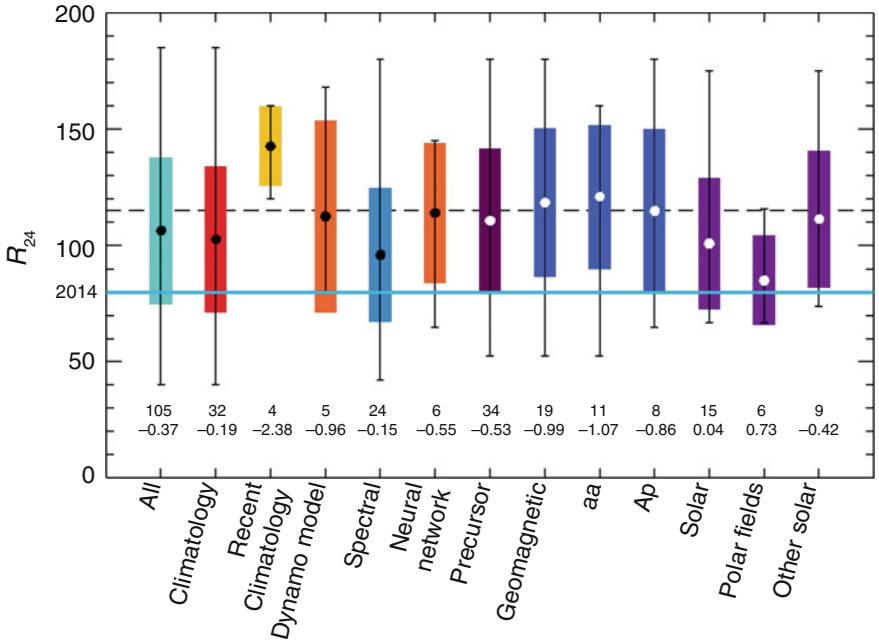

Fig. 12.10 Summary of cycle 24 predictions from various methods (Pesnell 2016). Sunspot relative number during maximum of solar cycle 24 is represented by a light blue line. Average over entire forecast is represented by a dashed line. Length of color bars represents 1σ width of predictions in each category, and error bars represent minimum and maximum predictions in each category. Numbers also represent the number of forecast results in that category and the forecast skill score. Note: Sunspot relative numbers are the values before revision

12.3.3 Relationship Between Magnetic Field of Polar Regions During Solar Minimum and Next Cycle Activity

As discussed in the previous section, the prediction of the next solar cycle activity using the solar polar magnetic field is currently the best prediction method. In this section, we present the observation evidence as well as an intuitive explanation of the relationship between the polar magnetic field during the solar minimum and the next cycle activity.

Sunspots appearing on the solar surface are bundles of magnetic fields, and the magnetic field on the solar surface changes as the sunspots decay and are transported by surface flows and diffusion processes. Figure 12.11 shows the time variations in the magnetic field on the solar surface averaged in the longitude direction (magnetic butterfly diagram), and several features are noticeable there. The region where sunspots frequently appear moves to mid-latitudes at the beginning of the cycle and to lower latitudes at the end of the cycle, as discussed in Sect. 12.1. Furthermore, we can see a striated structure extending poleward from the active region belt where

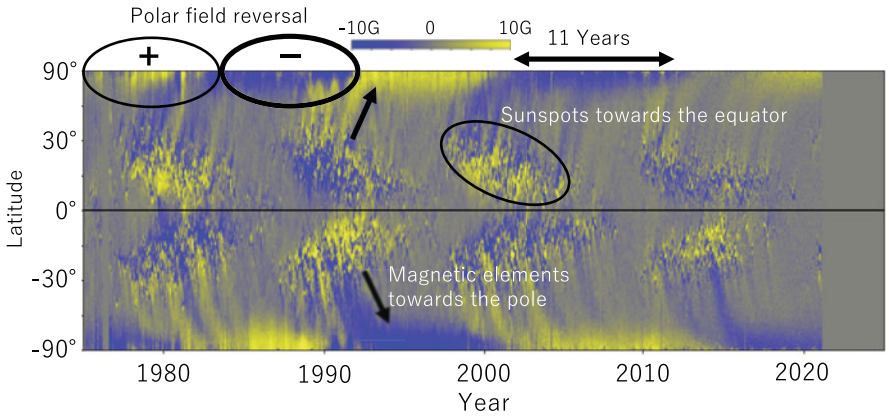

Fig. 12.11 Time variations in the latitudinal distribution of solar magnetic fields averaged in longitude direction (magnetic butterfly diagram; Hathaway 2015). Blue represents negative magnetic fields, and yellow represents positive magnetic fields

sunspots frequently emerge. As explained in Sect. 12.1, leading sunspots are known to appear at lower latitudes than trailing sunspots (Joy's law; Hale et al. 1919, Zirin 1988). This difference in the latitude of appearance allows more of the magnetic fields of the following sunspots to be transported to the polar regions. The polarity of the magnetic field of the following sunspots is opposite to that of the original polar field, and the polar field decreases as the cycle progresses, reversing during the solar activity maxima and becoming strongest during the solar minima. Figure 12.12 shows the observed polar magnetic fields during solar activity cycles 20–24. The strength of the polar field in the minimum period is indicated by orange arrows. It can be seen that the polar magnetic fields in the minimum periods of cycles 20 and 21 were stronger than those in the other two periods. In contrast, the polar field in the minimum period of cycle 23 was extremely weak compared to the others. Thus, how can the activity of the next cycle be explained? Cycles 21 and 22, in which the previous polar field was strong, were relatively active cycles, whereas cycle 24, the next cycle after cycle 23, in which the polar field was extremely weak, was the least active cycle in the last 100 years. Although we have only four examples of cycles, there is certainly a close relationship between the polar magnetic field during the solar minimum and the activity of the next cycle.

The relationship between the polar magnetic field or its equivalent index and the next solar cycle activity level during the solar minimum is shown in Fig. 12.13. Because the solar polar magnetic field determines the global structure of the solar wind in the heliosphere, it is correlated with the geomagnetic index, which indicates the degree of interaction between the solar wind and the magnetosphere of the Earth. There is a linear relationship between the polar magnetic field and the next solar cycle activity during the solar minimum.

We also attempted to gain an intuitive understanding of the importance of the polar magnetic field during the solar minimum for the next solar cycle activity (see

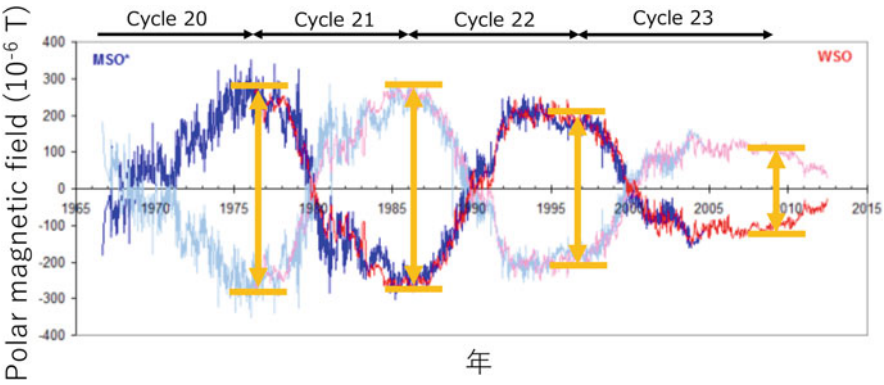

Fig. 12.12 Time variations in the solar polar magnetic field. [Adapted from Svalgaard et al. (2005)]

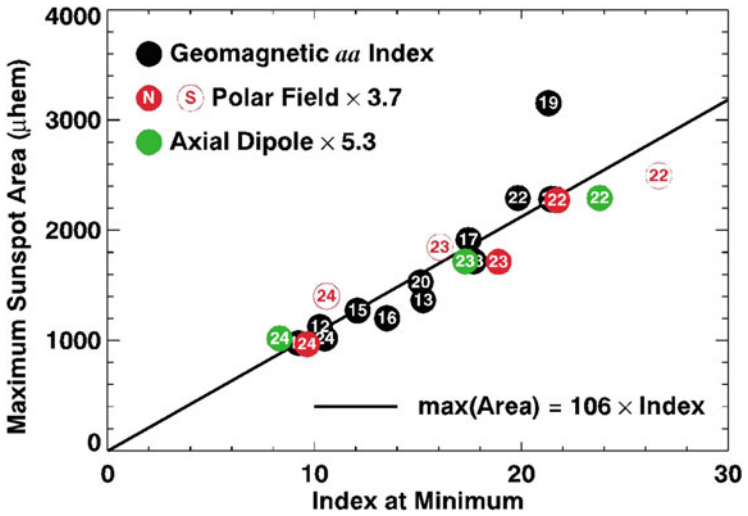

Fig. 12.13 Polar magnetic field or equivalent index (horizontal axis) versus next solar cycle activity (vertical axis) during solar minimum (Hathaway and Upton 2016). Note: Axial dipole is equivalent to a polar magnetic field; see Iijima et al. (2017) for details

Cameron and Schussler (2015) for the mathematical formulation and other details). As explained in Sect. 12.2, the standard model of the solar dynamo considers that a flow in the solar interior is important for the generation of the solar magnetic field. The dipole magnetic field that runs from the north to south poles of the Sun is stretched and amplified by the differential rotation in the solar interior, and is converted into a toroidal magnetic field, which forms sunspots (see Fig. 12.5). Specifically, the seed magnetic fields that produce sunspots are the dipole magnetic fields that run from the north pole to the south pole in the previous cycle. Moreover, a strong seed magnetic field implies a high solar activity in the next cycle.

12.3.4 Solar Surface Flux Transport Model

In the previous section, we explained that the next solar cycle activity can be predicted well by estimating the polar magnetic field in the solar minimum. In this section, we briefly present methods to estimate the polar magnetic field during the solar minimum; refer to Jiang et al. (2014) for details.

The magnetic field on the solar surface originating from sunspots is transported over the entire Sun by the flow and diffusion on the solar surface. This process can be described by the following time evolution equation of the magnetic field. This equation is derived from the MHD equation (for the B_r component in spherical coordinates of the induction equation),

$$\begin{aligned}\frac{\partial B_r}{\partial t} = &-\frac{1}{R_s \sin\theta}\frac{\partial}{\partial \phi}(uB_r) - \frac{1}{R_s \sin\theta}\frac{\partial}{\partial \theta}(vB_r \sin\theta) \\ &+\eta\left[\frac{1}{R_s^2 \sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial B_r}{\partial \theta}\right) + \frac{1}{R_s^2 \sin^2\theta}\frac{\partial^2 B_r}{\partial \phi^2}\right] + S(\theta, \phi, t).\end{aligned} \tag{12.7}$$

Here, θ and φ represent the latitude and longitude coordinates, respectively, and u and v represent the flows in the direction of rotation (differential rotation) and meridional direction (meridional circulation), respectively. η is the coefficient of turbulent diffusion. Furthermore, S represents the magnetic field supplied from the solar interior by the emerging magnetic flux. The time evolution equation of this magnetic field is in the form of an advection–diffusion equation in a spherical shell with the addition of a source term.

The velocities of the differential rotation and meridional circulation commonly used in solar surface flux transport (SFT) calculations are as follows (Jiang et al. 2014):

$$u(\theta) = 13.38 - 2.30\ \cos^2\theta - 1.62\ \cos^4\theta - 13.2 \quad \text{deg day}^{-1} \tag{12.8}$$

$$v(\theta) = 31.3\ |\sin\theta|^{2.5} \cos\theta \quad \text{m s}^{-1} \tag{12.9}$$

Most of the parameters used in the SFT model can be determined from observations; however, some are difficult to obtain, such as the location, date, time, size, field strength, polarity, and shape of the emerging magnetic flux (sunspots), which are related to the future characteristics of sunspots.

Figure 12.14 shows the time evolutions of the magnetic field at the solar surface reproduced by the SFT model (upper panel) and obtained by observation (lower panel). It can be seen that the SFT model calculation reproduces well the time variations in the solar surface magnetic field, including the transport of the magnetic field to the polar regions. This result shows that we can basically estimate a future polar magnetic field as long as information on the future surface magnetic field is available.

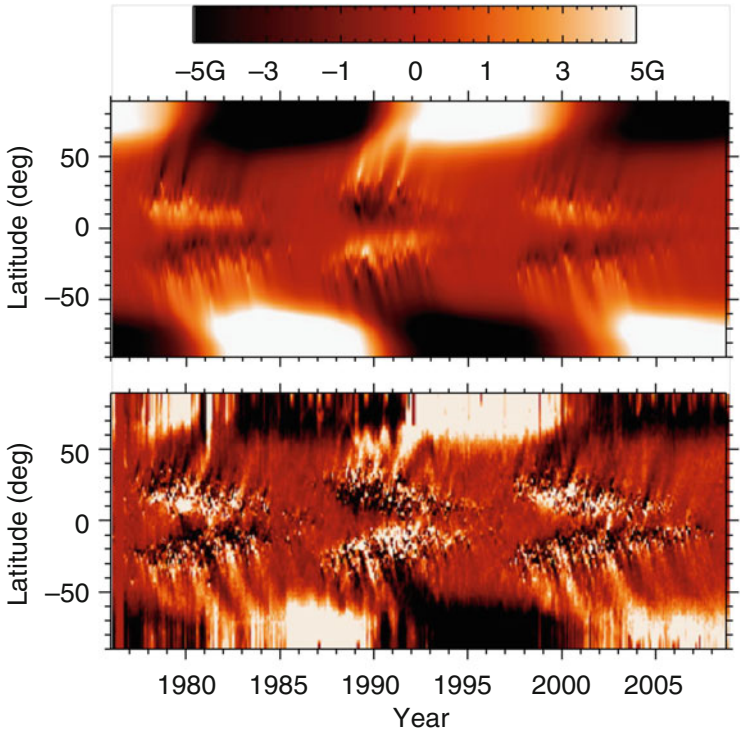

Fig. 12.14 Time variations in the latitudinal distribution of solar surface magnetic fields averaged in longitude direction (magnetic butterfly diagram; Jiang et al. 2014). Top row: calculated results using the SFT model; bottom row: observed results

How do we determine the parameters involved in the emergence of magnetic flux (sunspots) in the future? In most of the studies, these parameters have been modeled using statistical rules, e.g., Joy's law described in Sect. 12.1 (Cameron et al. 2016). Some studies have also used information from similar past cycles to determine the parameters of the emerging magnetic flux in the future. For example, Upton and Hathaway (2014) found that the temporal variations in the sunspot number during solar activity cycle 24 are similar to that of cycle 14 (Fig. 12.15), and therefore, used the one observed during cycle 14 as the future emerging flux parameter.

It should be noted that a few relatively large sunspots are actually sufficient to reverse the polar field. Of course, sunspots with positive and negative polarity appear in pairs, and it is difficult to transport only one of them to the polar regions. In comparison, sunspots that deviate from Joy's law are frequently observed, and one such distinct sunspot may have a strong influence on the polar field reversal. Therefore, the determination of the parameters specifying the emerging magnetic flux is very important.

Iijima et al. (2017) performed an SFT model calculation ignoring the future emerging magnetic flux ($S = 0$ is assumed in Eq. (12.7)) and compared the results

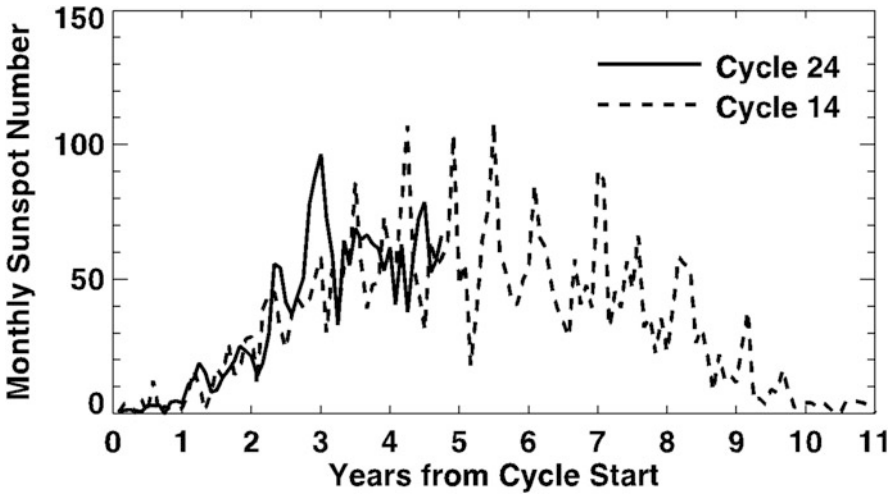

Fig. 12.15 Time variations in sunspot numbers for solar cycle 24 compared to those for solar cycle 14 (Upton and Hathaway 2014)

of observations and model calculations of past cycles. They identified the period of the cycle in which predictions independent of the future emerging magnetic flux can be made. Figure 12.16 compares the observed polar field with the results of the SFT model calculation without new sunspots. The thick line represents the observed polar magnetic field, and the thin line represents the time variations in the polar magnetic field by the SFT model calculation. It can be seen that there is almost no difference between the results of the SFT model calculation, which ignores the future emerging magnetic flux, and the results from the polar field observation from approximately 3 years before the minimum period. It can be inferred that the sunspots appearing in the last 3 years of the solar cycle do not affect the solar polar field variations for some physical reason. Specifically, the effect of sunspot emergence can be neglected from 3 years before the solar minimum, and we can predict the polar magnetic field from approximately 3 years before the solar minimum without sunspot information. There are various reasons why sunspots at the end of the cycle do not affect the polar field, which are still under exploration.

Subsequently, we used the polar magnetic fields estimated by the SFT model to predict the sunspot relative numbers in the solar activity cycle 25 maximum in around 2025. As mentioned above, there is a good correlation between the polar magnetic field in the solar minimum and the sunspot numbers in the next cycle. For the past three cycles, the relationship between the prediction of the polar magnetic field of the active minimum period approximately 3 years ago and the observed maximum sunspot relative numbers in the next cycle is shown by the diamonds in Fig. 12.17, indicating a good correlation. If we predict the maximum value of the sunspot relative numbers in solar cycle 25 using the predicted value of the polar field in the solar minimum of cycle 24 based on this relationship, it becomes like the plus

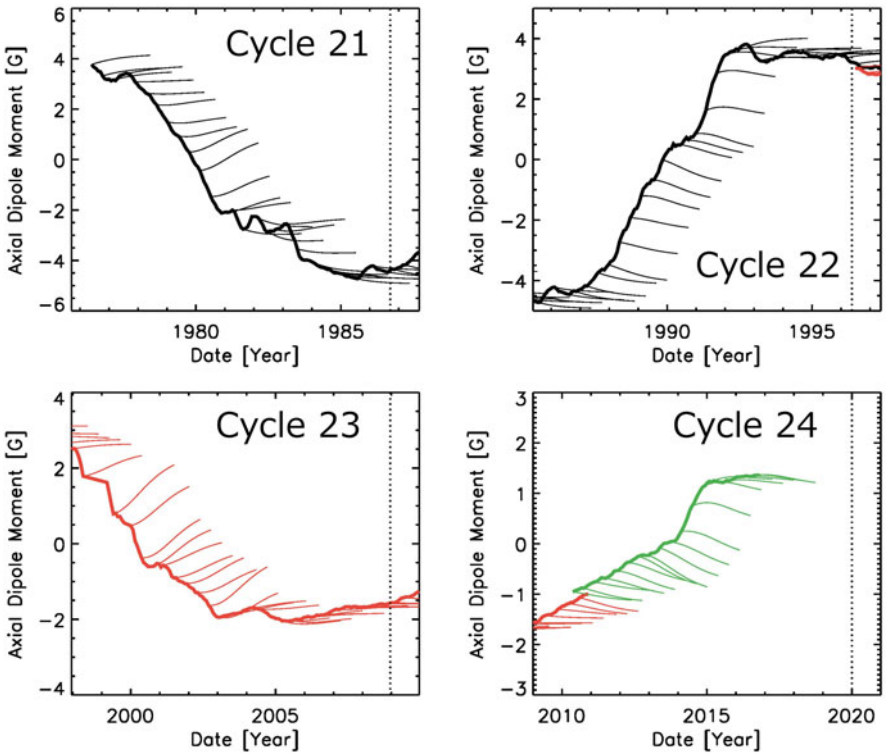

Fig. 12.16 Comparison of the observed time evolution of the polar field (thick line) and results of the SFT model calculation (thin line), ignoring the appearance of new sunspots (Iijima et al. 2017). Data from the Wilcox Solar Observatory at Stanford University: MDI on SOHO is in red, and HMI on SDO are in black, red, and green, respectively

sign in Fig. 12.17. Consequently, solar cycle 25 will be approximately 70% as active as solar cycle 24, which was the least active one in the last 100 years. Various studies have predicted the activity level of solar cycle 25 similarly, and most of them show that the activity level of solar cycle 25 will be as low as that of solar cycle.

12.3.5 Future Issues in Solar Cycle Activity Prediction Research

Finally, we discuss the future development of solar cycle prediction research. There are three major issues that need to be addressed: (1) What are the extents of the effects of the temporal and spatial variations in the physical parameters used in the SFT model on the prediction results? (2) Is it possible to predict cycles beyond

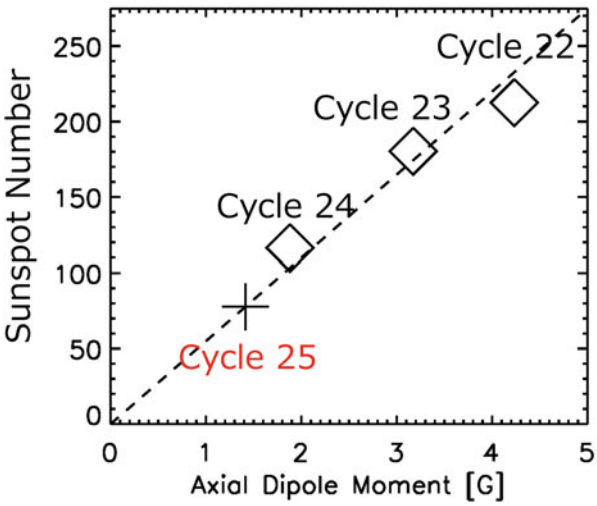

Fig. 12.17 Relationship between activity level of the next solar cycle and polar field (axial dipole moment) predicted years before activity minimum (marked using diamonds). Predicted results for solar cycle 25 are shown by + symbols (Iijima et al. 2017)

the next cycle? And (3) is it possible to predict extremely weak cycles such as the Maunder Minimum?

Most previous prediction studies have assumed that the differential rotation and meridional circulation flows are constant in time. However, observationally, these flows are known to vary over solar cycles, and even within the same cycle, they are known to vary between the maxima and minima (Hathaway and Rightmire 2011). How much these factors affect the predictions is an important problem to be considered. In addition, it has recently been shown that these flows are also affected by the strength of the magnetic field (Imada and Fujiyama 2018), and the effect of the feedback of the magnetic field strength on them is also a crucial issue.

Current forecasting studies are limited to predicting the next cycle. The largest uncertainties in the next solar cycle prediction studies are the characteristics and occurrence of sunspot emergence, which are difficult to predict more than 3 years before the minimum period, as explained by Iijima et al. (2017). If we can understand the statistical properties of the parameters of the emerging magnetic flux in conjunction with the dynamo, the predictions for cycles beyond the next cycle will become possible. Thus, improving the understanding of the emerging magnetic flux is important for predicting solar cycle activity.

The current solar cycle 24 is a unique cycle with the lowest number of sunspots in the past 100 years, and attention is focused on whether a cycle with very weak activity like the Maunder Minimum will follow in the future. However, the origin of a Maunder Minimum-like cycle is still not understood well. For the prediction of solar cycle activity, it is very critical to understand whether the extended minimum is due to the dynamo processes in the solar interior.

References

Babcock, H.W.: The solar magnetograph. Astrophys. J. **118**, 387 (1953)

Babcock, H.D.: The sun's polar magnetic field. Astrophys. J. **130**, 364 (1959)

Babcock, H.W.: The topology of the sun's magnetic field and the 22-year cycle. Astrophys. J. **133**, 572 (1961)

Babcock, H.W, Babcock, H.D.: The Sun's Magnetic Field, 1952-1954. Astrophys. J. 121, 349 (1955)

Bray, R.J., Loughhead, R.E.: Sunspots, The International Astrophysics Series, p. 1. Chapman and Hall, London (1964)

Cameron, R., Schussler, M.: The crucial role of surface magnetic fields for the solar dynamo. Science. **347**, 1333 (2015)

Cameron, R.H., Jiang, J., Schussler, M.: Solar cycle 25: another moderate cycle? Astrophys. J. Lett. **823**, L22 (2016)

Carrington, R.C.: On the distribution of the solar spots in latitude since the beginning of the year 1854; with a map. Monthly Notices R. Astron. Soc. **19**, 1 (1858)

Choudhuri, A.R., Schussler, M., Dikpati, M.: The solar dynamo with meridional circulation. Astron. Astrophys. **303**, L29 (1995)

Choudhuri, A.R., Chatterjee, P., Jiang, J.: Predicting solar cycle 24 with a solar dynamo model. Phys. Rev. Lett. **98**, 131103 (2007)

Christensen-Dalsgaard, J., Däppen, W., Ajukov, S.V., et al.: The current state of solar modeling. Science. **272**, 1286 (1996)

Clette, F., Lefèvre, L.: The new sunspot number: assembling all corrections. Solar Phys. **291**, 2629 (2016)

Cowling, M.A.: The magnetic field of sunspots. Monthly Notices R. Astron. Soc. **94**, 39 (1933)

Dikpati, M., Charbonneau, P.: A Babcock-Leighton flux transport dynamo with solar-like differential rotation. Astrophys. J. **518**, 508 (1999)

Dikpati, M., de Toma, G., Gilman, P.A., Arge, C.N., White, O.R.: Diagnostics of polar field reversal in solar cycle 23 using a flux transport dynamo model. Astrophys. J. **601**, 1136 (2004)

Du, Z., Du, S.: The relationship between the amplitude and descending time of a solar activity cycle. Solar Phys. **238**, 431 (2006)

Duhau, S.: An early prediction of maximum sunspot number in solar cycle 24. Solar Phys. **213**, 203 (2003)

Featherstone, N.A., Hindman, B.W.: The emergence of solar supergranulation as a natural consequence of rotationally constrained interior convection. Astrophys. J. Lett. **830**, L15 (2016)

Ghizaru, M., Charbonneau, P., Smolarkiewicz, P.K.: Magnetic cycles in global large-eddy simulations of solar convection. Astrophys. J. Lett. **715**, L133 (2010)

Gilman, P.A.: Nonlinear dynamics of Boussinesq convection in a deep rotating spherical Shell. I. Geophys. Astrophys. Fluid Dyn. **8**, 93 (1977)

Gizon, L., et al.: Meridional flow in the Sun's convection zone is a single cell in each hemisphere. Science. **368**, 1469 (2020)

Goldberg, L.: In: Kuiper, G.P. (ed.) The Sun, p. 3. University of Chicago Press, Chicago IL (1953)

Hale, G.E.: On the probable existence of a magnetic field in sun-spots. Astrophys. J. **28**, 315 (1908)

Hale, G.E., Nicholson, S.B.: Magnetic Observations of Sunspots 1917–1924. Carnegie Institution of Washington, Washington, DC (1938)

Hale, G.E., Ellerman, F., Nicholson, S.B., Joy, A.H.: The magnetic polarity of sun-spots. Astrophys. J. **49**, 153 (1919)

Hanasoge, S.M., Duvall Jr., T.L., Sreenivasan, K.R.: Anomalously weak solar convection. PNAS. **109**, 11928 (2012)

Hathaway, D.H.: The solar cycle. Living Rev. Sol. Phys. **12**, 4 (2015)

Hathaway, D.H., Rightmire, L.: Variations in the axisymmetric transport of magnetic elements on the sun: 1996-2010. ApJ. **729**, 80 (2011)

Hathaway, D.H., Upton, L.: Predicting the amplitude and hemispheric asymmetry of solar cycle 25 with surface flux transport. J. Geophys. Res. Space Phys. **121**, 10–744 (2016)
Hotta, H., Kusano, K.: Solar differential rotation reproduced with high-resolution simulation. Nature Astron. **5**, 1100 (2021)
Hotta, H., Rempel, M., Yokoyama, T.: Large-scale magnetic fields at high Reynolds numbers in magnetohydrodynamic simulations. Science. **351**, 1427 (2016)
Iijima, H., Hotta, H., Imada, S., Kusano, K., Shiota, D.: Improvement of solar-cycle prediction: plateau of solar axial dipole moment. Astron. Astrophys. Lett. **602**, L2 (2017)
Imada, S., Fujiyama, M.: Effect of magnetic field strength on solar differential rotation and meridional circulation. Astrophys. J. **864**, L5 (2018)
Jiang, J., Hathaway, D.H., Cameron, R.H., Solanki, S.K., Gizon, L., Upton, L.H.: Magnetic flux transport at the solar surface. Space Sci. Rev. **186**, 491 (2014)
Leighton, R.B.: A magneto-kinematic model of the solar cycle. Astrophys. J. **156**, 1 (1969)
Lord, J.W., Cameron, R.H., Rast, M.P., Rempel, M., Roudier, T.: The role of subsurface flows in solar surface convection: modeling the spectrum of supergranular and larger scale flows. Astrophys. J. **793**, 24 (2014)
Maunder, E.W.: Note on the distribution of sun-spots in heliographic latitude, 1874-1902. Monthly Notices R. Astron. Soc. **64**, 747 (1904)
Miesch, M.S., Elliott, J.R., Toomre, J., et al.: Three-dimensional spherical simulations of solar convection. I. Differential rotation and pattern evolution achieved with laminar and turbulent states. Astrophys. J. **532**, 593 (2000)
Parker, E.N.: Hydromagnetic dynamo models. Astrophys. J. **122**, 293 (1955)
Pesnell, W.D.: Predictions of solar cycle 24: how are we doing? Space Weather. **14**, 10 (2016)
Quassim, M.S., Attia, A.F., Elminir, H.K.: Forecasting the peak amplitude of the 24th and 25th sunspot cycles and accompanying geomagnetic activity. Solar Phys. **243**, 253 (2007)
Sakurai, T.: Helioseismology, dynamo, and magnetic helicity, Progress in Solar/Stellar Physics with Helio- and Asteroseismology (eds. H. Shibahashi, M. Takata, and A. E. Lynas-Gray). Astron. Soc. Pacific Conf. Ser. **462**, 247 (2012)
Schou, J., Antia, H.M., Basu, S., et al.: Helioseismic studies of differential rotation in the solar envelope by the solar oscillations investigation using the Michelson doppler imager. Astrophys. J. **505**, 390 (1998)
Schwabe, H.: Sonnenbeobachtungen im Jahre 1843. Astron. Nachr. **21**, 234 (1844)
Sekii, T.: Helioseismology to Probe the Interior of the Sun. In: Sakurai, T., et al. (eds.) Series of Contemporary Astronomy, Vol.10 "The Sun", p. 45. Nihon Hyoronsha, Tokyo (2009)
Sheeley Jr., N.R., Nash, A.G., Wang, Y.-M.: The origin of rigidly rotating magnetic field patterns on the sun. Astrophys. J. **319**, 481 (1987)
Spörer, G.: Beobachtungen der Sonnenflecken zu Anclam. Publ. Astron. Ges. **13**, 1 (1874)
Svalgaard, L., Cliver, E.W., Kamide, Y.: Correction of errors in scale values for magnetic elements for Helsinki. GRL. **32**, L01104 (2005)
Upton, L., Hathaway, D.H.: Predicting the sun's polar magnetic fields with a surface flux transport model. Astrophys. J. **780**, 5 (2014)
Warnecke, J., Rheinhardt, M., Tuomisto, S., et al.: Turbulent transport coefficients in spherical wedge dynamo simulations of solar-like stars. Astron. Astrophys. **609**, A51 (2018)
Wolf, R.: Sonnenflecken-Beobachtungen in der zweiten Hälfte des Jahres 1850. Mitt. Naturforsch. Ges. Bern. **207**, 89 (1851)
Wolf, R.: Quelques Résultats Déduits de la Statistique Solaire. Mem. Soc. Spettrosc. Ital. **10**, 61 (1881)
Yoshimura, H.: Solar-cycle dynamo wave propagation. Astrophys. J. **201**, 740 (1975)
Zhao, J., Bogart, R.S., Kosovichev, A.G., Duvall Jr., T.L., Hartlep, T.: Detection of equatorward meridional flow and evidence of double-cell meridional circulation inside the sun. Astrophys. J. **774**, L29 (2013)
Zirin, H.: Astrophysics of the Sun, p. 307. Cambridge University Press, Cambridge (1988)

Chapter 13
Solar Activity in the Past and Its Impacts on Climate

Hiroko Miyahara, Ayumi Asai, and Satoru Ueno

13.1 Long-Term Variations in Solar Irradiance and Ultraviolet Radiation and Their Estimation

The Sun is far away from the Earth, at a distance of approximately 150 million km. However, the Earth receives an enormous amount of sunlight and electromagnetic radiation from the Sun, under which life on Earth is nurtured. The total light energy received by the Earth is 1361 W/m^2 when all wavelengths of the radiation are integrated. This is called total solar irradiance (TSI). The top panel of Fig. 13.1 shows the variations in the TSI over the past 40 years (Hansen et al. 2013). There are several sharp decrements in the TSI variations; these are not due to observation errors, but rather to the appearance of a large sunspot or multiple sunspots. As shown in Fig. 12.1 and the bottom panel of Fig. 13.1, the sunspot number increases and decreases periodically in a cycle of approximately 11 years. Overall, the variations in the sunspot number and TSI are synchronized. Specifically, the Sun is brighter in the solar maximum than it is in the solar minimum. However, the range of fluctuation is only 0.1% (1/1000), which is at most approximately ±1 W/m^2 in one solar cycle (i.e., 11 years). The reason the TSI is called the "solar constant" is that the variations are very small.

In addition to visible light, which humans can see, the Sun emits light in a wide range of wavelengths, including short wavelengths, such as ultraviolet (UV) rays, X-rays, and γ-rays, and long wavelengths, such as infrared rays and radio waves. The intensity of the radiation at each wavelength is called the spectral solar irradiance (SSI) (Fig. 13.2a). Because the atmosphere of the Earth absorbs various wavelengths

H. Miyahara (✉)
Humanities and Sciences/Museum Careers, Musashino Art University, Tokyo, Japan
e-mail: miyahara@musabi.ac.jp

A. Asai · S. Ueno
Astronomical Observatory, Graduate School of Science, Kyoto University, Kyoto, Japan

K. Kusano (ed.), *Solar-Terrestrial Environmental Prediction*,
https://doi.org/10.1007/978-981-19-7765-7_13

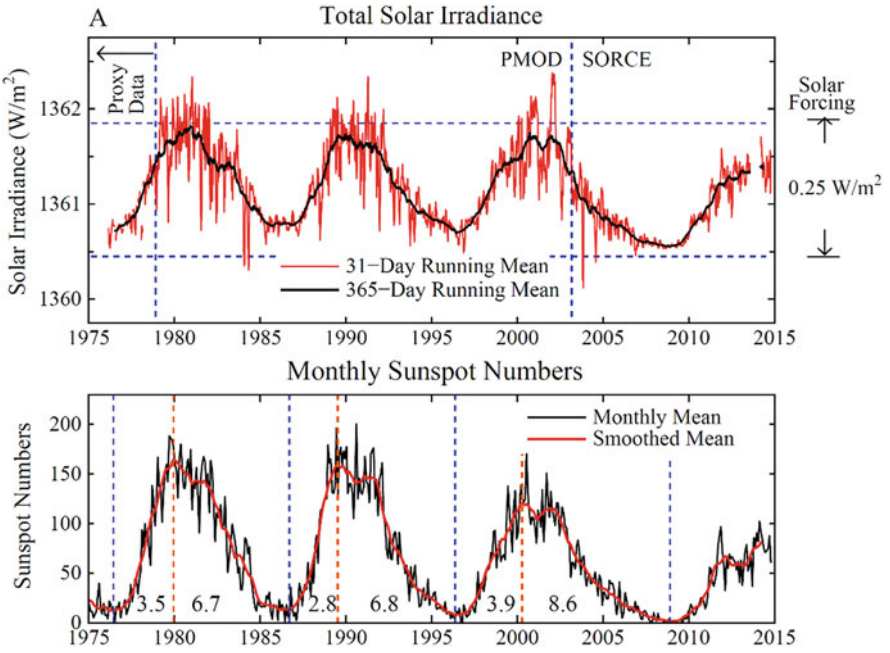

Fig. 13.1 TSI (top panel) versus sunspot number (bottom panel). (Hansen et al. 2013: https://commons.wikimedia.org/wiki/File:Changes_in_total_solar_irradiance_and_monthly_sunspot_numbers_1975-2013.png)

and the amount of absorption significantly varies depending on the atmospheric conditions, the TSI and the SSI must, in principle, be measured outside the atmosphere of the Earth using satellites. Although they have been measured using several instruments for a long time, the observed values of the TSI have been slightly different owing to the improvement in instrumentation. Therefore, to observe the long-term variability in the TSI, it is necessary to relate the TSI data from multiple satellites. However, even the variability in the TSI over the past 40 years is uncertain, ranging from almost no change to a gradual increase and a gradual decrease, depending on the calibration method for relating the data from different instruments (Solanki et al. 2013). During the period of extremely low solar activity (Maunder Minimum) between 1645 and 1715, the TSI has been also estimated to have declined, with estimates of the decrease ranging from 0.9 W/m^2 to 3 W/m^2, depending on the model (Schrijver et al. 2011; Krivova et al. 2010; Steinhilber et al. 2009; Solanki et al. 2013). If the TSI was extremely low to the extent of the upper limit of the above range, it could explain the cold climate (mini-ice age) around the time of the Maunder Minimum.

In contrast to the TSI, the interest in the variability of the SSI, particularly of the solar UV radiation, has been increasing recently. Most of the light with wavelengths shorter than visible light is absorbed by the atmosphere of the Earth. As shown in Fig. 13.2b (Ermolli et al. 2013), the altitude of the atmosphere of the Earth at which

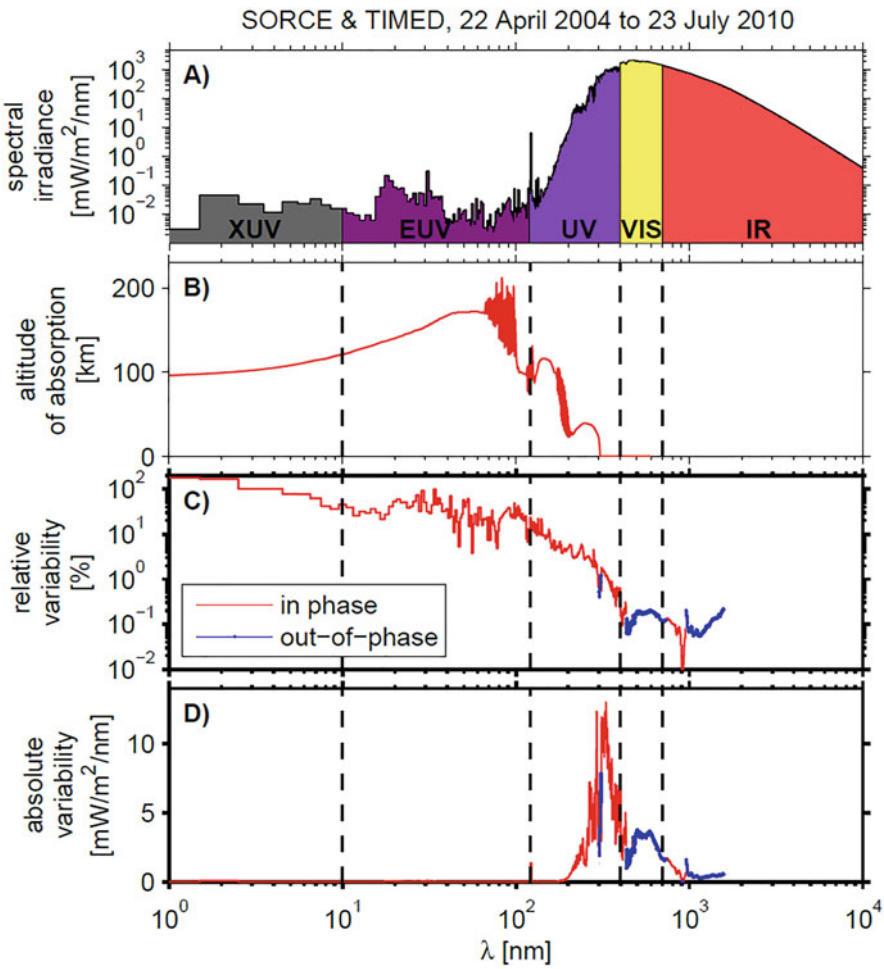

Fig. 13.2 (**a**) SSI, (**b**) characteristic altitude of absorption in the atmosphere of Earth for each wavelength, (**c**) relative variability in solar cycle variations inferred from measurements obtained between April 22, 2004, and July 23, 2010, (**d**) absolute variability. (Reprinted by Ermolli et al. (2013))

light is absorbed depends on its wavelength and corresponds to the thermosphere, mesosphere, and stratosphere. The solar UV radiation also varies synchronously with the solar activity cycle; however, its variability is much larger than that of the TSI (0.1%), ranging from 10% to 100% depending on the wavelength (Fig. 13.2c). The absolute amount of variability (Fig. 13.2d) also shows that the UV region accounts for a large proportion, suggesting that it is responsible for the variability in the TSI (e.g., Lean 1997). However, the long-term variability in the SSI also remains uncertain, because continuous satellite-based observations have been limited to recent years.

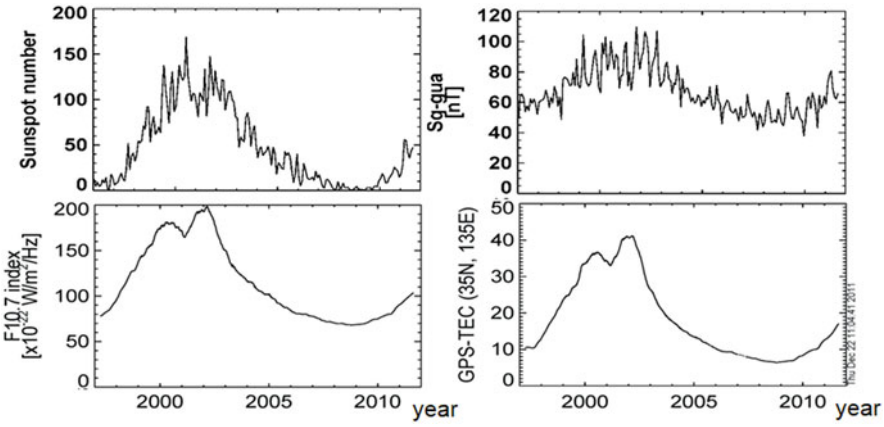

Fig. 13.3 Variations in sunspot number (upper left), solar radio F10.7 intensity (10.7 cm wavelength radio wave: lower left), geomagnetic Sq field (Guam station: upper right), and ionospheric total electron number over Japan by GPS satellite (lower right). (Courtesy of Dr. A. Shinbori)

Each layer of the atmosphere of the Earth that absorbs UV radiation responds sensitively to solar UV variations. Figure 13.3 shows the geomagnetic solar quiet daily variations (geomagnetic Sq field) (upper right) and the total electron content (TEC) (lower right) of the ionosphere. The geomagnetic Sq field is the amount of geomagnetic variations due to the electric current generated by the absorption of UV radiation of approximately 100–200 nm by the atmosphere of the Earth at an altitude of approximately 90–120 km. The TEC shows the variations in the number of electrons resulting from the ionization of the atmosphere of the Earth by UV radiation with a wavelength shorter than 50 nm in the upper ionosphere at high altitudes (approximately 200–500 km). The panels show the variations after subtracting specific variations of the Earth, such as seasonal changes, and it can be confirmed that both variations are synchronized with the variations in the sunspot number.

In the period when there is no (poor) direct observation data of solar UV by satellites, it is necessary to estimate its variations from other observation data. The solar UV radiation is emitted from the solar atmosphere—which is called the "chromosphere" and has a temperature of approximately 10,000 K—above the solar surface called the "photosphere." Therefore, the variations in the sunspot number can be indicators of magnetic activity to some extent; however, it is insufficiently precise to estimate the activity of the solar chromosphere. In contrast, solar radio intensity F10.7 (the total amount of radio waves with a wavelength of 10.7 cm emitted from the entire surface of the Sun; lower left of Fig. 13.3) has been frequently used (Tapping and Charrois 1994). The F10.7 intensity is known to be sensitive to the chromosphere and correlates well with the sunspot number and other physical parameters of the solar activity (particularly those attributed to the chromosphere). In addition, it has been used as an indicator of the long-term UV radiation variability because of the stable observation records accumulated since

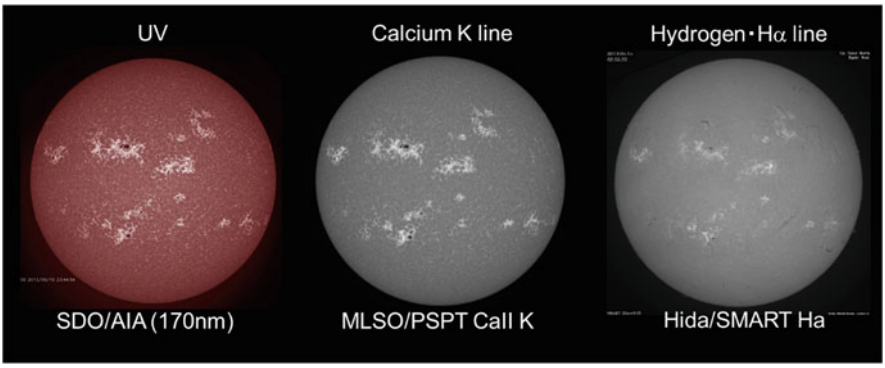

Fig. 13.4 Comparison of solar UV images (left: 171-nm image taken by SDO/AIA, National Aeronautics and Space Administration), calcium K-line image (middle: taken by Precision Solar Photometric Telescope at Hawaii Mauna Loa Solar Observatory), and Hα-line image (right: taken by SMART at Hida Observatory, Kyoto University) on May 15, 2013

1947. In Japan, the Nobeyama Solar Radio Observatory of the National Astronomical Observatory of Japan (NAOJ) has been continuously observing the Sun in the microwave band (frequency 1–10 GHz, wavelength 3–30 cm) since 1957 (Shimojo et al. 2017), and microwaves show similar variations to the F10.7 intensity.

The solar minimum around 2008 (the transition period from the 23rd to the 24th solar cycle) was an "anomalous" minimum period with a remarkably low activity level in recent years, and the behaviors of the geophysical parameters used as indicators of the solar activity were different (Fröhlich 2009). In particular, the TSI and the UV-SSI showed a significant decrease, whereas radio emissions such as F10.7 did not decrease as much (and were comparable to those in earlier minima). Therefore, the solar UV radiation using the F10.7 intensity overestimates the UV radiation intensity, at least in this solar minimum period.

As mentioned earlier, solar UV radiation is mainly emitted from the solar atmosphere, called the chromosphere, at a temperature of approximately 10,000 K (Lean 1987). The left part of Fig. 13.4 is an image of the solar UV radiation (170 nm) taken by a UV imager, the Atmospheric Imaging Assembly (AIA) onboard the Solar Dynamics Observatory (SDO) satellite of NASA. The middle and right parts are solar images at spectral lines formed by the solar chromosphere in the visible light region. They are the calcium K-line image taken by the Mauna Loa Solar Observatory in Hawaii and the Hα-line image by the Solar Magnetic Activity Research Telescope (SMART) at the Hida Observatory of Kyoto University, respectively. These images show very similar features, particularly their bright regions around sunspots (plages). A plage is a region of relatively high magnetic field strength (not as high as that of sunspots, but sufficiently higher than quiet regions) where magnetic field activity occurs.

Therefore, estimates of solar UV intensity based on solar chromosphere images are underway (e.g., Chatzistergos et al. 2019; Kakuwa and Ueno 2022). In the case of calcium and Hα images, observation data spanning approximately 100 years have

been accumulated. As examples, observations of the calcium K line using glass plates have been made at Kyoto University since 1928 (Kitai et al. 2014), at the NAOJ since 1917 (Hanaoka 2013), and at Kodaikanal, India, since 1907 (Hasan et al. 2010). Other instances are those at the Mt. Wilson Observatory, U.S., since 1915 (Bertello et al. 2010), and at the Arcetri Observatory, Italy, since 1931 (Ermolli et al. 2009). Although these also do not directly represent solar UV intensity, the similarities between the UV and chromospheric images presented in Fig. 13.4 have led to estimates of the solar UV radiation (Bertello et al. 2010, Chatterjee Chatterjee et al. 2016), particularly focusing on the plage region. Furthermore, the calcium K line observations at the above observatories have different bandpass widths around the center of the absorption line. In particular, observations with narrower bandpass widths (e.g., Mt. Wilson Observatory) show a higher atmospheric layer than those with wider bandpass widths (e.g., Kodaikanal, Kyoto University), and it has been reported that they may be highly correlated with shorter wavelengths of light in the UV region (Foukal et al. 2009). However, the conversion (calibration) of densities of glass plate data to calcium K line intensities is still under development (Kakuwa and Ueno 2021), and the estimation of the solar UV radiation is difficult.

Regarding the geomagnetic Sq field, because it responds to a limited wavelength range, we could estimate the solar UV radiation from its variations. However, Elias et al. (2010) investigated the long-term variations in the geomagnetic Sq field of the Earth, and found that it shows a monotonic increasing trend even after subtracting the "variation due to solar activity." Furthermore, they proposed that this increasing trend is related to the lower atmosphere, e.g., due to global warming. Comparatively, Shinbori et al. (2014) investigated the geomagnetic Sq field in more detail and over a longer period and reported a slightly decreasing trend. However, in the derivation, they used the sunspot number for the former (Elias et al. 2010) and the solar radio intensity F10.7 for the latter (Shinbori et al. 2014) to estimate the solar UV variability. Thus, we have not yet been able to accurately distinguish between the influence of solar UV radiation and the component of the variations specific to the atmosphere of the Earth, from the Sq variations. In this respect, more accurate SSI estimation in the UV region is also desired.

13.2 Detailed Analyses of Past Solar and Cosmic-Ray Variations

Concentrations of cosmic-ray-induced nuclides such as carbon-14 (half-life: 5730 years) in tree rings or beryllium-10 (half-life: 1.36 million years) in ice cores from polar regions are powerful tools to reconstruct the variations in the cosmic-ray intensity and the solar activity in the past. These nuclides are produced by the nuclear interactions between cosmic rays and the atmosphere of the Earth. Carbon-14 forms carbon dioxide, which is absorbed by trees via photosynthesis, whereas beryllium-10

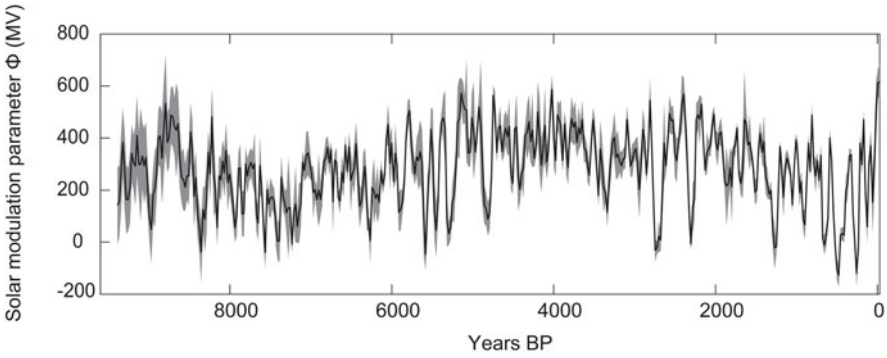

Fig. 13.5 Solar activity during the past 10,000 years reconstructed based on carbon-14 in tree rings and beryllium-10 in multiple ice cores. (Steinhilber et al. 2012)

attaches to aerosols and precipitates on ice sheets in the polar regions or on sediments.

Both nuclides have their strengths and weaknesses as tools for reconstructing past cosmic-ray and solar variations. For example, the signatures of cosmic-ray/solar variations in carbon-14 are significantly attenuated and distorted during its circulation in the carbon cycle, particularly those with a timescale of less than a decade. For precise reconstruction of signatures, the time profile of the production rate needs to be restored using a carbon cycle box model. In contrast, records of carbon-14 obtained from tree rings with definitive visible annual layers enable the reconstruction of the variations in the cosmic-ray intensity and the solar activity with accurate time stamps. This aspect is essential, particularly when the index is used for reconstructing the cycle lengths. Two methods are generally used to determine the absolute ages of tree rings: dendrochronology, which utilizes the temporal variations in tree-ring widths (Douglass 1937, 1946), and searching the carbon-14 concentration peak caused by nuclear bomb tests (Cain and Suess 1976). In recent years, there have also been attempts to utilize the temporal variations in oxygen isotope ratios to determine the ages of tree-ring samples.

In the case of beryllium-10, attenuation in the atmosphere is relatively small; however, its precipitation or deposition processes are occasionally significantly affected by climate variations. Therefore, it is necessary to correct these effects based on climate proxies to obtain an accurate cosmic-ray history. It should be noted that when using ice cores, the obtained records may contain dating uncertainties of several years. It is also possible to obtain beryllium-10 records using annually laminated lake sediments; however, they are highly susceptible to variations in climate and, thus, should be carefully examined.

Therefore, it is important to obtain and compare the records of multiple nuclides for accurate reconstruction of the cosmic-ray intensity and the solar activity. For the past 10,000 years, long-term variations in solar activity have been reconstructed by extracting common signatures from carbon-14 and beryllium-10 records (Fig. 13.5).

For older ages, reconstructions have been achieved mainly using beryllium-10 with a longer half-life (e.g., Horiuchi et al. 2016).

Beryllium-10 allows the reconstruction of cosmic-ray history for the past million years or more. However, when using ice cores, it becomes difficult to achieve annual resolution at high depths owing to the thinning effect of ice layers (Beer 2000). Thus far, annual layers have been traced to approximately 30,000 years ago in the case of Antarctic ice cores (Sigl et al. 2016) and to approximately 60,000 years ago in the case of Greenland ice cores (Svensson et al. 2008).

Recently, it was newly discovered that cosmic-ray intensity can also be reconstructed from the annual layers of carbonate deposits called travertines (Xu et al. 2019; Miyahara et al. 2020). Travertines frequently contain visible annual layers produced by the inflow of soils in rainy seasons. Different from ice sheets, travertines are barely compressed after their formation. Therefore, analyses with high temporal resolution can be achieved even at high depths. Until now, the existence of travertines up to one million years ago has been confirmed (Capezzuoli et al. 2014); therefore, beryllium-10 in travertines may enable extending high-resolution records further back in time.

The intensity of galactic cosmic rays (GCRs) arriving at the Earth is generally inversely related to the solar activity level, depending mainly on the intensity of the heliospheric magnetic field and the structure of the Parker spiral (Jokipii and Thomas 1981). As the solar activity increases, the solar wind becomes intensified, and thus, the attenuation level of GCRs increases. In addition, when the spiral structure of the heliospheric magnetic field is significantly developed, the orbit of GCRs, mainly consisting of protons, is altered and the efficiency of energy loss is increased. Owing to these effects, GCRs with energies below a few tens of GeV are modulated in the heliosphere. Based on neutron monitor data, the flux of GCRs varies by 20–25%, associated with solar activity variations. In addition, coronal mass ejections occasionally cause transient decrements in GCRs. This phenomenon is called the Forbush decrease.

In addition to the variations associated with solar activity, GCRs present a 22-year variation related to the reversal of the solar dipole magnetic field (Kota and Jokipii 1983). The large-scale motion of GCRs in the heliosphere is determined by the direction of the heliospheric magnetic field (drift effect), and thus, by the direction of the solar dipole magnetic field (Fig. 13.6). When the polarity of the solar dipole magnetic field is negative, GCRs tend to drift inward along the wavy heliospheric magnetic field. In contrast, GCRs tend to penetrate the polar regions of the heliosphere, when the polarity is positive. These effects cause the waveform of the cosmic-ray variation to be different every other 11-year cycle, resulting in a 22-year cycle. GCRs are more susceptible to the structure of the heliospheric magnetic field when the magnetic polarity of the Sun is negative than when it is positive. It has been suggested that the 22-year component in GCR variations may be particularly pronounced when, for example, the solar activity becomes extremely decreased or increased (Miyahara et al. 2008).

Beryllium-10 records from the Maunder minimum, when sunspots disappeared for several decades, have suggested that the spiral heliospheric magnetic field was

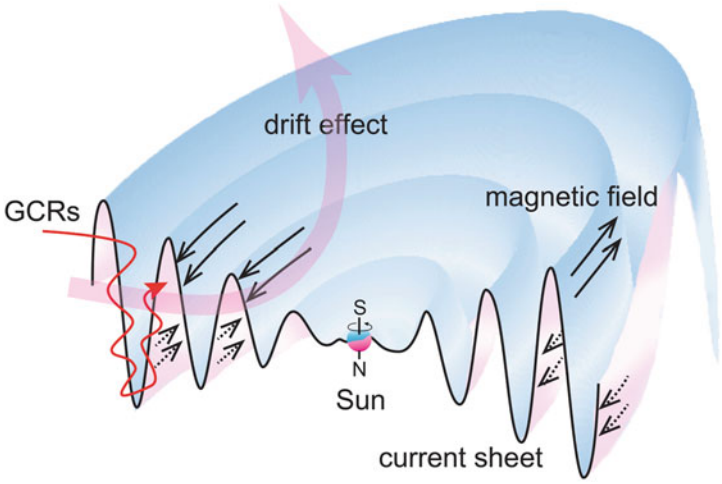

Fig. 13.6 Schematic of the spiral structure of the heliospheric current sheet and orbit of GCRs. Thick arrows indicate directions of meridional circulation of GCRs when the solar dipole magnetic field is negative. Directions of arrows are reversed when the dipole magnetic polarity of the Sun becomes positive

extremely flattened at the solar cycle minima and that the 22-year cycle might have been pronounced at the time (Miyahara et al. 2008; Yamaguchi et al. 2010; Kataoka et al. 2012). An approximately 40% increase in the GCR flux was found at the minima of negative polarity, in addition to decadal-scale variations. Along with the effects of the flattening of the heliospheric magnetic field, there may have been an impact from the outer heliosheath causing the above increase. A better understanding of the structures of the heliosphere and the heliosheath and a highly realistic model of the GCR modulation therein are needed to understand the variability in the GCR flux at the Earth.

Annually resolved records of cosmogenic nuclides provide information on the characteristics of the solar cycles in the past and, thus, are essential for understanding the mechanisms of the long-term solar activity level. It has been suggested that the length of the solar activity cycle may reflect the condition of the meridional circulation in the solar convection zone (Dikpati and Charbonneau 1999). Analysis of the carbon-14 content in tree rings from the Maunder minimum has indicated that the solar cycle might have been lengthened to approximately 14 years (Miyahara et al. 2004). It has also been suggested that this lengthening might have started three cycles before the onset of the Maunder minimum (Miyahara et al. 2021). A similar tendency has been also found for the Spoerer minimum (Miyahara et al. 2010; Moriya et al. 2019). Further examinations of the solar cycle lengths for the various types of grand solar minima based on high-precision carbon-14 analyses may elucidate the possible factors determining the magnitude of grand solar minima.

During the past decade, records of cosmogenic nuclides have also been utilized for searching past large-scale solar flare events. Significant increments have been

found in the production rate of carbon-14, e.g., in 774–775 C.E., 993–994 C.E., and 660 B.C.E (Miyake et al. 2012, 2013; O'Hare et al. 2019; Sakurai et al. 2020). Moreover, it has been suggested that they were caused by large-scale solar proton events. For 774–775 C.E. and 993–994 C.E. events, increments in both beryllium-10 and chlorine-36 have been reported based on ice-core analyses (Mekhaldi 2015). It is expected that these records are important not only from astrophysical and space weather perspectives but also as markers to determine the absolute ages of natural materials or the proxy records obtained from them.

13.3 Solar Influence on Climate Observed in Paleoclimate Data

Past climate changes can be examined based on several proxy records. For example, temporal variations in tree-ring widths are frequently used to reconstruct summertime temperatures. This is because the growth rates of trees are frequently correlated to the temperature in the growing season and are reflected in the widths of annual rings. Occasionally, trees may be more sensitive to the amount of precipitation, depending on the surrounding climatic environment. Therefore, it is essential to determine in advance the meteorological parameters the trees are most susceptible to. Past temperature variations can also be reconstructed by measuring the temperatures in boreholes (Huang et al. 2000). This method is based on the fact that temperature variations at the surface of the Earth are transported to the subsurface by heat conduction. In addition, temperature measurements in the Greenland ice sheet have revealed the temperature history for the past 1000 years (Dahl-Jensen et al. 1998). Although the temperature reconstructions based on drilled holes have a lower temporal resolution, they reflect more directly the thermal history of the surface of the Earth.

Another possible methodology for reconstructing past temperature variation is based on plant physiology. For example, the date of the full bloom of cherry trees is highly correlated with the average temperature in March, and accordingly, temperature fluctuations over the past 1200 years have been reconstructed using diaries preserved in Kyoto (Aono and Kazui 2008). In another case, the number of rainy days was examined based on the weather records in diaries, and the summertime temperature was reconstructed based on their correlation (Mikami 2008). In Europe, records of grape harvest for wine-making were used to reconstruct the temperatures during the past 600 years (García de Cortázar-Atauri et al. 2010). In other cases, variations in the sea surface temperature over hundreds of years have been reconstructed based on Sr/Ca ratios in coral skeletons (Kawakubo et al. 2017).

Long-term temperature variations can be reconstructed, e.g., from analyses of pollen in lake and marine sediments (Nakagawa et al. 2003; Hayashi et al. 2010; Kitaba et al. 2013) and from ice-rafted debris in marine sediments (Bond and Lotti

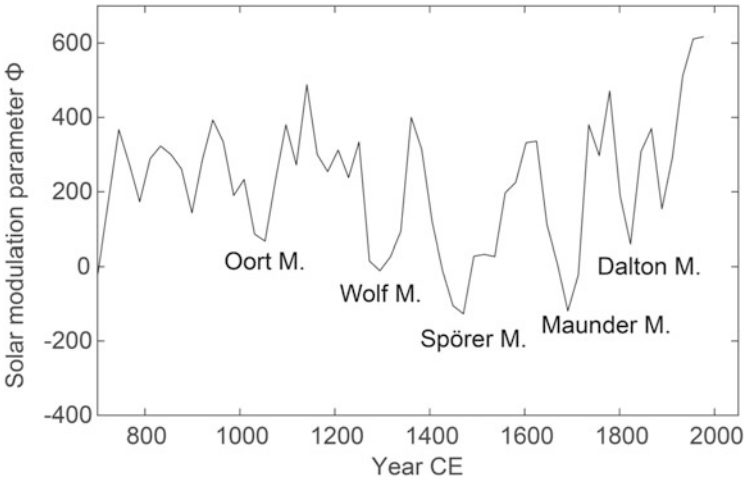

Fig. 13.7 Solar activity minima in the last 1000 years. (Based on Steinhilber et al. 2012)

1995). Oxygen isotope ratios in ice cores have also been commonly used as indicators of climate change (Grootes et al. 1993).

The trends in dry and wet environments can be reconstructed based on the oxygen isotope ratio in stalagmites (Wang et al. 2005). When precipitation increases, the oxygen isotope ratio decreases (amount effect). This tendency allows to reconstruct the amount of precipitation in the past. In recent years, the oxygen isotope ratio of annual tree rings has also been used to reconstruct dry and humid environments. In the case of tree rings, the oxygen isotope ratio in cellulose reflects the amount of precipitation and/or relative humidity (Li et al. 2015).

The relationship between climate and solar activity has long been studied. Various direct and indirect climate records such as temperature, precipitation, and droughts have suggested that periodic variations are synchronized to solar cyclic variations with timescales of a decade to a few millennia. One of the most notable examples is the Little Ice Age, which occurred synchronously with the series of grand solar minima since the late thirteenth century, including the Maunder minimum (Fig. 13.7). Before the Little Ice Age, the climate had been warm for several 100 years, corresponding to a period of Medieval solar activity maximum. It has been suggested that volcanic eruptions such as those of Mount Tambora in 1815 C.E. also contributed to the decrease in temperature during the Little Ice Age (Gao et al. 2008).

The relationship between the weakening of solar activity and cold climate has been further confirmed by research based on ice-rafted debris in deep-sea sediment cores from North Atlantic (Bond et al. 2001). The number of ice-rafted debris reflected a millennial-scale cyclic variation synchronization with 1000/2000-year solar activity cycles throughout the past 10,000 years. An increase in ice-rafted debris in deep-sea sediments implies an increase in the amount of iceberg discharge from the Arctic region and/or the southward expansion of the cold sea surface water.

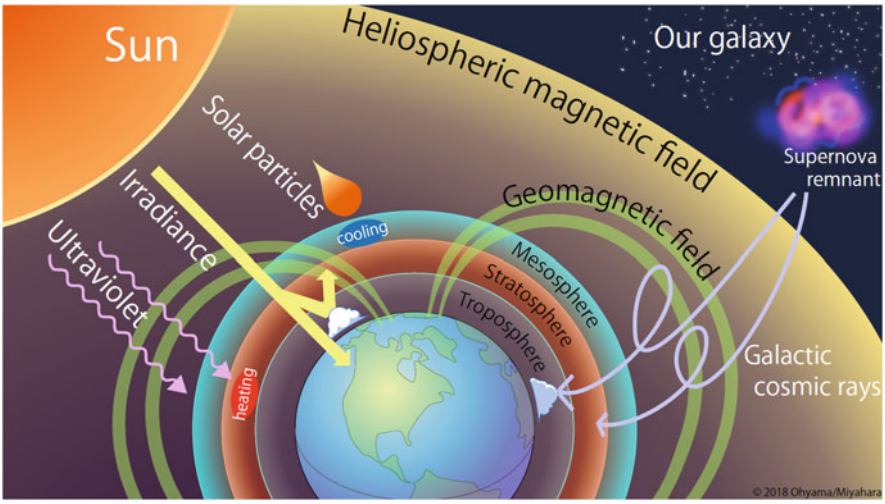

Fig. 13.8 Schematic of pathways of solar influence on the climate system

Similar 1000/2000-year cycles were also found in the last glacial period (Obrochta et al. 2012), suggesting that solar activity had a pronounced influence during the time.

It becomes relatively difficult to identify the solar influence on climate at short timescales; however, studies have suggested the impacts of the 11-year solar cycle (e.g., van Loon et al. 2004; Misios et al. 2019), and even the solar rotation cycle. The rotation of the Sun within a 27-day period causes quasi-cyclic variations in the TSI and UV associated with the migration of sunspots and faculae on the solar surface. The GCR flux also shows similar monthly variations, because coronal mass ejections causing the transient decrements in GCRs originate in sunspots on the solar surface. Impacts of the solar rotation cycle have been found, e.g., in global lightning activity monitored by the Schumann resonance (Sato and Fukunishi 2005), thunderstorm activity in Japan (Muraki et al. 2004; Miyahara et al. 2017) and the United Kingdom (Scott et al. 2014), and in equatorial cloud activity (Takahashi et al. 2010; Hong et al. 2011). The important feature of these impacts is that they are rarely detected during periods of low solar activity. They are only significantly detected when there are many sunspots and monthly variations in solar-related parameters are pronounced. A study on the thunderstorm activity for the past 150 years based on old diaries further confirmed that the impact of the solar rotation cycle becomes stronger as the number of sunspots increases (Miyahara et al. 2018).

The above examples demonstrate that correlations between solar activity and climate or meteorological phenomena have been found on various timescales, although the mechanisms behind the link are poorly understood. There are several possible pathways by which solar activity can influence climate (Fig. 13.8). The simple path is by the TSI variations; however, TSI observations since the late 1970s have revealed that the difference in the TSI between the solar activity maxima and

minima is only approximately 1 W/m^2. Therefore, it is not highly probable that the TSI variations can explain the observed temperature/precipitation variations, unless a strong positive feedback is identified. Another possible path is by UV radiation, which affects stratospheric temperatures. The other possible candidates are the impacts of solar wind particles on atmospheric chemical reactions and the influence of GCRs on the formation of cloud condensation nuclei and/or on the efficiency of the coalescence of aerosols and cloud particles.

There are a few possible approaches to discriminate the effects of GCRs from other solar-related parameters. For example, examination of the response of climate to the 22-year cycle allows identifying the effects of GCRs (Miyahara et al. 2008; Yamaguchi et al. 2010). It may also be possible to discriminate the impacts of GCRs by focusing on the solar rotation period. Although the amount of solar radiation fluctuates depending on the areas of the sunspots and faculae and their migrations on the solar surface, the variations in GCRs are mainly determined by the crossing of the coronal mass ejections accompanying the strong magnetic field around the Earth. Therefore, the characteristics of monthly-scale variations are slightly different from each other. Comparison of paleo-climate records with the history of the geomagnetic field intensity will also allow understanding the impacts of GCRs on the climate (Kitaba et al. 2013). Searching for the effects of GCR variations at even longer timescales, e.g., the variations with the change in nearby star formation rate, may also provide insight into its role (Shaviv 2002).

It should be noted that there are significant regional differences in the responses of the climate to solar activity. It has been suggested by studies on the Little Ice Age that there existed regions in which the temperature dropped by as much as 2.5 °C (Aono and Kazui 2008; Mann et al. 2009), although the globally averaged decrease was only approximately 0.6 °C. Japan appears to have been one of the most susceptible regions to solar activity variations.

Precipitation also shows significant regional differences in the responses to solar activity. For example, the amount of summertime precipitation increased around the South China Sea during the Little Ice Age, whereas the northern China region has drier condition (Yan et al. 2015). Such a regional pattern may reflect the changes in atmospheric circulation around the time. Oxygen isotope ratios in Japanese tree rings indicated that precipitation increased during the Little Ice Age in central Japan, particularly at the end of the event (Sakashita et al. 2017). A delay in the responses of the precipitation to solar activity implies that the mechanism of Sun–climate connection involves changes in the sea surface temperature. Examination of the spatial pattern in the responses of the climate to the solar activity would provide insight on the process of the solar influence on the climate system.

Acknowledgements The Precision Solar Photometric Telescope (PSPT) was maintained and operated at Mauna Loa Solar Observatory from 1998 to 2015 by the High Altitude Observatory/National Center for Atmospheric Research (HAO/NCAR). The data was processed at and is served to the community by the Laboratory for Atmospheric and Space Physics (LASP) at the University of Colorado, Boulder. The PSPT effort was supported by the National Science Foundation (NSF).

References

Aono, Y., Kazui, K.: Phenological data series of cherry tree flowering in Kyoto, Japan, and its application to reconstruction of springtime temperatures since the 9th century. Int. J. Climatol. **28**, 905–914 (2008)

Beer, J.: Long-term indirect indices of solar variability. Space Sci. Rev. **94**, 53–66 (2000)

Bertello, L., Ulrich, R.K., Boyden, J.E.: The Mount Wilson Ca ii K Plage index time series. Sol. Phys. **264**, 31–44 (2010)

Bond, G.C., Lotti, R.: Iceberg discharges into the North Atlantic on millennial time scales during the last glaciation. Science. **267**, 1005–1010 (1995)

Bond, G., et al.: Persistent solar influence on North Atlantic climate during the Holocene. Science. **294**, 2130–2136 (2001)

Cain, W.F., Suess, H.E.: Carbon 14 in tree rings. J. Geophys. Res. **81**, 3688–3694 (1976)

Capezzuoli, E., Gandin, A., Pedley, M.: Decoding tufa and travertine (fresh water carbonates) in the sedimentary record: the state of the art. Sedimentology. **61**, 1–21 (2014)

Chatterjee, S., et al.: A butterfly diagram and Carrington maps for century-long CA II K spectroheliograms from the Kodaikanal observatory. Astrophys. J. **827**, id87 (2016)

Chatzistergos, H., et al.: Analysis of full disc ca II K spectroheliograms. II. Towards an accu-rate assessment of long-term variations in plage areas. Astron. Astrophys. **625**, 69 (2019)

Dahl-Jensen, D., et al.: Past temperatures directly from the Greenland ice sheet. Science. **282**, 268–271 (1998)

Dikpati, M., Charbonneau, P.: A Babcock-Leighton flux transport dynamo with solar-like differential rotation. Astrophys. J. **518**, 508–520 (1999)

Douglass, A.E.: Tree rings and chronology. Univ. Arizona Bull. **8** (1937)

Douglass, A.E.: Precision of ring dating in tree-ring chronologies. Univ. Arizona Bull. **17** (1946)

Elias, I., et al.: Trends in the solar quiet geomagnetic field variation linked to the Earth's magnetic field secular variation and increasing concentrations of greenhouse gases. J. Geophys. Res. **115**, A08316 (2010)

Ermolli, I., et al.: The digitized archive of the Arcetri spectroheliograms. Preliminary results from the analysis of ca II K images. Astron. Astrophys. **499**, 627–632 (2009)

Ermolli, I., et al.: Recent variability of the solar spectral irradiance and its impact on climate modelling. Aomos. Chem. Phys. **13**, 3945–3977 (2013)

Foukal, P., et al.: A century of solar ca ii measurements and their implication for solar UV driving of climate. Sol. Phys. **255**, 229–238 (2009)

Fröhlich, C.: Evidence of a long-term trend in total solar irradiance. Astron. Astrophys. **501**, L27–L30 (2009)

Gao, C., Robock, A., Ammann, C.: Volcanic forcing of climate over the past 1500 years: an improved ice core-based index for climate models. J. Geophys. Res. **113**, D23111 (2008)

García de Cortázar-Atauri, I., et al.: Climate reconstructions from grape harvest dates: methodology and uncertainties. Holocene. **20**, 599–608 (2010)

Grootes, P.M., et al.: Comparison of oxygen isotope records from the GISP2 and GRIP Greenland ice cores. Nature. **366**, 552–554 (1993)

Hanaoka, Y.: Long-term synoptic observations of the sun at the National Astronomical Observatory of Japan. J. Phys. Conf. Ser. **440**, 012041 (2013)

Hansen, J., et al.: Assessing "dangerous climate change": required reduction of carbon emissions to protect young people, future generations and nature. PLoS One. **8**, e81648 (2013)

Hasan, S.S., et al.: Solar physics at the Kodaikanal observatory: a historical perspective. A historical perspective. Astrophys. Space Sci. Proc. **19**, 12 (2010)

Hayashi, R., Takahara, H., Hayashida, A., Takemura, K.: Millennial-scale vegetation changes during the last 40,000 yr based on a pollen record from Lake Biwa, Japan. Quater. Res. **74**, 91099 (2010)

Hong, P.K., et al.: Implications for the low latitude cloud formations from solar activity and the quasi-biennial oscillation. J. Atmos. Sol. Terr. Phys. **73**, 587–591 (2011)

Horiuchi, K., et al.: Multiple 10Be records revealing the history of cosmic-ray variations across the Iceland Basin excursion. Earth Planet. Sci. Lett. **440**, 105–114 (2016)
Huang, S., Pollack, H.N., Shen, P.-Y.: Temperature trends over the past five centuries reconstructed from borehole temperatures. Nature. **403**, 756–758 (2000)
Jokipii, J.R., Thomas, B.: Effects of drift on the transport of cosmic rays IV - modulation by a wavy interplanetary current sheet. Astrophys. J. **243**, 1115–1122 (1981)
Kakuwa, J., Ueno, S.: Investigation of the long-term variation of solar CaII K intensity. I. Density-to-intensity calibration formula for historical photographic plates. Astrophys. J. Suppl. Ser. **254**, id44 (2021)
Kakuwa, J., Ueno, S.: Investigation of the long-term variation of solar CaII K intensity. II. Reconstruction of solar UV irradiance. Astrophys. J. **928**, id97 (2022)
Kataoka, R., Miyahara, H., Steinhilber, F.: Anomalous 10Be spikes during the maunder minimum: possible evidence for extreme space weather in the heliosphere. Space Weather. **10**, S11001 (2012)
Kawakubo, Y., Alibert, C., Yokoyama, Y.: A reconstruction of subtropical western North Pacific SST variability back to 1578, based on a Porites coral Sr/Ca record from the northern Ryukyus, Japan. Paleoceanography. **32**, 1352–1370 (2017)
Kitaba, I., et al.: Midlatitude cooling caused by geomagnetic field minimum during polarity reversal. Proc. Natl. Acad. Sci. **110**, 1215–1220 (2013)
Kitai, R. et al.: Digital Archiving of Solar Synoptic Observation (1) Project Outline and Meta-Data Archiving. Technical Reports from Kwasan and Hida Observatories, Graduate School of Science, Kyoto University 2–2 (2014)
Kota, J., Jokipii, J.R.: Effects of drift on the transport of cosmic rays. VI - a three-dimensional model including diffusion. Astrophys. J. **265**, 573–581 (1983)
Krivova, N.A., Vieira, L.E.A., Solanki, S.K.: Reconstruction of solar spectral irradiance since the maunder minimum. J. Geophys. Res. **115**, A12112 (2010)
Lean, J.: Solar ultraviolet irradiance variations - a review. J. Geophys. Res. **92**, 839–868 (1987)
Lean, J.: The sun's variable radiation and its relevance for earth. Annu. Rev. Astrophys. **35**, 33–67 (1997)
Li, Z., Nakatsuka, T., Sano, M.: Tree-ring cellulose δ18O variability in pine and oak and its potential to reconstruct precipitation and relative humidity in Central Japan. Geochem. J. **49**, 125–137 (2015)
Mann, M.E., et al.: Global signatures and dynamical origins of the little ice age and medieval climate anomaly. Science. **326**, 1256–1260 (2009)
Mekhaldi, F.: Multiradionuclide evidence for the solar origin of the cosmic-ray events of AD 774/5 and 993/4. Nature Commun. **6**, 8611 (2015)
Mikami, T.: Climatic variations in Japan reconstructed from historical documents. R. Met. Soc. **63**, 190–193 (2008)
Misios, S., et al.: Slowdown of the Walker circulation at solar cycle maximum. Proc. Natl. Acad. Sci. **116**, 7186–7191 (2019)
Miyahara, H., et al.: Cyclicity of solar activity during the maunder minimum deduced from radiocarbon content. Sol. Phys. **224**, 317–322 (2004)
Miyahara, H., Yokoyama, Y., Masuda, K.: Possible link between multi-decadal climate cycles and periodic reversals of solar magnetic field polarity. Earth Planet. Sci. Lett. **272**, 290–295 (2008)
Miyahara, H., et al.: Is the sun heading for another maunder minimum? -precursors of the grand solar minima. J. Cosmol. **8**, 1970–1982 (2010)
Miyahara, H., et al.: Solar 27-day rotational period detected in a wide-area lightning activity in Japan. ANGEO Commun. **35**, 583–588 (2017)
Miyahara, H., et al.: Solar rotational cycle in lightning activity in Japan during the 18–19th centuries. ANGEO Commun. **36**, 633–640 (2018)
Miyahara, H., et al.: Measurement of beryllium-10 in terrestrial carbonate deposits from South China: a pilot study. Nucl. Instrum. Methods Phys. Res. B. **464**, 36–40 (2020)

Miyahara, H., et al.: Gradual onset of the maunder minimum revealed by high-precision carbon-14 analyses. Sci. Rep. **11**, 5482 (2021)
Miyake, F., et al.: A signature of cosmic-ray increase in AD 774–775 from tree rings in Japan. Nature. **486**, 240–242 (2012)
Miyake, F., Masuda, K., Nakamura, T.: Another rapid event in the carbon-14 content of tree rings. Nature Comm. **4**, 1748 (2013)
Moriya, T., et al.: A study of variation of the 11-year solar cycle before the onset of the Sporer minimum based on annually measured 14C content in tree rings. Radiocarbon. **61**, 1749–1754 (2019)
Muraki, Y., et al.: Effects of atmospheric electric fields on cosmic rays. Phys. Rev. D. **69**, 123010 (2004)
Nakagawa, T., Kitagawa, H., Yasuda, Y., Tarasov, P.E., Nishida, K., Gotanda, K., Sawai, Y.: Asynchronous climate changes in the North Atlantic and Japan during the last termination. Science. **299**, 688–691 (2003)
O'Hare, P., et al.: Multiradionuclide evidence for an extreme solar proton event around 2,610 B.P. (~660 BC). Proc. Natl. Acad. Sci. **116**, 5961–5966 (2019)
Obrochta, S.P., Miyahara, H., Yokoyama, Y., Crowley, T.J.: A re-examination of evidence for the North Atlantic "1500-year cycle" at site 609. Quat. Sci. Rev. **55**, 23–33 (2012)
Sakashita, W., et al.: Hydroclimate reconstruction in Central Japan over the past four centuries from tree-ring cellulose δ18O. Quat. Int. **455**, 1–7 (2017)
Sakurai, H., et al.: Prolonged production of 14C during the ~660 BCE solar proton event from Japanese tree rings. Sci. Rep. **10**, 660 (2020)
Sato, M., Fukunishi, H.: New evidence for a link between lightning activity and tropical upper cloud coverage. Geophys. Res. Lett. **32**, L12807 (2005)
Schrijver, C.J., et al.: The minimal solar activity in 2008–2009 and its implications for long-term climate modeling. Geophys. Res. Lett. **38**, L06701 (2011)
Scott, C.J., et al.: Evidence for solar wind modulation of lightning. Environ. Res. Lett. **9**, 055004 (2014)
Shaviv, N.J.: Cosmic ray diffusion from the galactic spiral arms, iron meteorites, and a possible climatic connection. Phys. Rev. Lett. **89**, 089901 (2002)
Shimojo, M., et al.: Variation of the solar microwave spectrum in the last half century. Astrophys. J. **848**, id62 (2017)
Shinbori, A., et al.: Long-term variation in the upper atmosphere as seen in the geomagnetic solar quiet daily variation. Earth Planets Space. **66**, id155 (2014)
Sigl, M., et al.: The WAIS divide deep ice core WD2014 chronology – part 2: annual-layer counting (0–31 ka BP). Clim. Past. **12**, 769–786 (2016)
Solanki, S., et al.: Solar irradiance variability and climate. Annu. Rev. Astrophys. **51**, 311–351 (2013)
Steinhilber, F., Beer, J., Fröhlich, C.: Total solar irradiance during the Holocene. Geophys. Res. Lett. **36**, L19704 (2009)
Steinhilber, F., et al.: 9,400 years of cosmic radiation and solar activity from ice cores and tree rings. Proc. Natl. Acad. Sci. **109**, 5967–1971 (2012)
Svensson, K., et al.: A 60 000 year Greenland stratigraphic ice core chronology. Clim. Past. **4**, 47–57 (2008)
Takahashi, Y., et al.: 27-day variation in cloud amount and relationship to the solar cycle. Atmos. Chem. Phys. **10**, 1577–1584 (2010)
Tapping, K.F., Charrois, D.: P.: limits to the accuracy of the 10.7-centimeter flux. Sol. Phys. **150**, 305–315 (1994)
van Loon, H., Meehl, G.A., Arblaster, J.M.: A decadal solar effect in the tropics in July–August.-J. Atmos. Sol.-Terr. Phys. **66**, 1767–1778 (2004)

Wang, Y., et al.: The Holocene Asian monsoon: links to solar changes and North Atlantic climate. Science. **308**, 854–857 (2005)

Xu, H., et al.: High-resolution records of 10Be in endogenic travertine from Baishuitai, China: a new proxy record of annual solar activity? Quat. Sci. Rev. **216**, 34–46 (2019)

Yamaguchi, T., Yokoyama, Y., Miyahara, H., Sho, K., Nakatsuka, T.: Synchronized northern hemisphere climate change and solar magnetic cycles during the maunder minimum. Proc. Natl. Acad. Sci. **107**, 20697–20702 (2010)

Yan, H., et al.: Dynamics of the intertropical convergence zone over the western Pacific during the little ice age. Nat. Geo. **8**, 315–320 (2015)

Chapter 14
Effects of Solar Activity on the Upper Atmosphere

Hitoshi Fujiwara, Yasunobu Miyoshi, and Hidekatsu Jin

14.1 Effects of Solar Activity on the Upper Atmosphere

This section provides an overview of the variabilities in the upper atmosphere (mainly in the thermosphere: altitudes ranging from about 85–95 km to about 350–800 km). First, we review the typical phenomena observed in the thermosphere and subsequently present the factors that cause these variations and discuss the magnitudes of the variations in the thermospheric parameters (e.g., temperature, density, and wind speed).

14.1.1 Variations in the Upper Atmosphere (Thermosphere)

The thermosphere, which is located at the upper end of the atmosphere, can be considered as the boundary region between the atmosphere and space, where auroras shine and low-altitude satellites and the International Space Station (ISS) fly at an altitude of about 400 km. Therefore, the thermospheric winds, temperature, density, and composition are significantly changed by the inflow of energy from space and the lower atmosphere.

Figure 14.1 illustrates the energetic coupling process between the thermospheric region and the lower atmosphere. In particular, it has been clarified by observations

H. Fujiwara (✉)
Seikei University, Tokyo, Japan
e-mail: h_fujiwara@st.seikei.ac.jp

Y. Miyoshi
Department of Earth and Planetary Sciences, Kyushu University, Fukuoka, Japan

H. Jin
National Institute of Information and Communications Technology, Tokyo, Japan

K. Kusano (ed.), *Solar-Terrestrial Environmental Prediction*,
https://doi.org/10.1007/978-981-19-7765-7_14

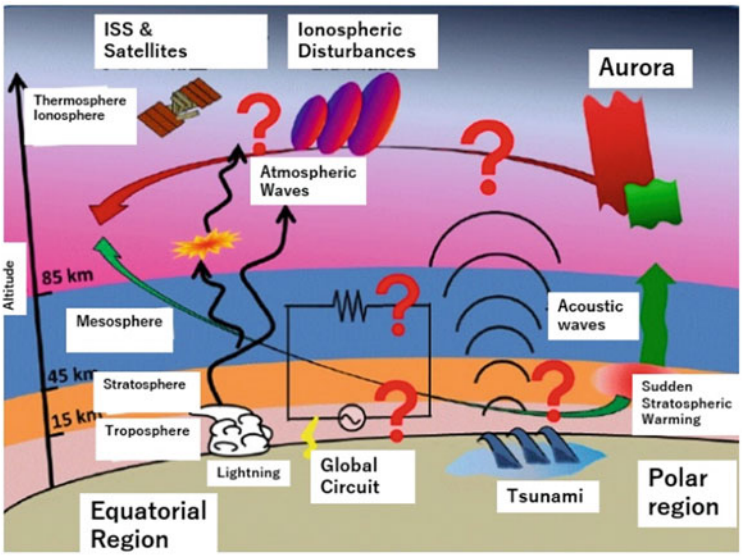

Fig. 14.1 Influence and latitudinal coupling from the lower atmosphere to the middle and upper atmosphere (Courtesy of Society of Geomagnetism and Earth, Planetary and Space Sciences. The original figure is seen in "Current Status and Future of Geomagnetism and Earth, Planetary and Space Sciences 2018")

and numerical simulations that atmospheric waves originating in the lower atmosphere have significant impacts on the thermospheric and ionospheric regions (e.g., Jin et al. 2011). Recently, it has become clear that various meteorological phenomena originating in the troposphere and climate change also affect the thermospheric region. An example of the former is sudden stratospheric warming, and of the latter is the cooling of the thermosphere due to the increase in carbon dioxide, which causes global warming in the troposphere. The effects of the lower atmosphere on the ionospheric variability are discussed in Sect 14.2, and the effects on the middle atmosphere variability are discussed in the next section.

In addition to the influence of the lower atmosphere, the energy inflow originating from the sun causes drastic changes in the thermosphere. In the long term, the temperature, composition, density, and wind speed of the thermosphere are changed by the variations in the solar extreme ultraviolet (EUV) and ultraviolet (UV) radiation, which are dependent on the solar activity cycle. In the short term, the thermospheric temperature, composition, density, and wind speed are changed by the changes in the energy inflow into the polar regions (auroral particles and electromagnetic energy inflows from the magnetosphere), which depend on the variations in the geomagnetic activity (auroral activity). In particular, thermospheric variations occur globally during magnetic storm periods. These thermospheric variations are occasionally called thermospheric storms, as well as magnetospheric and ionospheric storms. In the next section, we present the details of thermospheric variations and the factors that cause them.

14.1.2 Solar Radiation Variability Causing Changes in the Upper Atmosphere

Most of the solar radiation is visible light. The environment surrounding us is formed and maintained by the visible light energy that reaches the surface of the Earth. In comparison, in the upper atmosphere of the Earth, a small amount of the EUV/UV light in the solar radiation is absorbed and becomes the energy source for the atmosphere. The energy of the visible light reaching the Earth is 1.75×10^{17} W, whereas that of the EUV light ranges from about 10^{11} W (solar minimum) to about 10^{12} W (solar maximum). Because the air in the thermosphere is extremely thin, the absorption of the EUV/UV radiation results in a high temperature of approximately 1000 K. The temperature of the thermosphere is strongly dependent on the solar activity because the EUV/UV energy fluctuates by about one order of magnitude between the solar minimum and maximum periods. The UV radiation in the wavelength range of about 10–102.8 nm is important for the heating of the thermosphere and the generation of ionospheric plasma, and this wavelength range is called EUV. Solar radiation with wavelengths lower than 102.8 nm causes ionization of oxygen molecules, which are one of the main components of the atmosphere of the Earth. Solar radiation with wavelengths longer than 102.8 nm does not cause ionization of any other atmospheric molecules in the thermosphere of the Earth. Solar radiation in the visible region varies by about 0.1% during an 11-year cycle, whereas X-rays and EUV light with wavelength range shorter than 100 nm show variations of about 200% on average (Lean 1991).

In order to estimate various parameters in the thermosphere and the ionosphere, solar radiation spectral models (in particular EUV models) and absorption cross-section data of atmospheric molecules with the same wavelength resolution are indispensable. In general, solar EUV spectral models are frequently constructed using the F10.7 index as a proxy (the F10.7 index is the solar radio intensity at a wavelength of 10.7 cm with 1 sfu $= 10^{-22}$ W m^{-2} Hz^{-1} as a unit). Because of insufficient observations of the solar X-rays and EUV/UV radiation spectra and uncertainties in absorption cross-section data, we should continue to revise solar spectral models and cross-section data. In particular, models for estimating the solar EUV radiation when F10.7 is below 68 and past EUV radiations are important for reproducing and predicting climate change. Solar EUV/UV spectral models are described in detail by Lean (1991) and Lean et al. (2011).

14.1.3 Temperature Changes in the Thermosphere

Fig. 14.2 shows examples of typical altitude profiles of the thermospheric temperature. The profiles were created using NRLMSISE-00 (Picone et al. 2002), a standard empirical (statistical) model, during geomagnetically quiet periods at mid-latitude noon. Two examples (winter during the solar minimum and summer

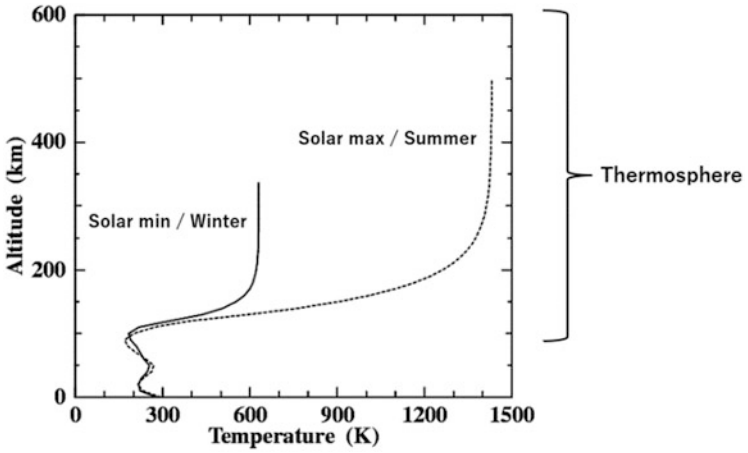

Fig. 14.2 Examples of altitude profiles of the thermospheric temperature during a geomagnetically quiet period by NRLMSISE-00

during the solar maximum) are shown in which the temperature difference is significant. In the examples shown here, the temperature difference is as much as 800 K at the upper end of the thermosphere (exosphere). In comparison with a case of geomagnetically quiet periods (shown in Fig. 14.2), more significant changes are expected in the thermospheric temperature during geomagnetically disturbed periods.

An example of the global distribution of the atmospheric temperatures and horizontal wind speeds on a constant-pressure surface at an altitude of about 300 km is shown in Fig. 14.3. Figure 14.3 is an example of a calculation based on an atmospheric general circulation model (Miyoshi and Fujiwara 2003) under conditions of a solar minimum and low geomagnetic activity on December 1 (when the Northern and Southern Hemispheres are in summer and winter, respectively). In the polar regions, high-temperature regions are caused by joule heating caused by the precipitating auroral particles and the electric field. There are also local high- and low-temperature regions, which are probably produced by the effects of the lower atmosphere. There is a temperature difference of about 500 K between the two poles and a day–night temperature difference of about 200–300 K.

The primary atmospheric heating mechanism in the thermosphere is solar EUV/UV heating. In addition, there is polar heating due to auroral (geomagnetic) activity. Various temperature distributions are formed by the dynamical heat transport and the adiabatic compression heating/expansion cooling. The atmospheric cooling mechanism in the thermosphere is the release of heat energy into space by infrared radiation from atmospheric molecules and the transport of heat from the thermosphere by heat conduction (e.g., Roble et al. 1987). Infrared radiation cooling is mainly caused by emissions of carbon dioxide (CO_2), nitric oxide (NO), and atomic oxygen (O), with wavelengths of 15 μm, 5.3 μm, and 63 μm, respectively. Until the 1970s, it was considered that the most effective infrared radiation was the

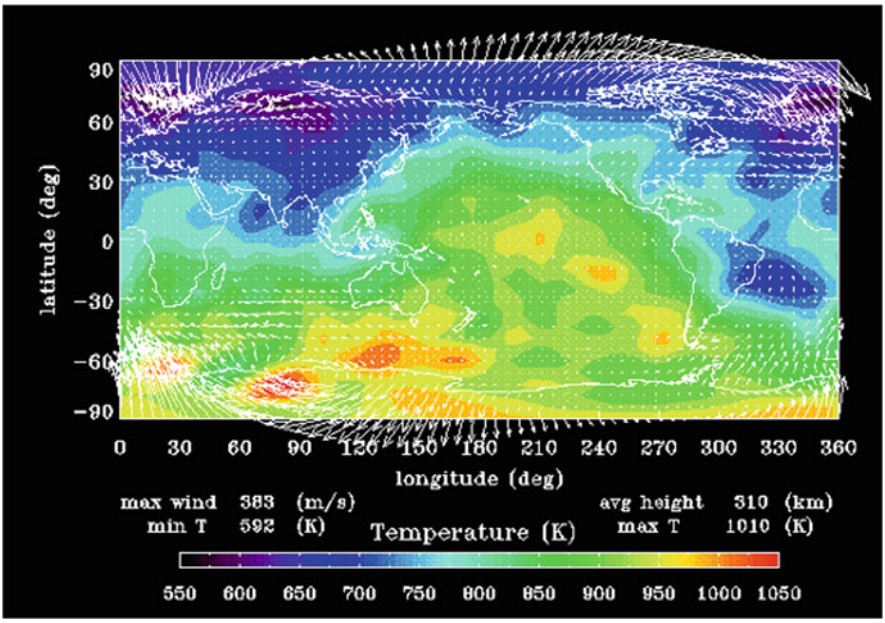

Fig. 14.3 Global distributions of the thermospheric temperatures and horizontal wind speeds at an altitude of about 300 km as calculated by the atmospheric general circulation model for solar minimum, geomagnetically quiet, and December conditions

63 μm radiation from O, which is abundant in the thermosphere (particularly in the upper thermosphere). Currently, it is known that the radiation from O was overestimated at that time and that the radiations from CO_2 and NO are more important. As mentioned earlier, the increase in CO_2 in the thermosphere is related to its increase in the troposphere (global warming). NO is produced by the dissociation of molecular nitrogen (N_2) caused by the solar EUV radiation, auroral particles, and solar protons, depending on the solar activity (details of the effects of NO are given in the next section).

14.1.4 Changes in Density and Composition of the Thermosphere

As mentioned above, the temperature of the thermosphere significantly varies depending on the solar activity. When atmospheric heating and temperatures are high, the air in the thermosphere expands and spreads to high altitudes. In contrast, when atmospheric heating and temperature are low, the air in the thermosphere shrinks and contracts to low altitudes. Therefore, the altitude of the upper end of the thermosphere rises and falls depending on the solar activity, and the mass density of

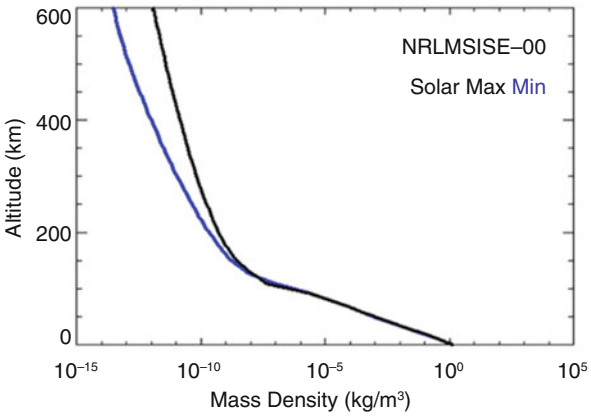

Fig. 14.4 Global mean altitude profiles of the atmospheric mass density from NRLMSISE-00

the thermosphere increases and decreases in a certain altitude range. Figure 14.4 shows the altitude distributions of the atmospheric mass density in the altitude range of 0–600 km, which are examples of the global mean profiles obtained using NRLMSISE-00 during geomagnetically quiet and solar minimum and maximum periods. For example, at an altitude of 400 km, the density during the solar maximum period is approximately ten times higher than that during the solar minimum period.

Figure 14.5 shows examples of the global distributions of the thermospheric mass density (March, June, September, and December) at an altitude of about 400 km during solar minimum and geomagnetically quiet periods calculated by the atmospheric general circulation model. The changes in latitude and local time are remarkable, and local increments and decrements are seen in the temperature distribution. The spatial distribution of and temporal variations in the thermospheric mass density affect the magnitude of the atmospheric drag force applied on satellites flying in low-altitude orbits and the International Space Station. Understanding the spatiotemporal variations in the atmospheric mass density is an important issue for the operation of these spacecraft (the atmospheric drag force acting on satellites is described in Chap. 2).

A major difference between the upper and lower atmospheres is that the atmospheric composition (mixing ratio) changes with altitude. The lower atmosphere is well-mixed, and its atmospheric composition (mixing ratio) is almost constant. However, in the upper atmosphere, heavier atmospheric molecules are distributed downward and lighter ones upward owing to gravity separation effects. Fig. 14.6 shows examples of plasma (electron and ion) and neutral atmosphere number density distributions in the thermosphere and the ionosphere obtained by the ionospheric empirical model, IRI2007 (Bilitza and Reinisch 2008), and the neutral atmosphere model, NRLMSISE-00. The profiles are obtained at the vernal equinox during geomagnetically quiet and solar minimum (solid line) and maximum (dotted line) periods at mid-latitude noon. At an altitude of about 300 km, where the electron density peaks, the electron density is approximately 0.1% of the neutral density, indicating that the neutral atmosphere is much more abundant than the ionospheric

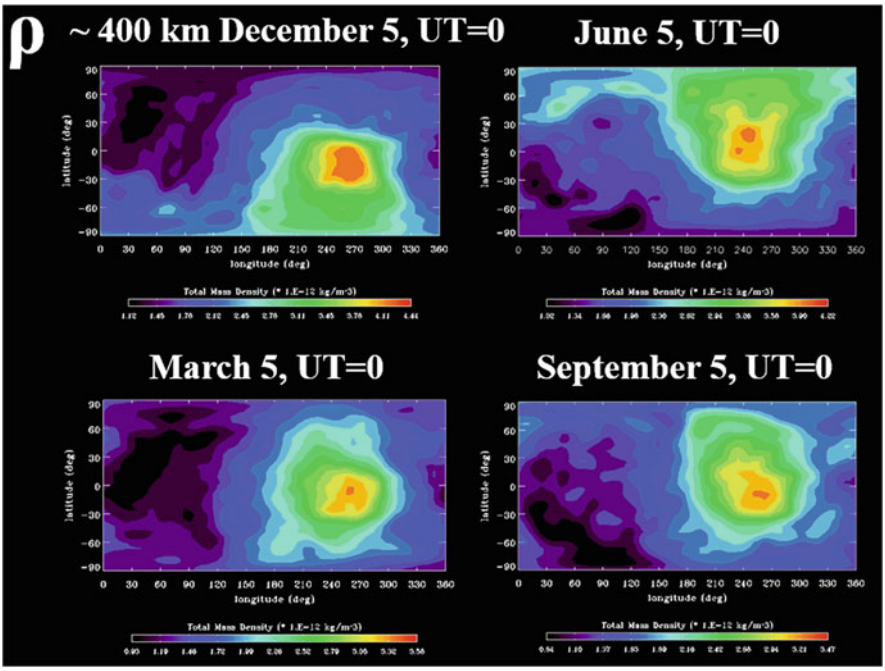

Fig. 14.5 Global distributions of the thermospheric mass densities at an altitude of about 300 km as calculated by the atmospheric general circulation model. Examples are shown for March, June, September, and December during solar minimum and geomagnetically quiet periods

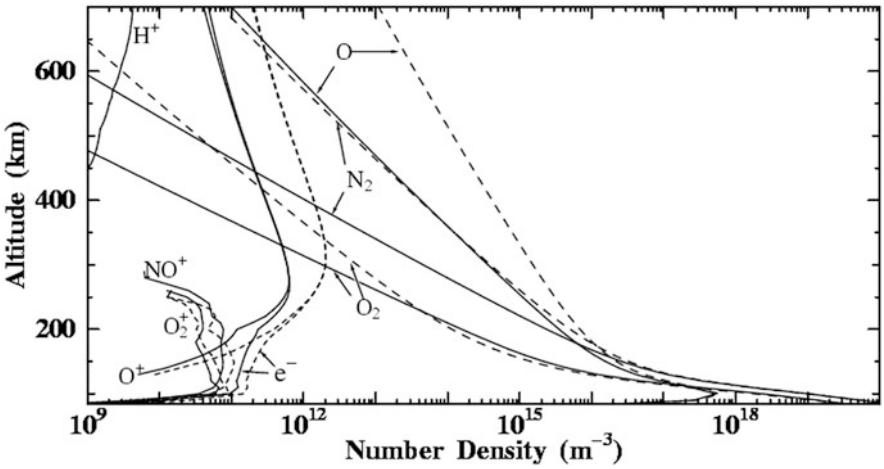

Fig. 14.6 Examples of plasma and neutral atmosphere number density profiles in the thermosphere and ionosphere obtained by IRI2007 and NRLMSISE-00

plasma in the ionospheric F region. In the upper thermosphere (above 200 km altitude), the solar activity-dependent changes in the composition ratio are larger than those in the lower region. The most abundant atmospheric components in the upper thermosphere are oxygen atoms, which contribute to the degradation of spacecraft surface materials (e.g., Tagawa et al. 1994).

14.1.5 Wind System Changes in the Thermosphere

In the troposphere, the Hadley, Ferrel, and polar circulations are characteristic, and in the stratosphere, the Brewer–Dobson circulation is characteristic, as is the meridional circulation. In the thermosphere, solar EUV/UV heating (which is larger at lower latitudes) and auroral heating (at higher latitudes) are the main atmospheric heat sources, and the thermospheric circulation is driven by the pressure gradient forces and ion drag at high latitudes.

A schematic of the thermospheric meridional circulation by Roble (1987) is shown in Fig. 14.7. Meridional circulation cells driven from low- and high-latitude heating regions can be seen. The sizes of the circulation cells change depending on the magnitude of the heating and the extent of the heating region in various conditions.

Similar to Figs. 14.7 and 14.8 is a schematic of the thermospheric meridional circulation by Forbes (2007) and the associated composition changes. In the upwelling region, the nitrogen-rich atmosphere in the lower layer is lifted, whereas in the

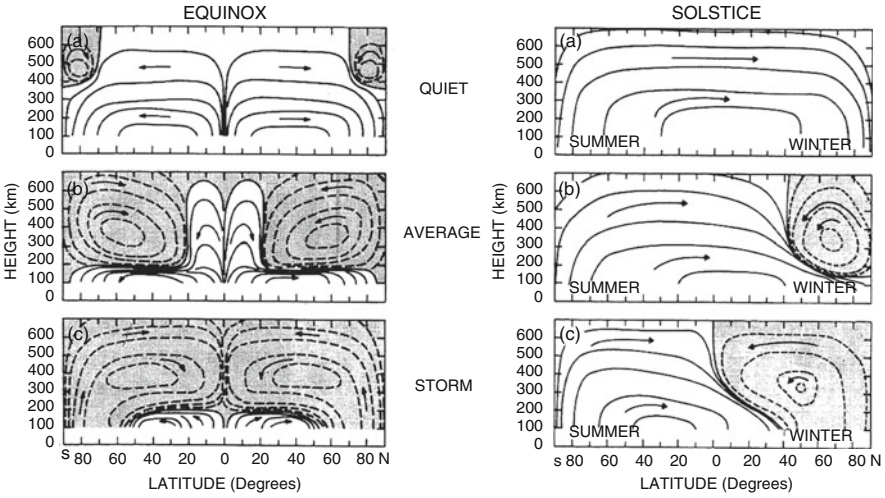

Fig. 14.7 Schematics of the thermospheric circulations. Left panels correspond to the vernal and autumnal equinoxes, and right panels correspond to the summer and winter solstices. From top to bottom, circulations are shown during geomagnetically quiet, averaged geomagnetic activity, and geomagnetically disturbed periods. (Roble 1987)

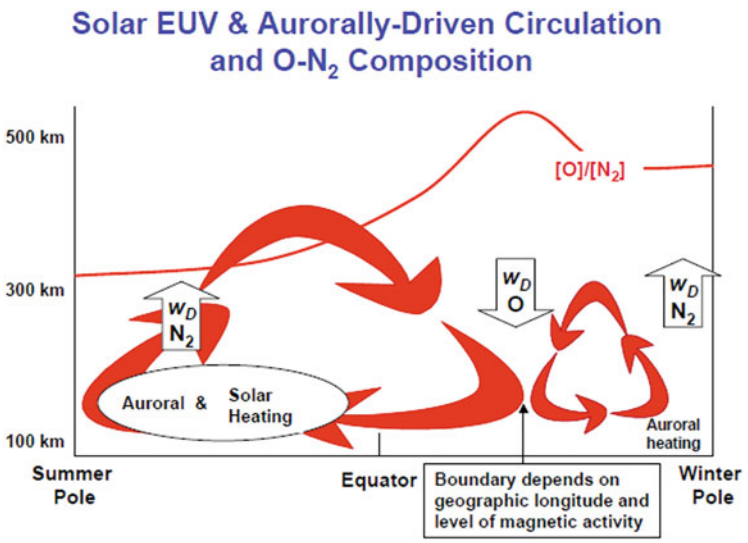

Fig. 14.8 Meridional circulation in the thermosphere and associated composition changes. (Forbes 2007)

downwelling region, the oxygen-rich atmosphere in the upper layer is lowered. N_2 and O_2 produce NO^+ and O_2^+ by charge exchange with O^+. Because NO^+ and O_2^+ recombine with electrons more rapidly than O^+, N_2 and/or O_2-rich air contribute to the reduction in the plasma. Thus, the ratio of N_2 to O (O/N_2) is an important parameter for understanding the atmospheric circulation and the formation and disappearance of the plasma.

The global wind system in the upper thermosphere is shown in Fig. 14.3, where high-speed thermospheric winds are noticeable in the polar region. It is understood that these are formed by the combination of the day-to-night flow (driven by the pressure gradient force) and the atmospheric motion due to the ion drag force. In the upper thermosphere, the geostrophic balance is no longer valid, and winds blow in the direction of the barotropic tilt (Fuller-Rowell 1995).

14.1.6 Thermospheric Changes During Geomagnetic Disturbances

During geomagnetically disturbed periods, various changes occur in the thermosphere with time scales ranging from several tens of minutes to several hours and days. The spatiotemporal scales of the thermospheric variability are highly diverse to the extent that it is extremely difficult to understand the causes as well as to predict and forecast the thermospheric weather. In this section, we present an overview of

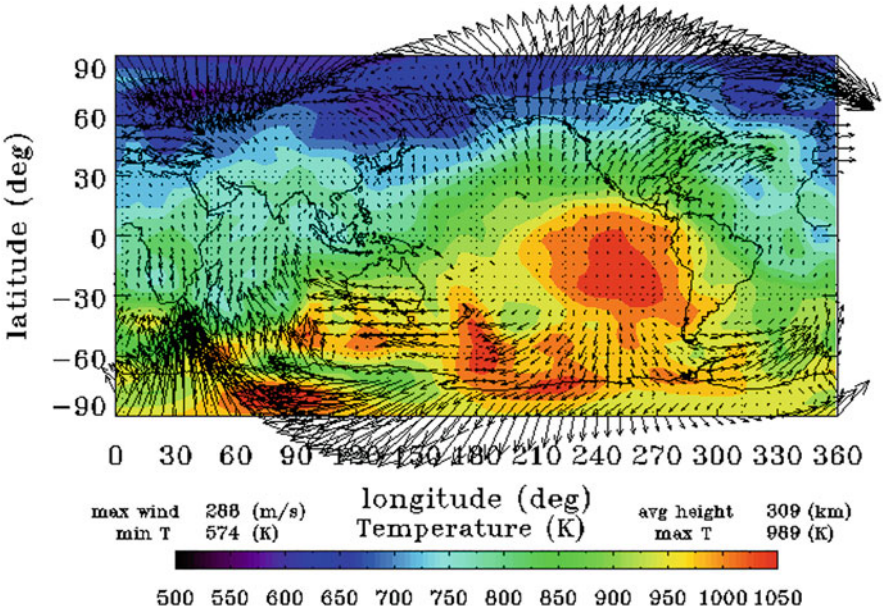

Fig. 14.9 Global distributions of the thermospheric temperatures and horizontal wind speeds at an altitude of about 309 km as calculated by the atmospheric general circulation model. Examples are shown for solar minimum, geomagnetically quiet (when the polar cap potential difference is 30 kV), and December conditions

the currently known thermospheric variations during geomagnetically disturbed periods and some of the future issues.

The global distributions of the thermospheric temperatures and horizontal winds calculated by the atmospheric general circulation model are shown under the condition of low solar activity and December during geomagnetically quiet periods (polar cap potential difference of 30 kV) in Fig. 14.9. The corresponding distributions when the energy inflow to the polar region is increased for 24 h (polar cap potential difference of 156 kV) are shown in Fig. 14.10. Both Figs. 14.9 and 14.10 show the values on a constant-pressure surface, and the global mean heights are 309 km and 384 km, respectively. In Fig. 14.9, the maximum temperature is 989 K and the maximum wind speed is 288 m/s, whereas in Fig. 14.10, these are 1573 K and 908 m/s, respectively. Note that the magnitude of NO infrared radiation may be underestimated in the calculation results shown in Fig. 14.10, and the values of both temperature and wind speed may be slightly larger. These changes in the geomagnetic activity (auroral activity) are caused by the solar wind variability associated with solar flares and other events. These geomagnetic activity changes contribute to thermospheric variabilities with short timescales. In the case of a strong geomagnetic storm, the amplitude of the thermospheric variability is comparable to and sometimes exceeds that associated with the solar activity cycle changes. Although changes in the geomagnetic activity (auroral activity) are caused by changes in the

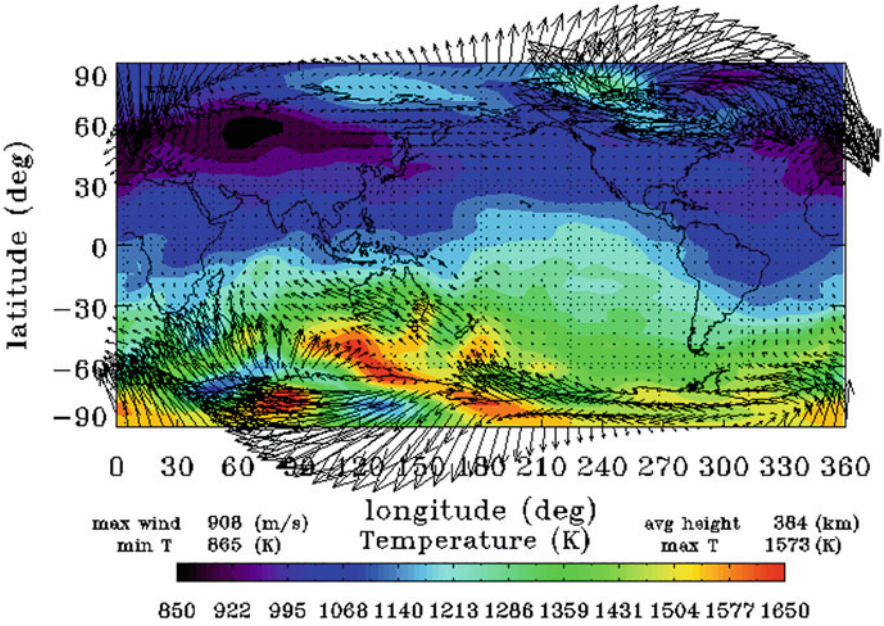

Fig. 14.10 Global distributions of the thermospheric temperatures and horizontal wind speeds at an altitude of about 384 km as calculated by the atmospheric general circulation model. Examples are shown for solar minimum, geomagnetically disturbed (when the polar cap potential difference is 154 kV for 24 h from 30 kV), and December conditions

amount of energy flowing into the polar regions, the thermospheric changes due to them are not limited to the polar regions but extend to mid- and low latitudes, resulting in global variations. In some cases of geomagnetic storms, the thermospheric circulation, as shown in Fig. 14.7, will significantly change over the course of a day.

Traveling atmospheric disturbances (TADs) are known to be one of the main causes of global thermospheric variability. There are various sources of TADs, and the details are still unknown. Large-scale TADs (LS-TADs) are frequently observed during geomagnetically disturbed periods (e.g., Fujiwara and Miyoshi 2006; Bruinsma and Forbes 2007). LS-TADs propagate from the polar regions to lower latitudes with speeds of about 600–700 m/s, causing changes in the thermospheric temperature, wind, mass density, and composition (e.g., Prölss 1995).

Measurements of the atmospheric mass density in the upper thermosphere (near 400 km altitude) by an accelerometer onboard the CHAMP satellite have been successful in observing various thermospheric variations (e.g., Liu et al. 2005; Bruinsma and Forbes 2007). During large geomagnetic storms, CHAMP observations have shown that the thermospheric mass density along the orbit can increase by a factor of up to five within a few tens of minutes. The recovery time for the increased mass density to the original state in the upper thermosphere after a geomagnetic storm is estimated to be about 12–15 h, based on current physical

models (e.g., TIE-GCMs). However, the CHAMP satellite observations have shown about 8 h as the recovery time for the mass density (Lei et al. 2011). The causes of the discrepancy between the model calculations and the CHAMP observations are still unknown. For predicting the thermospheric weather and for the safe operation of spacecraft, we should understand the mass density variations in the thermosphere.

14.2 Impacts of Variation in Solar Activity on the Mesosphere and Lower Thermosphere

In this section, we describe the general circulation of the atmosphere in the mesosphere and the lower thermosphere (altitudes from about 50 km to 120 km). First, we describe the general circulation during the solar minimum and geomagnetically quiet conditions. Subsequently, we show the impacts of the variations in the solar UV/EUV flux and the geomagnetic activity on the general circulation in the mesosphere and the thermosphere. We also present a study on the transport of nitric oxide (NO) from the thermosphere to the stratosphere, which is important to understand the effects of the upper atmosphere on the lower atmosphere.

14.2.1 Temperature Distributions in the Mesosphere and Thermosphere

The atmosphere is divided into the troposphere, stratosphere, mesosphere, and thermosphere, based on the vertical distribution of temperatures. The troposphere is from the surface to an altitude of about 10–15 km (the tropopause), and the stratosphere is from the tropopause to an altitude of about 50–65 km (the stratopause). The mesosphere is from the stratopause to an altitude of about 80–95 km (the mesopause), and the thermosphere is located above the mesopause. The height region from the stratosphere to the lower thermosphere (up to an altitude of about 120 km) is called the middle atmosphere. The condensation of water vapor plays an important role in the general circulation in the troposphere, whereas the interaction process between the neutrals and plasmas is important to study the general circulation in the thermosphere.

Figure 14.11 shows the latitude–height distributions of temperatures in June under a low solar UV/EUV flux (F10.7 = 70) and a geomagnetically quiet (Kp = 1) condition. This temperature distribution is obtained using the neutral atmosphere part of the coupled atmosphere–ionosphere model, the ground-to-topside atmosphere and ionosphere model for aeronomy (Miyoshi and Fujiwara 2003; Jin et al. 2011). The temperature distribution in the middle atmosphere is almost completely determined by the radiation (solar and infrared) and the heat transport by dynamics. Heating by solar UV absorption by ozone is dominant from the

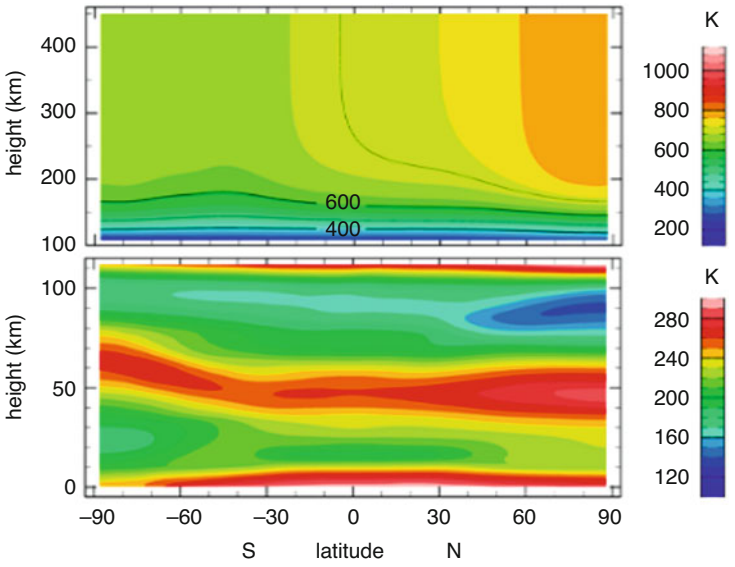

Fig. 14.11 Latitude–height distributions of the zonal mean temperatures averaged over June 21–25 during the low solar UV/EUV flux (F10.7 = 70) and geomagnetically quiet (Kp = 1) condition. Units are K

stratosphere to the lower mesosphere, whereas the solar UV/EUV absorption by oxygen and nitrogen molecules is prominent in the thermosphere. Therefore, the temperature in the Arctic region (high latitudes in the summer hemisphere) in the stratosphere and the thermosphere, where solar heating is high, shows maxima. In contrast, the high latitudes in the winter hemisphere in the stratosphere and the thermosphere, where there is no solar heating, have a low temperature. The temperature minimum is located near the mesopause region, where the absorption of solar radiation is minimum. The cooling rate due to infrared radiation is determined by the temperature and the concentrations of CO_2, NO, and O. For details, please refer to the infrared radiation observations by the SABER satellite (Mlynczak et al. 2010).

Figure 14.12 shows the heating and cooling rates due to radiation (the sum of the solar and infrared radiation). It is noticeable that in the summer hemisphere, the solar heating exceeds the infrared cooling. However, the infrared cooling exceeds the solar heating in high latitudes in the winter hemisphere. If there is no heat transport by dynamics, the resultant temperature in the summer (winter) hemisphere is higher (lower) than the temperature obtained in Fig. 14.11. This suggests that the temperature distribution shown in Fig. 14.11 is maintained by the balance between the radiation and the heat transport through dynamics. In particular, from the upper mesosphere to the mesopause (70–100 km altitude), the temperature in the high-latitude summer hemisphere is lower than that in the high-latitude winter hemisphere. This latitudinal distribution of temperature cannot be explained by radiative heating/cooling. The details are explained in the next section.

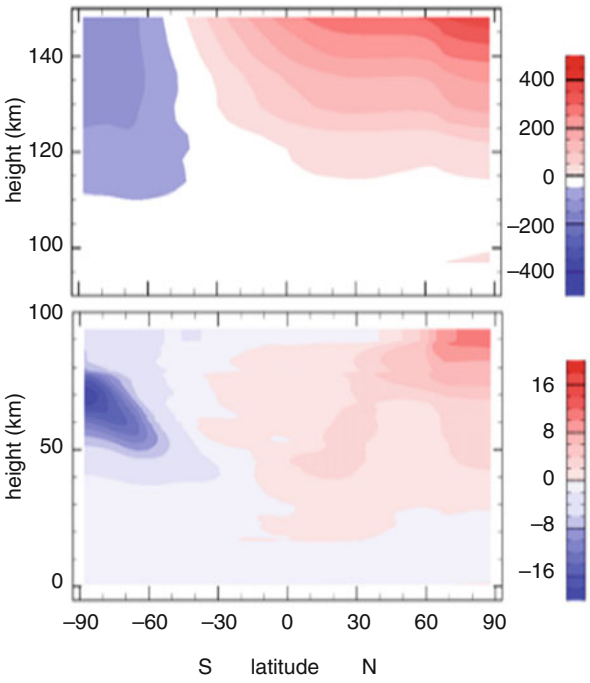

Fig. 14.12 Latitude–height distributions of heating/cooling rates due to the solar and infrared radiation (sum of the solar and infrared radiation). Units are K/day

14.2.2 Zonal and Meridional Wind Distributions in the Mesosphere and Thermosphere

This section describes the zonal and meridional wind distributions in the mesosphere and the thermosphere. Figure 14.13 (left figure) shows the latitude–height distributions of the zonal mean zonal winds in June under a low solar UV/EUV flux (F10.7 = 70) and a geomagnetically quiet (Kp = 1) condition. In the stratosphere, high temperatures in the high latitudes of the summer hemisphere and low temperatures in the high latitudes of the winter hemisphere produce high and low pressures, respectively. This temperature distribution results in easterly winds in the summer hemisphere and westerly winds (called polar night jets) in the winter hemisphere via the thermal wind balance. The magnitudes of the easterly and westerly winds in the mesosphere decrease with increasing height, and the zonal wind speed becomes weak in all latitudes at altitudes of 80–100 km. This weak wind layer is caused by the drag force generated by the breaking of atmospheric waves (such as gravity waves and planetary waves). In the easterly region of the summer hemisphere, eastward-propagating gravity waves can propagate vertically easily (westward-propagating gravity waves cannot propagate vertically). Moreover, the easterly winds are weakened by the release of the eastward momentum caused by the breaking of the westward-propagating gravity waves in the upper mesosphere. However, in the winter hemisphere, westward-propagating gravity waves attenuate the westerly

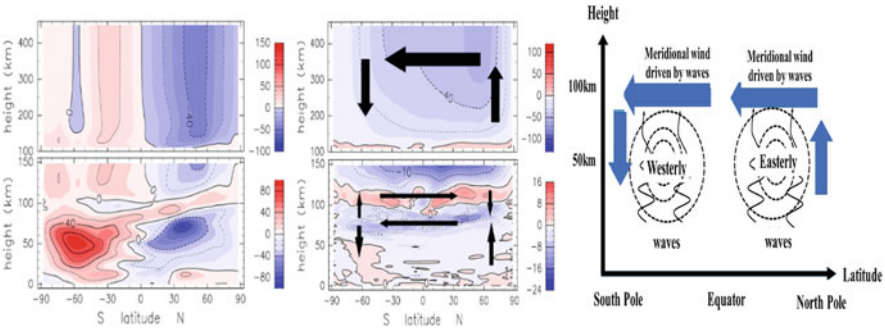

Fig. 14.13 (**left figure**) Latitude–height distributions of the zonal mean zonal winds averaged over June 21–25 during the low solar UV/EUV flux (F10.7 = 70) and geomagnetically quiet (Kp = 1) condition. Units are m/s. Positive (negative) values represent westerly (easterly) winds. (**middle figure**) Latitude–height distributions of the zonal mean meridional winds. Positive (negative) values denote southerly (northerly) winds. (**right figure**) Schematic of the effects of atmospheric waves on the zonal mean meridional circulation. Blue arrows indicate the meridional circulation driven by the breaking of atmospheric waves

winds in the upper mesosphere. In the stratosphere, the drag due to planetary waves is also important (Sato et al. 2018). If no drag is produced by atmospheric waves, the polar night jets will exceed 200 m/s. The drag caused by the breaking of atmospheric waves drives the meridional circulation (which shows an upward motion in high latitudes in the summer hemisphere), meridional winds from the summer hemisphere to the winter hemisphere (northerly winds in June), and downward motion in high latitudes in the winter hemisphere (Fig. 14.13 (middle)). These upward and downward motions produce adiabatic cooling and heating, respectively. Specifically, the adiabatic cooling (heating) leads to lower (higher) temperatures in high latitudes of the summer (winter) hemisphere. Therefore, the latitudinal distribution of the temperatures in the mesosphere is explained by the adiabatic heating/cooling due to the wave-induced meridional circulation (see Fig. 14.13 (right)).

In the thermosphere above 120 km, the upward motion in the summer hemisphere, meridional winds from the summer to the winter hemisphere (northerly winds in June), and downward motion in the winter hemisphere prevail. The meridional circulation in the thermosphere is not generated by the drag due to atmospheric waves, as described above; instead, it is a thermal circulation caused by the latitudinal distribution of solar heating. The meridional circulation shown here is diurnal and longitudinally averaged. Because thermospheric winds present significant diurnal variation, it is necessary to investigate their diurnal and longitudinal distributions (see Sect. 14.1).

The meridional circulation in the lower thermosphere at altitudes of 100–120 km is opposite to those in the stratosphere and the mesosphere (i.e., upward motion in high latitudes of the winter hemisphere, meridional winds to the summer hemisphere, and downward motion in high latitudes of the summer hemisphere). A numerical study (Miyoshi et al. 2015) suggested that this meridional circulation is

generated by the drag due to the breaking of gravity waves. However, global observations of the wind system in the lower thermosphere are restricted, and further modeling and observational studies are required.

In mid- and high latitudes of the mesosphere and the lower thermosphere, temperature and wind systems have notable annual cycles due to the seasonal changes in the solar radiation. However, in low latitudes, the seasonal variations in the solar radiation are small, and there are unique daily variations. For example, in the upper stratosphere and the lower mesosphere, the semi-annual oscillation (Hirota 1980), in which the directions of the zonal winds change every 6 months, is remarkable. In contrast, in the upper mesosphere, an intraseasonal oscillation, in which the zonal winds change every 30–60 days, is also observed (Lieberman 1998). These oscillations are considered to be generated by the drag due to the breaking of atmospheric waves propagating from the lower atmosphere. However, the momentum budget of these variations is still unclear, and future studies are necessary to clarify.

14.2.3 Temperature Changes Due to Variations in the Solar and Geomagnetic Activities

In previous sections, we have shown the general circulations in the mesosphere and thermosphere under the low solar UV/EUV flux and the geomagnetically quiet condition. In this subsection, we investigate the impacts of the variations in the solar UV/EUV flux and the geomagnetic activity on the temperatures in the mesosphere and the thermosphere. Figure 14.14a shows the latitude–height distribution of the temperatures under the high solar UV/EUV flux (F10.7 = 200) and the geomagnetically quiet (Kp = 1) condition. The temperature difference between the high and low solar UV/EUV fluxes is shown in Fig. 14.14 (right figure). Above an altitude of 100 km, the temperature difference is significant. For example, at an altitude of about 110 km, the temperature difference is 10–25 K. The temperature difference increases with altitude, and at an altitude of 200 km, the temperature difference is 300–500 K. However, the temperature difference below 100 km altitude is small (less than 2 K). The temperature difference at an altitude of 100 km is only a few Kelvin. In the stratosphere and the mesosphere, the temperature increase due to the increase in the UV/EUV flux is also observed; however, the temperature increase is limited to about 1–2 K. This temperature increase is due to the increase in the solar UV flux absorbed by the ozone in the lower stratosphere and the mesosphere.

To investigate the temperature change caused by geomagnetic activity, a geomagnetic disturbance simulation with the geomagnetically disturbed condition (Kp = 5) was conducted, which is referred to as EXP1. The temporal variation in Kp used in EXP1 is shown in Fig. 14.15 (left figure). A simulation with Kp = 1 was conducted until 00 Universal time (UT) on June 21, following which a simulation was conducted under the condition of Kp = 5 from 00 UT on June 21 to 00 UT on

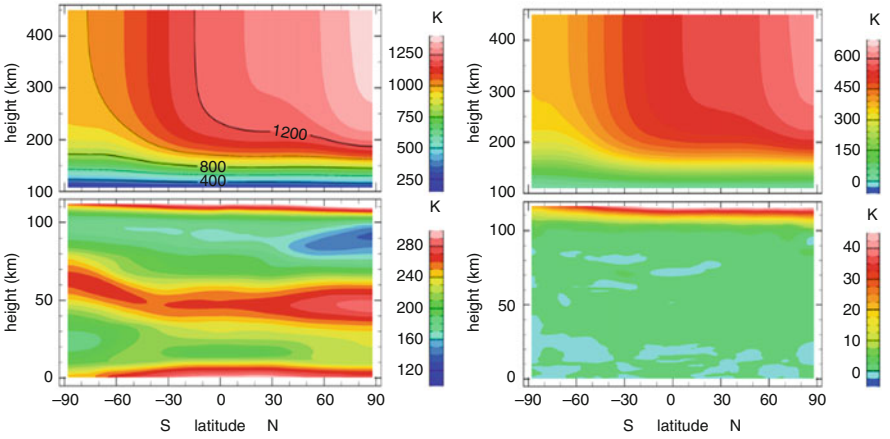

Fig. 14.14 (**left figure**) Latitude–height distributions of the zonal mean temperatures averaged over June 21–25 during the high solar UV/EUV flux (F10.7 = 200) and geomagnetically quiet (Kp = 1) condition. (**right figure**) Temperature difference between the high (F10.7 = 200) and low (F10.7 = 70) solar UV/EUV fluxes. Units are K

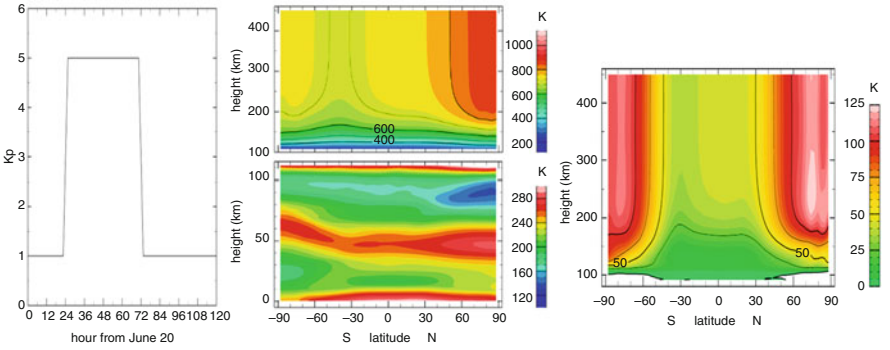

Fig. 14.15 (**a**) Temporal variation in Kp used in the numerical experiment. (**b**) Latitude–height distributions of the zonal mean temperatures during the geomagnetically disturbed and solar radiation minimum conditions. (**c**) Temperature difference between the geomagnetically disturbed and geomagnetically quiet (Kp = 1) conditions. Diurnal mean temperatures on June 22 are shown. The solar UV/EUV fluxes in both conditions are fixed at F10.7 = 70. Units are K

June 23. After 00 UT on 23 June, a simulation was conducted again under Kp = 1. The conditions other than Kp were the same as those shown in Fig. 14.11. A low solar UV/EUV flux (F10.7 = 70) was assumed during the simulation. Figure 14.15 shows the temperature distribution on the second day with Kp = 5 obtained by EXP1 and the temperature difference between Kp = 5 and Kp = 1. It is noticeable that the temperature increase is enhanced at high latitudes in the thermosphere. In particular, in the winter hemisphere, the temperature is higher at high latitudes than at middle latitudes (the north–south gradient of temperature is reversed). The temperature

increase due to the disturbance reaches 75–120 K above an altitude of 150 km, whereas that in the lower thermosphere is small. For example, the temperature rise is 20 K at an altitude of 120 km and a few Kelvin at an altitude of 100 km. Furthermore, the temperature increase below 100 km of altitude is negligible.

14.2.4 Wind Changes Due to Variations in the Solar and Geomagnetic Activities

We investigate the impacts of the variations in the solar UV/EUV flux and the geomagnetic activity on the zonal and meridional winds in the mesosphere and the thermosphere. First, we examine the wind changes caused by the variation in the solar UV/EUV flux (Fig. 14.16). As in the case of temperature, no change is observed in the easterly wind below 110 km altitude, and the difference between 110 and 130 km altitudes is about 2–4 m/s. Above 130 km altitude, the easterly (westerly) winds in the summer (winter) hemisphere and the meridional winds from the summer to the winter hemisphere above 130 km altitude are enhanced under the high solar UV/EUV flux.

There is no change in the meridional wind distribution below 110 km altitude. However, above an altitude of 120 km, the meridional wind is also strengthened. This enhanced meridional circulation during the high solar UV/EUV condition is caused by the enhanced north–south temperature gradient, which is due to the increase in solar heating.

We also examine the wind changes caused by the geomagnetic activity (Fig. 14.17). Figure 14.17 shows the zonal mean zonal and meridional winds during the low solar UV/EUV flux (F10.7 = 70) and geomagnetically disturbed (Kp = 5)

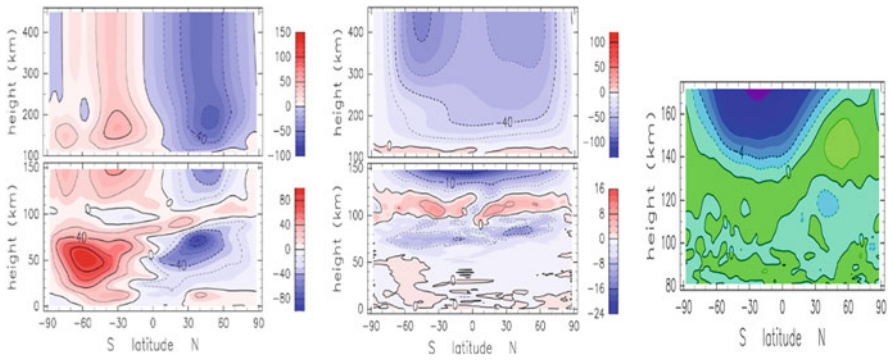

Fig. 14.16 (**left figure**) Latitude–height distributions of the zonal mean zonal winds averaged over June 21–25 during the high solar UV/EUV flux (F10.7 = 200) and geomagnetically quiet (Kp = 1) condition. Units are m/s. (**middle figure**) Meridional wind distributions averaged over June 21–25 during the high solar UV/EUV flux and geomagnetically quiet condition. (**right figure**) Meridional wind difference between the high and low solar UV/EUV fluxes

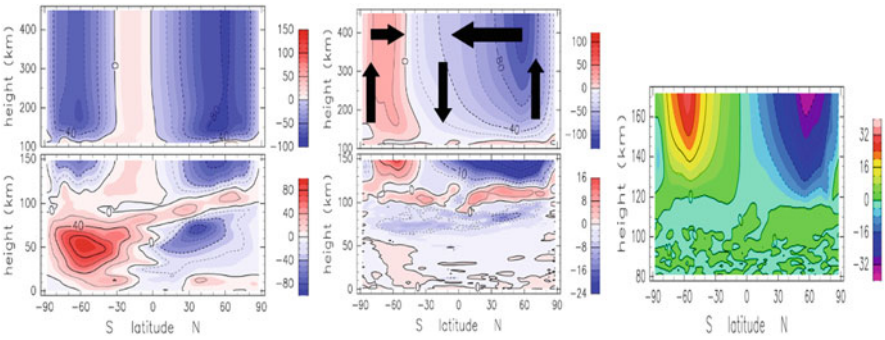

Fig. 14.17 (**left figure**) Latitude–height distributions of the zonal mean zonal winds under the low solar UV/EUV flux (F10.7 = 70) and geomagnetically disturbed (Kp = 5) condition. Diurnal mean on 22 June is shown. Units are m/s. (**middle figure**) Meridional wind distributions under the low solar UV/EUV flux and geomagnetically disturbed condition. (**right figure**) Meridional wind difference between the geomagnetically disturbed and quiet conditions

condition. No change is obtained in the zonal winds below 110 km altitude. Above 110 km altitude, the westerly winds in the winter hemisphere found in the low UV/EUV case are replaced by easterly winds, which is related to the reversal of the north–south temperature gradient caused by the geomagnetic disturbance. The meridional circulation below an altitude of 110 km also shows little change, whereas that above 110 km altitude in the winter hemisphere shows a significant change. The heating in the high latitudes of the winter hemisphere produces an upward motion in the polar regions and equatorward winds, as in the summer hemisphere. Specifically, during the geomagnetic disturbance, the meridional circulation in the winter hemisphere has the same characteristics as that in the summer hemisphere.

14.2.5 Changes in Minor Constituents Due to the Solar and Geomagnetic Activities

Although the general circulation is above 120 km altitude, the thermosphere is significantly affected by the solar UV/EUV variation, and the lower thermosphere is found to be little affected. In addition, it is found that the stratosphere and the mesosphere are almost unaffected. Thus, the direct effects of the solar UV/EUV flux and geomagnetic disturbances on the mesosphere and the lower thermosphere are considered to be small. However, the effects on the stratosphere and the mesosphere of variations in minor constituents (such as NO) have been discussed. For example, N (nitrogen atoms) are produced from N_2 (nitrogen molecules) by solar UV/EUV and auroral energetic particle precipitation from the magnetosphere, and NO is produced by the reaction with O_2 (oxygen molecules) (production peak of NO is located at about 110–130 km altitude). NO causes infrared cooling in the lower thermosphere and O_3 (ozone) depletion in the stratosphere and the mesosphere.

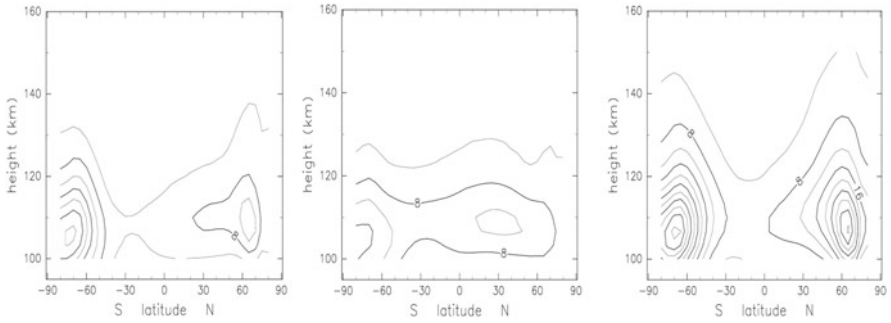

Fig. 14.18 (**left figure**) Latitude–height distribution of NO densities on 21 June during the low solar UV/EUV flux (F10.7 = 70) and geomagnetically quiet (Kp = 1) condition obtained by an empirical model. Units are $10^7/cm^3$. (**middle figure**) NO distribution during the high solar UV/EUV flux (F10.7 = 200) and geomagnetically quiet (Kp = 1) condition. (**right figure**) NO distribution during the low solar UV/EUV flux (F10.7 = 70) and geomagnetically disturbed (Kp = 5) condition

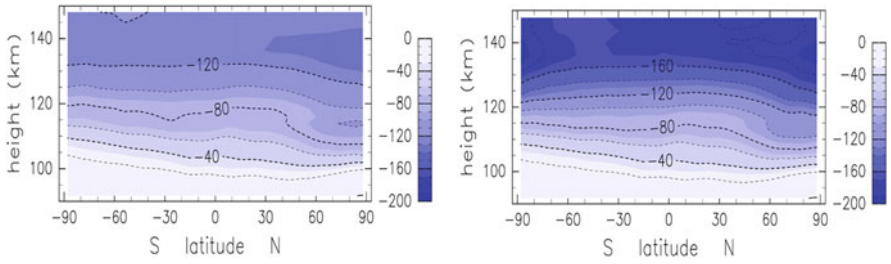

Fig. 14.19 (**left figure**) Latitude–height distribution of cooling rates due to the infrared radiation averaged over June 21–25 during the low solar UV/EUV flux (F10.7 = 70) and geomagnetically quiet (Kp = 1) condition. Units are K/day. (**right figure**) Distribution of cooling rates due to the infrared radiation during the low solar UV/EUV flux and geomagnetically disturbed (Kp = 5) condition

Therefore, changes in the NO concentration are considered to play important roles in the atmospheric general circulation in the middle atmosphere.

$$N_2 + h\upsilon \rightarrow 2N\left(^2D\right), N\left(^2D\right) + O_2 \rightarrow NO + O$$

The latitude–height distribution of the NO concentration in the thermosphere obtained from an empirical model is shown in Fig. 14.18 (Marsh et al. 2004). It can be seen that even during the geomagnetically quiet condition, the NO concentration in the low and mid latitudes increases with increasing the solar UV/EUV flux, whereas the NO concentration in the high latitudes increases during the geomagnetically disturbed condition.

First, we investigate the effects of infrared cooling by NO. Figure 14.19 (left) shows the cooling rates due to the infrared radiation during the geomagnetically quiet (Kp = 1) condition (left figure) and geomagnetically disturbed (Kp = 5) (right

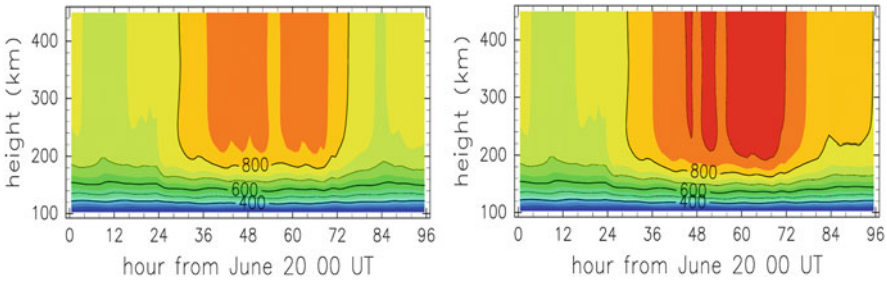

Fig. 14.20 (**left figure**) Time variation in the temperature in the Arctic region (averaged from 90°N to 70°N) with NO variation (EXP1; temporal variation in NO density is considered). (**right figure**) Time variation diagram of the temperature in the Arctic region without NO variation (EP2; NO density is fixed)

figure) conditions simulated in EXP1, respectively. It can be seen that the infrared cooling rate in the thermosphere increases with the geomagnetic disturbance. This increase in the cooling rate by the infrared radiation is considered to be a result of the increase in the NO concentration and temperature caused by the increase in joule heating and auroral energetic particle precipitation. To investigate the effect of the increase in NO concentration on the thermospheric temperature, we conducted an additional numerical simulation (EXP2). The temporal variation in Kp used in EXP2 is the same as that in EXP1. However, the NO concentration in EXP2 is fixed during the numerical simulation (EXP2 is identical to EXP1 except for the NO concentration). Figure 14.20 shows the time variations in the temperature in the polar region (averaged from 90°N to 70°N) simulated in EXP1 and EXP2. In both experiments, the temperature increases on 21 June (24 h in the figure), when the geomagnetic disturbance occurs. In EXP1, the temperature decreases rapidly after 00 UT on 23 June (72 h in the figure) when Kp decreases from 5 to 1, and it returns to the original value after 12 h (84 h in the figure). In contrast, in EXP2, the temperature remains high even at 00 UT on 23 June (72 h in the figure), indicating that the temperature recovery (temperature decrease) after the geomagnetic disturbance is delayed. Thus, it can be inferred that the change in the NO concentration associated with a geomagnetic disturbance is important for the temperature change (heat balance) in the thermosphere.

14.2.6 Impacts of the Solar Activity on the Stratosphere and Mesosphere

The NO produced in the thermosphere by the solar UV/EUV flux and the auroral energetic particle precipitation from the magnetosphere is transported to the mesosphere by dynamical transport processes and affects the stratospheric ozone concentration. In this subsection, we consider the transport of NO from the thermosphere to

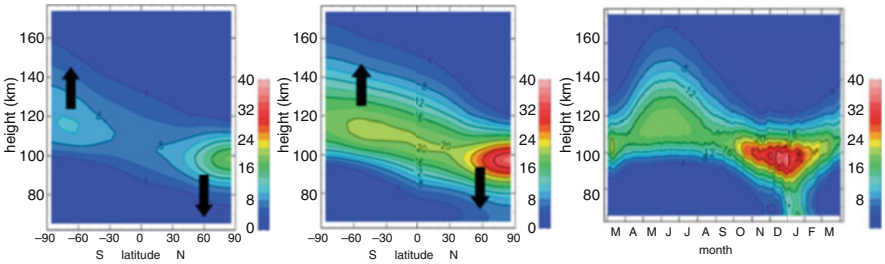

Fig. 14.21 (**left figure**) Latitude–height distribution of NO densities obtained by the numerical simulation. NO density distribution in December under the low solar UV/EUV flux (F10.7 = 70) and geomagnetically quiet (Kp = 1) condition. Units are $10^7/cm^3$. (**middle figure**) NO density distribution in December under the high solar UV/EUV flux (F10.7 = 200) and geomagnetically quiet (Kp = 1) condition. (**right figure**) Month–height distribution of NO densities in the Arctic region

the lower atmosphere. We present the results of a whole atmosphere general circulation model (Miyoshi and Fujiwara 2003) that incorporates the photochemical processes related to NO production and the NO transport process through dynamics. Figure 14.21 (left) shows the latitude–height distribution of NO in December under the low solar UV/EUV flux (F10.7 = 70) and geomagnetically quiet (Kp = 1) condition. Although NO production occurs in the lower thermosphere of both hemispheres, the peak height of the NO concentration in the summer hemisphere is higher than that in the winter hemisphere. This difference between the peak heights is explained by the transport process through dynamics. Specifically, the downward (upward) motion in the winter (summer) hemisphere transports NO downward (upward). In particular, the increase in NO in the mesosphere in the Arctic region is significant, where strong downward motion prevails.

Figure 14.21 (middle) shows the distribution of NO in December under the high solar UV/EUV flux (F10.7 = 200) and geomagnetically quiet condition. The increase in NO concentration is found not only at low latitudes, where NO is mainly produced, but also at high latitudes. In particular, the NO concentration in the Arctic region below 100 km altitude during the high solar UV/EUV condition increases significantly, because NO-rich air moves from the lower thermosphere to the mesosphere by the predominant downward motion. To clarify the seasonal variation in NO transport, Fig. 14.21 (right) shows the time–height distribution of the NO concentration in the Arctic region. The NO concentration in the mesosphere increases (decreases) in winter (summer) when the downward (upward) motion is dominant. Thus, the NO concentration in the mesosphere is closely related to the meridional circulation. The NO transported from the thermosphere to the mesosphere is considered to be related to stratospheric ozone depletion. Because the present model does not include the photochemical process of stratospheric ozone, it is impossible to estimate the effects of NO transport on the stratospheric ozone. However, the model indicates that variation in the solar UV/EUV flux may affect the

stratospheric ozone and the stratosphere–troposphere general circulation through NO transport.

References

Bilitza, D., Reinisch, B.W.: International reference ionosphere 2007: improvements and new parameters. Adv. Space Res. **42**(4), 599–609 (2008). https://doi.org/10.1016/j.asr.2007.07.048

Bruinsma, S.L., Forbes, J.M.: Global observation of traveling atmospheric disturbances (TADs) in the thermosphere. Geophys. Res. Lett. (2007). https://doi.org/10.1029/2007GL030243

Forbes, J.M.: Dynamics of the thermosphere. J. Meteorol. Soc. Jpn. **85B**, 193–213 (2007)

Fujiwara, H., Miyoshi, Y.: Characteristics of the large-scale traveling atmospheric disturbances during geomagnetically quiet and disturbed periods simulated by a whole atmosphere general circulation model. Geophys. Res. Lett. **33**, L20108 (2006). https://doi.org/10.1029/2006GL027103

Fuller-Rowell, T.J.: The Dynamics of the Lower Thermosphere. In: Johnson, R.M., Killeen, T.L. (eds.) The Upper Mesosphere and Lower Thermosphere: A Review of Experiment and Theory, Geophys. Monogr. Ser., vol. 87, pp. 23–36. AGU, Washington, DC (1995)

Hirota, I.: Observational evidence of the semiannual oscillation in the tropical middle atmosphere-a review. Pure Appl. Geophys. **118**, 217–238 (1980)

Jin, H.Y., Miyoshi, Y., Fujiwara, H., Shinagawa, H., Terada, K., Terada, N., Ishii, M., Otsuka, Y., Saito, A.: Vertical connection from the tropospheric activities to the ionospheric longitudinal structure simulated by a new Earth's whole atmosphere-ionosphere coupled model. J. Geophys. Res. **116**, A01316 (2011). https://doi.org/10.1029/2010JA015925

Lean, J.: Variations in the sun's radiative output. Rev. Geophys. (1991). https://doi.org/10.1029/91RG01895

Lean, J., Woods, T.N., Eparvier, F.G., Meier, R.R., Strickland, D.J., Correira, J.T., Evans, J.S.: Solar extreme ultraviolet irradiance: present, past, and future. J. Geophys. Res. (2011). https://doi.org/10.1029/2010JA015901

Lei, J., Thayer, J.P., Lu, G., Burns, A.G., Wang, W., Sutton, E.K., Emery, B.A.: Rapid recovery of thermosphere density during the October 2003 geomagnetic storms. J. Geophys. Res. (2011). https://doi.org/10.1029/2010JA016164

Lieberman, R.S.: Intraseasonal variability of high-resolution doppler imager winds in the equatorial mesosphere and lower thermosphere. J. Geophys. Res. **103**(D10), 11221–11228 (1998)

Liu, H., Lühr, H., Köhler, W.: Global distribution of the total mass density derived from CHAMP. J. Geophys. Res. **110**, A04301 (2005). https://doi.org/10.1029/2004JA010741

Marsh, D.R., Solomon, S.C., Reynolds, A.E.: Empirical model of nitric oxide in the lower thermosphere. J. Geophys. Res. **109**, A07301 (2004). https://doi.org/10.1029/2003JA010199

Miyoshi, Y., Fujiwara, H.: Day-to-day variations of migrating diurnal tide simulated by a GCM from the ground surface to the exobase. Geophys. Res. Lett. **30**(15), 1789 (2003). https://doi.org/10.1029/2003GL017695

Miyoshi, Y., Fujiwara, H., Jin, H., Shinagawa, H.: Impacts of sudden stratospheric warming on the general circulation of the thermosphere. J. Geophys. Res. Space Physics. **120**, 10897–10912 (2015). https://doi.org/10.1002/2015JA021894

Mlynczak, M.G., et al.: Observations of infrared radiative cooling in the thermosphere on daily to multiyear timescales from the TIMED/SABER instrument. J. Geophys. Res. **115**, A03309 (2010). https://doi.org/10.1029/2009JA014713

Picone, J.M., Hedin, A.E., Drob, D.P., Aikin, A.C.: NRLMSISE-00 empirical model of the atmosphere: statistical comparisons and scientific issues. J. Geophys. Res. **107**(A12), SIA–S15 (2002). https://doi.org/10.1029/2002JA009430

Prölss, G.W., Ionospheric F-region storms, In: Handbook of Atmospheric Electrodynamics, 195–248, Volume II, CRC Press Boca Raton, FL (1995)

Roble, R.G.: In: Akasofu, S.-I., Kamide, Y. (eds.) The Earth's Thermosphere, The Solar Wind and the Earth, pp. 245–264. Terra Scientific Publishing Company (TERRAPUB), Tokyo (1987)

Roble, R.G., Ridley, E.C., Dickinson, R.E.: On the global mean structure of the thermosphere. J. Geophys. Res. **92** (1987). https://doi.org/10.1029/JA092iA08p08745

Sato, K., Yasui, R., Miyoshi, Y.: The momentum budget in the stratosphere, mesosphere, and lower thermosphere. Part I: contributions of different wave types and in situ generation of Rossby waves. J. Atmos. Sci. **75**, 3613–3633 (2018). https://doi.org/10.1175/JAS-D-17-0336.1

Tagawa, M., Tomita, M., Umeno, M., Ohmae, N.: Atomic oxygen generators for surface studies in low earth orbit. AIAA J. **32**, 95–100 (1994)

Chapter 15
Impacts of Solar Activity Variations on Climate

Shigeo Yoden and Kohei Yoshida

15.1 Climate Impact Assessment of Solar Activity Variations Using a Hierarchy of Numerical Models

The development of climate models using electronic computers has followed the history of the development of numerical weather prediction (NWP) models (e.g., Lynch 2008 and references therein). Figure 15.1 shows the advancement in the processing speeds of the most advanced computers in the world since the first electronic computer, Electronic Numerical Integrator And Computer (ENIAC). Studies on experimental NWP using ENIAC began with Charney et al. (1950). Specifically, time integrations of a highly approximated prognostic equation (i.e., a horizontal two-dimensional vorticity equation in a rotating system) were performed with an initial condition based on observations in a limited domain over the North American continent. The objective was to forecast a large-scale flow field in the future (1-day ahead). With the limited computational power available at the time, those procedures required 5 weeks of computational time and did not "predict" actual weather changes in advance. However, they demonstrated the basic computational principle of forecasting the temporal evolution of atmospheric motion based on the physical law of fluids. In contrast, climate model experiments attempted to reproduce the atmospheric general circulation in a forced-dissipative system, including meridional variations in diabatic heating and friction, instead of the initial value problem of NWPs. These experiments began with the study by Phillips (1956), who used a two-layer mid-latitude atmospheric model with periodic boundary conditions in the zonal direction and two boundary walls to the north and south. Accordingly,

S. Yoden (✉)
Institute for Liberal Arts and Sciences, Kyoto University, Kyoto, Japan
e-mail: yoden.shigeo.53r@st.kyoto-u.ac.jp

K. Yoshida
Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Ibaraki, Japan

K. Kusano (ed.), *Solar-Terrestrial Environmental Prediction*,
https://doi.org/10.1007/978-981-19-7765-7_15

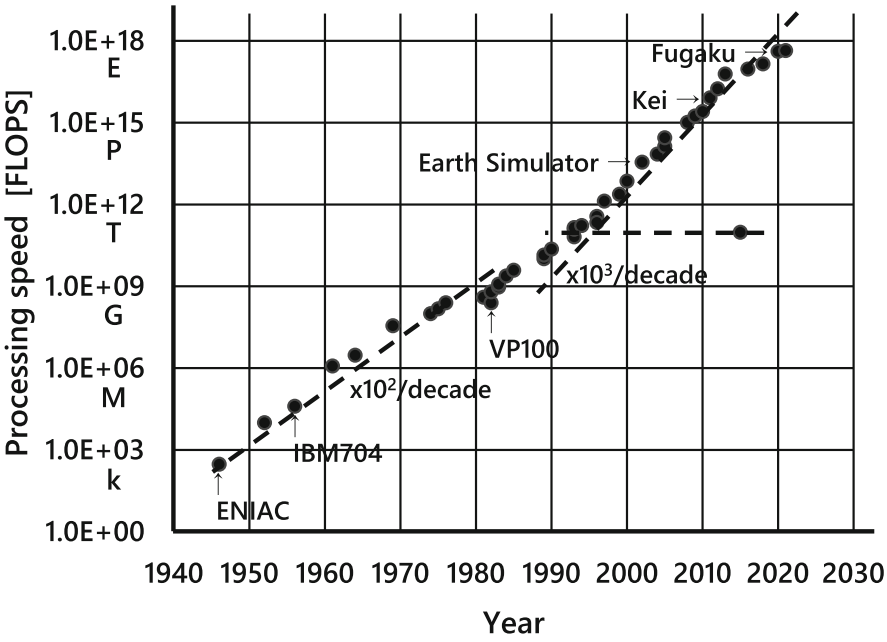

Fig. 15.1 History of processing speed improvement of the most advanced computers in the world; basically, exponential growth of processing speed has continued since the birth of ENIAC, accelerating from about 2 digits per decade until the 1980s to about 3 digits per decade since then. This is largely due to the development of vector computers and massively parallel computers. One point of difference in 2016 is the processing speed of my laptop computer, which is equivalent to the fastest level in the world about 20 years ago

he performed a 1-month time integration from the initial condition of a motionless atmosphere to show the appearance of a westerly jet and the superimposition of high- and low-pressure disturbances. The poleward heat transport caused by the disturbances was quantitatively evaluated from the output data of the time integration. However, it was reported that numerical errors rapidly increased and longer time integrations could not be performed. It was not until the 1960s that a finite difference scheme that ensured the conservation of energy and enstrophy for a two-dimensional incompressible flow was developed.

Since then, the processing speed of computers has increased exponentially, and the advancement of computers has promoted studies on weather "prediction" as an initial value problem using NWP models. In addition, it has motivated climate "simulation" to obtain a time-mean state in the forced-dissipative climate system and climate "projection" to investigate the system responses to future change scenarios of carbon dioxide and others. In the 1960s, global three-dimensional atmospheric general circulation models incorporating visible solar radiation and infrared terrestrial radiation processes and climate models that include water cycle processes and combine ocean and land surfaces were already developed. The advent

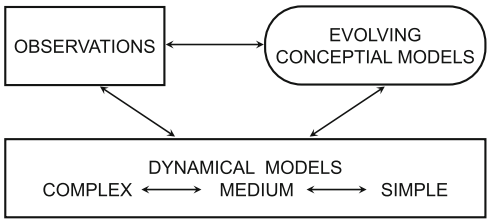

Fig. 15.2 A schematic illustration of the optimum situation for meteorological research. (Redrawn from Hoskins (1983))

and development of computers have had a tremendous influence on meteorology and climatology, even though the processing speed has always limited the total number of computations by an upper bound.

However, the progress of meteorology and climatology as computational sciences has faced some complex situations. At the time when vector supercomputers began to appear on the market (and when one of the authors, SY, obtained his degree and began numerical model studies as a full-fledged researcher), Hoskins (1983), in his New Year's address in the Quarterly Journal of the Royal Meteorological Society, recognized the unhealthy situation in meteorological research. The fields of meteorology and climatology, and earth sciences, in general, have consisted of observation (data analysis), theory, and numerical model studies utilizing computers. Coupled with the drastic advances in observation technology, including remote sensing from satellites, a large amount of atmospheric observation data began to accumulate. In addition, in every era, advanced numerical models produced numerous computer outputs. However, those large amounts of data were only diagnosed using classical conceptual models that had not changed in decades. Concurrently, theoretical studies were conducted without much relation to data from the real atmosphere or complex numerical models, contributing little to the diagnosis of specific phenomena.

In response to this unhealthy situation at that time, Hoskins showed Fig. 15.2 to stress the potential advantages of the optimum situation for meteorological research using a hierarchy of dynamical models. Here, sophisticated *complex* numerical models for weather prediction and climate simulation/projection fall at the high end of the model hierarchy with different complexities and are indispensable for quantitative descriptions. In comparison, *simple* models are useful for proposing new concepts and providing qualitative descriptions. In addition, *medium* models between them are frequently useful for understanding mechanisms. To provide a concrete example of a complex model, currently, a typical global NWP model has a horizontal resolution of about a few tens of kilometers and is a system of simultaneous ordinary differential equations with several billion dimensions (degrees of freedom) by discretizing each dependent variable. For accurate forecasting with small forecast errors, a refined NWP model with high resolution is necessary. In contrast, the three-variable ordinary differential equations of Lorenz (1963) form a simple model example, which was the basis of the discovery of chaos. It was used to present a new concept of the limit of prediction (i.e., predictability) by showing the sensitive dependence of irregular nonperiodic solutions on the initial condition. Examples of the medium class of atmospheric models include dry atmosphere

models excluding moist and precipitation processes and aqua-planet models, in which the entire globe is covered by oceans excluding land surface processes. Although these models are idealized and simplified model settings that do not exist in reality, they are effective in understanding the essence of the mechanism when used in combination with model results from other classes. Here, the classification is based on the model's degrees of freedom, which change with the advancement of computer performance; therefore, it should be based on the role of the model in each era. In the proposed optimal situation, theoretical conceptual models are evolved by the successful interaction of all classes of model research with each other and with the analysis of real atmospheric observations. Such theoretical concepts are useful for diagnosing the behaviors of the atmosphere and complex models. Observational, theoretical, and numerical model studies are expected to develop in close collaboration.

In the remainder of this section, we introduce the basic recognition of the climate impacts of solar activity variations from the viewpoint of the multiplicity of stable solutions in nonlinear systems, using the most simplified zero-dimensional energy balance model (Yoden 1996, 2011). Climate simulations using high-end state-of-the-art Earth system models and climate impact assessment studies of solar activity variations are described in detail in the subsequent sections.

Let T be the average temperature of the Earth–atmosphere system averaged over a year, and consider the energy balance of this system, focusing on the ice-albedo feedback. Let the net solar radiation absorbed by the system be $Ri\ (T)$ and the infrared radiation emitted from the system to the space be $Ro\ (T)$, then,

$$C\ dT/dt = Ri\ (T) - Ro\ (T)$$

where C is the heat capacity of the system, t is a time longer than the order of a year. Here, Ri is given by $Ri = So\ (1 - \alpha)/4$, where the solar constant $So = 1.37 \times 10^3$ Wm^{-2} and α is the albedo of the system. Although the albedo is considered to depend on T via cloud and surface conditions (e.g., snow and ice cover and vegetation distribution), we focus on the feedback from the increase or decrease in the snow and ice cover, which is considered to have the largest impact, in a simple form. It is assumed that the entire globe is covered with snow and ice, resulting in a high albedo ($\alpha = 0.7$) if the temperature is sufficiently low, whereas when the snow and ice disappear, the albedo is low ($\alpha = 0.3$) if the temperature is sufficiently high. Thus, $Ri\ (T)$ is assumed to be represented by a piecewise linear function, which is shown by a solid line in Fig. 15.3a. The cases where the solar constant decreases to 0.8 times and increases to 1.2 times the present value are shown by dotted and dash-dotted lines, respectively. Concurrently, the thermal radiation, Ro, from the system is assumed to be σT^4, where σ is the Stefan–Boltzmann constant for blackbody radiation. Here, Ro is linearly approximated as $Ro\ (T) = A + BT$, considering the effects of atmospheric absorption and emission and adopting values $A = -363$ Wm^{-2} and $B = 2.1$ $\mathrm{Wm}^{-2}\ \mathrm{K}^{-1}$ based on the analysis of satellite observation data (dashed line in Fig. 15.3a). The intersections of the solid and dashed lines in Fig. 15.3a are the steady solutions, and there are three equilibrium climates shown by ①–③ for the

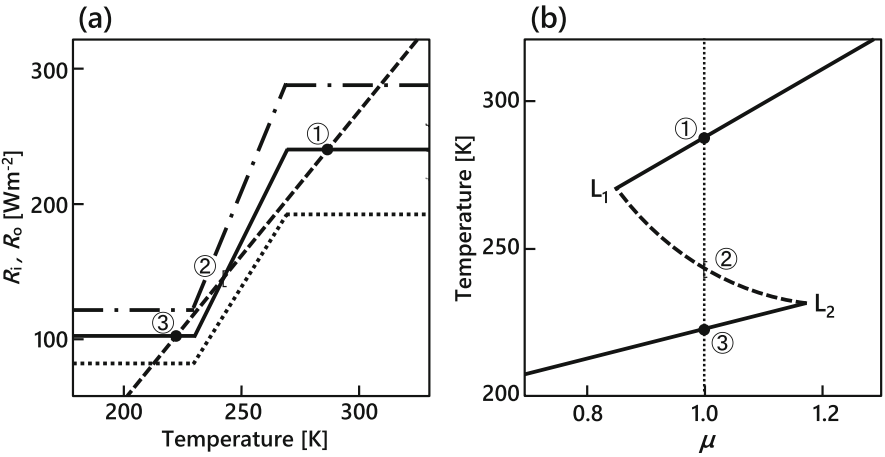

Fig. 15.3 (**a**) Temperature dependence of energy balance, with *Ri* (*T*) shown as a solid line and *Ro* (*T*) as a dashed line. Intersections ①–③ provide steady solutions. When the solar constant is set at 0.8- and 1.2-times its present value, *Ri* (*T*) is shown by dotted and dash-dotted lines, respectively. (**b**) Bifurcation diagram showing the dependence of equilibrium climate on the solar constant (μ*So*). Stable steady solutions are shown by solid lines and unstable ones by dashed line; ①–③ at μ = 1 correspond to the steady solutions in (**a**). (Redrawn from Yoden (1996, 2011))

present solar constant. Solution ① corresponds to the present climate ($\bar{T}$= 287 K), ③ is for the climate completely covered by snow and ice at $\bar{T}$= 222 K, and ② is a climate intermediate between these cases ($\bar{T}$= 243 K).

The linear stability of these steady solutions can be studied by considering a small perturbation, T', from each steady solution $\bar{T}$. At point ①, ②, or ③ in Fig. 15.3a, if the slope of the solid line is smaller than that of the dashed line, the solution is stable ($|dT'/dt| < 0$), whereas if it is larger, the solution is unstable ($|dT'/dt| > 0$). Solutions ① and ③ corresponding to the current climate and the globally freezing climate are stable, whereas solution ② is unstable. When a positive (negative) perturbation is applied to solution ① or ③ to increase (decrease) the temperature compared to the steady solution, negative feedback is applied to reduce the temperature perturbation. This is because the infrared radiation of the system is above (below) the absorption of solar radiation, resulting in net cooling (heating). In contrast, solution ② applies the opposite positive feedback, and the perturbation grows with time.

Here, we consider climate change when the solar activity changes and the solar constant becomes different (μ*So*) from the present value (μ ≠ 1). When the solar constant is set as a smaller value than the present (μ < 1) in Fig. 15.3a, the graph of *Ri* (*T*) moves downward from the solid line (the dotted line is for μ = 0.8). As the solar radiation decreases, stable solutions ① and ③ equilibrate at lower temperatures, whereas unstable solution ② equilibrates at higher temperatures. Solutions ① and ② approach each other and disappear simultaneously, and only solution ③ exists. In contrast, if the solar radiation increases (μ > 1), the stable solutions equilibrate at higher temperatures, solutions ② and ③ disappear, and only solution

① remains (the dash-dotted line is for $\mu = 1.2$). Figure 15.3b shows the dependence of the steady solutions on the solar constant, μSo, with μ as the bifurcation parameter. Branches of the stable steady solutions are shown by solid lines, and the branch of the unstable one is shown by a dashed line. At the limit point, L_1 ($\mu = 0.84$, $\bar{T} =$ 270 K), stable solution ① and unstable solution ② are generated in pairs. In addition, at another limit point L_2 ($\mu = 1.17$, $\bar{T} =$ 230 K), unstable solution ② and stable solution ③ vanish in pairs. In the range of $0.84 < \mu < 1.17$, there are three steady solutions, two of which are stable, which means that there is a multiplicity of stable equilibrium climates in this parameter range. Stable solutions ① and ③ equilibrate at higher temperatures with μ, and the gradient, $|d\bar{T}/d\mu|$, is larger for ① than that for ③. The results presented in this section, including the above figure, were obtained with a simple model, and are solely for qualitative conceptual descriptions. Their validity and quantitative examination require studies using medium and complex models, e.g., a one-dimensional energy balance model that retains the latitudinal dependence of the processes and three-dimensional atmospheric general circulation models and climate models, respectively (Yoden 1996, 2011).

15.2 Earth System Model Simulation of Climate Impacts of Solar Activity Variations

The radiative heating from sunlight must be included in numerical simulations for daily weather forecasts, long-term seasonal forecasts, and longer-term climate projections. This is because, apart from terrestrial heating, it is essentially only the solar radiation that heats the Earth, and without it, the Earth would cool down immediately. The heat balance of the Earth is maintained by absorbing solar radiation and emitting infrared rays (as seen with a simple model in the previous section). Moreover, greenhouse gases such as carbon dioxide and water vapor absorb the infrared rays and re-emit, thereby warming the lower layers of the atmosphere, and the Earth maintains the present temperature at which living things can survive. In addition, global warming, which has become a severe problem in recent years, is caused by the increase in greenhouse gas concentrations, which changes the distribution of atmospheric heating by the solar radiation and warms the lower layers of the atmosphere. Therefore, the solar radiation is indispensable for the simulation of the weather and climate on Earth. In addition, variations such as the solar 11-year cycle, 27-day cycle, and grand solar minimum (e.g., the Maunder minimum) can cause short- and long-term changes in the atmosphere and oceans. For example, simulations have been conducted to investigate how the climate would be affected by a significant solar minimum in a warming climate (see the next section). In addition, there have been recent studies on how the atmosphere is affected by the energetic particles accelerated by solar flares and coronal mass ejections (e.g., solar protons) and by galactic cosmic rays, the amount of which falling on the Earth

depends on the solar activity. Furthermore, if such solar activity variations can be predicted to some extent, it may improve weather/seasonal forecasts and climate projections. Thus, the incorporation of the solar activity variability into global weather and climate simulations is important and has various advantages. However, it is not as simple as prescribing the value of the total solar irradiance (TSI). In the next section, we explain how to incorporate solar activity variations into a simulation.

To incorporate solar radiation heating into a model, it is necessary to consider the spectral solar irradiance (SSI) for each wavelength, instead of only providing the TSI. This is because the radiative properties of each atmospheric component, such as absorption, scattering, and transmission, are wavelength-dependent. For example, the ultraviolet (UV) radiation is absorbed by ozone, and the near-infrared radiation is absorbed by water vapor and carbon dioxide. To simulate the top-down mechanism, in which the effects of solar activity variations on the climate reach the surface of the Earth through the stratosphere and the mesosphere, it is important to appropriately deal with the UV radiation heating. This heating is strongly affected by the atmospheric composition such as ozone in the stratosphere and the mesosphere. Although the line-by-line method, which is based on detailed wavelength-by-wavelength calculations of individual atmospheric components, is ideal, it requires an extremely large computational resource and cannot be used in climate models that resolve the entire globe over the long term. The radiative process is one of the costliest processes in the climate model, and thus, it is insufficient to only focus on accuracy; the computational efficiency should also be considered. A method such as the k-distribution method, which converts complex wavelength-specific absorption coefficient distributions into smooth probability densities and cumulative probability distributions for each absorption coefficient, is necessary to improve the efficiency and simplicity of the calculation (e.g., Sekiguchi 2011; Yabu et al. 2005). In addition, the radiative processes include the reflection, scattering, transmission, and absorption of the solar radiation by clouds and aerosols as well as the reflection and absorption on the ground surface, such as by snow, glaciers, and sea ice.

Moreover, solar activity variations affect not only heating by the radiative processes but also the photochemical reaction processes by UV and visible light. The photochemical processes involved in the production and loss of ozone in the stratosphere and the mesosphere (ozone chemistry) are typical examples of such processes requiring detailed wavelength-dependent solar radiation. If the ozone concentrations in the stratosphere and the mesosphere are not reproduced correctly, the distributions of the radiative heating and the associated temperature will be biased, and the climatic effects of solar activity variations would be reproduced inaccurately by the top-down mechanism. In addition, the model needs to reproduce the concentrations of many atmospheric trace constituents related to the ozone chemistry and the various chemical reactions among them, as well as the atmospheric temperature and circulation. The latter two are important for reproducing chemical reactions, mass transport, and even the emission of precursors such as ozone-depleting substances.

In recent years, solar activity variations have begun to be incorporated into climate models, including solar protons, medium-energy electrons (here, we consider the 30–1000 keV range, which has a strong impact at altitudes of 60–90 km), and the chemical effects of cosmic rays. To deal with these factors appropriately, it is necessary to use numerical models that implement not only the photochemical processes in the neutral atmosphere but also the complex chemical processes including ion species that are particularly important at higher altitudes, including the thermosphere. However, it is difficult to deal with these processes in climate simulations, which have some fine spatial resolution and long timescale calculations on the order of 10–100 years, owing to the high computational load. Currently, the commonly used methods for atmospheric chemical processes in climate models are simplified methods such as parameterizations to generate nitrogen oxide radicals NO_x (NO, NO_2) and hydrogen oxide radicals HO_x (OH, HO_2) using the ion-pair production rate of the atmosphere as input (e.g., Mathes et al. 2017). According to Porter et al. (1976), an ion-pair generated by some energetic particle produces nitrogen 1.25 N, approximately 45% of which is in ground-state N (4S) and approximately 55% of which is in an excited state. N (4S) can produce NO_x and its chemical family including its reservoir NO_y by the chemical reaction processes incorporated in the model. It is also assumed that N (2D) is rapidly converted into NO. Because the model requires the inclusion of very complex ion chemical processes to deal with HO_x in detail, we use a simple parameterization that provides the HO_x production rate as a function of the altitude and the ion-pair production rate (Solomon et al. 1981; Jackman et al. 2005). The typical top altitude of recent climate models is near the mesopause (80 km), and these models cannot represent the downward advection of NO_y produced in the lower thermosphere into the mesosphere. This problem can be addressed using a numerical model that incorporates detailed ionic chemistry processes or by prescribing NO_y concentrations or NO_y fluxes generated separately by an empirical model as the upper boundary conditions of the climate model.

An example of a climate impact simulation of solar activity variations is presented. Figure 15.4 shows the climate responses to 11-year cycle solar variations using numerical experiments that reproduce the historical climate variability from 1850 to 2005 (Mitchell et al. 2015). The numerical experiments are conducted using models that participated in the fifth phase of the Coupled Model Intercomparison Project (CMIP5), an international project that compares the results of state-of-the-art climate models at that time. The 13 models used here are high-top models (i.e., model top is equivalent to or higher than the stratopause at approximately 50 km), which are considered to accurately represent the upper atmosphere. The lower right panel of Fig. 15.4, labelled "reanalysis," corresponds to a model experiment assimilated by observations and provides additional physically consistent information that cannot be covered by observations. In the reanalysis, a positive temperature anomaly of approximately 1 °C appears over the entire stratopause (1 hPa). Significant positive temperature anomalies are also simulated in most climate models; however, they are smaller in magnitude (note that this includes the effects of the differences in the data period). In addition, because the latitudinal gradient of the temperature

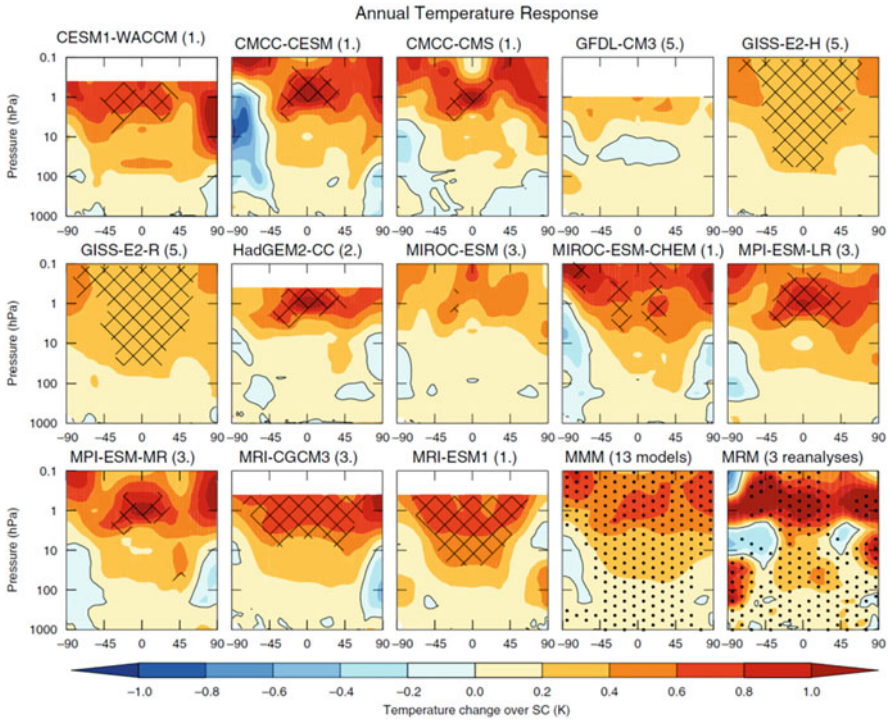

Fig. 15.4 The annual temperature response due to the 11-year solar cycle for each high-top model participating in the CMIP5 and the multi-model mean (MMM). Data used are from 1850 to 2005. Hatched areas for the individual models show 95% significance. Bottom right panel shows the multiple reanalysis mean (MRM) for 1979–2009. Stippled areas on MRM represent that all reanalyses have responses of the same sign. For MMM, stippled areas denote where more than 80% of the models have regression coefficients of the same sign. (Reprinted under the Creative Commons Attribution 4.0 License from Mitchell et al. 2015 © Author(s) 2015)

anomalies differs between models, there are also differences in the responses of the zonal winds, and no agreement is found between the models.

The degree of accuracy with which the above-mentioned processes are incorporated depends on the model. For example, the number of wavelength divisions of solar radiation significantly varies from 6 to 23, only approximately seven models treat the atmospheric ozone chemistry in detail by interacting it with atmospheric circulation and radiation, and only four models reproduce the stratospheric quasi-biennial oscillation. The latter is considered important for the climatic effects of solar activity variations. However, the precision of the handling of these processes does not necessarily correspond to the reproducibility of the climate responses of the models. Thus, the basic performance of a climate model, such as the reproducibility of the atmosphere–ocean circulation, is also considered to be an important factor. Figure 15.5 shows the polar ozone variability calculated by the Earth system model, MRI-ESM2.0, of the Meteorological Research Institute, which participated in the

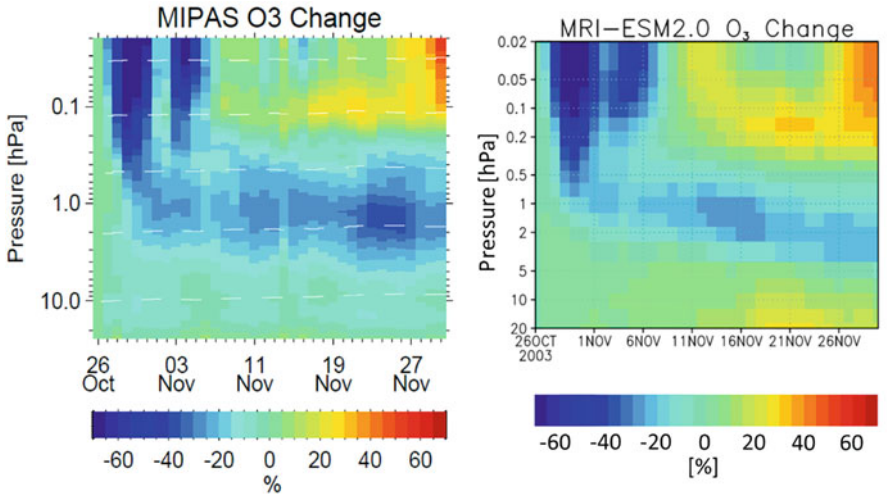

Fig. 15.5 Arctic (70–90°N) ozone change rate during the 2003 Halloween event, showing difference from October 26. (left) MIPAS satellite observation (reprinted under the Creative Commons Attribution 3.0 License from Funke et al. 2011 © Author(s) 2011), (right) three-member ensemble average of results from simulation (CMIP6-AMIP) given observed sea surface temperatures, by MRI-ESM2.0

sixth phase of the CMIP (CMIP6), during the Halloween event at the end of October 2003, when a large energetic particle precipitation occurred. On the left side of the figure, the ozone variations observed by the MIPAS satellite are presented (Funke et al. 2011), which show a large ozone depletion at the end of October. The right panel shows the results of MRI-ESM2.0, using the simple parameterization of NO_x and HO_x described in the previous section, which reproduces well the ozone variations observed by the satellite, including the magnitude, during the Halloween event.

15.3 Climate Impact Assessment of Solar Activity Variations by Recent Studies

The Intergovernmental Panel on Climate Change (IPCC) was established in 1988 by the World Meteorological Organization and the United Nations Environment Programme to comprehensively assess anthropogenic climate change, impacts, adaptation, and mitigation measures from scientific, technological, and socioeconomic perspectives. The IPCC publishes assessment reports approximately every 5–6 years as well as special reports as appropriate. The fifth assessment report (AR5) was published in 2013, and the sixth one is scheduled to be published. Associated with the IPCC assessment reports, the World Climate Research Programme and the

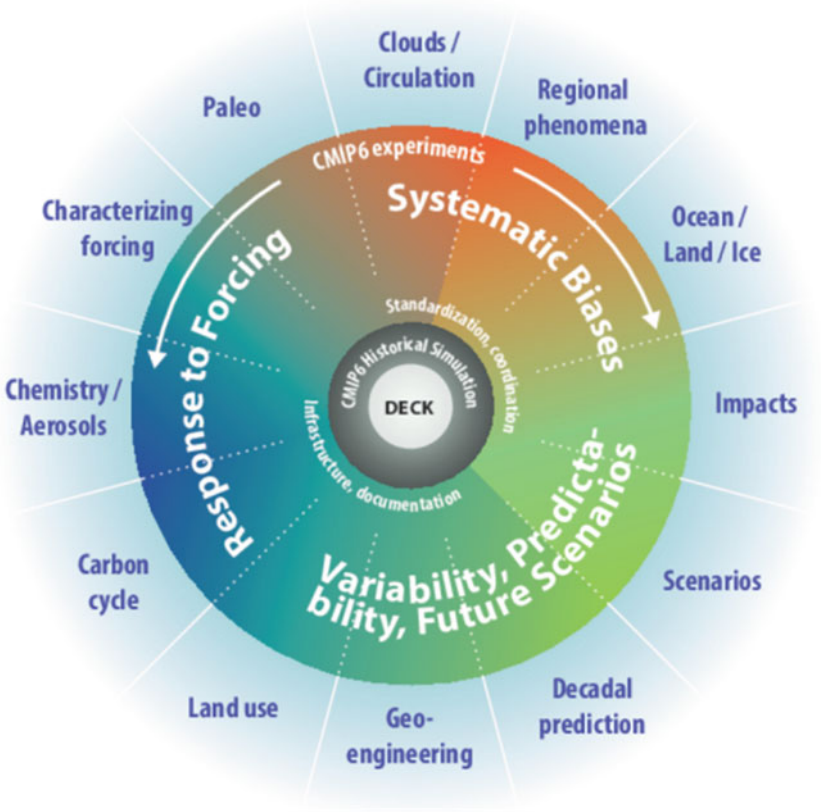

Fig. 15.6 Schematic of CMIP6 experimental design with basic CMIP, DECK, and historical experiments shown inside and the utility of DECK experiment in white text around it. More specific scientific topics to be addressed by detailed model intercomparison are shown in blue text outside. White text outside denotes each scientific topic (forced response, variability/predictability, and systematic bias in models) intended to be addressed. (Reprinted under the Creative Commons Attribution 3.0 License from Eyring et al. 2016© Author(s) 2016)

Working Group on Coupled Models are leading the CMIP, in which climate model centers from around the world are participating. In addition, experiments to assess climate change using state-of-the-art climate models are conducted under a common protocol (Fig. 15.6). By conducting similar climate change assessment experiments with multiple models, the differences in the results caused by the models (model uncertainty) are carefully evaluated by model intercomparison to enhance the reliability of the results. The results of the CMIP experiments are actively used in the IPCC assessment reports. As of 2020, CMIP6 is underway, and experiments have almost been completed in each model center and are being analyzed by researchers globally.

The AR5 contains summaries of the findings from previous studies on the relationship between climate change and solar activity variations. The following are excerpts from the AR5.

The changes in the global mean surface temperature during the solar maximum and minimum periods are estimated to be of the order of 0.1 °C based on regression studies and forcing estimates. However, some argue that this value is actually even smaller owing to estimation methodologies, and in fact, climate models estimate this change to be less than half the value. Therefore, it is extremely unlikely that the contribution of the solar forcing to the observed global warming since 1950 has been larger than that of greenhouse gases. In addition, although the uncertainty of future changes in the solar activity levels is high (very low confidence), the impacts of solar activity variations on anthropogenic climate change over the range of the solar activity maxima and minima are quite small (high confidence).

From a regional perspective, the variability associated with the 11-year solar cycle has been shown to cause short-term and seasonal climate anomalies, particularly in the Indian Ocean, Pacific Ocean, north Asia, and North Atlantic regions (medium evidence). However, some uncertainty remains owing to the difficulty in distinguishing it from other variability signals such as the El Niño/Southern Oscillation (ENSO). In addition, Ineson et al. (2015) suggest from climate model sensitivity experiments that the negative phase of the North Atlantic Oscillation (NAO) is more frequent during solar minimum periods. Moreover, there is some observation evidence including an increased frequency of blocking anticyclone events over Europe in winter.

Regarding the cloud formation associated with the production of new aerosol particles by cosmic rays, the AR5 stated as follows: “Cosmic rays enhance new particle formation in the free troposphere, but the effect on the concentration of cloud condensation nuclei is too weak to have any detectable climatic influence during a solar cycle or over the last century (medium evidence, high agreement). No robust association between changes in cosmic rays and cloudiness has been identified. In the event that such an association existed, a mechanism other than cosmic ray-induced nucleation of new aerosol particles would be needed to explain it.”

The status of the latest research since the AR5 is described below.

For future climate projections, climate model simulations have been used to investigate the extent of the global warming modification caused by grand solar minimum events, such as the Maunder Minimum (e.g., Meehl et al. 2013; Ineson et al. 2015; Maycock et al. 2015). Although the results depend on the assumed weakening of solar activity, a 0.25% reduction in the TSI in the U.S. Earth system model, CESM1, which includes interactive atmospheric chemistry processes, caused a global mean surface temperature decrease of approximately 0.2 °C. However, the temperature decrease only occurred when the minimum period was given, and after the recovery, the temperature was similar to the warming scenario without the minimum period (Meehl et al. 2013). In addition, a study using the UK climate model, HadGEM2-CC, found that a 0.12% reduction in the TSI resulted in an annual average reduction in the global mean temperature of 1.2 °C near the stratopause, whereas it caused only approximately 0.1 °C decrease near the surface (Maycock

et al. 2015). These changes depend on the wavelength. Comparing no change in UV at all to almost exclusively UV, for the same −0.13% TSI weakening, a change in the temperature reduction in northern winter Europe from −0.4 to −0.8 °C was shown. This corresponds to approximately 20–30% of the temperature difference in northern winter Europe at the end of the twenty-first century between the maximum greenhouse gas emission scenario (RCP8.5) and the medium stabilization scenario (RCP4.5) (Ineson et al. 2015). Cloud charging effects due to cosmic rays are not mentioned in the IPCC AR5; however, they have been reviewed by Mironova et al. (2015), who show, for example, that cloud edge charging has been observed. Moreover, although charging effects on water droplets with small particle sizes are theoretically expected, it was not concluded that these changes make a quantitatively important contribution to cloud formation and rainfall. It is necessary to monitor this research trend in the future.

Recently, two paleoclimate reconstruction experiments for the mid-Holocene (6000 years before) and the Last Glacial Maximum (21,000 years before) were conducted by introducing the variations in the orbital elements of the Earth due to the Milankovitch cycle and the carbon dioxide concentrations in those periods. The Meteorological Research Institute Earth System Model, MRI-ESM1, was used, and the model's performance was evaluated by comparing its results with the corresponding geological climate proxy data (Noda et al. 2017, 2018). It was shown that atmospheric chemical reaction processes, mainly stratospheric ozone photochemistry, significantly affect climate reproducibility. Moreover, it was recommended that paleoclimate reconstruction experiments use a distribution of the chemical composition (particularly ozone) that is consistent with the solar irradiance and the carbon dioxide concentration. The above is preferred over prescribing the same chemical composition distribution as a pre-industrial control as in the CMIP5 and CMIP6 protocols.

In addition, a recent study (Chiodo et al. 2019) has disputed the result that 11-year solar cycle variation significantly affects the NAO. The NAO is a major mode of the atmospheric circulation variability in the Northern Hemisphere winter mid- and high latitudes and was previously considered to be relatively reliable. The lack of such a signal prior to the mid-1960s and the stochastic appearance of periodic variations similar to the solar variation signal in the NAO without solar variations in simulations such as the global chemistry-climate model were the basis for the contradiction. This claim is under debate. It is possible that there was a problem with the model used or the method of analysis; however, this will be clarified by future research.

In CMIP6, many experiments have been conducted to evaluate climate change, for the IPCC AR6, the latest climate change assessment. These include a pre-industrial experiment (piControl), which assumes the pre-warming year of 1850, a 1pctCO_2 experiment, which increases the CO_2 concentration by 1% per year from piControl, and a CO_2 rapid increase experiment, which quadruples the CO_2 concentration. Other examples are a historical experiment, which reproduces the climate from 1850 to the present, and future experiments in various scenarios (the Shared Socioeconomic Pathways), which investigate the responses of the climate to multiple scenarios of future greenhouse gases and anthropogenic aerosol emissions

according to the level of global warming countermeasures. There are many other experiments related to the effects of solar activity variations, such as a historical natural-only experiment, in which only the natural variations in large-scale volcanic eruptions and solar activity variations are imposed. Another example is a historical solar-only experiment (hist-sol), in which only solar activity variations are imposed, which is expected to use a direct assessment of solar activity variations. There are two methods to evaluate the climatic effects of solar activity variations: assessment of the difference between the solar activity maximum and minimum periods, and extraction of periodic variations close to 11 years. However, these methods must be aware of the possibility of false detection of long-term large-scale variability signals such as the ENSO and large-scale volcanic eruptions. If the number of samples can be increased using long-term experimental data in which only solar activity variations are turned on and off, the possibility of false positives can be reduced. Paleoclimate experiments are also useful for solar impact assessment. As of 2020, the multi-model intercomparison analysis of CMIP6 is being actively conducted worldwide, which should add new knowledge on the climate impacts of solar activity variations.

References

Charney, J.G., Fjørtoft, R., von Neumann, J.: Numerical integration of the barotropic vorticity equation. Tellus. **2**, 237–254 (1950)

Chiodo, G., Oehrlein, J., Polvani, L.M., et al.: Insignificant influence of the 11-year solar cycle on the North Atlantic oscillation. Nat. Geosci. **12**, 94–99 (2019). https://doi.org/10.1038/s41561-018-0293-3

Eyring, V., Bony, S., Meehl, G.A., et al.: Overview of the coupled model intercomparison project phase 6 (CMIP6) experimental design and organization. Geosci. Model Dev. **9**, 1937–1958 (2016). https://doi.org/10.5194/gmd-9-1937-2016

Funke, B., Baumgaertner, A., Calisto, M., et al.: Composition changes after the "Halloween" solar proton event: the high energy particle precipitation in the atmosphere (HEPPA) model versus MIPAS data intercomparison study. Atmos. Chem. Phys. **11**, 9089–9139 (2011). https://doi.org/10.5194/acp-11-9089-2011

Hoskins, B.J.: Dynamical processes in the atmosphere and the use of models. Quart. J. R. Met. Soc. **109**, 1–21 (1983)

Ineson, S., Maycock, A., Gray, L., et al.: Regional climate impacts of a possible future grand solar minimum. Nat. Commun. **6**, 7535 (2015). https://doi.org/10.1038/ncomms8535

Jackman, C.H., DeLand, M.T., Labow, G.J., et al.: Neutral atmospheric influences of solar proton events in October–November 2003. J. Geophys. Res. **110**, A09S27 (2005). https://doi.org/10.1029/2004JA010888

Lorenz, E.N.: Deterministic nonperiodic flow. J. Atmos. Sci. **20**, 130–141 (1963)

Lynch, P.: The origins of computer weather prediction and climate modeling. J. Comp. Phys. **227**, 3431–3444 (2008). https://doi.org/10.1016/j.jcp.2007.02.034

Matthes, K., Funke, B., Andersson, M.E., et al.: Solar forcing for CMIP6 (v3.2). Geosci. Dev. **10**, 2247–2302 (2017). https://doi.org/10.5194/gmd-10-2247-2017

Maycock, A.C., Ineson, S., Gray, L.J., et al.: Possible impacts of a future grand solar minimum on climate: stratospheric and global circulation changes. J. Geophys. Res. Atmos. **120**, 9043–9058 (2015). https://doi.org/10.1002/2014JD022022

Meehl, G.A., Arblaster, J.M., Marsh, D.R.: Could a future "grand solar minimum" like the maunder minimum stop global warming? Geophys. Res. Lett. **40**, 1789–1793 (2013). https://doi.org/10.1002/grl.50361

Mironova, I.A., Aplin, K.L., Arnold, F., et al.: Energetic particle influence on the Earth's atmosphere. Space Sci. Rev. **194**, 1–96 (2015). https://doi.org/10.1007/s11214-015-0185-4

Mitchell, D.M., Misios, S., Gray, L.J., et al.: Solar signals in CMIP-5 simulations: the stratospheric pathway. Quart. J. R. Met. Soc. **141**, 2390–2403 (2015). https://doi.org/10.1002/qj.2530

Noda, S., Kodera, K., Adachi, Y., Deushi, M., Kitoh, A., Mizuta, R., Murakami, S., Yoshida, K., Yoden, S.: Impact of interactive chemistry of stratospheric ozone on southern hemisphere paleoclimate simulation. J. Geophys. Res. Atmos. **122**, 878–895 (2017). https://doi.org/10.1002/2016JD025508

Noda, S., Kodera, K., Adachi, Y., Deushi, M., Kitoh, A., Mizuta, R., Murakami, S., Yoshida, K., Yoden, S.: Mitigation of global cooling by stratospheric chemistry feedbacks in a simulation of the last glacial maximum. J. Geophys. Res. Atmos. **123**, 9378–9390 (2018). https://doi.org/10.1029/2017JD028017

Phillips, N.A.: The general circulation of the atmosphere: a numerical experiment. Quart. J. R. Met. Soc. **82**, 123–164 (1956)

Porter, H.S., Jackman, C.H., Green, A.E.S.: Efficiencies for the production of atomic nitrogen and oxygen by relativistic proton impact in air. J. Chem. Phys. **65**, 154–167 (1976). https://doi.org/10.1063/1.432812

Sekiguchi, M.: Modeling of radiative processes in GCMs. Meteorological Research Notes. **223**, 1–23 (2011)

Solomon, S., Rusch, D.W., Gérard, J.-C., et al.: The effect of particle precipitation events on the neutral and ion chemistry of the middle atmosphere. II - odd hydrogen. Planet. Space Sci. **29**, 885–892 (1981)

Yabu, M., Murai, T., Kitagawa, H.: Clear-sky radiation scheme. *Numerical Prediction Division Report, Supplement*, Forecast Division, Japan Meteorological Agency, **51**, 53–64 (2005)

Yoden, S.: Mathematical models of climate and climate variations (in Japanese), "*Climate Change Theory*", Iwanami Koza Earth and Planetary Sciences Vol. 11, Iwanami Shoten, 221–266 (1996)

Yoden, S.: Mathematical models of climate and climate variations (in Japanese), "*Climate Change Theory*", Iwanami Koza Earth and Planetary Sciences Vol. 11, Iwanami Shoten, 221–266 (2011)

Afterword

Kanya Kusano
Institute for Space-Earth Environmental Research, Nagoya University, Nagoya, Aichi, Japan

I am very pleased to publish this textbook with the collaboration of many people despite the various difficulties due to the pandemic of COVID-19 during this period. I would like to express my gratitude to them for their cooperation.

This textbook summarizes the achievements of the Project for Solar–Terrestrial Environment Prediction (PSTEP). However, it is also a message for the next generation of researchers who will address the problems that remain unsolved by the PSTEP. The PSTEP was planned based on the ideas discussed at the Geospace Environment Modelling System for Integrated Studies (GEMSIS) workshop held at Nagoya University in the summer of 2013 and subsequently implemented for 5 years starting in 2015. In the summary of the GEMSIS workshop, I raised the following points:

- The understanding of various sciences has advanced with the evolution from "empirical models" to "physical models" for predicting phenomena. Therefore, the synergy between "improvement in predictability" and "deepening of understanding" is the right direction of research that we should aim for.
- Space weather and space climate contain abundant sources of predictive issues, and therefore, they are excellent research topics as complex dynamical systems. To solve these problems, we should bridge the "valley of death" between forecasting operations and scientific research and strongly promote the synergetic interaction between the two. Furthermore, we should evaluate predictability objectively and quantitatively and reason scientifically regarding the inability to predict unpredictable phenomena.

K. Kusano (ed.), *Solar-Terrestrial Environmental Prediction*,
https://doi.org/10.1007/978-981-19-7765-7

– By promoting such activities, the predictability of the complex phenomena in space weather and even extreme space weather events should be significantly improved in the next decade.

These messages became the basic principles supporting the activities of the PSTEP, and I believe that they have resulted in a range of outcomes that are also described in this textbook.

However, our predictive power of the solar–terrestrial environment is still insufficient, and extreme space weather events such as the Carrington event in 1859 and beyond, which are certain to occur in the future, are potential risks that threaten modern society. The questions posed in 2013 are still valid as a guide for the future, and new research must be developed to resolve them.

I hope that this textbook will serve as a milestone in that direction. I sincerely anticipate that this textbook will help young people who are interested in various fields related to the solar–terrestrial environment to pioneer new research for the development of prediction and understanding.