ADVANCES IN GEOSCIENCES

Editor-in-Chief: Wing-Huen Ip (National Central University, Taiwan)

A 5-Volume Set

Volume 1: Solid Earth (SE)  
ISBN-10  981-256-985-5

Volume 2: Solar Terrestrial (ST)  
ISBN-10  981-256-984-7

Volume 3: Planetary Science (PS)  
ISBN-10  981-256-983-9

Volume 4: Hydrological Science (HS)  
ISBN-10  981-256-982-0

Volume 5: Oceans and Atmospheres (OA)  
ISBN-10  981-256-981-2

A 4-Volume Set

Volume 6: Hydrological Science (HS)  
ISBN-10  981-270-985-1

Volume 7: Planetary Science (PS)  

Volume 8: Solar Terrestrial (ST)  
ISBN-10  981-270-987-8

Volume 9: Solid Earth (SE), Ocean Science (OS) & Atmospheric Science (AS)  
ISBN-10  981-270-988-6
EDITORS

Editor-in-Chief: Wing-Huen Ip

Volume 6: Hydrological Science (HS)
Editor-in-Chief: Namsik Park
Editors: Chunguang Cui
Eiichi Nakakita
Simon Toze
Chulsang Yoo

Volume 7: Planetary Science (PS)
Editor-in-Chief: Anil Bhardwaj
Editors: C. Y. Robert Wu
Francois Leblanc
Paul Hartogh
Yasumasa Kasaba

Volume 8: Solar Terrestrial (ST)
Editor-in-Chief: Marc Duldig
Editors: P. K. Manoharan
Andrew W. Yau
Q.-G. Zong

Volume 9: Solid Earth (SE), Ocean Science (OS) & Atmospheric Science (AS)
Editor-in-Chief: Yun-Tai Chen
Editors: Hyo Choi
Jianping Gan
REVIEWERS

The Editors and Editor-in-Chief of Volume 8 (Solar Terrestrial) would like to thank the following referees who have helped review the papers published in this volume:

Graham Barnes
David Boteler
Iver Cairns
Marc Duldig
Ted Fritz
John Humble
Y. Kamide
Karl-Ludwig Klein
Axel Korth
Bo Li

P. K. Manoharan
Yoshizumi Miyoshi
Takahiro Obara
Bruce Tsurutani
J.-N. Tu
Gherardo Valori
Andrew W. Yau
Hui Zhang
Q.-G. Zong
This page intentionally left blank
CONTENTS

Editors v
Reviewers vii
Preface xi

From IGY to eGY: The Importance of Real-Time Data in Space Physics 1
Y. Kamide

A New Perspective on the Relationship Between Substorms and Magnetic Storms 25
B. T. Tsurutani and W. D. Gonzalez

Storm–Substorm Relationship: Controversies and Recent Development 47
T. Hori

Temporal Development of Dayside TEC Variations During the October 30, 2003 Superstorm: Matching Modeling to Observations 69
O. P. Verkhoglyadova, B. T. Tsurutani and A. J. Mannucci

Cutoff L-Values of Solar Protons in Comparison with Ring Current Protons During a Storm: NOAA/POES Observations 79

Geomagnetic Activity and Auroras Caused by High-Speed Streams: A Review 91
F. L. Guarnieri, B. T. Tsurutani, E. Echer and W. D. Gonzalez

Development of Pulsation Index for Space Weather 103
K. Kitamura, S. Watari and M. Kunitake
Contents

Atmospheric Neutral Analyzer for In-Situ Neutral Mass Composition and Velocity Distribution Measurements in Ionosphere–Thermosphere Coupling Studies 113
P. V. Amerl, E. P. King and A. W. Yau

Reconstruction of Nonlinear Force-Free Fields and Solar Flare Prediction 123
M. S. Wheatland

Solar Radio Fine Structures Detected with Super-High Temporal Resolution in Decimeter Waveband 139
S. J. Wang, Y. Y. Liu, Y. H. Yan and Q. J. Fu

Comments on the Observed Galactic Cosmic Ray Modulations in the Heliosphere 147

The Inverse Problem for Galactic Cosmic Ray Propagation in the Heliosphere Based on Neutron Monitor and Satellite Data 167
L. Dorman

Radiation Hazard from Large SEP Events for Aircraft: Monitoring and Forecasting by Using On-Line One-Min Cosmic Ray Data 189
L. Dorman

Monitoring and Forecasting of Radiation Hazard for Aircraft from Galactic Cosmic Rays 209
L. Dorman

Comparative Measurements of Cosmic Radiation Monitors for Aircrew Exposure Assessment 223
I. L. Getley, L. G. I. Bennett, M. L. Boudreau, B. J. Lewis, A. R. Green, A. Butler, M. Takada and T. Nakamura

Modeling of Aircrew Radiation Exposure from Galactic Cosmic Rays and Solar Particle Events 233
M. Takada, B. J. Lewis, M. Boudreau, H. Al Anid and L. G. I. Bennett
In order to honor the ground-breaking contributions to space physics and fundamental research in science by Prof. W. Ian Axford, the Asia Oceania Geosciences Society has dedicated its first named Distinguished Lecture Series to Prof. Axford. It is our great fortune that Ian has agreed to lend his name to grace the Society. Looking back, Ian’s distinguished scientific career was highlighted by his pioneering work on the magnetospheric convection system in connection with the viscous interaction between the solar wind and magnetic field reconnection. This early achievement was followed by the proposal of the wind-shear mechanism in the formation of the ionospheric sporadic E layer, the far-reaching theoretical work on the origin of the solar wind, and last but not least, the transport of the solar and galactic cosmic rays in the heliosphere.

The origin and acceleration of the high-energy galactic cosmic rays are problems still very dear to Ian’s heart. Frankly speaking, any one of these achievements would have won a scientist ever-lasting fame. And Ian has done them all. In addition, Ian’s foresight and leadership had brought him to Europe from sunny California. From that point on, Ian had helped the European Space Agency to initiate a number of very interesting space and planetary projects on which basis Europe has grown to be a major player in this arena. His effort in injecting new life to the European Geophysical Society is a well-known story. True to his character, since returning to New Zealand, Ian has pushed forward the idea of establishment an Asia-based organization for Earth Science with a view to share the limelight with American Geophysical Union (AGU) and European Geosciences Union (EGU). The outcome is the society called Asia Oceania Geosciences Society as we know it now. At the very beginning, this possibility existed only in the mind of one other person, namely, Professor Y. Kamide who is himself an outstanding space physicist of unsurpassed international reputation. As far as I can tell, their dialogue on AOGS started in 2002 when Ian visited the Nagoya University where Prof. Kamide served as the Director of the famous Solar-Terrestrial Laboratory. It is therefore most appropriate that
Preface

the first Axford Lecture, which appears as the first paper in this volume, was given by Prof. Kamide on a topic chosen by Ian himself. It is our wish that, from now on, the Society can call upon the Axford Lecturers to serve as the role models for our young scientists. From now on, AOGS will share such legendary figures and pioneering spirit, and call them our own.

Professor Wing Ip
Inaugural President
Asia Oceania Geosciences Society
Advances in Geosciences
Eds. Marc Duldig et al.
© World Scientific Publishing Company

FROM IGY TO eGY: THE IMPORTANCE OF REAL-TIME DATA IN SPACE PHYSICS∗

Y. KAMIDE
Solar-Terrestrial Environment Laboratory, Nagoya University
Nagoya, Aichi 464-8601, Japan
kamide@stelab.nagoya-u.ac.jp

Reviewing the history of space physics, particularly of solar-terrestrial relationships, over the last 50 years from its birth at the International Geophysical Year (1957–1958), this paper emphasizes the importance of collecting data from space and from around the world and processing these data, all in real time. These real-time data are essential to run computer simulations under realistic boundary conditions and thus to make space weather predictions. The availability of real-time data combined with real-time simulations not only changes the practical style of research but also changes the concept of solar-terrestrial physics in such a way that it becomes increasingly imperative to understand the whole spectra of processes, especially coupling between different plasma regions and coupling between large-scale and small-scale processes. It is quite timely that the scientific community is planning to conduct the Electronic Geophysical Year, because this international enterprise is harmonized with the rapid growth in modern networking technologies and high-quality observation techniques.

1. Introduction

When the author entered the graduate school of the University of Tokyo, one of the first subjects he learned about the Earth’s magnetosphere and its dynamics was through the Axford/Hines theoretical model of plasma convection. Figure 1 shows a schematic illustration of the convection model Axford and Hines1 published in 1961. They proposed that plasma convection in the magnetosphere must be driven through viscous-like interactions between the solar wind and magnetospheric plasmas, resulting in two large-scale vortices centered in the dawn and dusk hours of the

∗Based on the First Axford Lecture, delivered on July 10, 2006 at the occasion of the Third Annual Meeting of the Asia and Oceania Geosciences Society held in Singapore.
Fig. 1. Streamlines of plasma convection in the equatorial plane of the magnetosphere, proposed by Axford and Hines.\textsuperscript{1} After Axford.\textsuperscript{2}

magnetosphere. It was suggested that the driving force of large-scale convection motion is associated with the tangential stress exerted by the solar wind on the magnetospheric surface, which was called viscous-like interactions between solar wind and magnetospheric plasmas.

Figure 2(a) is a projection of the convection pattern onto the polar ionosphere, consisting again of two vortices.\textsuperscript{3} Figure 2(b) is the convection pattern in the ionosphere calculated primarily from ground-based magnetometer data using a magnetogram-inversion technique (see also Fig. 9). In both plots the so-called twin-vortex patterns consisting of high- and low-potential vortices in the morning and evening sectors,
Fig. 2. (a) Projection of magnetospheric convection onto the polar ionosphere. Points A and B can be found in Fig. 1. After Axford.² (b) Example of the electric potential distribution calculated from real-time ground magnetometer data through the KRM algorithm.

respectively, can clearly be identified, although some deformations in Fig. 2(b) represent individual complications. Some of the deformations originate from inhomogeneities in the ionospheric conductivity, indicating that the ionospheric convection is not a simple geometrical projection of magnetospheric convection along magnetic field lines. In a sense, what many scientists in the field have been doing over the last 40 years is chasing the Axford/Hines convection model by different means, such as radar, satellite, and ground-based measurements.

The present author is one solar-cycle younger than Axford. That is, when Axford began his doctoral studies in 1957 at the beginning of the International Geophysical Year (IGY), he was an elementary school boy in Sapporo, Japan. Fifty years ago, as he was about to enter a junior high school, he was fascinated by spectacular natural phenomena and was trapped by a number of questions surrounding them, such as Why ocean waves get excited even without strong winds? What is the cloud formation mechanism generating different shapes? How light from the Sun travels the long distance to the Earth? and How auroras are formed? During the course of his “elementary school and junior high school” years, which coincided with the IGY, he faced many interesting articles in newspapers and science magazines about space and Antarctic expeditions: TV was not popular at all at that time. Exciting to a young boy, these did bring much new
knowledge about solar-terrestrial relationships into his world. For example, he was very curious about auroral displays. Why are auroras most active on the nightside when the Sun is on the dayside? It was about that time that he, more or less, decided that he wanted to be an Earth and space research scientist when he grew up.

Looking back over the history of space physics, this paper presents the author’s personal view on the importance of conducting well-organized space-ground observations and of handling the resulting data products in real time. The author does not pretend that this paper is unbiased toward the author’s own experience, but he tries to recount some of the major accomplishments which have led to our present ability of specifying, as well as forecasting, the state of the magnetosphere–ionosphere system on the basis of solar wind observations.

2. Before IGY

“Before IGY” may include many different meanings, depending on how far we go back before the IGY. Of course the author has no intention of taking the readers back to Aristotle’s discovery of auroral breakup or the finding of sunspots in China. These occurred well prior to the birth of science in the present sense.

It may be quite interesting to seek what the beginning of solar-terrestrial relationships studies was. In tracing the history of the field, it can be found that studies of geomagnetic records and solar and auroral observations were conducted in parallel with discoveries of the important laws of electromagnetism (see Fig. 3). In other words, the study of geomagnetic disturbances was one of the areas which contributed significantly to the development of electromagnetism in general, leading to Maxwell’s equations, which were systematically organized in 1873.

It is well known that Graham in London, and Hiorter and Celsius in Uppsala, Sweden, noticed that large disturbances in the geomagnetic field occurred simultaneously at these two locations. Gauss and Weber later showed these magnetic disturbances to be a worldwide phenomenon, such that magnetic “weather” is much less local than ordinary weather on the Earth’s surface or in the troposphere. It was in September 1859, following Carrington’s observation of a group of giant white light flares on the Sun, that a great geomagnetic storm began. These are some of the examples which indicate that the root of studies of solar-terrestrial relationships goes back to the 18th and 19th centuries.
The Beginning of Solar-Terrestrial Physics

1716
The discovery of connection between geomagnetic disturbances and auroras by Halley

1724
The discovery of geomagnetic storms (later term) by Graham

1807
Examination of magnetic thunderstorms by Humboldt

1821
The law of Ampère’s force

1820
The discovery by Oersted that electric currents produce magnetic forces

1831
Electromagnetic induction by Faraday

Fig. 3. Some of the early important findings and laws in electromagnetism in the 18th and 19th centuries. These led to Maxwell’s equations of electromagnetism.

Who introduced the terminology “geomagnetic storm”, and when, into the scientific community? It is known that Alexander von Humboldt (1769–1859) used “magnetisches Ungewitter” (magnetic thunderstorms in English) to describe the variability of geomagnetic needles, which were associated with the occurrence of “light meteor” (i.e., auroras). Humboldt realized that geomagnetic disturbances and auroras are two manifestations of the same phenomenon. He maintained a lifelong interest in geomagnetic disturbances, establishing magnetic stations around the world through his diplomatic contacts at, for example, Bombay, Toronto, and Sitka.

It was during the First Polar Year (1882–1883) that scientists defined “geomagnetic storms” as “intense, irregular variabilities of the geomagnetic field which occur as a consequence of solar disturbances”. It was still unclear how intense was intense and how irregular was irregular, but it was worth noting that the Sun was identified as the cause of geomagnetic storms.

3. Research Style Near and During IGY

To obtain more complete knowledge of planet Earth, the member countries of IGY worked together to set up a number of stations on the Earth’s
Y. Kamide

surface, and even to establish World Data Centers, with the notion that data obtained from the observations during IGY are precious assets common to humankind. All Earth Science disciplines including space physics were involved with this worldwide campaign. The IGY officially adopted the use of artificial satellites to explore the Earth’s upper atmosphere. In fact, the Soviet Union and the United States of America successfully launched Sputnik 1 and Explorer 1 into orbits around the Earth in 1957 and 1958, respectively, opening the so-called space era. The younger generation may not believe, however, that at the time of IGY, and even throughout the author’s graduate studies, scientists used microfilmed data instead of digital data: thus making copies of data and scientific papers in journals meant for the scientists to take their photographs, of course not in color. Computers were not available, so that numerical calculations had to be made by hand, even for quite complicated ones.

Comparing what scientists knew 50 years ago about how geomagnetic disturbances were generated with what we know today about the dynamics of the magnetosphere within the entire Sun–Earth system, we find that there is a great difference between the two sets of pictures. This difference results partly from progressive changes in methodology over the last decades. The old technique scientists had long relied on was to measure magnetic disturbances on the Earth’s surface, monitoring the state of the electromagnetic environment before the concepts of the magnetosphere, solar wind–magnetosphere coupling, and magnetosphere–ionosphere coupling were introduced. After the 1960s, in-situ measurements of the plasma structure in the near-Earth space environment became available. Adding to our wealth of knowledge has also been the constant trial-and-error theoretical analysis of the space between the Sun and the Earth by our frontier scientists, such as Axford, Dungey, and Parker.

To realize the dramatic advance in technology for data handling over the last 50 years, two examples are shown in which it can easily be guessed how much time and energy scientists at that time must have spent for what we would now need only a few seconds. Figure 4, taken from Akasofu\(^6\) shows the distribution of auroral displays, which was constructed by using records from 11 all-sky cameras of the IGY network. For obtaining this type of “world” map, it is required first that all or most of the all-sky camera stations be cloudless. One all-sky camera is able to cover only a very small fraction of the entire auroral distribution (see the small circles in Fig. 4). Second, all or most of the all-sky cameras must be operating properly under freezing Arctic conditions. It was heard from engineers,
who were active during IGY, that polar bears at times knocked out their instruments. Finally, each of the auroral forms was manually mapped to the ionospheric level whose altitude was assumed to be 110 km. This was not a trivial work at all without digital data and computers. In this way, the world aurora maps were generated for a number of days by repeating the procedure for data obtained every 1-min, having led to the concept of auroral substorms. It is now possible for a satellite to picture the entire polar region almost instantly, regardless of the surface temperature and weather.

Figure 5, taken from Fukushima, is another 50-year-old example where the so-called equivalent current system in the ionosphere is mapped. Ground-based magnetometer records were used to obtain current streamlines in the ionosphere that can account for the observed geomagnetic perturbations at all stations. This is a good example of “easy to say but difficult to do practically”. That is, through a trial-and-error method
Fig. 5. Example of hand-drawn equivalent current system obtained for the maximum epoch of an intense substorm, which was then called as an elementary storm. A circle indicates the location of the Sun. After Fukushima.

by using a pencil and a rubber eraser, it was required to draw two-dimensional, divergence-free streamlines to satisfy the following conditions: the density of the streamlines is proportional to the magnitude of the magnetic perturbation and they are perpendicular to the observed magnetic vectors at all the points. This time-consuming process is equivalent to solving a second-order differential equation for magnetic potentials, from which one can obtain the equivalent current system in the ionosphere. It is, at present, a matter of seconds by a computer to complete the entire process, but at least several hours were needed to construct just one world map of the equivalent current system before Boström and Kamide et al. devised automated algorithms.

Without in-situ data from satellites, scientists of 50 years ago relied heavily on ground-based observations of magnetic fields and auroras, from which they only inferred physical parameters in space. The two examples clearly demonstrate that deducing useful data products or physical parameters from magnetic and auroral observations was extremely laborious and often required special personal techniques or talent.
Fig. 6. Two auroral image data from the Polar spacecraft. Dynamic features of auroras can be monitored, including the sudden creation of bright spots and their latitudinal and longitudinal expansion. Courtesy of M. J. Brittnacher and G. K. Parks.

Presently, not only ground magnetometer data but also data from various spacecraft are available in real time or near real time. The entire high latitudes can now be viewed from above the poles in different wavelengths, so that the global distribution of auroras and its dynamical changes can be monitored in real time. Figure 6 is one such example, in which two auroral photographs taken from the Polar spacecraft are shown. We can see progressive changes in the global distribution of auroras, particularly how bright spots generated in the midnight sector soon become “seeds” for a sudden expansion of auroral activity in 10 min.

Not only data from near-Earth satellites but also those from the upstream solar wind are available in real time. Figure 7 shows an example of 24-h real-time data from the ACE satellite, located at the L1 point between the Sun and the Earth. Note that in these real-time data, the last point at each quantity is not exactly of real-time observations but represents observations of 5 s ago, for the light speed to travel from the satellite to the Earth.

4. From Discoveries to Predictions

Progress in any fields of natural science begins with “discovery” and ends with “proof”. Space physics is not an exception. “Proof” in our research area, however, may be reworded by “prediction or reproduction” of solar-terrestrial phenomena under study. The entire process in research can further be classified into several practical steps. For example, the
Fig. 7. Example of real-time data of the solar wind and the interplanetary magnetic field, from the ACE satellite. Adapted from the NOAA/SEC homepage.

research procedure for individual solar-terrestrial phenomena, such as magnetospheric substorms, which are the basic energy process in the solar wind–magnetosphere–ionosphere system, normally consists of the following four sub-steps:

1. Ascertaining characteristics of a phenomenon through observations,
2. Introducing a hypothesis to account for the phenomenon,
3. Testing repeatedly the hypothesis (or model) against observations,
4. Evaluating the model quantitatively.

In the field of solar-terrestrial relationships, the predictability of solar-terrestrial phenomena is one of the key factors for evaluating a hypothesis. It is possible to predict, at present, the occurrence of an intense magnetospheric substorm with perhaps more than 80% accuracy if (1) the solar wind variations are relatively simple and (2) accurate information about the solar wind variability is given in real time. This does not mean at all, however, that we are currently 80% successful in predicting the
occurrence of magnetospheric substorms, because the solar wind variation is almost always complicated.

The advance in research of the Earth’s magnetosphere thus far has also been achieved through repeating these four sub-steps. The magnetosphere, in which the Earth’s magnetic field dominates, was first discovered through spacecraft observations. The magnetotail and then the plasma sheet and other plasma domains were subsequently discovered, and the nature of these plasma regions was ascertained. Concurrent with these discoveries, extensive efforts were made to search for the mechanism of the entry of solar wind energy into the magnetosphere and to locate the energy reservoir for substorms. A number of theoretical models about the generation mechanism of substorms, i.e., the sudden release of accumulated energy, were proposed, and repeatedly tested against new data from spacecraft and ground-based observations. Computer simulations were also conducted to try to reproduce the main features of substorms. Prediction codes were evaluated according to their efficiency of predicting the occurrence and intensity of substorms.

Table 1 summarizes the development of magnetospheric physics over the last 50 years in three phases. Major discoveries are identified for each phase. Note that the list for each phase comes from the point of view of substorm research, and that different lists would result if we deal with

<table>
<thead>
<tr>
<th>Phase 1: Explorer and discovery of the magnetosphere</th>
</tr>
</thead>
<tbody>
<tr>
<td>Property of the solar wind</td>
</tr>
<tr>
<td>Average configuration of the magnetosphere</td>
</tr>
<tr>
<td>Magnetopause, bow shock, and magnetotail</td>
</tr>
<tr>
<td>Plasma regions in the magnetosphere</td>
</tr>
<tr>
<td>Plasma sheet, lobe, ring current, and acceleration region of auroral particles</td>
</tr>
<tr>
<td>Dynamics of the magnetosphere</td>
</tr>
<tr>
<td>Plasma convection and substorm-related changes</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase 2: Understanding of solar-terrestrial processes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solar wind control of substorms</td>
</tr>
<tr>
<td>Intensity, location, and probability of substorms</td>
</tr>
<tr>
<td>Configuration changes of the magnetotail</td>
</tr>
<tr>
<td>Plasma sheet thinning/thickening</td>
</tr>
<tr>
<td>Ring current formation</td>
</tr>
<tr>
<td>Electrodynamics of the ionosphere</td>
</tr>
<tr>
<td>Computer simulations</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Phase 3: Predictions — Space weather applications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Now-cast of the state of the magnetosphere</td>
</tr>
<tr>
<td>Modeling of the entire the solar wind–magnetosphere system</td>
</tr>
</tbody>
</table>
different research disciplines, such as ionosphere–thermosphere coupling. Note also that substorm research is a result of a convergence of multidisciplinary sciences, which developed from several traditional fields of research, such as geomagnetism, auroral physics, and aeronomy.

It was during the first phase when the solar wind and the magnetosphere were observed. Explorer 10 explored the nightside magnetosphere, i.e., the magnetotail boundary, as early as 1961. Without repeated crossings over the boundary, it would be impossible to discern the entire magnetotail. It was the IMP 1 spacecraft, which discovered that the Earth’s magnetic field is consistently distorted by the solar wind. The bow shock, the magnetosheath, and the magnetopause were later identified. Figure 8, from an early paper by Ness, shows the initial results of IMP 1, mapping the positions of the magnetopause. IMP 1 also surveyed the lobe structure and length of the magnetotail, as well as the response to substorm activity. This series of important discoveries about the nature of the magnetosphere occurred in the midst of extensive theoretical discussions among theorists.

Subsequently, the configurations of both the dayside and nightside magnetosphere were modeled. The presence of solar wind particles within the magnetosphere was confirmed. Changes of each plasma region at the expansion onsets of substorms were identified. Along with ground-based observations of auroral dynamics and magnetic perturbations in the polar region, satellites began to assemble data for changes in the magnetotail,

Fig. 8. IMP 1 and Explorer observations of the magnetospheric boundaries, generated by solar wind–magnetosphere interactions. After Ness.
such as the large-scale topology changes in fields and plasma flow associated with substorms.

In the second phase, which may be called the phase of understanding of solar-terrestrial processes, integration of satellite and ground-based observations became essential. It was found that nonlinear coupling occurs in each of the plasma regions. More and more, computer simulations became a powerful tool for quantitative understanding of such complicated processes. Substorms involving both suddenly initiating and rapidly growing auroral breakups are typical examples of such nonlinear processes. Our dream is to input current solar wind data into the computer, which will then tell us how, when, and where substorms of what magnitude will take place.

We are now entering the third phase, where physical relationships among various phenomena occurring in different regions, described in detail by basic equations, must be established. This is quite important not only toward a complete understanding of the entire solar-terrestrial system in all scale sizes, but also for space weather predictions, which are presently societal needs.

5. Real-Time Observations and Calculations

Over the last 50 years, the scientific community has made drastic changes in the amount of data that space physicists must deal with. The rate of the increase in the data still keeps increasing at this moment. Are we able to continue to handle such an increase for years to come? The answer is: “Probably yes”, or “Yes, at least up to the near future”. Fortunately, the rapid increase in the quantity of data we must handle has paralleled the rapid growth in modern computing technologies. The development in technologies has coincided with the urgent need to integrate observations and computer simulations in solar-terrestrial physics. Observations need solid backup from theories, while theories need realistic boundary conditions resulting from observations to run models.

While observations from the Earth’s surface can be considered as a type of “remote sensing” for solar-terrestrial processes and thus “indirect”, they nevertheless generate high spatial/temporal resolutions. On the other hand, satellite observations, being in-situ and thus “direct”, provide only “point” measurements along satellite orbits. Clearly, these are complementary to each other.

The Solar-Terrestrial Environment Laboratory (STEL), installed a computer system, called Geospace Environment Data Analysis System
(GEDAS), to promote integrated studies of ground- and satellite-based observations as well as modeling and simulation research. GEDAS is not a simple display system for real-time data. In the GEDAS system, real-time data are used as initial and boundary conditions for computer simulations to execute in real time, permitting us to specify the state of the near-Earth plasma environment and to predict space weather events that may occur or how the substorm just begun grows in the next several minutes. The correct answer will be known in several minutes when real-time data flow into the system. If this type of test of a particular simulation model is not satisfactory, i.e., failing to predict what is coming next, one must modify the existing code and/or change the parameters within a realistic degree, such that the revised simulation provides better predictions for the next space weather events. This revision cannot be ad hoc to try to account for only the next real-time observations, but must be universal, leading to deepening of our basic knowledge of the solar-terrestrial processes.

As one of the active projects currently underway at GEDAS, we collect ground magnetometer data on a near real-time basis in an attempt to compute the instantaneous, two-dimensional distribution of ionospheric parameters, such as ionospheric electric fields and currents and field-aligned currents. This joint effort of STEL and NOAA’s National Geophysical Data Center (NGDC) uses operationally updated versions of the KRM and AMIE programs.\textsuperscript{18–20}

In this real-time project, the uneven distribution of ground magnetometers is one of the inevitable problems. The AMIE code, along with solar wind observations, is first used to estimate the global distribution of the electric potential, which represents a statistical pattern of the ionospheric potential, commensurate with the solar wind condition. Once the global pattern is calculated, that pattern is used as the boundary condition to calculate more detailed structures of ionospheric parameters in a limited region, where ground magnetometer data are available from a number of observatories in real time, on the basis of the KRM algorithm.

Figures 9(a) and 9(b) present examples of the distribution of the electric potential and the corresponding currents in the ionosphere, calculated in near real time. It is clearly seen that deformations in the usual twin-vortex potential pattern, which Axford envisaged, and strong auroral electrojets are reproduced. Comparing these real-time weather maps with Fig. 5, it is noticeable that there is an enormous
change in our knowledge about the ionospheric quantities over the last 50 years. All one could obtain 50 years ago was the equivalent ionospheric current system from ground magnetometer data, not “true” currents.

We calculate presently these ionospheric parameters every 10 min (see http://gedas22.stelab.nagoya-u.ac.jp/index.j.html). GEDAS provides the scientific community with specification of the geospace environment well beyond what is available from conventional geomagnetic activity indices, such as the Kp and AE indices. Further, since the data products this particular GEDAS program provides are based on real-time recordings of magnetic perturbations from a number of stations, the output should be more realistic than what empirical potential patterns using “point” measurements of the solar wind estimate.

A weak point of our system is that to run the KRM algorithm for local regions, the ionospheric conductance must be plugged in as an input parameter. As a near-future project, the ionospheric conductance can be normalized using other sets of real-time observations of the global distribution of auroras by polar-orbiting satellites. Radar data in real time are also quite useful to estimate the ionospheric conductances although they are basically “point” measurements. This is an advantage of real-time data, making mutual calibrations possible. It is important to point out

Fig. 9. Example of real-time calculations of the distributions of (a) electric potential and (b) the corresponding ionospheric current vectors for disturbed periods.
that our output from this project is valuable for understanding the status of the auroral electrojets, which are important for forecasting the strength of induced currents.\textsuperscript{21,22}

Figure 10 shows another example of real-time calculations at GEDAS.\textsuperscript{23} This is a front-view snapshot of the magnetic field configuration and the related plasma temperature distributions in the magnetosphere just before and immediately after an interplanetary shock arrived at the front of the magnetosphere for a unique space weather event of October 24, 2003, during which several serious satellite anomalies, including the Japanese environmental satellite ADEOS-2, were reported. For this MHD model, one-min data of interplanetary magnetic field (IMF) and solar wind parameters from the ACE spacecraft were used. It is possible to illustrate such magnetospheric configurations from different view angles as well. In a sense, this simulation system is a virtual magnetosphere in the eGY concept.

Figure 11 shows an example of changes in the neutral density at an altitude of 300 km as a result of ionosphere–thermosphere coupling calculations.\textsuperscript{24} This coupling model uses, as the input parameters, not only the convection electric field and ionospheric conductivities, but also the output from a solar wind–magnetospheric MHD model, shown in Fig. 10 whose input is real-time solar wind data. Instead of the ionospheric electric field mapped down from the magnetosphere obtained from the MHD

---

Fig. 10. Front view of magnetic configuration just before (left) and after (right) of the shock arrival during the intense interplanetary disturbance of October 24, 2003, calculated from a solar wind-magnetosphere coupling MHD simulation. Different lines (different colors in the original diagrams) are used to distinguish open and closed field lines, and field lines which are not connected to the earth. Solar wind data from ACE were used. Courtesy of T. Ogino.
Fig. 11. Variations in the neutral density at an altitude of 300 km calculated by a global thermosphere–ionosphere model. The region in northern Canada indicates an increase in the neutral density, and the numbers for the intensity contour are in the unit of 0.01%. The input parameters of the model are the convection electric field and conductivities, and are taken from an MHD model for the condition of the steady solar wind with southward IMF. Courtesy of H. Shinagawa.

simulation, it is possible to use the output from the KRM magnetogram-inversion as well. We can treat this type of real-time computing system as a virtual thermosphere in eGY. It is clear that mutual calibrations of different real-time simulations can be conducted.

6. Issues for Future Studies

The Earth’s magnetosphere is a vast laboratory for plasma physics in which unique processes occur, monitoring one of the aspects of Sun–Earth relationships. It is bounded by the solar wind on the outer side and by the ionosphere on the inner side. These boundaries represent dynamic borders within which the magnetosphere is influenced. As is evident from Table 1, during the past 50 years we have seen significant advances in studies of magnetospheric physics. These changes are primarily because of advances in various techniques, such as satellites and radars, as well as computer simulations dealing with real-time data.

As demonstrated, studies of the solar-terrestrial environment began with inspecting geomagnetic disturbances. Following the discoveries of
the secrets of the interaction between the solar wind and the Earth’s magnetic field, researchers have almost reached a consensus about the average configuration of the magnetosphere. In the following, several major areas in which further clarifications are required are discussed.

6.1. **Global versus local**

There are still many unknowns about details of magnetospheric processes that are very different in different substorms, representing perhaps different preconditions and local ionospheric conditions. This variability seems to be quite similar to local weather in the troposphere, which is driven not only by large-scale air circulation resulting from the global pressure/temperature distribution, but also by local boundary conditions, such as three-dimensional structures of near-by mountains. Just suppose that these local conditions, for example, the plasma pressure in the magnetosphere, vary constantly. It would then be hopeless to find some consistent features in small-scale auroral motions. It may not be completely unlikely that by now, we already know what we are supposed to know about large-scale substorm phenomenology, although it is still extremely important to separate large-scale and small-scale phenomena in real-time observations, which are basically local measurements.

6.2. **Macroscopic versus microscopic**

To fully solve the substorm problem, both microscopic and macroscopic processes must be accounted for together. “Microscopic” processes do not necessarily mean “local” processes in Sec. 6.1, but represent those of particle scales. This is simply because the two types of processes, such as acceleration of auroral particles, i.e., a microscopic mechanism, and the generation of field-aligned currents, i.e., a macroscopic mechanism, are taking place in the same high-latitude region at the same time. Individual electrons, for instance do not aware whether they are participating in macroscopic or microscopic processes. One of the typical examples is an auroral breakup at substorm expansions. To understand these processes properly for the magnetosphere–ionosphere system, a kinetic approach is necessary. On the other hand, macroscopic processes occur as a result of large-scale energy conversion in the nearly entire magnetosphere, responding to changes in the solar wind.
6.3. **Average configuration versus nonsteady process**

It is comparatively easy to assume that disturbed times are substorm times, and the current system grows and decays as a whole in the entire ionosphere, but we all know that individual substorms are much more complex than this assumption implies. Substorm expansions that take place in local regions near midnight are short-lived. Thus, by observing an increase in the current intensity at one point, we can never be certain whether the increase is the result of a global enhancement of the electric fields generated from the solar wind–magnetosphere coupling or of a local enhancement in the ionospheric conductivities. Further, substorms are a multi-dimensional phenomenon consisting of the directly-driven process and the unloading process. Their relative importance varies from substorm to substorm, and it even depends on the phases within a single substorm.

It is also important to realize that averaging individual data tends to smear out important local aspects. Therefore, by employing averaging methods, we cannot obtain more than the well-known twin-vortex pattern of the electrostatic potential that Axford predicted 45 years ago. The average picture often misleads us about how nonlinear processes occur within such a large-scale system as the magnetosphere, particularly when the solar wind is highly variable. We are not allowed to consider that disturbed times are equivalent to substorm expansion times.

6.4. **The solar wind and the boundary condition to the magnetosphere**

The Sun does not determine all details of the processes in the entire solar-terrestrial system, particularly those inside the magnetosphere. Therefore, using data from the solar wind and the IMF alone, we cannot determine all processes that occur within the magnetosphere. In other words, the solar wind only gives the boundary condition to the magnetosphere, under which various types of disturbances, such as substorms and convection enhancements, take place internally. It is not too difficult to realize this by simply considering that the Sun itself cannot decide where (what latitude and local time) the next auroral breakup of what intensity will take place in the polar ionosphere. It is the time history for, at least, several hours, of the inner magnetosphere and the polar ionosphere that seems to be critical in determining detailed processes in the magnetosphere–ionosphere system.
It would not be highly recommended, in this respect, to try to reproduce all the fine-scale variations of geomagnetic indices by using only solar wind parameters. If a paper is published in which correlation between the solar wind and geomagnetic indices are close to 1.0, arguably something must be wrong with its analysis technique.

7. Conclusions

The final goal of Sun–Earth relationships research is to become to predict very accurately geoeffective storm/substorm events when the “present” condition of the Sun is given. For this to happen in reality, however, we will have to fully understand all the basic processes in the Sun–Earth system, such that all physical details can accurately be coded in computers. Also, in order to achieve this degree of accuracy in the prediction of space weather, a super computer of extremely high speed and an infinite memory size would be necessary.

It is quite obvious that we are currently far from this final goal. Therefore, we are trying to understand details of multi-time/spatial-scale processes in the solar-terrestrial system, and to use large, instead of infinite, computer memories. We also make use of as many different types of observations as possible at a large number of points. Computer simulations and modeling are extremely powerful in filling gaps where no observation points exist on the Earth’s surface and in space. It is very useful to rely not only on physics-based theories of the processes but also on empirical formulas.

This paper has attempted to demonstrate the importance of real-time data in studies of solar-terrestrial relationships. The importance of real-time data can be summarized in terms of the following specific meanings.

7.1. Space weather predictions

This paper has not explicitly addressed the usefulness of real-time data in space weather predictions, but there is no doubt that real-time data are a must to improve predictions of space weather.

7.2. Finding yourself in the global framework

In conducting any type of observations, it is important to identify the precise location of your instrument within the entire solar-terrestrial system. For example, without identifying the location of what a particular satellite is
measuring in the large-scale energy flow from the Sun to the Earth, one is not able to discuss the physics of these measurements self-consistently. It is not sufficient at all to know only that your radar is located accurately at geomagnetic latitude 64.9° and MLT = 23.6, measuring the southward electric field in the region of the westward electrojet. It is crucial to understand whether you are currently inside the convection electrojet system or a substorm expansion, both giving the same westward electrojet.

7.3. Test of your idea in real time

Real-time data can be used in computer simulations as the initial and boundary conditions to run a model calculation. For example, with solar wind data as input, one can run an MHD simulation, providing an output which predicts the structure of the magnetospheric plasma regions for the next moment. The most exciting point in such a scheme of integration of observations and simulations in real time is that any researcher in the world can be a project leader, using real-time observations to drive models in his/her own research projects in his/her own office. In this way fresh ideas can be tested almost instantly against real data.

7.4. Output becoming input to other studies

An example has been shown of real-time calculations of two-dimensional distribution of ionospheric electrodynamic quantities at high latitudes. Once the ionospheric parameters have been computed, the output can be sent to institutions around the world, where this output will possibly become input for other modeling studies. For example, our electric field distribution in the ionosphere obtained from GEDAS can be mapped to the magnetosphere and is therefore useful for tracing particles in the magnetosphere. Joule heating from our calculations can be used as input for calculating neutral winds in the thermosphere, which will modify the electric field pattern in several minutes.

7.5. Education and outreach

It is evident that real-time data system such as GEDAS can be utilized as a live classroom, in which students can see data directly from satellites and ground stations in real time. They can share the excitement of pursuing
real-time research and of predicting geomagnetic storms and substorms along with tremendous auroral displays.

It is important to rephrase that real-time data combined with real-time simulations are not only capable of changing the practical style of research but also change the concept of research in solar-terrestrial physics. While most of the previous studies in this field have attempted to account for one cross-section of one phenomenon in one plasma region within the entire Sun–Earth system, it will become increasingly crucial to understand the whole spectra of processes, particularly coupling between different plasma regions and coupling between large-scale and small-scale processes. This trend will be more accelerated by the new international program called the eGY that the scientific community is planning to undertake (see http://www.egy.org). It is an ICSU-endorsed initiative of the International Union of Geodesy and Geophysics (IUGG), driven by the International Association of Geomagnetism and Aeronomy (IAGA). The eGY program is quite timely in the sense that it uses the advantage of modern networking technologies that are harmonized with high-quality observation technologies. Promoting the development of virtual observatories is a central feature of eGY.

No one was predicting during the IGY that 50 years later, the scientific community would discuss this new international enterprise, eGY, for collecting data from space as well as from around the world and for processing these data and using these in computer simulations, all in real time. The present paper has discussed how important it is, for further progress of solar-terrestrial physics, to conduct well-coordinated observations, to handle various data and data sets, and to process data products on a real-time basis. Although the author admits that examples and the discussion subjects given in the present paper are strongly biased toward his own experience, it is hoped that the general conclusions are applicable to other areas in Earth Sciences as well.

Acknowledgments

It was a great honor for me to deliver the First Axford Lecture. When Professor Ian Axford was a visiting professor at our institute, we discussed seriously the need of a new scientific society in Geosciences in the Asia/Oceania region. I valued his extensive experience in establishing the European Geophysical Society (EGS), now the European Geosciences Union (EGU). While he was staying at our institute, we started to write
a memorandum of intent. It was six years ago that Professors Ian Axford and Wing Ip and myself began to establish formally AOGS. This work was supported in part by “A Cooperative Research Program Between Japanese and US Space Weather Centers of Excellence” (grants AOARD-02-4013 and 00-4022) of the Asian Office of Aerospace Research and Development (AOARD), and in part by the Grant-in-Aid for Creative Scientific Research “The Basic Study of Space Weather Prediction” (grant 17GS0208, PI: K. Shibata) of the Ministry of Education, Culture, Sports, Science and Technology (MEXT) of Japan.

References
A NEW PERSPECTIVE ON THE RELATIONSHIP BETWEEN SUBSTORMS AND MAGNETIC STORMS

BRUCE T. TSURUTANI
Research Institute for Sustainable Humanosphere (RISH)
Kyoto University, Gokasho, Uji, Kyoto 611-0011, Japan
Jet Propulsion Laboratory MS 169-506
California Institute of Technology
4800 Oak Grove Drive, Pasadena, CA 91109, USA

WALTER D. GONZALEZ
Brazilian National Space Research Institute (INPE)
Av dos Astronautas 12227-010
Sao Jose dos Campos, SP, Brazil

Recent findings concerning the occurrence of substorms within magnetic storms and within high-intensity, long-duration, continuous AE activities (HILDCAAs) are reviewed and discussed. For this study, the Akasofu [Planet. Space Sci. 12 (1964) 273] definition of a substorm is used. A magnetic storm is defined as an event with a significant ring current with Dst/SYMH < −50 nT. It has been shown that storms without substorms exist. These storms have some properties similar to (large) convection bays, and some that are different. HILDCAA events are convection events that appear to be different than either substorms or convection bays. We group magnetic storms without substorms, HILDCAAs, and convection bays into a new category which we call “convection events” (CEs). Magnetospheric convection is a process present in all three of these types of events. Extending this thought one step further, it is asked whether all magnetic storms, all HILDCAAs, and all convection bays are phenomena different from substorms? From this point of view, substorms are a separate entity which may or may not occur simultaneously with CEs. If substorms do occur simultaneously with CEs, the substorm magnetospheric electric fields should be superposed with the CE electric fields. Further observations should be able to confirm or deny this hypothesis.

1. Introduction

There have been two important recent review articles dealing with the relationship between substorms and magnetic storms, one discussing the nature and interplanetary cause(s) of magnetic storms¹ and the other...
discussing the relationship between storms and substorms. Each article defined the state of knowledge at the time of publication. The Kamide et al. article (see also Ref. 3) was an important stimulus for this present work. Kamide et al. issued a challenge for researchers to find a magnetic storm main phase that did not also have substorms. Until that time, all magnetic storms examined appeared to have substorms. Kamide et al. concluded that although substorms were present during storm main phases, they were not an essential element. This was in contrast to the earlier supposition of Chapman and Akasofu and Chapman that substorms were an integral and necessary part of a magnetic storm. The name “substorm” was coined by the latter authors because they thought that substorms were sub-elements of magnetic storms, i.e., magnetic storms were composed of many substorms.

One work that was partially responsible for this new direction of thinking was that it was found that a unique interplanetary condition was necessary and sufficient for the occurrence of magnetic storm main phases. The eastwardly directed interplanetary electric fields equivalent to a southward interplanetary magnetic field (IMF) had to be intense ($E_{sw} > 5 \text{ mV/m}$) for a long duration of time ($T > 3 \text{ h}$) for a major storm ($\text{Dst} < -100 \text{ nT}$) to occur. An interpretation of this condition is that intense, continuous, and long-lasting magnetospheric convection is necessary for a storm-time ring current to form. It was also found that cases with a series of short duration southward IMF events did not lead to the formation of magnetic storms. These southward fields caused injection events of which high-intensity, long-duration, continuous AE activities (HILDCAAs) are the ionospheric manifestations of the injections. Very little ring current activity is found during such phenomena. The interrelationship between HILDCAAs and storms (and substorms) will be discussed later in this paper.

The organizers of the Asia Oceania Geosciences Society (AOGS) 2006 special session (ST-20: storm–substorm relationships) have asked us to express our thoughts on the current status on the relationship between storms and substorms in light of work done after 1998. Below are some ideas that we hope the reader will find stimulating.

We should note that our discussion is unfortunately not complete. Due to limited time we have not covered the topics of “sawtooth” events and “bursty bulk flows” occurring in the tail. To our knowledge these phenomena have not been intercompared with imaging data and substorms. We leave this to the interested reader to address at a future date.
2. Definition of Terms

We must first start with a definition of terms. Without this, it will be difficult, if not impossible, to make any real progress. We need to define what “magnetic storms” and “substorms” are, and also ask ourselves if there are other global auroral/magnetospheric features that exist? If there are such features, we should then ask how they might or might not be related to storms and substorms. This will be the first step in our discussion.

2.1. What is a magnetic storm?

Magnetic storm events were recently defined by Gonzalez et al.\(^1\) as events in which there are “significant” ring current enhancements as measured by Dst. This is a definition following the lines of Chapman and Bartels.\(^8\) It was Chapman’s leadership which led to the establishment of a chain of near-equatorial magnetic stations that were/are used to measure the global diamagnetic effect of the magnetospheric ring current particles (see Ref. 9). The average of the H-component fields of four nearly equidistantly spaced magnetometers (in longitude) comprises the Dst index [the high time resolution SYMH indices were established much more recently by the Kyoto University Geophysical Institute (KUGI) — and is an excellent replacement for the hourly average Dst index; see Ref. 10]. Later Dessler and Parker\(^11\) and Sckopke\(^12\) proved mathematically that the magnetic decrease detected by the near-equatorial stations was equivalent to the total kinetic energy of the ring current particles, thus the index is a quantitative measure for intensity of the storm (There are currently debates on the influence of other current systems on the Dst/SYMH measurements, but this is beyond the scope of this present discussion. We direct the reader to Feldstein et al.\(^13\) for more detailed discussion).

The difficulty for this present work is to define the word “significant” used in the Gonzalez et al.\(^1\) definition. Sugiura\(^14\) and Gonzalez et al.\(^1\) used a working definition of Dst < −30 nT for magnetic storms. However, we note that isolated substorms almost always cause some measurable magnetic effects at the geomagnetic equator. Clearly the substorm process injects electrons and ions into the magnetosphere,\(^15\) and these particles experience gradient and curvature drift around the Earth, creating a mini-ring current. Substorm injections can at times cause Dst to decrease to values more negative than −30 nT. For the following discussion, we will select a slightly higher value of Dst < −50 nT for our “working” definition for a magnetic storm.
2.2. What is a substorm?

There are many different definitions of a substorm that exist in the literature. Of these, many focus on either ground magnetic or magnetospheric measurements. We think that these definitions can be misleading and also at times could/would be contradictory to each other. We have thus chosen the original definition of a substorm made by Akasofu.\textsuperscript{16} This is based on visual auroral imagery and auroral morphology. Figure 1 is a schematic from his seminal work. Basically, a substorm

![Substorm Diagram]

Fig. 1. The temporal evolution of auroral forms during a substorm. Source: Ref. 16.
is defined as having the following features: (1) a breakup of the most equatorward arc, (2) the expansion of auroral forms to the north, west, and east, and (3) reformation of quiet arcs. This sequence can be as short as 10 min and as long as several hours.

2.3. Are there other collective auroral and/or magnetospheric events?

One major large-scale phenomenon different from substorms is convection bays. These events have been identified and described by ground magnetic measurements. These measurements have been used as proxies for the location and intensity of ionospheric currents. Little or no auroral images have been reported to the authors’ knowledge. Convection bays have several identifying features. One is that they are magnetic bays that have particularly smooth intensity–time profiles. A second feature is that the eastward electrojet current intensity measured by the AU index is often larger than the westward electrojet current intensity (AL) (Y. Kamide, personal communication, 2006). It should be noted, however, that when using the latter criteria, this comparison must be done during northern hemispheric summer. Ahn et al. have pointed out that if the auroral zone ionosphere is in solar darkness, it will have insufficient conductivity to allow the true intensity of the eastward electrojet to be determined from ground magnetic measurements.

There have been magnetospheric particle events that have “sawtooth” intensity–time profiles. The relationship of these events to other collective events is still under debate. However for the purpose of this review, we will consider these as a form of substorm. To our knowledge simultaneous auroral imaging has not been analyzed to confirm this or not.

3. Results

3.1. “Triggering” of substorms?

It is instructive to first briefly discuss what external (to the magnetosphere) or internal features are thought to trigger isolated substorms. That may provide some insight into the further topic of substorms that occur or do not occur during storms.

This topic of substorm triggering is quite controversial in magnetospheric research and it will not serve the reader well to provide an encyclopedic
B. T. Tsurutani and W. D. Gonzalez

review of all of the various ideas that have been presented to date. However, a brief discussion of some of the basic proposed mechanisms will be useful. Southward turnings of the IMF as substorm triggers is one of the earliest ideas based on the Dungey’s\textsuperscript{20} theory of magnetic reconnection. Some early references (that are nowhere near complete) are: Tsurutani and Meng,\textsuperscript{21} Meng \textit{et al.}\textsuperscript{22} Caan \textit{et al.},\textsuperscript{23} and Iyemori.\textsuperscript{24} More recently, northward IMF turnings following southward magnetic fields have been suggested.\textsuperscript{25,26} An additional mechanism is solar wind ram pressure increases\textsuperscript{27–31} due to either density increases, velocity increases or both (such as for fast forward shocks). There are also clear cases where no external (interplanetary) triggers are apparent. In these cases, it is generally assumed that internal magnetosphere/magnetotail stress buildup was the cause.

The reason that there is so much controversy on the topic of substorm triggering is that there are serious spacecraft coverage limitations to adequately address the problem. Typically only the beginning and end points of the sequence of physical processes/phenomena are observed. The linkage in between is typically not observed/detected. With an interplanetary monitor, one can note the southward or northward turning of the IMF or a pressure pulse. Then with either a polar orbiting satellite or ground imagers/magnetometers, one can observe auroral (or ground magnetic) displays in the ionosphere. The magnetic reconnection that is hypothesized to occur at the magnetopause and the plasmasheet response to the enhanced convection are typically not observed. Further complications are the apparently variable time delays from the interplanetary event to the auroral event. This may or may not depend on the present state of the magnetosphere/magnetotail and thus raises issues of “magnetospheric priming”. Also the variety of different suspected causes of substorm triggering makes the identification of the trigger mechanism for a specific substorm onset somewhat moot. Thus, because of all of these problems, there is uncertainty and even doubt about the cause–effect relationship between the purported substorm trigger and the substorm. All of these features make the determination of the cause of triggering highly problematic.

The most intense potential substorm triggers are large ram pressure pulses such as fast forward shocks. There are both large density increases and large velocity increases across such fast shocks, so the ram pressure increases can be a factor of 10. When shocks impinge upon the magnetosphere, the response is abrupt and thus easily identifiable. We show an example of a fast forward shock in Fig. 2. The top six panels are the IMF magnitude, the $B_z$ component, the proton thermal speed, the proton
Fig. 2. The relationship between an interplanetary fast forward shock and AE and AL indices.

density, the solar wind speed, and the ram pressure. The shock front is denoted by a vertical dashed line. The bottom two panels are the AE and AL indices. A vertical dashed line across the latter two indices indicates the apparent “substorm” onset. The time delay between the shock detection and the substorm onset is due to solar wind propagation between ACE and the magnetosphere. There appears to be a clear causative relationship between the shock and the later AE/AL event.

Figure 3 shows a sequence of UV auroral images taken from the polar UV instrument. These images were taken in the Lyman–Birge–Hopfield long wavelength band at $\sim 1700 \, \text{Å}$. The images clearly show auroral features related to both the ram pressure of the shock and also a substorm onset and development. The dayside aurora appears first and the nightside substorm later. This is due to the shock first compressing the dayside magnetosphere.
Fig. 3. Same event as in Fig. 2, on September 23, 1998. The polar UV images showing the dayside aurora created by shock compression and the nightside substorm aurora.

and the nightside tail later. A sudden appearance of dayside aurora is noted in the 2345:57 UT image. This is caused by the shock compression of previously existing dayside magnetospheric outer zone electrons, stimulating the loss cone instability, which then leads to wave–particle scattering and particle precipitation to the ionosphere. Since the dayside aurora was present in this image, the shock hits the magnetosphere sometime between this image and the earlier one (the image cadence was 1 min 13 s). The onset of an intense substorm is noted in the 2347:11 UT image. The auroral forms expand and reach a maximum intensity at 2352:05 UT or 2359:19 UT. What is particularly interesting/curious is that the auroral onset occurs prior to the electrojet intensification onset (measured by AE/AL shown previously in Fig. 2). The AE and AL event onsets occur at ~ 23:53 UT, near the maximum of intensity/expansion of the auroral displays.

Details on the mechanisms that lead to this substorm triggering is beyond the scope of this paper. However, interested readers can find some further discussion in Ref. 34.

There have been many works on shock triggering of auroras dating all the way back to the 1960s. All works are in agreement that substorms are not always triggered by fast forward shocks. Zhou and Tsurutani have suggested that IMF $B_{\text{south}}$ “priming” must occur for a fast forward shock to trigger a substorm. Further research on this important topic is warranted.
3.2. Are there magnetic storms without substorms?

Stimulated by the Kamide et al.\(^2\) challenge, some researchers have attempted to search for magnetic storms that did not have substorms or had a paucity of substorms. It had been earlier noted that magnetic clouds/driver gases are characterized by a lack of Alfvén waves and discontinuities.\(^35\) These features had previously been used as a method to identify magnetic clouds/driver gases. The IMF southward or northward turnings discussed earlier are interplanetary discontinuities. Thus if there were any magnetic storms that might have a lack/paucity of substorms, it might be magnetic cloud-induced magnetic storms. This was the thought behind the effort.

Figure 4 shows a magnetic cloud detected in the WIND data on February 10, 1997. The important panels for this discussion are the top two and the bottom three. The top two panels contain the solar wind ram pressure, the magnetic field magnitude, and the \(B_z\) component. The magnetic field magnitude and the \(B_z\) component are shown together in the second panel. Starting with the third from the bottom panel are: the plasma

![Figure 4](image-url)
beta (the plasma thermal pressure divided by the magnetic pressure), the AE, and Dst indices.

The magnetic cloud is identified by a smooth and intense field magnitude (second from top panel) and the low plasma beta (third from bottom panel).\textsuperscript{36,37} The extent of the cloud has been indicated in the top panel. In this case, the current technique (low plasma beta) was used as the identifying feature.

The storm peak Dst was reached at $\sim1200$ UT with a value of $-68$ nT. The southward component of the IMF was approximately $-7$ nT for about 7 h. The $B_z$ component was relatively smooth and the corresponding AE index was smooth as well.

An interval when polar imaging was available is indicated in the shading from $\sim06:40$ to $\sim11:40$ UT. This corresponds to a portion of the main phase of the storm (bottom panel). Figure 5 shows the polar UV images throughout the entire $\sim5$-h interval at a $\sim6$-min cadence. Akasofu-type\textsuperscript{16} substorms were not detected in this interval. The AE values shown previously in Fig. 4 was reasonably smooth, i.e., there was no evidence of rapid auroral electrojet intensifications. Further description of the auroral forms that were present can be found in Ref. 38.

### 3.3. Are storms without substorms different?

Not all magnetic cloud-induced magnetic storms studied by Tsurutani \textit{et al.}\textsuperscript{38} were without substorms. Only about half fell into this category. Out of 11 storms studied, five events were devoid of substorms when polar images were available. We call particular attention to the magnetic storms that occurred on January 10–11, 1997 and August 3–4, 1997, respectively. What is particularly intriguing about these events is that the interplanetary electric field was $>5.4$ mV/m for 4.5 h for the January event, and 6.1 mV/m for 3 h for the August event. These two events met the interplanetary criteria of Gonzalez and Tsurutani ($E_{sw} > 5$ mV/m and $T > 3$ h). However, the peak Dst of these two storms were only $-78$ and $-49$ nT (the ram pressure corrected Dst or $D_{st}^*$ value was only slightly different), breaking the one-to-one relationship between major interplanetary events and major magnetic storms ($D_{st} < -100$ nT) mentioned earlier. For some reason, the storm intensities were abnormally weak. The authors speculated that the lack of occurrence of substorms is an obvious possible cause of the weak storms. Substorms cause an energization of plasmasheet particles,
Relationship Between Substorms and Magnetic Storms

Fig. 5. The same event as in Fig. 4. The corresponding polar UV auroral images for \( \sim 5 \) h of the main phase of the February 10, 1997 storm. There are no substorms identified in this interval.

preheating the ring current plasma. The lack of substorms will therefore lead to a weaker ring current intensity. The authors have discussed other possible explanations as well. To our knowledge no further work has been done in this area at the time of this writing.
3.4. Are there other auroral macroscale events different from substorms, convection bays, and storms?

A type of auroral event that has been previously mentioned is HILDCAA events. The properties of HILDCAAs will be reviewed. The pertinent question is “are HILDCAAs a series of substorms or convection bay events, or perhaps something entirely different?”

Figure 6 shows four days of the magnetic field component of Alfvén waves present in a high-speed solar wind stream. The dates of the interval are May 15–18, 1974, a time when the IMP-8 spacecraft was being tracked. IMP-8 was in orbit about the Earth, a condition necessary to observe the relationship between the sometimes short scale (Alfvén wave) interplanetary events and resultant geomagnetic activity. The top three panels in the plot are the solar wind speed, the magnetic field magnitude, and the $z$-component of the magnetic field (in GSM coordinates). The bottom two panels contain the AE and Dst indices. The feature to note is that for every major southward $B_z$ interval, there is an AE increase and a Dst decrease.

Soraas et al.\cite{39} have studied other HILDCAA events using NOAA low altitude satellite particle detectors and have shown that there are shallow injections of plasma to $L > 4$ during such intervals. This finding

![Graph showing magnetic field components and indices during May 15-18, 1974.](image)

Fig. 6. An example of four days of a HILDCAA event. The AE increases are related to the IMF $B_z$ components of Alfvén waves. The AE increases are associated with Dst decreases, maintaining the Dst value at $\sim -25$ nT for almost the entire four-day interval.
is consistent with magnetic reconnection during the short southward component magnetic fields. The slight Dst decreases are also consistent with this scenario. However, the main question remains: What are the AE increases? Are they substorms or convection bays or something else?

To answer the first part of the question, we examine Fig. 7. This is an interval of AE increases caused by Alfvén wave $B_z$ fluctuations. The AU and AL indices are also shown (note that the event occurred on January 12, 1997, a period during local winter, so the eastward electrojet intensities might be underestimated). Indicated in shading are substorm intervals, identified by the polar UV imaging. The imaging figures are not shown to minimize space. However for the interested reader, he/she could refer the original paper. What is surprising is that the AE and AL peaks do not correspond to substorm intervals. Whether this is due to the global auroral/current system (see Ref. 41) masking the nightside auroral zone system or not, is not known at this time. But certainly the peak AE values

Fig. 7. The AE, AU, and AL indices during a HILDCAA interval. Substorms identified in the polar UV images are indicated by vertical shading. The peak AE/AL values are not well correlated with substorms.
of the HILDCAA event do not correspond to substorms, at least by the Akasofu\textsuperscript{16} definition.

Somewhat puzzling was the previously mentioned lack of correspondence between the start of substorms indicated by imaging and that by the geomagnetic indices, AE and AL. For many years now researchers have simply “assumed” that the start of AE increases or AL decreases were an indication of the substorm onset. This same issue may be pertinent to the identification of HILDCAAs (discussed here) as well. We encourage the interested reader to study these related phenomena in more detail.

To attempt to answer the other part of the question, are HILDCAAs convection bays, we examine the relationship between the auroral electrojet indices during summer months. Figure 8 shows an example of a HILDCAA event occurring from July 23 to 24, 1998, a summer interval. The AU, AL, and AE indices are given. It can be noted that AL is greater than the AU value throughout the \( \sim 2+ \) day long HILDCAA interval. Thus, from this second criterion for convection bay events, \( \text{AU/AL} > 1 \), HILDCAAs do not appear to qualify as convection bays.

It appears that HILDCAAs are neither substorms nor convection bays. We will place them in the “other” category for the present. However, we do know that it is probable that magnetic reconnection is responsible for the solar wind energy transfer to the magnetosphere and that HILDCCAs are some form of convection events.

Gonzalez et al.\textsuperscript{42} have shown that when the fluctuating IMF \( B_z \) fields of the Alfvén waves are in the negative domain, there is a higher probability of having substorms. Presumably this is due to the greater energy transfer from the solar wind to the magnetosphere during such intervals.

3.5. Are magnetic cloud-induced storms huge convection bays?

We examine the AE, AU, and AL data for the two magnetic storm events without substorms discussed earlier, event 1 on January 10–11, 1997 and event 7 on August 3–4, 1997. Figure 9 shows the storm event on January 10–11, 1997. From top to bottom are the: AU, AL, AE, \( -\text{AU/AL}, -\text{AL/AE} \) and SYMH indices. For the magnetic storm, the peak Dst reached \( -78 \) nT. The one-min average SYMH values shown here have a similar value (\( \sim -82 \) nT). The main phase of the magnetic storm is quite long, lasting from \( \sim 04:00 \) to \( \sim 09:00 \) UT on day 10. The main phase amplitude profile
Fig. 8. A short duration high-speed stream. Large amplitude Alfvén waves are present in the stream. The $B_{\text{south}}$ component of the Alfvén waves create geomagnetic activity as denoted by a HILDCAA event.

was very smooth. During this interval there were modest AE increases and AL decreases. Little or no AU signatures are noted (but it is local winter). The ratio of AL to AE was very close to 1.0.

The August 3, 1997 magnetic storm is shown in Fig. 10. The format is the same as for Fig. 9. Because this event occurs in summer, it is much
better suited to examine the AU/AL ratio. The minimum SYMH value is $\sim -50$ nT, placing this as a small magnetic storm. What is very interesting about this event is that the AU, AL, and AE index profiles are generally quite smooth, similar to the requirement for lesser intensity convection bays. However, the AU/AL ratio is typically $\sim 0.8$ during the storm main phase. This value is less than 1.0, indicating that it is somewhat different than convection bays.
Fig. 10. The format is the same as in Fig. 9. This storm occurred during northern summer. The AU/AL ratio during the storm main phase is $\sim 0.8$.

4. Discussion

Where do we presently stand on understanding the relationship between substorms, convection bays, HILDCAAs, and magnetic storms? We think that a “logic diagram” showing the overlap between various phenomena would be useful. This is shown in Fig. 11.

The figure shows a schematic of the substorm–storm relationship just after the discovery that some storms could exist without the presence
Fig. 11. A schematic showing the relationship between storms and substorms. Before 1998, all storms were thought to contain substorms. We now know that some storms are either devoid of substorms or have very few of them. This is denoted by the storm CE events that do not overlay the substorm area.

of substorms. We call the latter events CEs to distinguish them from convection bays. At the time of the Kamide article, the magnetic storm domain would have been entirely within the substorm domain. So the change is only a small one. There are some storms (CEs) that do not coexist with substorms. They are now separate entities.

We now redraw Fig. 11 to include convection bays, storm convection events, and HILDCAAs as well. This is shown as Fig. 12. On the left are CEs. In this category, we include magnetic storms, HILDCAA events, and convection bays. On the right we have substorms. There is an area of overlap where storms and substorms coexist and HILDCAA events and substorms coexist. It is also possible that some convection bays have substorms. However, studies examining the latter hypothesis have not been performed to our knowledge. Thus, this is left as a question mark at the present time.

Let us go one step further with this line of thought. Since all magnetic storms are driven by large scale southward IMFs, perhaps all magnetic storms are global convection events. Sometimes the interplanetary drivers contain discontinuities or pressure pulses that trigger substorms, but this is incidental to the storm (other than making its intensity greater). This idea has a testable prediction: that the storm magnetospheric convection electric field and the substorm convection electric fields are separate entities due to different physical processes. If they are indeed different, then when substorms occur during magnetic storm main phases, the driving electric fields would be
Fig. 12. Figure 11 redrawn to include HILDCAA events and convection bays. There is some overlap between substorms and storms and HILDCAA events. There is also possible overlap between substorms and convection bays (indicated by a question mark in the figure). An extension of this idea is to consider CEs (storms, HILDCAAs, and convection bays) as a separate entity from substorms. Substorms may or may not occur simultaneously with CEs. Their physical drivers are different but are present at the same time.

superposed. Tests could be made by computer modeling or by electric field measurements.

This latter concept has an additional effect in interpreting Fig. 12. The CEs would then be entirely separate phenomena from substorms. Substorms may or may not occur simultaneously with CEs. An important physical effect is that the internal magnetospheric drivers for two or more phenomena can and do occur at the same time.

5. Final Comments
We have shown that magnetic storms can occur without substorms. We have also shown that continuous, sporadic plasma injections into the magnetosphere (and substorms) occur without significant ring current development (magnetic storms). It is thus concluded that storms and substorms are separate magnetospheric/ionospheric dynamical processes.
Acknowledgments

Portion of this work was performed at the Jet Propulsion Laboratory, California Institute of Technology under contract with NASA. We thank O. P. Verkhoglyadova and F. L. Guarnieri for preparing and critically reading the manuscript.

References

Relationship Between Substorms and Magnetic Storms

This page intentionally left blank
STORM–SUBSTORM RELATIONSHIP: CONTROVERSIES
AND RECENT DEVELOPMENT

TOMOAKI HORI
National Institute of Information and Communications Technology
4-2-1 Nukui-Kita-Machi, Koganei, Tokyo 184-8795, Japan

The storm–substorm relationship is one of the oldest and most controversial issues in solar-terrestrial physics. A considerable amount of arguments have existed regarding this issue since 1960s when the idea, a magnetic storm consists of successive occurrence of multiple substorms, came out. Later on a different paradigm, called enhanced convection paradigm, was proposed to explain the ring current development during a storm main phase. In this paper, the results of the studies addressing these questions are reviewed carefully for an attempt to clarify how our current understanding on the ring current injection is summarized. We also discuss some recent results regarding the storm–substorm relationship argument which can be important clues to elaborating the current model of the storm-time ring current evolution.

1. Introduction

It is well known that the long-lasting controversies on the storm–substorm relationship were initiated in the early 1960’s by Akasofu and Chapman,\(^1\) which indicated that very energetic auroral activities are often seen during the main phase of a magnetic storm. The coincidence of the auroral activities with the storm development gave birth to an idea described in Chapman\(^2\) represented by Chapman’s own words: “A magnetic storm consists of sporadic and intermittent polar disturbances, lifetime being usually one or two hours. These I call polar substorms”.

The characteristic signature of a magnetic storm is a depression in the H component of the ground magnetic field lasting over some tens of hours, which was considered as the effect of a ring current.\(^3\) Figure 1 shows schematically the variation of the total energy content of ring current protons corresponding to various fashions of substorm occurrence, proposed by Akasofu.\(^4\) Akasofu thought that a substorm feeds the ring current content and a substorm of larger intensity causes a larger enhancement of
the ring current. Akasofu’s hypothesis is that frequent occurrence of intense substorms causes the ring current buildup, forming a storm main phase, then followed by the recovery phase resulting from less frequent/intense substorm occurrence, as shown as the case (3) in Fig. 1. Later on, satellite observations showed that energetic particles are actually injected to geosynchronous orbit in association with substorm activities.\textsuperscript{5} Such a particle flux enhancement, referred to as substorm injection, has been considered to be a direct source for the ring current particles. Recently, a statistical study using a large data set of geosynchronous substorm injections showed that the storm size is well correlated with a product of the occurrence number and the intensity of substorm injections during the corresponding storm main phase, supporting the above hypothesis.\textsuperscript{6}

The storm–substorm relationship proposed by Akasofu.\textsuperscript{4} promoted a number of detailed correlation studies between storms and the associated substorms. Before the systematic observation of substorm injection was available, such studies were often performed by comparing substorm activities measured by the AE index with the variation of the Dst index.
Interestingly, however, some of those studies provided results that were not consistent with those expected from the hypothesis of successive substorm occurrence, leading to a different hypothesis to explain the storm-time ring current growth.

One of those is the work done by Kamide and Fukushima. They showed that the variation of the DRS index, essentially equivalent to Dst, for a storm can be expressed as a function of the simultaneously observed AE value, as seen in Fig. 2. Their new and important finding is that, to reproduce the DRS variation by AE, they need to introduce a factor expressing the efficiency of a substorm for the ring current growth which has a larger value for the main to early recovery phase than for the late recovery phase, even for a substorm of the same intensity. Later this result led to the suggestion by Kamide that the frequent occurrence of substorms alone may not be sufficient to generate a storm. Moreover, detailed correlation studies between Dst, AE, and the solar wind parameters revealed that Dst can be expressed better as a function of the solar wind electric field \[ V_B \times a \text{ product of solar wind speed, } V, \text{ and southward (IMF) component, } B_s \] than that of AE. They implied that the reproduction of the Dst variation mathematically does not need the component due to a substorm. These results were synthesized into a hypothesis, called enhanced convection hypothesis, that the particle injection of the storm-time ring current is caused by enhanced convection directly driven by the enhanced dayside reconnection due to a large \( V_B s \), rather than through the substorm processes.

The heart of the storm–substorm relationship dispute is to understand how substorms cause a storm main phase in which the ring current keeps intensifying by being fed with energetic ions. As stated above, historically controversies regarding this issue have been made over the two hypotheses. In this paper, our discussion is focused on the ring current development.

Fig. 2. An example of the DRS index reproduced on the basis of the AE variation shown by Kamide and Fukushima.
Section 2 is devoted to the discussion on how the controversies between the two hypotheses have turned out to be by far. In situ measurement of the ring current particles shed light on a possibility that substorms play a crucial role in forming an oxygen-abundant ring current, which is reviewed in Sec. 3. In Sec. 4, we examine some of the latest important topics relating to the storm–substorm relationship. Section 5 provides the summary discussion.

2. Substorm Injection and Enhanced Convection as an Agent of Ring Current Particle Transport

After the direct measurement of the ring current by CRRES in early 1990s, no satellite mission has succeeded so far in a systematic observation in the equatorial region of the inner magnetosphere. Instead, extensive numerical simulations of the ring current dynamics have been performed to understand the ring current evolution from a theoretical point of view. A comprehensive review of the recent ring current simulations can be found in, e.g., Ebihara and Ejiri.13

One of their most important achievements is the reproduction of the actually observed temporal variation of Dst based on physical models treating particles’ motion in a kinetic manner in more or less realistic electric and magnetic field models. The top panel of Fig. 3 shows that the observed Dst variation is quite similar to that predicted by the ring current–atmosphere interaction model performed by Liemohn et al.14 Most of the ring current simulations including the one by Liemohn et al.15 employed the convection electric field models, whose intensity varies in close correlation with the intensity of the observed southward IMF. This means that the electric field driving the particles’ motion is the one expected from the enhanced convection hypothesis. Therefore, it proves that the particle injection driven by enhanced convection can reproduce physically the observed Dst variation.

Another important conclusion from the ring current simulations is that the ring current during a storm main phase is not closed around the Earth. The enhanced sunward convection is strong enough for the particles’ \( \mathbf{E} \times \mathbf{B} \) drift to overcome their grad-\( \mathbf{B} \) and curvature drift in the longitudinal direction. As a result, the ring current particles injected from the nightside drift westward and reach the dayside magnetopause, forming the asymmetric shape of the high pressure region, corresponding
Fig. 3. The result of the ring current simulation for the September, 1998 storm done by Liemohn et al.\textsuperscript{14}

to an asymmetric ring current as shown in the upper dial plots for time 1 (early main phase) and 2 (late main phase). It should be noted that particles forming the asymmetric ring current are not trapped but keep being supplied from the nightside by convection and escaping from the dayside magnetopause, which is fundamentally different from those implied by the successive substorm hypothesis assuming that intermittent injections make a trapped, most likely symmetric, ring current population. In addition to the simulations, the observation of the local time distribution of the ground magnetic field perturbation\textsuperscript{15} and of the ring current shape by the energetic neutral atom (ENA) imaging\textsuperscript{16,17} favor the former, namely the asymmetric ring current during a storm main phase.

However, it should be noted that many studies suggesting the dominant role of enhanced convection in the ring current buildup do not necessarily
exclude additional roles of substorms in promoting efficient particle transport into the inner region. Fok et al.\textsuperscript{18} performed a numerical simulation of particle injection for the ring current by incorporating the substorm effect as the magnetic field variation and its associated inductive electric field corresponding to field line stretching of the growth phase and subsequent depolarization of the expansion phase. To examine the effect of a substorm, they compared the simulated result without a substorm and that with a substorm. The result on the temporal variation of total $H^+$ energy content within $R < 12R_E$ (solid lines) and $R < 6.6R_E$ (dashed lines) is shown in Fig. 4. In the case when weak convection is applied to the model magnetosphere (upper panel), the substorm occurrence makes no significant difference in total $H^+$ energy between before and after the substorm onset at $R < 6.6R_E$. However, in the case of strong convection corresponding to a storm situation (lower panel), the simulation run with a substorm gives a much larger energy content inside the geosynchronous distance than that with no substorm, even though the resultant energy content is almost the same at $R < 12R_E$ regardless of substorm occurrence. Their interpretation is that strong convection itself

![Fig. 4. The variations of the total energy of simulated $H^+$ ions in association with substorm onset under the weak and the strong convection condition.\textsuperscript{18}](image-url)
Storm–Substorm Relationship

could fill in the outer region of the model magnetosphere with particles even without a substorm, while the particle injection into the inner region would be achieved by the coexistence of strong convection and substorm processes.

A recent observational study of substorm injections also claimed the importance of the cooperation of strong convection and the substorm effect. Reeves and Henderson\textsuperscript{16} compared geosynchronous proton flux variations between isolated substorm injections and storm-time substorm injections, as shown in Fig. 5. Their result indicates that isolated substorms cause flux increases lasting for only tens of minutes (black lines in the figure), whereas under the storm condition substorms cause the initial rise of proton flux and then the enhanced flux is sustained by the particle supply driven by enhanced convection, resulting in high proton fluxes for longer period (\(~\)hours) at the geosynchronous distance, especially at the energies of typical ring current particles of tens to a few hundred keV.\textsuperscript{19} Nevertheless, a case study for an intense storm\textsuperscript{20} showed that substorms do not depress the Dst index but rather increase it, even though such a long-lasting particle injection do occur. The discrepancy of the variation between Dst and the geosynchronous flux may be attributed to the fact that the main part of storm-time ring current lies in the range of radial distance of 4–6 $R_E$\textsuperscript{21} and thus the geosynchronous particles do not represent well the majority of the actual ring current causing the Dst perturbation.

Fig. 5. The variation of energetic protons at geosynchronous orbit in association with storm-time substorm injections and isolated substorm injections.\textsuperscript{16}
3. Another Role of Substorms: $\text{O}^+$ Supply to the Ring Current

In addition to the correlation studies based on the geomagnetic indices, \textit{in situ} measurement of the ring current particles had become available and contributed greatly to the understanding of their physical characteristics. In particular, the information on the composition of ring current ions obtained by the AMPTE/CCE satellite provided a striking result that the oxygen ions, most likely singly charged oxygen ions ($\text{O}^+$) of ionospheric origin, make a significant contribution to the total energy content of the storm-time ring current.\textsuperscript{22} Figure 6 shows the radial structure of the energy density for each of major ion species near the Dst peak of a storm. It was found that the energy density of $\text{O}^+$ is as large as that of hydrogen ions ($\text{H}^+$) at radial distances where the main body of ring current lies. Their study also showed that the rate of the $\text{O}^+$ contribution to the ring current energy content depends on the storm phase: the $\text{O}^+$ abundance increases as a storm main phase develops and then drops rapidly during an early recovery phase. This result was reconfirmed by the inner magnetosphere observations with a more comprehensive set of instruments made by the CRRES satellite during late 1980s to early 1990s. Moreover, the studies of the ring current composition based on the CRRES observation revealed that $\text{O}^+$ ions dominate the total energy content of the ring current especially during intense storms.\textsuperscript{23}

![Fig. 6. The radial profile of the energy densities of ring current ions near a storm peak observed by AMPTE/CCE.\textsuperscript{22}](image-url)
Storm–Substorm Relationship

Fig. 7. The energy density ratio of H\textsuperscript{+} and O\textsuperscript{+} to the total ion energy density during an intense storm observed by CRRES.\textsuperscript{28}

Figure 7 shows the temporal variation of the energy density ratio of H\textsuperscript{+} (top panel) and O\textsuperscript{+} (second panel) to the total energy density of all ions during an intense storm event for which the Dst variation is shown in the bottom panel. It is seen from the figure that the O\textsuperscript{+} energy density ratio generally increases from the storm onset until the Dst peak (approximately −300 nT), then gradually decreases during the storm recovery phase. The O\textsuperscript{+} contribution to the total energy content is as much as 80% at the storm peak, indicating that the O\textsuperscript{+} population dominates the ring current. In addition to this case study, the systematic observations made by AMPTE/CCE and CRRES revealed that the same variation of the O\textsuperscript{+}/H\textsuperscript{+} ratio is found for all storms, while the O\textsuperscript{+}/H\textsuperscript{+} ratio is small for small to moderate storms and becomes larger with storm intensity.\textsuperscript{24,25}
This finding of the $\text{O}^+$ dominance shed light on a possibility that substorms play an essential role in the ring current buildup through the supply of $\text{O}^+$ ions to the ring current. It had been well known that ionospheric ions flow out of the auroral region to the magnetosphere during substorms. Yau et al.\textsuperscript{26} made a statistical study on the ion outflows on the basis of the low altitude satellite observation and found that the amount of $\text{O}^+$ outflow increases exponentially with increasing AE and always is greater than that of $\text{H}^+$ outflow, indicating that $\text{O}^+$ ions may be preferentially transported into the magnetosphere during substorms. In addition to the supply process, some theoretical works suggested that dipolarization of the magnetic field associated with substorm onset may accelerate $\text{O}^+$ ions preferentially due to their long gyroperiod comparable to the time scale of the field variation, which may contribute to the $\text{O}^+$-dominance in the ring current, e.g., Ref.\textsuperscript{27}. Considering all the observational facts, Daglis and Kamide\textsuperscript{28} claimed that substorm occurrence is essential to forming an $\text{O}^+$-rich ring current through both the supply and the subsequent acceleration of $\text{O}^+$ ions to the typical energies of ring current particles ($\sim$ several $10^1$ to a few $10^2$ keV).

Interestingly, some works using the same data (from CRRES) gave opposite results on the role of substorms discussed above. Grande et al.\textsuperscript{29} performed a superposed epoch analysis on the variation of Dst, the number density, and average energy of several ion species in association with substorm injections. Their detailed analysis revealed that the increase in $\text{O}^+$ energy density on substorm injection is due to the increase in its number density, not to the energization of the pre-existing $\text{O}^+$ population. This result implies that substorms work mainly as a supply process of $\text{O}^+$ ions, although some acceleration processes are likely to occur concurrently to accelerate fresh, cold $\text{O}^+$ ions from the ionosphere to the typical ring current energies. Nonetheless, their result of the superposed epoch analysis on $\text{He}^{2+}$ ions showed an increase in number density to the same degree as $\text{O}^+$. Therefore, they concluded that the composition ratio between $\text{O}^+$ of ionospheric origin and $\text{He}^{2+}$ of solar wind origin does not change so much on a time scale of a substorm of $\sim$ hour. Their analysis on the composition change on a time scale of a storm showed that the preferential supply of $\text{O}^+$ ions occurs at rather lower L-shells ($L < 5$), while the injection of $\text{He}^{2+}$ ions takes place mostly at $L = 5$ or above. Thus it is implied that substorm injection itself is not capable of transporting particles deep into the lower L-shells to increase the $\text{O}^+$ abundance there.
Moreover, in regard to the O\(^+\) injection to the ring current, Korth et al.\(^{30}\) claimed the importance of enhanced convection. They examined the local time distribution of different ion species observed by CRRES during storms and showed that the local time evolution is almost the same for different ions including H\(^+\) and O\(^+\) indicating their injection is governed by the same mechanism, most likely, by enhanced convection. In addition, Korth et al.\(^{25}\) examined in detail the temporal variation of the O\(^+\)/H\(^+\) energy density ratio for several storms and derived the same conclusion. Figure 8 shows the O\(^+\)/H\(^+\) ratio (top panel) together with the AE and the Dst variation in the second and bottom panel, respectively, during an intense storm on July 13–15, 1991. They noted that the O\(^+\)/H\(^+\) ratio variation consists of large but spiky fluctuations with time scales of tens of minutes to a few hours which appear to be associated with AE activities, and a slowly-varying trend quite similar to the overall trend, precisely speaking inverse trend, of Dst. They concluded that substorms represented as AE activities increase the O\(^+\) abundance significantly but such an increase is temporary and does not last longer than \(\sim\) hours. Whereas, the overall trend of the O\(^+\)/H\(^+\) ratio variation seems to vary in close correlation with Dst, suggesting that both the O\(^+\)/H\(^+\) and the Dst variations are governed by the same mechanism, most likely by particle injection from the plasma sheet controlled by enhanced convection.

![Figure 8](image-url)

Fig. 8. The variation of the O\(^+\)/H\(^+\) ratio, AE, and Dst observed by CRRES during an intense storm on July 13–15, 1991.\(^{25}\)
4. Recent Topics on the Storm–Substorm Relationship

In this section, we discuss some of the very recent works regarding the transport and acceleration processes of ring current particles which are important from the perspective of the storm–substorm relationship.

4.1. Actual profile of particle transport by enhanced convection

As discussed in Sec. 2, the electric field models based on the enhanced convection hypothesis have been employed by many ring current simulations. Their success in reproducing some fundamental properties of the storm-time ring current is one of the reasons why enhanced convection has been generally accepted as the main driver of the ring current particle injection. Whereas, not so many studies were conducted so far to examine detailed properties of the assumed enhanced convection or its associated convection electric field itself. Recently, some studies have been done to examine it and obtained the results which may pose questions on the convection features expected straightforwardly from the hypothesis.

Most of the kinetic simulations of ring current injection solve the trajectories of particles in a given static magnetic field. Unlike those simulations, the rice convection model (RCM)\textsuperscript{31} is one of the methods which can treat both the particle motion and the electric/magnetic field self-consistently. The rice convection model-equilibrium (RCM-E)\textsuperscript{32} numerically simulated the adiabatic motion of flux tubes with a self-consistently computed magnetic field. Despite the success by the former version,\textsuperscript{31} the RCM-E had difficulty in reproducing the pressure structure corresponding to a storm-time ring current. The reason is that, under the self-consistent evolution of the magnetic field, a region filled with flux tubes of high pressure is formed on the nightside as a result of their earthward convection and then prevents further earthward convection toward the inner magnetosphere, as demonstrated in Fig. 9. Their result indicated that some nonadiabatic processes are needed for the reduction of $PV$ of flux tubes in the near-Earth tail to allow continuous earthward convection. In other words, theoretically such a nonadiabatic process is essential to the enhanced convection generating a storm-time ring current.

Another issue regarding the enhanced convection was raised by Hori et al.\textsuperscript{33,34} which showed the characteristics of the actually observed convection electric field in the near-Earth plasma sheet during storms. Figure 10 shows the dependence of the duskward electric field strength...
Storm–Substorm Relationship

Fig. 9. The pressure structure of the inner magnetosphere simulated by the RCM-E.\\n
Fig. 10. The correlation of the storm-time duskward electric field observed by Geotail with those predicted by the empirical polar cap potential model.

on the model values calculated by dividing the modeled polar cap potential by the modeled tail width. It is seen that the duskward electric field, driving earthward convection, shows poor dependence on the predicted values, implying that the tail convection electric field is somehow independent of
the strength of the storm driver, which is not consistent with the enhanced convection hypothesis. In addition, the fact that no systematic difference in the electric field strength between the main phases (red circles and triangles) and the recovery phases (blue circles and triangles) is seen also supports the above implication.

What these studies on enhanced convection tell us is that the physical process of particle transport based on enhanced convection hypothesis has not yet been described fully in a theoretical point of view. Also the actually observed electric field in the storm-time plasma sheet does not seem to behave straightforwardly as expected from the hypothesis. For full understanding of the particle injection by enhanced convection, ring current simulations should be improved to be more comprehensive so that they can solve particles and fields self-consistently and then the simulated results need to keep being tested by the observation.

4.2. \( O^+ \) dynamics based on the ENA observation

Recently ENA images have been available from the IMAGE satellite\textsuperscript{35} as a new tool to diagnose the ring current profile. Although it is difficult to convert an ENA flux value to the real particle flux at a certain point in the inner magnetosphere, the global ENA imaging enables us to not only watch the spatial evolution of the ring current but also estimate a total energy content of the ring current for each instantaneous moment as long as the satellite is located in such a way that its instrument has a good field of view to look at the entire ring current distribution.

As for the storm–substorm dispute regarding the \( O^+ \) issue, in particular, Ohtani et al.\textsuperscript{36} made a superposed epoch analysis on the variation of the hydrogen and oxygen ENA fluxes in association with substorm onset during storm times. Figure 11 shows the resultant average variations of the 27–60\,keV hydrogen (third panel), 60–119\,keV hydrogen (fourth panel), and oxygen ENA flux of <160\,keV before 2001 and of 52–180\,keV after 2001. Those ENA flux variations are sorted by the time of dipolarization associated with substorms observed by the GOES satellites at the geosynchronous orbit, as shown in the top panel. Their ENA observation showed that there is no net increase in hydrogen ENA flux after substorm onset, while the oxygen ENA flux increases significantly (by \( \sim 20\% \) of the pre-onset level on average). This result can be interpreted as more direct evidence for the preferential acceleration of oxygen ions (mostly \( O^+ \))
in the magnetosphere) in association with dipolarization in the course of substorms.

However, it should be noted that Ohtani et al.\textsuperscript{36} focused on the variation on a time scale of substorm cycle and did not provide the ENA flux variation on a time scale comparable to a storm phase (\(\sim \) several hours). As discussed with the CRRES observation, a significant enhancement of \(\text{O}^+\) flux does occur in association with substorm onset but its persistence on a longer time scale was questioned by Korth et al.\textsuperscript{25} A case study done by Brandt et al.\textsuperscript{37} provided an example of such a long variation of the oxygen ENA flux during a moderate storm. The result shows that the oxygen ENA flux keeps increasing during the first half of the storm main phase, while it does not show any net increase during the second half of the main phase.
Unfortunately, it is difficult to draw any conclusion from this result about the storm-phase-scale ENA variation, because the temporal variation of ENA flux on time scales of several hours can be affected heavily by the orbital motion (inbound or outbound) of IMAGE during the measurement, as discussed by Ohtani et al.\textsuperscript{36}

5. Summary Discussion

One of the important issues to be addressed is the meaning of the existence of O\textsuperscript{+} ions in the ring current during intense storms. A number of observational evidences provided by CRRES and AMPTE/CCE show that the abundance of O\textsuperscript{+} ions in the ring current becomes higher during the storm main phase for more intense storms. Daglis and Kamide\textsuperscript{28} postulated that the O\textsuperscript{+} supply from the ionosphere and subsequent inward injection caused by a substorm during storm times would create a condition to cause a next substorm at a position closer to the Earth. In the course of such a substorm process, O\textsuperscript{+} ions are energized preferentially (relative to H\textsuperscript{+} ions) due to their slow gyrofrequency comparable to the field fluctuations associated with dipolarization. As a result, the positive feedback between the O\textsuperscript{+} supply and the substorm occurrence could facilitate a more inward injection of high-energy O\textsuperscript{+} ions in the inner magnetosphere, resulting in an intense ring current. In this mechanism, O\textsuperscript{+} ions are considered to be essential to generating an intense storm as a catalyst for the progressively inward-invasion of substorm injection as well as a major constituent of the ring current. On the contrary, Korth et al.\textsuperscript{25} claimed on the basis of the CRRES observation that the O\textsuperscript{+} enhancement caused by substorm injection is temporary and returns to the pre-substorm level in a few hours after each substorm. They claimed that the O\textsuperscript{+}/H\textsuperscript{+} ratio is primarily governed by enhanced convection varying on a longer time scale than that of a substorm cycle. In their hypothesis, O\textsuperscript{+} ions are supplied significantly from the ionosphere by substorms and then are essential as a constituent of the intense ring current, but do not contribute much to the particle acceleration and transport processes themselves in terms of the net growth of the ring current content, as suggested by Grande et al.\textsuperscript{29} In addition, Greenspan and Hamilton.\textsuperscript{38} showed that a simple O\textsuperscript{+}-rich condition does not lead to an intense storm, indicating that an intense storm is created primarily by an intense storm driver, that is, a large southward IMF.\textsuperscript{39}
Storm–Substorm Relationship

Viewed in the above light, the issue of the substorm role may eventuate in a question of whether the existence of substorms yields the net growth of O\(^+\) abundance in the ring current beyond the substorm time scale, more than just as an O\(^+\) source. However, it is very difficult to evaluate observationally the pure effect of substorms in that sense. That is because we have to examine the variation of O\(^+\) content on longer time scales than a substorm cycle, say, ~ several hours, which is already comparable in time scale to that expected from enhanced convection. For this reason, virtually we cannot examine separately the effect of substorm injection on the O\(^+\) content variation on such a time scale. It should be noted that basically we cannot avoid the same problem even if using the O\(^+\) variation based on the ENA observation, as long as time series of the O\(^+\) variation are simply examined. Therefore, it is desired to use a numerical simulation of the ring current including the O\(^+\) injection process to keep tracks of the injected O\(^+\) in the inner magnetosphere so that we can attribute quantitatively a variation in O\(^+\) flux at a certain point and time to the transport due to substorm injection or to enhanced convection.

In contrast to the O\(^+\) injection, the H\(^+\) injection seems to be achieved for the most part by enhanced convection. As shown by studies on substorm injection based on the CRRES observation, e.g., Ref. 40, substorm injection does transport H\(^+\) but usually to \(L \sim 6\), still well outside of the main body of the storm-time ring current. Moreover, the ENA observation shown by Ohtani et al.\textsuperscript{36} clarified the fact that the H\(^+\) content measured by the hydrogen ENA does not show any net increase in association with substorm expansion onset. Thus it is suggested that substorms do not play a significant role in the ring current H\(^+\) injection during storm times.

Another issue that has not been addressed well here but is crucial to the storm–substorm dispute is the caveats about Dst in using as a measure of the ring current energy content. In particular, Dst does not trace the true intensity of the ring current in the course of substorm expansion phase, because the Dst variation is heavily contaminated by the effect of the tail current disruption.\textsuperscript{41,42} In addition to the post-onset interval, McPherron\textsuperscript{12} suggested that Dst could begin to decrease concurrently with the start of a substorm growth phase, almost an hour before substorm expansion onset. Taking all these facts into account, it is suggested that Dst can no longer be used for discussing the ring current content within a few hours from each substorm onset, otherwise the result might be misleading. In analogy to the above argument on the substorm role on the O\(^+\) injection, it is suggested that we cannot examine the effect of each substorm on the ring current.
content on the basis of Dst. The Dst variation on longer time scales, say, \( \sim \) several hours may still correspond to the true variation of the ring current. However, the Dst variation on such a long time scale can also be interpreted as a result of particle injection driven by enhanced convection and thus its cause cannot be clearly specified, between the effect of substorm injection and that of enhanced convection.

In conclusion, evidence accumulated by studies during the past decade indicates that particle injection due to substorms and enhanced convection both contributes to the ring current growth during storm times. The transport driven by enhanced convection plays a main role particularly in the \( \text{H}^+ \) injection to the ring current region. Whereas, storm-time substorms play a crucial role in supplying the ionospheric \( \text{O}^+ \) ions to the magnetosphere and have great potential to transport and energize them to form a ring current with significant abundance of \( \text{O}^+ \) especially for intense storms. However, enhanced convection is also capable of injecting \( \text{O}^+ \) ions that preceding substorms had fed and thus the relative contribution to the \( \text{O}^+ \) abundance in the intense ring current between substorm injection and enhanced convection is still controversial to be examined quantitatively. Future works to address the storm–substorm relationship need to break away with Dst and then have to find an alternative measure for monitoring the net variation of the ring current on not only the substorm time scale but also time scales of storm phase. We know that the global ENA imaging is a promising tool to diagnose such a variation in detail. Also crucial for the quantitative analysis is a numerical simulation of the ring current evolution including not only enhanced convection but also the substorm effects, such as the supply of ionospheric \( \text{O}^+ \) ions, and the subsequent energization and injection driven by the time-varying fields associated with substorm expansion.

Acknowledgments
The author thanks all participants in Session ST20 “Storm–substorm relationship”, AOGS 2006, for providing interesting topics and stimulating discussions regarding the theme.

References
Storm–Substorm Relationship


This page intentionally left blank
TEMPORAL DEVELOPMENT OF DAYSIDE TEC VARIATIONS DURING THE OCTOBER 30, 2003 SUPERSTORM: MATCHING MODELING TO OBSERVATIONS

OLGA P. VERKHOGLYADOVA
IGPP, University of California at Riverside
900 University Ave., CA 92521, USA
RISH, Kyoto University, Uji 611-0011, Japan

BRUCE T. TSURUTANI and ANTHONY J. MANNUCCI
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive, Pasadena, CA 91109, USA

Strong electron density enhancements are formed in the dayside low-latitude ionosphere by promptly penetrating electric fields (PPEFs) during intense interplanetary electric field (IEF)/superstorm events. Although some of the basic physical ideas of this newly recognized phenomenon are qualitatively understood and have been implemented in various existing ionospheric (and plasmaspheric) models, these models have consistently underestimated the satellite and ground total electron content (TEC) measurements by a wide margin. As one example, no model has been able to duplicate the TEC increase of the October 30, 2003 IEF/superstorm event to within a factor of 2. We present an effort first to determine the ionospheric electric field from magnetic measurements of the Equatorial Electrojet and then to simulate TEC enhancements for this superfountain event based on a modified SAMI2 model, called SAMI2*. We find our results in a reasonable quantitative agreement with the CHAMP measurements. It is noted that dynamic dynamo (storm-time) winds are not necessary to replicate the CHAMP TEC observations, only the PPEFs and the resultant superfountain effects.

1. Introduction

Although the phenomenon of promptly penetrating electric fields (PPEFs) from interplanetary space has been known for some time,1–3 it has often been assumed that Region-2 field-aligned currents would lead to the shielding of these external electric fields after ~30–60 min.4 Earlier theoretical work suggested that shielding in the daytime could take hours to
develop. More recently, it has been shown that during intense storms and superstorms, very intense PPEFs are observed at the dayside equatorial and near-equatorial ionosphere. In some cases, very intense PPEF effects are believed to be present for up to 5 or 6 h. In their empirical study of low-latitude electric fields during moderately disturbed conditions, Fejer and Scherliess noted the occurrence of extended periods of weak shielding.

The PPEFs \( E \times B \) lift the dayside/dusk equatorial and near-equatorial ionosphere to heights much greater than 500 km (see above references), creating strong electron density enhancements. Convection of plasma to midlatitudes as high as \( +30^\circ \) and \( -30^\circ \) magnetic latitude (MLAT) have been reported. After the PPEFs lift the dayside ionosphere to greater heights where the electron–ion recombination time scales are considerably (hours) longer, solar photoionization will replace the uplifted plasma at lower altitudes, giving a net result of a highly enhanced column density of electrons or total electron content (TEC). This process is called the daytime superfountain effect.

The current study focuses on the dynamics of the daytime TEC during a PPEF associated with October 30th, 2003 superstorm event. At the time of the event, CHAMP was at an altitude of \( \sim 400 \) km and crossed the dayside magnetic equator at \( \sim 13:00 \) local time. An ionospheric electron density versus height model by Mannucci et al. was used to “verticalize” (correct for the slant-path) the TEC values. The temporal evolution of the TEC was traced during three sequential dayside CHAMP passes. It was shown that \( \sim 2h \) after the interplanetary electric field (IEF) event onset, the dayside Appleton anomaly moved to \( \sim \pm 30^\circ \) MLAT. The peak TEC above CHAMP had values of \( \sim 300 \) TECU. Since the quiet-time value for the ionospheric TEC at \( \pm 30^\circ \) MLAT (measured from the ground) is \( \sim 50 \) TECU, this disturbed-time value represents at least an \( \sim 600\% \) increase in the TEC column density at these latitudes. As a point of reference, the typical peak TEC value measured from the ground is \( \sim 100 \) TECU, where a TECU is \( 10^{16} \) electrons/m\(^2\) column density. This peak TEC occurs at \( \sim 14:00 \) local time at the Appleton anomalies.

We will first estimate the equatorial and near-equatorial dayside ionospheric electric field from magnetic measurements of the storm-time enhanced Equatorial Electrojet (EEJ). We will use this electric field (PPEF) as input to a modified SAMI2 ionospheric code. To match the CHAMP
measurements\(^{15}\), e.g., dayside TEC values above 400 km altitude, we will concentrate on simulations at \(\sim 22^\circ\) latitude.

2. Computer Simulations

The basic NRL SAMI2 model\(^{18-20}\) has been used as a baseline model and has been adapted for our specific modeling. We call this the SAMI2* model. SAMI2 is a low-latitude ionospheric model which describes dynamics and chemical evolution of seven ion species and, correspondingly, seven neutral species. Collisions between electrons, ions, and neutrals are taken into account. SAMI2 solves collisional two-fluid equations for electrons and ions along the Earth’s dipole magnetic field lines, taking into account photoionization of neutrals, recombination of ions and electrons, and chemical reactions. Drift of magnetic flux tubes defines the ionospheric plasma transport in a perpendicular direction to the magnetic field lines. The \(\mathbf{E} \times \mathbf{B}\) vertical drift is caused by the eastward polarization electric field\(^2\) superimposed on the Earth’s background magnetic field. The SAMI2 diurnal variation electric field “sine” model is assumed. For this electric field, the drift velocity \(V_d\) is proportional to \(\sin[(t - 7)/24]\), where \(t\) is the local time in hours. This drift describes a rise of magnetic flux tubes from dawn to dusk (local time) and their subsidence from dusk to dawn. We will use the “sine” model for calculating the background or undisturbed electron density altitude profiles. The assumed \(V_d\) peak value is taken as 15 m/s, which corresponds to a polarization electric field of 0.53 mV/m.

A second electric field needs to be added to introduce a superposition PPEF during IEF and substorm events. This represents the zonal electric field in the low-latitude ionosphere. We will initially assume that the magnitudes of the electric fields in the ionospheric E- and F-regions are the same.\(^{21}\) Different values can be inputted into SAMI2* if for some reason the fields are found to differ. The Kyoto University ionospheric model is used to obtain ionospheric conductivity values. For October 30, 2003 at an altitude of 105 km at the equator at noon, the Pederson conductivity \(\sigma_P\) is \(5 \times 10^{-5}\) S/m and the Hall conductivity \(\sigma_H\) is \(9.8 \times 10^{-4}\) S/m. Thus the Cowling conductivity, \(\sigma_H^2/\sigma_P\), is \(1.9 \times 10^{-2}\) S/m. The CHAMP magnetic field perturbation due to the EEJ intensification during the October 30, 2003 PPEF event was \(\sim 90\) nT (H. McCreadie, personal communication, 2005). Assuming a ground reflectance of \(\sim 11\%\) (Ref. 22; A. Richmond, personal
communication, 2006) and an infinite line current $I (I = \Delta B 2\pi r/\mu_0)$ centered at $\sim 105$ km altitude for the EEJ, an electric field value of approximately $4 \text{mV/m}$ is obtained.\textsuperscript{16} In the above expression, $\Delta B$ is the magnetic perturbation detected at CHAMP, $r$ is the distance from the line current to the observation point, and $\mu_0$ is $4\pi \times 10^{-7}$ in mks units.

In our model, a PPEF causes ionospheric plasma drift in a perpendicular direction to the magnetic field lines. An opposite viewpoint suggests that plasma flow drives electric field in MHD plasma.\textsuperscript{23} However, in our case the electric field is externally driven and thus arguments on internal electric field decay in plasma do not apply.

Some initial results on the ionospheric electron density simulations for the October 30, 2003 IEF/PPEF event are discussed below. $O^+$ ion densities have also been examined by Tsurutani \textit{et al.}\textsuperscript{16} as well as proton densities for this event, but are not discussed to conserve space. For this first-cut simulation, we simply assume a step function PPEF of $4.0 \text{mV/m}$ (corresponding to a vertical drift speed of $\sim 115 \text{m/s}$). The simulation is performed for the region from $-35^\circ$ to $+35^\circ$ MLAT and extending in altitude from 85 to 3000 km.

The results of SAMI2\textsuperscript{*} simulations after the PPEF has been applied for almost 2 h (13:51 local time) are shown in Fig. 1. The quiet-time dayside Appleton anomaly peak in the northern hemisphere is located at $\sim +10^\circ$ (not shown here). At this local time this peak has electron density of $\sim 3.9 \times 10^6 \text{cm}^{-3}$ at $\sim 340$ km altitude. After application of a “step”-like PPEF of $4 \text{mV/m}$ and the background electric field, from noon to 14:00 pm local time, the peak moves to higher altitude and latitude, and becomes enhanced. The resulting electron profile is shown in Fig. 1. The latitude range for the profile shown in the figure is about $22^\circ \pm 1.5^\circ$, which is the approximate position of the displaced dayside Appleton anomaly peak. The maximum electron density reaches $5.6 \times 10^6 \text{cm}^{-3}$ at $\sim 620$ km altitude. The electron density with only the sine electric field for the same latitude and altitude is $4 \times 10^5 \text{cm}^{-3}$, thus showing about 14 times density enhancement. This simulation clearly shows ionospheric “uplift” of the density daytime maximum from $\sim 340$ to 620 km and poleward expansion from $+10^\circ$ to $+22^\circ$. A continuous photoionization production at lower altitudes contributes to the density enhancement. These are clear features of the PPEF $\mathbf{E} \times \mathbf{B}$ convection effects alone. It is also noted that the peak electron densities are higher in the step PPEF case.

After the PPEF is turned off, the plasma starts falling down along the Earth’s magnetic field lines of force,\textsuperscript{16} descending to lower altitudes
Temporal Development of Dayside TEC Variations

Fig. 1. Electron density profiles for the October 30th, 2003 superstorm event at a position of the displaced dayside Appleton anomaly peak (see the text) at 22° ± 1.5° latitude and at about 14:00 pm local time after a PPEF of 4 mV/m has been applied for 2 h. Simulation results obtained by using SAMI2 model ("sine" electric field, shown by stars) and SAMI2* model (PPEF superimposed on the background "sine" electric field, shown by triangles) are presented.

and higher latitudes. The results are shown in Fig. 2 after the electric field has been turned off for ~3 h. The dayside Appleton anomaly peak is located at +24.5° ± 1.0°, at a height of ~430 km, and with a peak intensity of ~7.0 × 10^6 cm^-3. Corresponding background intensity given by the "sine" model electric field is ~2.3 × 10^6 cm^-3 at about 380 km altitude. For comparison, the quiet-time dayside Appleton anomaly peak has 4.1 × 10^6 cm^-3 of electron density value, and a position at 380 km in altitude and about 11° in latitude.

Next step is matching the TEC values observed by CHAMP. We used SAMI2* model to simulate TEC values in a column extending from the CHAMP altitude (400 km) to 1000 km altitude. Results for the 22° ± 1.5° latitude and 14:00 pm local time (the same as in Fig. 1) are shown in Fig. 3. The TEC values for the background ("sine") model and PPEF plus the background model start to diverge from the noon when the storm-time electric field is introduced. Notice about six times jump in the TEC values.
Fig. 2. Electron density profiles for the October 30th, 2003 superstorm event at a position of the displaced dayside Appleton anomaly peak (see the text) at 24.5° ± 1.0° latitude and at about 17:00 pm local time after the PPEF has been turned off for 3 h. Simulation results obtained by using SAMI2 model ("sine" electric field, shown by stars) and SAMI2* model (PPEF superimposed on the background "sine" electric field, shown by triangles) are presented.

(from about 40 to 250 TECU) due to the PPEF effect. After the storm-time electric field is turned off, electron density starts to recover to the background values.

3. Conclusions

We performed simulations of the ionospheric electron density dynamics during the October 30th, 2003 superstorm event using the modified SAMI2 model (or SAMI2*). Our results show strong uplift of the ionosphere due to the PPEF and displacement of the dayside Appleton anomaly peak in the northern hemisphere to higher latitudes and altitudes and continuous photoionization production at lower altitudes. Estimated TEC values show an ~ 600% increase in TEC column density at the 22° ± 1.5° latitude, which quantitatively agrees with the CHAMP measurements. After the storm-time electric field is turned off, plasma starts falling down the magnetic field lines and the ionosphere slowly recovers to the undisturbed state. This
scenario agrees with the superfountain effect first proposed by Tsurutani et al.\textsuperscript{6} Our first simulation results show that, by introducing a realistic PPEF model a reasonable agreement with the measurements during a superstorm event can be reached, and the actual event can be modeled at low and middle latitudes based on the basic principles of the ionosphere dynamics incorporated into the SAMI2* model.

It should be noted that the CHAMP TEC observations were matched by only using an accurate PPEF applied to the SAMI2* model. Electric fields associated with winds originating from auroral heating were not used and were not necessary to match simulations to the observations. Further work should be done to clarify the role of dynamo electric fields at local dayside noon.

Acknowledgments

Portion of this work was performed at the Jet Propulsion Laboratory, California Institute of Technology under contract with NASA. Authors
appreciate discussions with T. Araki, A. Saito, H., McCreadie, A. D. Richmond, and D. N. Anderson. This work uses the SAMI2 ionosphere model written and developed by the Naval Research Laboratory.

References
CUTOFF L-VALUES OF SOLAR PROTONS IN COMPARISON WITH RING CURRENT PROTONS DURING A STORM: NOAA/POES OBSERVATIONS

K. T. ASAI∗, T. NAGATSUMA, H. SHIMAZU, T. HORI and K. KITAMURA
National Institute of Information and Communications Technology
4-2-1 Nukuikita, Koganei, Tokyo 184-8795, Japan
∗asakt@nict.go.jp

Y. S. MIYOSHI
Solar-Terrestrial Environment Laboratory, Nagoya University
Aichi 464-8601, Japan

During solar proton events, the observation made by the polar-orbiting NOAA/POES satellites (N15, N16, N17, and N18) shows a clear cutoff in L-value of solar protons with energy of > 0.8 MeV. The solar proton event of August 2005 covered the entire time interval of the initial, main, and recovery phases of a storm with the minimum Dst of −216 nT. The cutoff L-values of the solar protons vary with the storm phase. The cutoff L-values of the solar protons are compared with those of the precipitating ring current protons (30–100 keV). It is found that the cutoff L-values of the solar protons essentially agree with those of the precipitating ring current protons during the storm. These cutoff L-values are well correlated with the geomagnetic indices. The result suggests that the depression of the magnetospheric field caused by the ring current controls the entry of the solar protons into the magnetosphere. The result implies that the varying position of the cutoff of the solar protons can be used as a new measure to diagnose the structure of the storm-time magnetosphere.

1. Introduction

Solar energetic particles are injected into the Earth’s magnetosphere but do not reach the inner magnetosphere where the field lines connect to the low latitudes according to the well-known Störmer theorem. Geomagnetic cutoff rigidity and cutoff latitudes of energetic particles have been investigated in the past studies and were reported to be controlled by geomagnetic activities. Obayashi and Hakura1 estimated solar particles from the radio
blackout observed in the polar region during geomagnetic storms. One of the earliest studies on geomagnetic cutoff of solar particles was reported by Hakura. As reviewed by Smart and Shea, many studies on the geomagnetic cutoff have been made on the basis of the observation of cosmic rays (>~100 MeV), while solar energetic particles of ∼MeV have been recently studied on the basis of satellite observation and particle simulation.

Correlation between the geomagnetic activity and the geomagnetic cutoff of high-energy particles was investigated by some past studies. Flückiger et al. discussed the effect of the ring currents and the magnetopause currents included in the Dst variation based on cosmic ray cutoff rigidity observed at geomagnetic latitude of ∼55° on the ground. Leske et al. compared 8–15 MeV proton cutoff latitudes with Dst and Kp indices for six cases (1992–1998) at the International Space Station (ISS). Birch et al. compared cutoff latitudes of 35 MeV protons with Kp (3 h), Dst (1 h), SYM-H (1 min) indices for storms in September and November 2001 using NOAA observations. These geomagnetic indices were derived from ground geomagnetic observations. Therefore, these studies did not directly discuss the relation between the entry of the high-energy particle and the magnetospheric phenomena, e.g., particles of magnetospheric origin.

The present study investigates the cutoff L-values of solar protons during a geomagnetic storm and compares them with those of particles originating in the magnetosphere using satellite observations. The polar-orbiting NOAA/POES satellites measure both the solar protons and the magnetospheric ring current protons simultaneously. Solar protons in the energy range from 0.8 to ∼500 MeV and ring current protons in the energy range of 30–100 keV are analyzed in this study.

2. Data Analysis
2.1. Solar proton event of August 2005

Fifteen solar proton events from 2002 to 2006 were identified. The events were selected by the following criteria on the NOAA/POES data: the protons flux with energies >16 MeV is larger than 10/s-cm^2-str and that for 2.5–6.9 MeV is larger than 100/s-cm^2-str. In the event of August 2005, the solar protons were detected from the initial and subsequent main phase through the recovery phase of the geomagnetic storm.
Figure 1 shows the overview for August 20–30, 2005: the geomagnetic indices in the top panel and the solar protons from the ACE observation in the second and third panels. The MeV solar proton flux was enhanced from day 22 and a geomagnetic storm occurred around the time of peak solar proton fluxes. The minimum Dst was \(-216\) nT at 11 UT on day 24.

### 2.2. NOAA data

The NOAA/POES satellites (N15, N16, N17, and recently launched N18) have been observing charged particles in different local time regions at altitudes of 810–890 km in the polar orbit. Since the launch of N17 in July 2002, the data from at least three satellites are available at the same time period. The combined data obtained by all of these satellites cover all local times and have the time resolution of 1–1.5 h due to the orbital period of about 100 min.

The onboard charged particle instruments cover a very wide energy range, from 50 eV to \(\sim 500\) MeV. The omni-directional instruments cover energies above 16 MeV and the directional instruments cover energies below 6.9 MeV. The instrumental details are given by Evans and Greer.\(^{12}\) The present paper shows only the directional proton data with local pitch angles of 0–45 (135–180°) in the northern (southern) hemisphere because the data
with pitch angles of 45–135° include errors due to the radiation belt particles at around $L \sim 3–4$. It has been confirmed that the flux intensities of perpendicular protons and those of precipitating protons are fairly the same.

The NOAA satellites observed solar protons in the present interval in August 2005. The top three panels of Fig. 2 show the $L$-value versus time ($L$–$T$) diagrams for the solar protons of $>16$ MeV (P6 channel), 2.5–6.9 MeV (P5), and 0.8–2.5 MeV (P4) detected by the NOAA satellites from August 22 (DOY 243) to 25 (DOY 237). Each vertical axis shows the McIlwain $L$-values using the corrected geomagnetic coordinates.\textsuperscript{13,14} The $L$-value can be roughly converted to the corrected geomagnetic latitude with the relation $L \sim \cos^{-1/2} \lambda$, where $\lambda$ is the geomagnetic latitude. The latitudes are shown on the vertical axis on the right side of each plot, for reference. The color code shows the flux intensity in unit of log/s-cm$^2$-str averaged for all local times.

For reference, the geomagnetic indices, interplanetary magnetic field (IMF), and solar wind velocity and density, in hourly values are plotted in the bottom panels of Fig. 2. It is found that the solar proton event continues for about three days over the geomagnetic storm triggered by a southward IMF.

The cutoff $L$-values of the solar protons are clearly seen and vary dynamically for all energy ranges as shown in Fig. 2. The cutoff $L$-values are about 5 before the storm and less than 3 at the peak of the storm. The difference of $L$-values corresponds to 8° or greater in corrected geomagnetic latitude.

These solar protons are compared with ring current proton precipitation. The fourth panel in Fig. 2 shows the precipitating protons with energy of 30–100 keV (P1 channel) detected by the NOAA satellites. The instrumental status reports (http://www.ngdc.noaa.gov/stp/NOAA/docs/) mention that the nominal energy range of the P1 channel is 30–80 keV but its energy range on the N15 satellite changes to 40–100 keV since late 2001 because of radiation damage. Therefore, our combined data of this energy channel from all the satellites should be regarded as a flux for the energy range of 30–100 keV.

Asai et al.\textsuperscript{15} using the NOAA observation reported that the protons of 30–100 keV originate in the plasma sheet and that they make the largest energy contribution to the ring current since their flux increases during storms. It is found from the comparison of the $L$–$T$ diagrams that the lower cutoff $L$-value of the ring current proton precipitation varies in close correlation with that of the solar protons.
Fig. 2. $L-T$ diagrams of particles observed by the NOAA satellites for four days of August 22–25, 2005: solar protons of $>16$, 2.5–6.9, and 0.8–2.5 MeV in the upper three panels, ring current protons of 30–100 keV in the fourth panel. The bottom panel shows the Dst and Kp indices, IMF, and solar wind velocity and density from the OMNI2 database.
2.3. Variation of cutoff $L$-values

In order to compare the cutoff $L$-values of the protons in different energy ranges, thresholds are set in such a way that they are greater than the background levels of each detector. The $L$-value of the cutoff was determined by the lower side at which the count rate drops below the threshold: $10/s\cdot cm^2\cdot str$ for $>16$ MeV solar protons, $100/s\cdot cm^2\cdot str$ for solar protons of the other energies, and $10^4/s\cdot cm^2\cdot str$ for the ring current protons. Figure 3 shows the variations of the cutoff $L$-values for the different energy ranges for the August 2005 event.

We compare the cutoff $L$-values before the storm with those during the storm. Before the storm (before about 4 UT of August 24), the protons with higher energy can penetrate into lower $L$-value regions. The cutoff $L$-value of the ring current protons (solid circles) is higher than those of solar protons (crosses, triangles, and open circles). During the storm (after about 4 UT on August 24), the cutoff $L$-values are almost the same. The solar protons of $>16$ MeV (crosses) disappeared before the storm peak at 11 UT on the day 24.

Fig. 3. Cutoff $L$-values of solar protons of $>16$ MeV (crosses), 2.5–6.9 MeV (solid triangles), 0.8–2.5 MeV (open circles), and ring current protons of 30–100 keV (solid circles) for the same interval. The bottom panel shows the Dst (curve) and Kp (steps) indices.
Figure 4 shows the correlations between the cutoff $L$-values of the ring current protons and those of the solar protons. The interval before the storm is from 22 UT on August 22 to 3 UT on August 24 and the storm interval is from 4 UT on August 24 to 9 UT on August 25. The correlation before the storm (solid diamonds and solid line) and that during the storm (open diamonds and dashed line) are clearly different. The regression coefficients shown on the right side of the figure indicate that the difference between the cutoff $L$-values of the ring current protons and those of the solar protons is about 3.4 on average before the storm and about 0.6 on average during the storm.

Figure 5 shows the correlations between the cutoff $L$-values and the geomagnetic indices for the present event from 22 UT on August 23 to 8 UT on August 25. The correlation coefficients, $\sqrt{R^2}$, of the ring current proton cutoff for Dst and Kp are both 0.84 and those of the solar proton cutoff are 0.82 Dst and 0.93 for Dst and Kp, respectively.

3. Discussion

Cutoff rigidity of energetic particles indicates the restriction of its penetration into the magnetosphere. The rigidity can be converted with almost a linear relation to particle energy as shown in Table 1 of Smart et al.\textsuperscript{16} The left panel of Fig. 6 shows the relation between particle energy and cutoff $L$-value based on the table by Smart et al.\textsuperscript{16} and the relation between vertical cutoff rigidity and cutoff $L$-value, $R_{cv} = 14.5/L^2$ from
Fig. 5. Correlation between cutoff $L$-values and geomagnetic indices. The upper panels are for ring current protons and the lower panels are for solar protons of 2.5–6.9 MeV. For each panel, the regression line is described as the equation on the right in which $x$ ($y$) is the value on the horizontal (vertical) axis. $R$ is the correlation coefficient for each.

Fig. 6. Relation of cutoff $L$-values to particle energies on the left panel. The right panel is the same but for the latitudes converted from the $L$-values. Open symbols are based on Smart et al.\textsuperscript{16} and solid rectangles are estimates for the energy ranges of the NOAA instrument.

Smart and Shea.\textsuperscript{3} This relation is the estimate for a static and dipole-like magnetosphere on the basis of the Störmer theorem using the IGRF geomagnetic field model. The right panel of this figure is the same as the left panel but with the vertical axis of the latitude converted from the $L$-value using $L \sim \cos^{-1/2} \lambda$.

It is noted that the ring current protons of 30–100 keV analyzed in the present study originate in the magnetosphere, particularly in the plasma sheet.\textsuperscript{15} In fact, as shown in the fourth $L$–$T$ diagram of Fig. 2, these protons steadily exist in the region of high $L$-values ($L > 6$ corresponds to $> 65^\circ$...
Cutoff L-Values of Solar Protons

in corrected geomagnetic latitude) regardless of the storm. The cutoff L-values of these protons during nonstorm period for August 22 and 23 (doy 234–235) are \( L \sim 5–6 \) which is quite different from the estimate shown in Fig. 6 (\( L > 35 \) and corrected geomagnetic latitude \( > 80^\circ \) for 30–80 keV protons). It should be noted that the Störmer theorem is based only on the magnetic drift. However, these ring current protons are transported from the plasma sheet to the inner magnetosphere under the significant effect of the magnetospheric convection electric field drift.\(^{17,18}\) Therefore, it is expected that the observed cutoff L-value of the ring current protons is smaller than the estimated cutoff L-value based on the Störmer theorem.

The observed solar protons of \( > 0.8 \text{ MeV} \) came from outside the magnetosphere and appeared after about 10 UT on August 22 (day 234). The cutoff L-values shown in Fig. 3 were 4.7 (\( > 16 \text{ MeV} \)), 5.1 (2.5–6.9 MeV), and 5.3 (0.8–2.5 MeV) on average before the storm. They are also smaller than the estimated cutoff L-values, 4–9, 11–15, and 15–20, respectively. In addition, the L-values decreased by about 5 or more during the storm and reached 2.5–2.8 at the storm peak. These variations coincide with that of the ring current protons.

This result can be interpreted as follows. The entry of the solar protons (\( > 0.8 \text{ MeV} \)) is governed by the geomagnetic field configuration because the motion of particles with such high energy is largely controlled by magnetic field drift rather than electric field drift. The ring current depresses the magnetic field intensity around the equator particularly during storms. Therefore, the magnetic field configuration is significantly changed in such a region that the trajectories of the solar protons may be deformed, allowing for the penetration into the lower L-values.

Such interpretation can explain the relation between the cutoff L-values of the solar protons and the geomagnetic activity. Past studies reported that the cutoff L-value of the solar protons is well correlated with geomagnetic activity but did not discuss the reason in detail.\(^9–11\) The present study confirmed this relation and additionally showed that the cutoff L-values of the ring current proton precipitation is also well correlated with geomagnetic indices, such as Kp and Dst. Some statistical studies indicate that the Kp index reflects the magnitude of the magnetospheric convection electric field which supplies ring current particles of the plasma sheet origin.\(^{19}\) The Dst index is a result of the magnetic field depression caused by the ring current. These indices are both related to the ring current protons which cause the variation of the magnetospheric field configuration leading to the penetration of the solar protons. Therefore,
the good correlation between the cutoff $L$-values of the solar protons and the geomagnetic indices is reasonable.

There are remaining problems on the difference before and during the storm. Future studies are needed to perform numerical tracing of particle trajectories in the time-varying field. Moreover, the present analysis used the average for all local times but the data before averaging suggested the local time dependence during the storm. It is necessary to further analyze the local time dependence in detail.

As an expectation for future study, the present result implies that the cutoff variation of the solar protons can be used as a new measure to diagnose the actual magnetic field structure of the storm-time magnetosphere.

4. Summary

The cutoff $L$-value of the solar protons ($>0.8$ MeV) during a storm was compared with that of the ring current proton precipitation (30–100 keV) on the basis of the NOAA observation. During the storm’s main and recovery phases, it was found that the cutoff $L$-value of the solar protons corresponds to the inner boundary of the ring current proton precipitation. An explanation is that the geomagnetic depression caused by the ring current of the magnetospheric origin controls the cutoff $L$-value of the solar protons from outside the magnetosphere. The result suggests that the variation of the cutoff $L$-values of the solar protons can be used as a new measure to diagnose the structure of the storm-time magnetosphere.

Acknowledgments

We thank GSFC/SPDF for the OMNI2 data (http://omniweb.gsfc.nasa.gov), WDC-C2 Kyoto for the geomagnetic indices (http://swdckw.jki.kyoto-u.ac.jp/), NOAA/SEC for the ACE data (http://www.sec.noaa.gov/ace/ACErtsw_home.html), and specially NOAA/NGDC for the POES data (http://www.ngdc.noaa.gov/stp/NOAA/noaa_poes.html).

References

Cutoff \( L \)-Values of Solar Protons

Geomagnetic activity and auroras on Earth are directly influenced by interplanetary conditions in the solar wind at 1AU. Throughout the 11-year solar (half) cycle, this solar wind is permeated by very different interplanetary structures. During the solar maximum, the dominant structures are interplanetary remnants of coronal mass ejections (ICMEs). In the descending and minimum phases, corotating high-speed streams play a major role. In this paper, we review the characteristics of high-speed streams observed near Earth and their geomagnetic consequences, contrasted with the well-known geomagnetic activity caused by ICMEs. Emphasis is given to the high-intensity long-duration continuous AE activity (HILDCAA) during the recovery phase of corotating interaction region (CIR) magnetic storms/high-speed streams. This activity, caused by the Alfvénic fluctuations embedded in the high-speed streams, is weaker but more continuous than geomagnetic storms. Some characteristics of HILDCAAs are described as well as the auroras caused by these events. The auroral shapes during HILDCAAs are compared with those that occur during typical substorms and great magnetic storms. These events are of extreme importance in Space Weather studies, since the dominance of high-speed streams in the descending and minimum solar cycle phases lead to conditions that increase the amount of relativistic electrons in the outer radiation belts.

1. Introduction

The solar and interplanetary causes of geomagnetic storms/activity are different during different parts of the solar cycle. Around solar maximum,
the main features present on the Sun are sunspots and active regions. These structures are related to sporadic coronal mass ejections (CMEs) and their interplanetary counterparts (ICMEs). During the descending and minimum solar cycle phases, coronal holes become the dominant solar features leading to geomagnetic activity. From these coronal holes, high-speed streams continuously flow with average velocities from $\sim 750$ to 800 km/s.\textsuperscript{1–4} If these coronal holes extend equatorward and persist for more than a solar rotation, the high-speed streams may reappear every $\sim 27$ days at Earth’s orbit, leading to the term corotating high-speed stream.\textsuperscript{5,6} When these fast streams overtake the slow solar wind ($\sim 300$–400 km/s), an interaction region between the two streams (fast and slow) is produced, leading to magnetic field and plasma compression. These structures are known as corotating interaction regions (CIRs).\textsuperscript{7} The storms produced by CIRs are relatively weak due to the fluctuating $B_z$ fields within the structures. Since this is not the main topic of this paper, we refer the interested reader to Tsurutani \textit{et al.}\textsuperscript{8} for further details. ICME storms have been widely studied in the last decades, so this subject will be included here only for comparative purposes. For the reader interested in this topic, we refer them to the articles by Gonzalez \textit{et al.},\textsuperscript{9} Kamide \textit{et al.},\textsuperscript{10} Echer \textit{et al.},\textsuperscript{11} and Tsurutani \textit{et al.}\textsuperscript{8}

The topic of this review is auroras produced in the high-speed streams following the CIR. This appears to occur in the “recovery phase” of the CIR magnetic storm, but we will show that it has nothing to do with the storm itself.

2. Results

2.1. \textit{High-speed streams and their effects on Earth}

The solar source of the high-speed streams are the coronal holes.\textsuperscript{12} These coronal holes appear as dark regions on the soft X-ray images of the Sun due to the absence of hot electrons. The mechanism for plasma heating in coronal holes is not well established, but some possible candidates are described in Parker,\textsuperscript{13} and Hollweg and Isenberg.\textsuperscript{14}

In the solar polar regions, there are large coronal holes from which high-speed streams flow, with speeds of $\sim 750$–800 km/s at 1 AU. At lower latitudes, the dominant structures are the helmet streamers in the solar corona. This latter region is the origin of the slow solar wind, with velocities from $\sim 300$ to 400 km/s. During the descending and minimum phases of the 11-year solar (half) cycle, these polar coronal holes expand or migrate
Fig. 1. Illustration of a high-speed stream flowing from a solar coronal hole and generating a CIR (adapted from Ref. 11).

Fig. 2. An example of a CIR observed near Earth’s orbit is shown in Fig. 2. The panels are, from top to bottom, the solar wind velocity (km/s), density (cm$^{-3}$), ram pressure (nPa), and magnetic field magnitude and the components $B_x$, $B_y$, and $B_z$ (nT). The remaining panels show geomagnetic indices Dst, AU/AL, and AE (all of them in nT). The interplanetary data are from the ACE spacecraft (located at the L1 point, $\sim$ 1.5 million km away from the Sun).

Another important feature of these coronal holes is the presence of large-amplitude fluctuations in the magnetic field and solar wind components in the high-speed streams. These fluctuations are nonlinear Alfvén waves ($\Delta B/|B| = 1$ to $2$) propagating outward from the Sun.$^{18-21}$

An example of a CIR observed near Earth’s orbit is shown in Fig. 2. The panels are, from top to bottom, the solar wind velocity (km/s), density (cm$^{-3}$), ram pressure (nPa), and magnetic field magnitude and the components $B_x$, $B_y$, and $B_z$ (nT). The remaining panels show geomagnetic indices Dst, AU/AL, and AE (all of them in nT). The interplanetary data are from the ACE spacecraft (located at the L1 point, $\sim$ 1.5 million km away from the Sun).
Fig. 2. Geomagnetic activity caused by a high-speed solar wind stream in July, 1998. The HILDCAA activity (indicated by AE, AU, and AL indices) is associated with Alfvénic fluctuations in the high-speed stream (adapted from Ref. 11).

upstream of the Earth) in the GSM coordinate system (x points radially outward from the Earth toward the Sun, y = Ω × x/|Ω × x|, where Ω is the south magnetic pole of the Earth, and z completes the right-hand system).

The high-speed stream reaches its peak at ~15:00 UT on July 23 (DOY 204), with a velocity ~750 km/s. During the high-speed stream, the density drops to very low values. The field magnitude decreases, and fluctuations
are seen simultaneously in $B_x$, $B_y$, and $B_z$. These fluctuations in the field components were verified as being Alfvén waves, using the method of Belcher and Davis.\textsuperscript{18}

After the peak Dst, the recovery phase begins, which is particularly different for CIR storms. This is the main point of this review. The Dst does not recover to the pre-storm values in a few hours as occurs in ICME storms. It can take several days of a slow “recovery”. Particle injections occurring during these slow “recoveries” were studied by Soraas \textit{et al.}\textsuperscript{22} and Sandanger \textit{et al.}\textsuperscript{23} They found that the slow Dst recovery is caused by freshly injected particles in the ring current. These injections are concurrent with the dissipation process, and thus keep the small ring current energized for a long time. During this time, the auroral electrojet can show an enhanced activity visible by the AE index. This is the case of the event shown in Fig. 2.

\textbf{2.2. High-intensity long-duration continuous AE activity events}

Some CIR storms present a recovery phase longer than two days, and during this time, the AE index shows a persistent high activity. Observing events with these features, Tsurutani and Gonzalez\textsuperscript{24} found some events they called high-intensity long-duration continuous AE activities (HILDCAAs). The event name contains itself the conditions for an interval to be considered a HILDCAA: the event must last longer than two days; during this time the peak AE must be higher than 1000 nT; and the AE index must not decrease below 200 nT for more than two hours at a time. Additionally, the event must occur outside the main phase of a geomagnetic storm.

These HILDCAA events in the Earth magnetosphere are well correlated with the presence of Alfvén waves in the interplanetary medium. The mechanisms associated with the energy transfer from these waves to the Earth’s magnetosphere are still to be determined.

Tsurutani \textit{et al.}\textsuperscript{25} and Guarnieri\textsuperscript{26} used POLAR UVI images to analyze these events at the Earth’s North Pole. Their goal was to verify whether the auroral forms observed during HILDCAAs were similar to those observed during substorms. The Akasofu\textsuperscript{27} criteria for an auroral substorm were employed, and the substorm events found were compared with the AE intensifications. The conclusions of those studies showed a lack of a one-to-one relationship between the AE intensifications and the substorm events.
identified. There are some AE peaks without associated substorm expansion phases, and on the other hand there are also some substorm occurrences during low AE activity. We refer the reader to Tsurutani et al.,\textsuperscript{25} where some intervals of geomagnetic indices were plotted and overlapped to intervals with high auroral activity to show the lack of correlation.

Guarnieri\textsuperscript{26} showed that during HILDCAA events, there are weak or moderate auroras, but distributed over the entire auroral oval. Further, there were images which showed the presence of auroras even in the dayside region. Figure 3 shows a POLAR UVI image taken at 09:38 UT on July 24, 1998, during a HILDCAA event (the same event shown in Fig. 2).

The grayscale in this image qualitatively indicates the photon fluxes (the darker the region, the more intense the photon flux). The auroral intensity during these HILDCAA events is only weak or moderate, with photon fluxes in the range from \( \sim 30–60 \) photons cm\(^{-2}\)/s. This aurora is located mainly between 60\(^\circ\) and 70\(^\circ\) MLAT. All the analyses performed for several events indicated that HILDCAA events are a new form of auroral activity, distributed over wide regions and lasting for several days or weeks, and different from continuous substorms.\textsuperscript{26,28,29}

Several auroral images during magnetic storms, substorms, and HILDCAAs were analyzed. In Fig. 4, we show one image representative of typical cases for each type of magnetic activity. The grayscale used in the images in Fig. 4 is different from that used in Fig. 3, where the scale was changed to make the aurora qualitatively more evident.

![POLAR UVI image showing the auroral forms distributed along the whole auroral oval during a HILDCAA event. The interplanetary parameters and geomagnetic indices for this event are shown in Fig. 2 (adapted from Ref. 28).](image-url)
Fig. 4. Auroral shapes and distribution comparison for storms, substorms, and HILDCAAs. The three types of activity are also different in the time scales and event duration (adapted from Ref. 28).

The image chosen for the storm event was taken on July 15, 2000 at 20:25:30 UT, during the main phase of the “Bastille Day” storm. This is a very intense storm that reached a negative peak Dst of $-301$ nT. The substorm event occurred at 07:01:42 UT on January 27, 2000 (unfortunately the instrument field of view does not include most of the dayside region for this event). The HILDCAA aurora image was taken at 09:36:19 UT on July 24, 1998 during the event shown in Fig. 2. This image was taken only a few minutes before the image in Fig. 3 (the grayscale is different for the figures).

Magnetic storms usually have the most intense auroral photon fluxes, with intensities ranging from $\sim 150$ to $\sim 600$ photons cm$^{-2}$/s. The most intense fluxes are generally located near local midnight with a range of approximately $\pm 3$ hours. During magnetic storms, auroras can be noted at all local times. For extremely intense storms such as this event, auroras expand to middle latitudes.

Substorm auroral intensities can frequently reach peak values as high as $\sim 100$ to $\sim 150$ photons cm$^{-2}$/s. Cases where they were as intense and broad as the “Bastille Day” storm auroras were not found. Substorm auroras are confined to a small region, usually located in the midnight sector or close by (see Ref. 27). A two or three hours local time span would be typical. POLAR UVI substorm auroras last from 15 minutes to several hours, consistent with the Akasofu$^{27}$ visible aurora scenario.

HILDCAAs, on the other hand, are characterized by much lower auroral fluxes, from $\sim 30$ to $\sim 60$ photons cm$^{-2}$/s. Some short-duration
intensifications may reach $\sim 100$ photons cm$^2$/s, but this is not a typical case. As mentioned previously, HILDCAA auroras cover the entire auroral oval, giving them a much greater longitudinal span than for substorm auroras.

For auroral event durations, magnetic storms last from a few hours to a few days (if only the storm main phases are considered, without HILDCAA occurrences in their recovery phases; see Ref. 9). Substorm timescales are tens of minutes to up to a few hours.$^{27}$

In summary, magnetic storm auroras are extremely intense but sporadic and short-duration events. In contrast, HILDCAA auroras are low-intensity events that are relatively constant, global, and can last for weeks.

Recently, Guarnieri$^{26,28}$ used POLAR UVI images to compare the relative photon fluxes observed during 14 HILDCAAs and two very intense storm events, which reached Dst peaks of $-173$ and $301$ nT. The goal of those studies was to verify what is more effective: the short-duration but high-intensity photon fluxes observed during storms, or the mild but sustained for several days fluxes, like those observed during HILDCAAs. For the studies, only a limited region in the nightside sector over the north pole (located between 21 and 03 LT and between 50$^\circ$ and 80$^\circ$ MLAT) was used. The images were integrated in time to obtain the relative photon fluxes for each event analyzed. It was found that during the main phases of the two storms the auroral photon fluxes were higher than those during any HILDCAA interval studied. This result is not surprising, considering the “explosive” nature of magnetic storms. The actual differences may be much larger, since very intense storms (as in the case of the “Bastille Day” event) have auroral displays at midlatitudes, outside the sampling region. However, during the recovery phases of magnetic storms, the auroral intensities are much lower than during the main phase. It is noted that HILDCAA auroral intensities are higher than those of magnetic storm recovery phases.

One more fact must be taken into account: while storm main phases last for a few hours and the recovery phase for at most a day (unless another storm takes place or a HILDCAA event occurs in its recovery phase), HILDCAA events can last for weeks or even as long as 27 days.$^{8,24,31}$ So, the HILDCAA events can produce much larger accumulated photon fluxes due to its long-duration character.

As a consequence for space weather, during these prolonged intervals of HILDCAA activity, higher fluxes of relativistic electrons are observed, which can have significant impact over satellite systems. These events are called killer electrons.$^{32-35}$ High-speed streams produce more relativistic
electrons in the outer radiation belts than do ICME-related storms.\cite{8,31,36–38}

The mechanism for electrons acceleration is presently unknown, but two classes of theories involve plasma waves. One mechanism involves electron radial diffusion within the magnetosphere due to long-period PC5 oscillations that break the particles' third adiabatic invariant.\cite{33,39–44}

A second mechanism is energy diffusion by cyclotron-resonant interactions with electromagnetic whistler mode waves, called chorus.\cite{45–49}

This interaction breaks the particles' first adiabatic invariant. It is important to note that both mechanisms are not related to the CIR itself, but with the southward-oriented $B_z$ components of Alfvén waves present in high-speed streams, which cause plasma injections into the magnetosphere.\cite{8,31}

In general, ICME-related storms are more dangerous for ground-based systems and technologies, while the effects associated with CIR-related storms are very harmful to space devices, due to the high relativistic electron fluxes and the high spacecraft charging levels observed during these events.\cite{37,38,50}

3. Conclusions

This paper describes high-speed streams in the interplanetary medium near Earth’s orbit, and its related geomagnetic activity. The geomagnetic effects of these high-speed streams were contrasted with those of ICME storms. The recovery phases of these CIR-magnetic storms or high-speed stream intervals were the main focus of this review. HILDCAAs are the cause of these long “recoveries” and are in turn caused by the southward interplanetary magnetic field components of Alfvén waves in the high-speed streams. These HILDCAA events were shown to be a new form of geomagnetic activity, different from continuous substorms. The auroras during HILDCAAs are distributed along the whole auroral oval, even in the dayside region. These events were shown as capable of producing more photon fluxes than those observed during the recovery time of very intense storms, and this effect can be more dramatic due to the long duration of these events, which could extend for as long as one month. Since these events occur more frequently in the descending phase of the solar cycle, this means that at times more energy may be transferred from the solar wind to the magnetosphere/ionosphere during the declining minimum phase of the solar cycle than in some strong events typical of the solar maximum.
Acknowledgments

FLG would like to thank CAPES (Brazil), project 2547/02-3, and also to FAPESP (Brazil), project 04/14784-4. EE would like to thank the Brazilian FAPESP (2005/03501-4) and CNPq (300104/2005-7 and 470706/2006-6) agencies for financial supports. Portions of this research were performed at the Jet Propulsion Laboratory, California Institute of Technology under contract with NASA.

References


DEVELOPMENT OF PULSATION INDEX FOR SPACE WEATHER

K. KITAMURA, S. WATARI and M. KUNITAKE
National Institute of Information and Communications Technology
4-2-1 Nukui-kitamachi, Koganei, Tokyo 184-8795, Japan

In this paper, we show the result of the statistical analysis using magnetic data from the auroral to low latitude stations from 2000 to 2005, in advance of the actual development of the Pulsation Index. As a result, the latitudinal and local time dependences of the Pc5 power show a good consistency with the previous studies, which revealed the generation and propagation mechanisms of the Pc5 pulsations. It indicates that our data and method used in this analysis are enough accurate to adapt to the actual development of the Pc indices. As another aspect of an importance of the Pc5 Index, we compared the temporal variations of Pc5 power with that of the >2 MeV electron flux observed by GOES10 satellite. The result of the analysis indicates that the enhancement of the >2 MeV electron flux at the geosynchronous orbit has time delay of 2–3 days and could be related with the toroidal and poloidal oscillation of the Pc5 pulsations at the inner magnetosphere.

1. Introduction

The ultra-low-frequency (ULF) pulsations (period: 0.2–600 s) have been studied by many investigators from 1950s and its generation and propagation mechanisms were advanced by leaps and bounds in 1970s and 1980s due to the many ground and satellite observations.1,2 Since the generation and propagation mechanisms are well understood by the previous attentive studies, the study of the ULF wave in terms of the space physics are now thought to be a convergent subject.

However, a recent increase in the importance of the space weather study makes the ULF pulsations useful tools to give significant information to the various kinds of users who need the information about specific magnetic disturbances. For example, the aeromagnetic survey needs the information of the geomagnetic disturbance to avoid an inaccuracy of the survey, because the spatial resolution depends on the background fluctuations of the
geomagnetic fields and the speed of the aircraft. In contrast, the magneto-
telluric method needs specific magnetic fluctuations as an input, because the period of the ULF is a significant parameter to obtain the skin depth of the resistivity. Thus the periodic magnetic disturbances such as a geomagnetic pulsation come to play an important role in the space weather.

In addition, Pc5 range (period: 150–600s) of ULF pulsations recently attracts many researcher’s attentions as a prevalent candidate of an acceleration mechanism of the relativistic electron in the outer radiation belt. The large increase of the relativistic electron causes the dose and bulk charging effect in spacecraft in most Earth orbit, so that the monitoring and prediction study on the relativistic electron at satellite orbit becomes a more important subject on the space weather study.

We start a development of the pulsation indices using the ground magnetic data obtained by National Institute of Information and Communications Technology (NICT) in order to provide the information as one of the space weather services to various kind of users. In this paper, we show the result of the statistical analysis of the Pc5 pulsations using magnetic data from the sub-auroral to low latitude stations from 2000 to 2005, in advance of the actual development of the indices. Then, we compare the present results with the previous works for the Pc5 and make some considerations to confirm the long-term variations of the magnetic data and the assurance of our method.

2. Data Analysis

The magnetic data from KSM (King Salmon, magnetic latitude = 58.01°, magnetic longitude = 257.88°), PTK (St. Paratunka, magnetic latitude = 45.58°, magnetic longitude = 221.13°), OKI (Okinawa, magnetic latitude = 16.87°, magnetic longitude = 198.41°), and YAP (Yap Island, magnetic latitude = 0.38°, magnetic longitude = 209.21°) in the interval from 2000 to 2005 are used in this study. The sampling time of the data is 1 min and the clock of the magnetometer are corrected by GPS with time accuracy of 1 s. Kozyreva et al. demonstrated the new method to detect the narrow-band component of the power spectral density (PSD) for the Pc5 pulsation. We adopt this method to the fast fourier transform data calculated for every 1 h. Then, we average the hourly PSD value obtained by the above procedure over 24 h to make a daily value of the PSD of Pc5 at each station. Hereafter, we denote the logarithm of PSD of the Pc5 by Pc5 power.
Figure 1 shows the occurrence frequency of the Pc5 power at four ground stations. The horizontal and vertical axes represent common logarithm of the power spectral density and the number of count, respectively. This figure shows a clear latitudinal dependence. The Pc5 power at maximum occurrence increases with increasing latitude, that is, the Pc5 powers at maximum occurrence are about \(-1 \log_{10}(nT^2/Hz)\) at KSM, \(-1.6\) at PTK, and \(-2.2\) at OKI. The distribution of the occurrence frequency at YAP is clearly different from the other stations, that is, the distribution seems to be broader than that at other stations and the Pc5 power at maximum occurrence is about \(-1.8\) which is larger than OKI though the latitude is lower than OKI. However, the half-width of the distribution at YAP is \(\sim 2(\log(nT^2/Hz))\), almost the same as that at KSM and PTK, while the apparent distribution at YAP seems to be broader. This indicates that the smaller number of total count than other stations makes the apparent distribution of the occurrence frequency broader, though the distribution of the occurrence probability at YAP is not broader. On the other hand, the singularity of the Pc5 power at maximum occurrence can be caused by the equatorial enhancement of Pc5 power due to the Cowling conductivity at the dip-equator.

The MLT dependence of the Pc5 power at four ground stations is shown in Fig. 2. The MLT dependence of the Pc5 power is obviously different in each station. The background level of the Pc5 power increases with increasing latitude, except for the YAP. The MLT dependences at KSM and PTK, which are middle latitude stations, show dual peaks at 8 and 13–14 h MLT (indicated thick arrows). The peak at 8 h MLT is more distinct than that at 13–14 h MLT. In contrast, the MLT dependences of Pc5 power at OKI and YAP which are low and equatorial stations have single peak at 10 h MLT. Moreover, the Pc5 power at YAP strongly enhances in the daytime sector (from 6 to 18 h MLT). This enhancement can be explained by the Cowling conductivity in the equatorial ionosphere and causes the singularity in the latitudinal dependence shown in Fig. 1. Such a difference between the low and middle latitude may be caused by the difference of the source of the Pc5 wave. We will discuss this in detail in the next section.

We calculated the cross-correlation coefficients (CCs) between the Pc5 power and the logarithm of the > 2 MeV electron flux obtained by GOES10 satellite. In order to investigate a time delay in the temporal variations between the Pc5 power and the electron flux, we calculated the CCs with the electron flux data which were shifted in time from \(-8\) to 8 days. The result is shown in Fig. 3. The vertical axis and horizontal axis represent
the CCs and time delay, respectively. The maximum CCs are 0.47 (at KSM), 0.46 (PTK), 0.42 (OKI), and 0.38 (YAP), respectively. It means that the maximum CCs increase with increasing latitude even though the CC of less than 0.5 does not represent a good correlation. In general, the compressional mode wave of Pc5 is predominant in the low latitude, whereas the Alfvén mode waves which are generated by the field line resonance near the plasmapause are predominant in the high latitude. Thus, the above result indicates that the temporal variations of the electron flux observed by GOES satellite is accompanying the Alfvén mode waves of the Pc5 rather than the compressional mode wave. Figure 3 also shows the distinct time delay in the maximum CCs at all the stations. Since the sampling rate of the data is one day, it is difficult to determine the details of the time delay, but time delays of two or three days are obvious in each panel of Fig. 3.
Development of Pulsation Index for Space Weather

MLT dependence of the Log10(Pc5 Power)

Fig. 2. The MLT dependence of the Pc5 power at four ground stations. The horizontal and vertical axes represent magnetic local time (h) and common logarithm of the power spectral density $[\log_{10}(nT^2/Hz)]$, respectively.

3. Discussion and Summary

The Pc5 pulsations observed on the ground are thought to be a superposition of two wave sources, which are Alfvén wave predominant in the higher latitude and the compressional wave propagating at the equatorial plane in the magnetosphere. The electric field accompanying the Alfvén wave is imposed on the polar ionosphere and instantaneously transmits to the equator. The ionospheric currents caused by such electric fields produce the magnetic perturbation whose amplitude decreases with decreasing latitude. However, the amplitude of the Pc5 associated with the ionospheric current strongly enhances at the equatorial region because of Cowling conductivity. On the other hand, Motoba et al.\textsuperscript{12,13} reported that the Pc5 pulsations, which are generated by the oscillation of the solar wind dynamic pressure, could be observed not only at high latitude but also at low latitude and indicated that their propagation mechanism could be explained by the same model as the geomagnetic sudden commencement.
Fig. 3. Time delay of the maximum correlation coefficients (CCs) between the Pc5 power and >2 MeV electron flux at GOES10 satellite. The horizontal and vertical axes represent time delay (days) and CCs, respectively.

In this case, the low-latitude Pc5 on the ground could be caused by the compressional waves and the power of wave is less likely to have longitudinal dependence. The latitudinal dependence of the Pc5 power shown in Fig. 1 can be explained in consequence of the superposition of the above two modes of the wave source.

The difference of the source wave of Pc5 can be seen in Fig. 2, which represents the local time dependence of the Pc5 power at four different stations. The local time dependence of the Pc5 power at two higher latitude stations (KSM and PTK) have dual peak at 8 and 13–14 h MLTs, whereas that shows the single peak near 8 h MLT at the two lower latitude stations (OKI and YAP). Furthermore, the dual peaks of the local time dependence of Pc5 power at KSM and PTK have remarkable dawn-dusk asymmetry, that is, the amplitude of the peaks at pre-noon sector are much larger than that at post-noon sector. These signatures are generally consistent with the previous works based on the ground and satellite observations.
Yumoto et al.\textsuperscript{20} considered that the dawn-dusk asymmetry of the Pc5 power must be associated with the dawn-dusk asymmetry of the dominant-mode occurrence in the outer magnetosphere. They thought that the toroidal Pc5 is predominant in the dawnside, whereas the poloidal and compressional Pc5 is predominant in the duskside. The compressional Pc5 observed in the duskside on the ground reduces its amplitude due to the screening effect by the atmosphere and the ionosphere. The present result well corresponds to the interpretation by Yumoto et al.\textsuperscript{20} Thus, we think that the toroidal and poloidal oscillations are predominant at two higher latitude stations, whereas the compressional mode waves are predominant at the two lower latitude stations.

Another aspect of the Pc5 pulsation to improve the space weather study is the relationship with the relativistic electron flux at the geosynchronous orbit. The Pc5 waves are thought to be a prevalent proxy of the flux of the relativistic electrons by several satellite observations and theoretical studies, though there are some physical models as candidates of the acceleration.

Figure 3 shows some important features about the relationship between the Pc5 and relativistic electron flux. The maxima of the CCs are in the range of 0.4–0.5 which are not so good correlation. However, the clear peak of the CCs as a function of the time delay of the electron flux to the Pc5 power indicates that the affection of the Pc5 on the electron flux at geosynchronous orbit is not instantaneous but has somewhat time constant of 2–3 days. Mathie and Mann\textsuperscript{21} reported that the time delay of the increase of the electron flux to the Pc5 is about 1–2 days in the storm-time event. The time delay of 2–3 days in our result is longer than their result. We think that the longer time delay is due to the difference of the event selection, that is, our result is obtained by total days from 2000 to 2005 including the quiet days, whereas the result of the Mathie and Mann\textsuperscript{21} is obtained only by the storm-time events. Moreover, Kataoka and Miyoshi\textsuperscript{22} recently indicated the characteristics of the increase of relativistic electron flux at the geosynchronous orbit have significant difference between the storms caused by the corotating interaction regions (CIRs) and by the coronal mass ejections (CMEs). They reported that the rise time of the electron flux in the CME events are shorter than that in the CIR events. This difference may affect the inconsistency of the result of our study and Mathie and Mann\textsuperscript{21}.

The various kinds of acceleration mechanisms of the relativistic electrons by Pc5 were presented by several authors. Summers and Ma\textsuperscript{5}
indicated that the fast-mode wave in the Pc4 to Pc5 frequency range, which typically observed wave amplitudes, can accelerate the seed electrons to energies of order MeV in a period of a few hours. On the other hand, there is another mechanism, that is, the electron acceleration could be due to a drift resonant interaction with Pc5.\(^3\,4\,23\) In this model, the radial electric fields generated by the toroidal and the poloidal mode oscillations at geosynchronous altitudes play an important role for the accelerations.

The maximum CCs show the slight but obvious latitudinal dependence in Fig. 3, that is, the maximum CCs increase with increasing latitude. This signature would give us a significant suggestion about the acceleration process of the relativistic electron at geosynchronous orbit. As we denoted above, the Pc5 at higher latitudes are generated mainly by the toroidal and poloidal oscillations at the outer magnetosphere whereas the compressional wave is dominant at the low latitude region. The increase of the CCs with latitude indicates the relationship between the relativistic electron flux and the Alfvén mode wave of Pc5. This fact indirectly supports the drift resonant interaction with the toroidal and poloidal oscillations to accelerate the seed electron to the energy range of MeV. However, we need more attentive studies as future works to clarify the acceleration mechanism of the relativistic electrons.

We could summarize the present study as follows:

1. We started the development of the pulsation indices, especially for the Pc5 pulsation, by using the ground magnetic data obtained by NICT.
2. The latitudinal dependence of the Pc5 power shown in this study is basically consistent with the previous studies so that we could confirm the long-term variations of the magnetic data and the assurance of our method for Pc index.
3. The MLT dependence of the Pc5 power indicates that the toroidal and poloidal oscillations are predominant at the high latitude stations and the compressional mode waves are predominant at the low latitude stations.
4. The cross-correlation analysis between the Pc5 power and the > 2 MeV electron flux at geosynchronous orbit shows that the increase of the electron flux have time delay of 2–3 days to the increase of Pc5 power. The latitudinal dependence of the CCs suggests that the acceleration of the relativistic electrons could be related to the toroidal and the poloidal oscillations of the field lines at geosynchronous orbit rather than the compressional mode waves.
Development of Pulsation Index for Space Weather

Acknowledgment

The electron flux data obtained by GOES10 satellite were provided by the NOAA through the anonymous ftp site.

References

This page intentionally left blank
ATMOSPHERIC NEUTRAL ANALYZER FOR IN-SITU NEUTRAL MASS COMPOSITION AND VELOCITY DISTRIBUTION MEASUREMENTS IN IONOSPHERE–THERMOSPHERE COUPLING STUDIES

P. V. AMERL*, E. P. KING and A. W. YAU

Department of Physics and Astronomy
Institute for Space Research, University of Calgary
Calgary, Alberta T2N 1N4, Canada
*amerl@phys.ucalgary.ca

The atmospheric neutral analyzer (ANA) is a new instrument for measuring in-situ the mass composition and velocity distributions of neutrals in the upper atmosphere. It combines radio frequency (RF) mass analysis and charge-coupled device (CCD) particle imaging to provide images of detailed two-dimensional velocity distribution functions of individual mass species. It comprises an entrance aperture (EA), electron source (ES), ion accelerator (IA), RF ion mass analyzer (MA), and an imaging particle detector (PD). The EA permits neutrals to enter the analyzer but inhibits ion and electron entry. The ES ionizes a collimated fraction of the incident neutral particles while preserving their velocities. The IA accelerates all ionized neutrals to a fixed energy in the perpendicular direction to the entrance aperture plane. The MA further accelerates those of a specific velocity (and hence mass) for detection, using an RF electric field of specific amplitude and frequency. In the imaging PD, the detected ions produce charges on the surface of a micro-channel plate detector, and a phosphor screen converts these charges to a visual image of the incident velocity distribution and registers the image on a CCD detector.

We present the design of ANA and discuss its development in preparation for flight in ionosphere–thermosphere missions.

1. Introduction

The atmospheric neutral analyzer (ANA) is a new instrument being developed for measuring in-situ the composition and detailed two-dimensional (2D) velocity distributions of neutrals in the upper atmosphere. Measurements of nonradiating upper atmospheric and thermospheric neutral constituents can be made using in-situ techniques. Measurements for radiating species can also be made using remote-sensing techniques.
Over the past decades, remote-sensing measurements of the neutral composition, density, and in some cases velocity were performed from satellites and rockets.\textsuperscript{1–4} In-situ measurements of nonmass resolved composition and density were made using a variety of mass spectrometers.\textsuperscript{5} In-situ measurements of nonmass resolved density and cross-track acceleration have also been made using accelerometers.\textsuperscript{6,7} The spatial resolution of these measurements was generally of the order of or greater than 10 km.

The coupled ionosphere–thermosphere system is one of the least studied geospace regions important for understanding the energy and momentum transfer from the magnetosphere to the thermosphere via the ionosphere.\textsuperscript{8,9} Because of the large variations in neutral number density and ion gyroradius with altitude in this region, the ion–neutral collision rate is strongly altitude dependent, as is the spatial scale of ion–neutral interactions.\textsuperscript{10} Thayer \textit{et al.}\textsuperscript{11} demonstrated that Joule heating and cooling rates are altitude dependent. Early measurements by DE-2, show the ion–neutral interactions to be dependent not only on instantaneous magnetospheric forcing but also on the history of the ionosphere, thermosphere, and magnetosphere. The heating is also expected to have a high spatial variability with scale sizes as small as 10 km, as it is believed to occur in localized regions of field aligned currents and parallel electric fields at the edges of auroral arcs.\textsuperscript{12} These are expected to give rise to vertical neutrals to speeds of 100 m/s or greater above the F-region. Indeed, horizontal and vertical speeds of this magnitude have been observed over the last few decades.\textsuperscript{13} In addition, the interaction between electrons and neutrals may lead to enhancements in molecular ion temperatures as seen during subauroral ion drifts.

The scientific aim of ANA is to study the small-scale structure and dynamics of the coupled ionosphere–thermosphere. Specifically, to study temperature enhancements in the top-side ionosphere and in the auroral zone, horizontal and vertical winds, and to determine space weather effects on the neutrals often observed as anomalous atmospheric drag. Of particular interest are the velocity responses of the thermosphere to auroral substorms and the relationship between the density and temperature enhancements of neutrals and those of ions and electrons. To meet these objectives, ANA will measure the composition, density, temperature, and velocity of both radiating and nonradiating neutrals at 1-s temporal and < 10-km spatial resolutions. Figure 1 shows the instrument design with some of its important components.
Fig. 1. ANA instrument comprising an entrance aperture, electron source, ion accelerator, RF analyzer, and an imaging detector. The neutrals enter the entrance aperture, are ionized by the electron source, are accelerated downward into the RF analyzer where their flight path is folded by DC electric fields while simultaneous RF fields are used to selectively accelerate ion of specific mass-to-charge ratios. Lastly the accelerated ions are imaged on the imaging particle detector.

2. Instrument Design and Operation

The ANA sensor combines the techniques of radio frequency (RF) mass analysis and charged-coupled device (CCD) particle imaging. It comprises an entrance aperture (EA), electron source (ES), ion accelerator (IA), RF ion mass analyzer (MA), and an imaging particle detector (PD), as shown in Fig. 2. The EA permits neutrals to enter the analyzer but inhibits ion and electron entry. The programmable electron source ionizes a fraction of the incident neutral particles while preserving their horizontal velocities. The ion accelerator draws ionized neutrals into the MA and accelerates them to a specific energy-per-charge. In the MA, ions of specific velocity (and therefore mass-per-charge) are selectively accelerated using specific RF voltages toward the sensor grids. The imaging PD employs a pair of micro-channel plates (MCP) and a phosphor screen (P) to convert the charges carried by the ions at the surface of the MCP to a visual image that is registered on the CCD detector, producing an image incident 2D velocity distribution of a selected neutral mass species.
2.1. Electron source

The ES uses the technique of electron impact ionization and is adapted from the source design of Erdman and Zipf.\textsuperscript{14} It generates a continuous, focused electron beam of fixed energy in the 70–130eV range, normal to the EA (upward in Fig. 2). The beam ionizes a small fraction of the incident neutrals. By virtue of the negligible electron-to-neutral mass ratio, the ions retain their incident velocities. The Faraday cup (FC) collects the emitted electrons and serves to calibrate the electron beam intensity in-flight. Figure 3 shows a cross-sectional view of the cylindrical electron source. The electron source is constructed of non-magnetic stainless steel and ceramic insulating spacers. It has a length of 44 mm (including the filament leads), a diameter of 28 mm, and a mass of 88.5 g. Its filaments are made of tungsten–rhenium alloy (W75Re25), chosen in favor of tungsten or rhenium because of its better structural strength than pure tungsten and higher electron emission compared with
Fig. 3. Cross-sectional view of the stainless-steel electron source along its cylindrical axis. The electrons are generated by a hot filament near the top, accelerated and focused by the electrodes until they exit at the bottom through a 2-mm aperture.

pure rhenium. The operating power for a 0.125-mm diameter tungsten–rhenium filament was typically between 3 and 4.5 W and operating at temperatures of 2200 K, well below the 3100 K melting temperature of the filament and generating electron beams in the range from 10 to 30 µA.

For most species in the thermosphere, including He, N, O, O₂, N₂, NO, and Ar, the ionization cross-section near 70 eV lies in the range of 0.6 × 10⁻¹⁶ to 2.7 × 10⁻¹⁶ cm². This variation, combined with the neutral density variations between the altitudes of 500 and 150 km, results in ion production rates in the range of 10⁵ to 10¹⁰ ions/s. Accelerated time-of-life laboratory testing has demonstrated that the source filament life exceeds 220 h with e-folding times of 130 h.
Figure 4 shows an example of the focusing achieved in laboratory test. The total beam cross-section is 1.4 mm high by 1 mm wide and approximately elliptical in shape.

2.2. Radio frequency mass analyzer

The radio frequency ion mass analyzer design is modified from that of the Akebono suprathermal mass spectrometer instrument.\textsuperscript{20} The IA shown in Fig. 2 deflects the ionized neutrals into the MA through grids AG1 and T1, which accelerate all the ions to the same energy-per-charge. Thereafter, the ions enter the MA comprising three RF regions, (R3, R4, and R5), where a DC electric field folds the trajectory of all the ions and a superposed RF electric field selectively accelerates only the accelerated ions in each region: these are ions that are in phase with the RF field whose vertical $y$-velocity components match the phase velocity of the RF field.

The MA achieves mass-per-charge selection simply by selectively accelerating only ions of a particular velocity for subsequent detection, since all the ions entering the RF section have the same energy-per-charge.
in the $y$-component. Figure 5 shows the residual energy of ions with a mass-per-charge of 14, 16, and 18 amu/e, respectively, that have been resonantly accelerated into the imager. The adjacent ions are well separated in frequency and thus resolved. Setting the RF at the centre of each peak will permit mass-resolved measurement of the velocity distribution of each species separately.

### 2.3. Particle imager

The imaging PD is comprised of a pair of MCP, a phosphor screen, and a reducing fiber-optic taper that is coupled to a CCD detector. Figure 6 shows a schematic diagram of the detector. As a resonant ion reaching the detector strikes the MCP surface, the MCP plates amplify the ion signal by generating the secondary electrons of order of $10^6$. The secondary electrons are then accelerated toward the phosphor screen and interact with the phosphor to create a 2D image of the particle arrival positions, which is equivalent to an image of the 2D incident velocity.
3. Summary

We present in this paper the design of the ANA, a new imaging instrument under development for measuring in-situ the composition and detailed 2D velocity distributions of neutrals in the upper atmosphere. The ANA sensor combines the techniques of RF mass analysis and CCD particle imaging. It consists of an EA section, ES, IA, RF ion MA, and an imaging PD. The EA permits neutrals to enter the analyzer but inhibits ion and electron entry. The programmable electron source ionizes a fraction of the incident neutral particles while preserving their horizontal velocities. The ion accelerator draws ionized neutrals into the mass analyzer and accelerates them to a
Fig. 7. Simulated image of a Maxwellian $N_2$ distribution with a temperature of 1000 K and a spacecraft ram speed of 7 km/s. The simulation assumes that each pixel corresponds to a square surface area of 1 mm$^2$ on the MCP. The intensity represents the number of ions received at each 1 mm$^2$ square.

specific energy-per-charge. In the mass analyzer, ions of specific velocity (and therefore mass-per-charge) are selectively accelerated using specific RF voltages toward the sensor grids. The imaging PD employs a pair of MCP and a phosphor screen to convert the charges carried by the ions at the surface of the MCP to a visual image that is registered on the CCD detector, producing an image of the incident 2D velocity distribution of the neutral particles in the plane of the EA.

Acknowledgments

We wish to acknowledge the support for the development of ANA from the Canadian Space Agency (CSA) Space Technology Development Program (STDP) and the Natural Science and Engineering Research Canada (NSERC) Industrial Research Chair (IRC) Program. We also thank A. Russell and J. M. Noel for fruitful discussions.
References

RECONSTRUCTION OF NONLINEAR FORCE-FREE
FIELDS AND SOLAR FLARE PREDICTION

M. S. WHEATLAND
School of Physics, The University of Sydney
Sydney, NSW 2006, Australia

A brief review is presented of methods for calculating nonlinear force-free fields, with emphasis on a new, fast current-field iteration procedure. The motivation is to reconstruct coronal magnetic fields using high-resolution vector magnetic field boundary data from a new generation of spectro-polarimetric instruments. Methods of solar flare prediction are also reviewed, with focus on the need to reproduce observed solar flare statistics. The event statistics method is described, as well as an extension of the method to incorporate additional information, based on Bayesian predictive discrimination.

1. Introduction

Coronal magnetic fields around sunspots provide the energy for large-scale solar activity, in particular solar flares and coronal mass ejections, so there is considerable interest in accurate descriptions of that field. Space weather effects associated with flares and CMEs motivate solar flare prediction. Flare prediction is currently in a nascent state, with existing prediction methods being imprecise and of limited practical use. Improved knowledge of the coronal magnetic field holds the promise of improved flare prediction.

It is difficult to infer the magnitude and direction of the coronal magnetic field. Measurements of the polarization state of certain magnetically sensitive spectral lines permit inference of the vector magnetic field in the low solar atmosphere (at the photosphere or chromosphere) for regions on the solar disk, subject to a number of uncertainties. There are substantial instrumental errors (in particular in the measurements associated with the components of the field transverse to the line of sight) and uncertainties due to approximations in the inversion of spectro-polarimetric measurements to give magnetic field values.\textsuperscript{1–3} Also, the magnetic field transverse to the line of sight is determined up to a 180°
ambiguity, which is usually resolved by an ad hoc method. Despite these problems, photospheric and chromospheric vector magnetic field values provide the most detailed information available on the state of the magnetic field in solar active regions.

A new generation of spectro-polarimetric instruments will soon provide a wealth of photospheric vector magnetic field determinations with high spatial resolution. The National Solar Observatory’s ground-based Synoptic Long Term Solar Investigations of the Sun Vector Spectromagnetograph (SOLIS/VSM), as well as the space-based Solar Dynamics Observatory Helioseismic and Magnetic Imager (SDO/HMI) and Solar-B Solar Optical Telescope (Solar-B/SOT) feature detectors with thousands of pixels on a side. In principle, measurements with these instruments may be used to better understand magnetic energy storage in solar active regions, and to improve flare prediction. However, the use of the data for these purposes hinges on the ability to accurately “reconstruct” the magnetic field in the solar corona, based on the field values at the photosphere. This problem presents a considerable challenge.

This paper reviews the approach to the problem based on the assumption that the magnetic field in the corona (and at the boundary) is force-free, i.e., has a vanishing Lorentz force. A new, fast approach to the solution of the nonlinear force-free equations is described in Sec. 2, as well as the prospects for application of the method to solar data. In Sec. 3, the state of flare prediction is discussed, with emphasis on the problem of reproducing observed flare statistics. The event statistics method of prediction is described. This is a simple approach which uses only the past history of flare occurrence to make a prediction. A general approach for incorporating additional information into this method is also outlined.

2. Reconstructing Coronal Magnetic Fields

2.1. Nonlinear force-free magnetic fields

A force-free magnetic field \( \mathbf{B} \) satisfies

\[
(\nabla \times \mathbf{B}) \times \mathbf{B} = 0
\]

as well as \( \nabla \cdot \mathbf{B} = 0 \). Force-free fields provide a simple model for the coronal magnetic field. The justification is that in most situations the Lorentz force density \( \mathbf{J} \times \mathbf{B} = \mu_0^{-1}(\nabla \times \mathbf{B}) \times \mathbf{B} \) is expected to dominate over other forces so that in the static situation the Lorentz force must be close to zero.
The force-free equations may be re-written in the form:

$$\nabla \times \mathbf{B} = \alpha \mathbf{B}, \quad \mathbf{B} \cdot \nabla \alpha = 0,$$

where \( \alpha(\mathbf{r}) \) describes the current distribution. The restrictive cases in which \( \alpha \) is zero and \( \alpha \) is a constant describe potential and “linear” force-free fields, respectively. The general case in which \( \alpha \) varies with position describes “nonlinear” force-free fields, which provide a minimal model for solar coronal magnetic fields.

A key question is whether the force-free equations may be solved for boundary conditions derived from solar spectro-polarimetric measurements. This provides an approach to the problem identified in Sec. 1, i.e., the reconstruction of the coronal field from lower boundary values. The appropriate boundary conditions for Eq. (2) are the normal component of \( \mathbf{B} \) in the boundary, and the value of \( \alpha \) prescribed over one polarity of \( \mathbf{B} \), i.e., over one sign of the normal component of \( \mathbf{B} \). In principle, \( \alpha \) may be estimated from differencing of observationally inferred vector magnetic field values. For example, assuming the vector field is available at the plane \( z = 0 \), then

$$\alpha|_{z=0} = \mu_0 \frac{J_z|_{z=0}}{B_z|_{z=0}},$$

where

$$\mu_0 J_z|_{z=0} = \left( \frac{\partial B_y}{\partial x} - \frac{\partial B_x}{\partial y} \right)|_{z=0}.\quad (4)$$

A basic difficulty to be met is that the field at the photosphere, the height of the measurements, is not force-free. The field is believed to become force-free in the chromosphere, typically at a height of around 500 km.\(^8\) One possible solution to this problem is “preprocessing”, i.e., alteration of the boundary conditions to make them more consistent with necessary force-free conditions.\(^9\)

A variety of methods of solution of nonlinear force-free equations have been investigated, including current-field iteration,\(^10\) magneto-frictional relaxation,\(^11\) and the optimization method.\(^12\) A recent test of methods\(^13\) examined their performance on a directly calculable axially symmetric nonlinear force-free field.\(^14\) The best performing method was a specific implementation of optimization.\(^15\)

Although the different methods have been demonstrated to work on test cases,\(^13\) they are computationally intensive, and hence slow. One simple
measure of the speed of a method is the time taken, as a function of the size of the problem. Since we are concerned with three-dimensional calculations we consider the time taken as a function of \( N \), for a calculation performed on a grid with \( N^3 \) points. As reported in Ref. 13, the time taken by specific implementations of the three methods mentioned above scaled as \( N^5 \) (optimization, magneto-frictional) and \( N^6 \) (current-field iteration).

If full resolution data from the new instruments are to be used, then the methods must cope with \( N \approx 1\text{–}2k \). Based on the reported scalings, calculations of this size may be unfeasible, although it should be noted that most of the instruments provide full-disk data, and data extracted for individual active regions will then be smaller in size. In any case, there is considerable interest in faster implementations of nonlinear force-free methods. Recently, Ref. 16 reported implementations of both optimization and current-field iteration which scale as \( N^4 \).

### 2.2. A faster current-field iteration method

Various implementations of current-field iteration\(^{10} \) have been devised. The general approach consists of a Picard iteration solution of Eq. (2). Specifically, at iteration \( k \) the equations

\[
\nabla \times \mathbf{B}^{k+1} = \alpha^k \mathbf{B}^k
\]

\[
\mathbf{B}^{k+1} \cdot \nabla \alpha^{k+1} = 0
\]

are solved, subject to the boundary conditions

\[
\mathbf{z} \cdot \mathbf{B}^{k+1} \bigg|_{z=0} = \mathbf{z} \cdot \mathbf{B}^{\text{obs}} \bigg|_{z=0} \tag{7}
\]

and

\[
\alpha^{k+1} \bigg|_{z=0,B_z>0} = \alpha^{\text{obs}} \bigg|_{z=0,B_z>0}, \tag{8}
\]

where \( \mathbf{B}^{\text{obs}} \) and \( \alpha^{\text{obs}} \) denote the observed values of these quantities at the boundary \( z = 0 \). For simplicity, we restrict attention to a problem in the half space \( z > 0 \). We also specify the boundary conditions using \( \alpha^{\text{obs}} \) on the positive polarity, although either polarity may be used.

The different implementations of current-field iteration differ in their methods of solution of Eqs. (5) and (6), as well as in their handling of the boundary conditions both at the lower boundary, and at the side and top boundaries of the computational grid.
Recently a particularly simple and fast current-field iteration procedure has been described. The method involves separating the field at a given iteration into $B^{k+1} = B_0 + B^k c + B^k c$, where $B_0$ is the potential field satisfying

$$\hat{z} \cdot B_0 |_{z=0} = \hat{z} \cdot B^{obs} |_{z=0}. \tag{9}$$

The nonpotential component $B^k c$ may be constructed via solution of

$$\nabla \times B^k c = \mu_0 J^k c, \tag{10}$$

where

$$J^k c = \begin{cases} \alpha^k B^k \mu_0 & \text{for } z \geq 0, \\ [-J^k c_x(x, y, -z), -J^k c_y(x, y, -z), J^k c_z(x, y, -z)] & \text{for } z < 0. \end{cases} \tag{11}$$

The construction (11), which is due to Ref. 17, ensures that

$$\hat{z} \cdot B^k c |_{z=0} = 0, \tag{12}$$

as required by Eqs. (7) and (9).

The vector potential $A^{k+1} c$ corresponding to $B^k c$ may be obtained by solving the vector Poisson equation via the Fourier transform method. Since the current density $J^k c$ is specified in all space, no explicit boundary conditions are required, which makes this step particularly fast — the time taken scales as $N^3 \log N$. For comparison, two earlier implementations of current-field iteration using an integral solution to Ampere’s law and solution by finite differences both perform this step with a $N^6$ scaling.

The other equation in the current-field iteration procedure (Eq. (6)), may be solved by field line tracing. For each point $r_i$ in the computational grid, the field line threading the point is traced in both directions. If the field line leaves the grid via the sides or top, then $\alpha(r_i)$ is set to zero. If the field line connects to the lower boundary at both ends, then $\alpha(r_i)$ is assigned based on the boundary values $\alpha^{obs}$ at one end of the field line. The time taken by field line tracing for all points on the grid scales as $N^4$. Hence solution of Eq. (6) is slower than solution of Eq. (5), and determines the overall time taken by the implementation.

The starting point for the method is the potential field $B_0$, i.e., $B^{k=0} = B_0$. The potential field is obtained from the boundary condition $\hat{z} \cdot B^{obs} |_{z=0}$ and needs to be calculated just once. More complete details of the method are provided in Ref. 5.
The new method has been tested\textsuperscript{5} by application to a known nonlinear force-free field,\textsuperscript{14} and to a simple bipole. The tests have confirmed the accuracy of the method, and the $N^4$ scaling of the time taken by the method. Typically the method converges in about 10 iterations.

Figure 1 presents an example of two calculations performed with the new method. The boundary conditions on $\hat{z} \cdot \mathbf{B}|_{z=0}$ are the same in each case, and consist of two Gaussian patches of field with positive polarity and two with negative polarity, representing two nearby bipoles. The color in the background indicates the value of the normal component of the field in the boundary, with light green showing regions with positive polarity and dark green showing regions with negative polarity. The boundary conditions on $\alpha$ are different in the two cases. In the upper case, there are two patches of positive $\alpha$ centered on the positive poles of each bipole, and $\alpha$ on the positive polarity is otherwise zero. In the lower case, there is a positive patch of $\alpha$ centered on the positive pole at lower right, and a negative patch of $\alpha$ centered on the positive pole at upper left, and $\alpha$ on the positive polarity is otherwise zero. The values of $\alpha$ on the negative polarity are an outcome of the calculation.

The two cases shown in Fig. 1 represent nearby bipoles with the same, and with opposite, sign of current helicity ($\mathbf{J} \cdot \mathbf{B}$), and the figure illustrates that the connectivity of current carrying bipoles depends on the currents which flow. The figure also illustrates a breaking of symmetry. For the upper case, the boundary conditions have a $180^\circ$ rotational symmetry, which is reflected in the symmetry of the calculated nonlinear force-free field. In the lower case, the boundary conditions no longer have the $180^\circ$ symmetry, and consequently the field is nonsymmetric.

Preliminary tests of the application of the new method to solar boundary data have also been performed. When the method is applied to solar vector magnetic field data, it is found that the current-field iteration does not converge. This may be due to the large errors in the estimated boundary values of $\alpha$, which can lead to large (spurious) localized values of the current density. It also may be due to the nonforce-free nature of the boundary data, as mentioned in Sec. 2.1. When the boundary data are preprocessed, or just smoothed, the localized currents are reduced and the method is found to converge. However, in this case the boundary conditions have been altered, so it is unclear how accurate the resulting coronal field model is. A basic problem with the application of nonlinear
Fig. 1. Examples of nonlinear force-free fields calculated with a new, fast current-field iteration procedure. This case involves two current-carrying bipoles. In the upper panel, the boundary conditions on the current are that the two bipoles have the same sign of current helicity. In the lower panel, the bipoles have the opposite sign of current helicity.
force-free methods to solar boundary data is that there is no unambiguous measure of the accuracy of the result. The application of the new method to solar boundary data continues to be investigated.

3. Solar Flare Prediction

3.1. Existing methods of flare prediction

A variety of properties of solar active regions are known to correlate with flare occurrence. For example, active regions with certain categories of magnetic complexity\(^{18}\) and certain sunspot classes\(^{19}\) are more likely to produce flares, in particular large flares. Other properties associated with flaring include the length of the sheared neutral line,\(^{20}\) flux emergence,\(^{21}\) moments of quantities derived from vector magnetic field maps,\(^{22}\) and the power-law index of the spectrum of line-of-sight magnetic field values.\(^{23}\)

Although these (and other quantities) correlate with flaring, there is no certain indicator that an active region will flare, and existing methods of flare prediction are probabilistic. The US National Oceanic and Atmospheric Administration uses an “expert system” to issue flare predictions. This system incorporates sunspot classification, as well as a variety of observations and rules of thumb. Predictions are made for the probability of occurrence of at least one flare within a day with a peak 1–8 Å flux between 10\(^{-5}\) and 10\(^{-4}\) W m\(^{-2}\) as measured by the Geostationary Observational Environmental Satellite (an M class flare), and for the probability of occurrence of at least one flare within a day with a peak 1–8 Å flux greater than 10\(^{-4}\) W m\(^{-2}\) (an X class flare). These probabilities may be labelled \(\epsilon_M - X\) and \(\epsilon_X\), respectively.

Observed solar flare statistics provide an important constraint on flare prediction. It is well known that flares follow a power-law peak flux distribution.\(^{24}\) In other words the number \(N(S)\) of events per unit time and per peak flux \(S\) is described by \(N(S) = \text{const} \times S^{-\gamma}\), or equivalently by

\[
N(S) = (\gamma - 1)\lambda_i S_i^{\gamma - 1} S^{-\gamma},
\]

where \(\lambda_i = \lambda_i(t) = \int_{S_i}^{\infty} N(S) dS\) is the total rate of events above size \(S_i\). The power-law index \(\gamma\) is typically found to be slightly less than two.\(^{25}\)

Assuming Poisson occurrence in the prediction interval \(T = 1\) day, the probabilities \(\epsilon_M - X\) and \(\epsilon_X\) may be related to corresponding rates \(\lambda_M - X\) and \(\lambda_X\):

\[
\epsilon_M - X = 1 - e^{-\lambda_M - X T}, \quad \epsilon_X = 1 - e^{-\lambda_X T}.
\]
Using Eq. (13), the rates $\lambda_{M-X}$ and $\lambda_X$ are given by

$$\lambda_{M-X} = \lambda_i S_i^{\gamma-1} \left( S_M^{\gamma+1} - S_X^{\gamma+1} \right)$$

and

$$\lambda_X = \lambda_i S_i^{\gamma-1} S_X^{-\gamma+1},$$

where $S_M$ and $S_X$ are the peak fluxes associated with M and X events, respectively. From Eqs. (14)–(16) the quantity

$$R = \frac{\ln(1 - \epsilon_{M-X})}{\ln(1 - \epsilon_X)}$$

is equal to

$$R = \left( \frac{S_M}{S_X} \right)^{-\gamma+1} - 1,$$

which is independent of $\lambda_i$ and hence constant in time. This provides a simple check of the compatibility of predictions with the power-law size distribution.

Figure 2 plots the quantity $R = \ln(1 - \epsilon_{M-X})/\ln(1 - \epsilon_X)$ for the predictions made by NOAA for the year 2005. The power-law index for

![Graph showing values of $R = \ln(1 - \epsilon_{M-X})/\ln(1 - \epsilon_X)$ for daily NOAA predictions for 2005. The horizontal line shows the value of $R$ predicted by Eq. (18).]
the GOES peak fluxes was estimated for all events in 2005 above size $M$ using a maximum likelihood method, and found to be $\gamma = 1.88 \pm 0.08$. The corresponding value of $R$ predicted by Eq. (18) is shown in Fig. 2 by the horizontal line. The NOAA predictions are rounded, i.e., are restricted to the values 0.01, 0.05, 0.10, 0.15, etc., and some of observed variation is due to this rounding. However, there is more variation than expected on this basis, and hence the NOAA predictions are inconsistent with power-law statistics.

3.2. Event statistics method

The event statistics method of flare prediction is a particularly simple approach which is consistent with observed flare statistics. A prediction is made just on the basis of events already observed.

The method requires an estimate $\lambda_1$ of the rate of small events (events above size $S_1$) and of $\gamma$, and then predictions are made for the occurrence of big events, i.e., events above size $S_2$. According to Eq. (13) the rate of big events is given, in terms of the rate of small events, by

$$\lambda_2 = \lambda_1 \left(\frac{S_1}{S_2}\right)^{\gamma^{-1}}. \quad (19)$$

This estimate may be made even if no big events have been observed. According to Poisson statistics, the probability of at least one big event in a time $T$ is then

$$\epsilon = 1 - e^{-\lambda_2 T}. \quad (20)$$

Equations (19) and (20) provide the required prediction.

The advantage of the method is that, if many small events are observed, the prediction (20) will be accurate. In particular, if $M$ events are involved in the estimation of the rate $\lambda_1$, then it is easy to show that the fractional error in the prediction is

$$\frac{\sigma_\epsilon}{\epsilon} \approx M^{-1/2}. \quad (21)$$

In Ref. 6, a Bayesian version of the event statistics method is developed. Given a sequence of events with sizes $s_1, s_2, \ldots, s_M$ above size $S_1$ which occur at times $t_1, t_2, \ldots, t_M$, the Bayesian version permits calculation of a posterior distribution $P(\epsilon)$ for the quantity $\epsilon$. If the rate of small events is
determined based on a recent interval $T'$ during which $M'$ events occurred with a constant mean rate, then the posterior distribution is given by

$$P(\epsilon) = C \left[ -\ln(1 - \epsilon) \right]^{M'} (1 - \epsilon)^{(T'/T)(S_2/S_1)^{\gamma-1}-1}$$

$$\times \Lambda \left[ -\ln(1 - \epsilon) \frac{S_2}{S_1}^{\gamma-1} \right],$$

(22)

where $\Lambda(\lambda_1)$ is the prior distribution for $\lambda_1$, and $C$ is the normalization constant, determined by the requirement $\int_0^1 P(\epsilon) d\epsilon = 1$.

Reference 7 describes an implementation of the method for whole-Sun prediction of GOES events, and a test of the method on the GOES record for 1976–2003. For each day in this period predictions $\epsilon_{M-X}$ and $\epsilon_X$ were made based on preceding events, and the results were compared with the historical record of whether M and X events did or did not occur. Comparison was also made with the corresponding NOAA predictions for 1987–2003. The event statistics method was found to outperform the NOAA method for prediction of X class flares. However, on average the predictions made by the method (as well as by the NOAA method) were slightly too large, and the predictions were also conservative: for all days in 1976–2003, $\epsilon_X$ was less than 0.5.

3.3. Incorporating additional information

The event statistics method uses only the past history of occurrence of small flares. It neglects all of the flare indicators discussed above, including sunspot classification and magnetic complexity. It should be possible to improve the method by incorporating additional information.

A systematic method for classifying whether an active region is flare-producing or not has recently been presented. The method uses discriminant analysis, an orthodox statistical technique for classification. In Ref. 22, discriminant analysis was applied to moments of quantities derived from vector magnetic field maps to classify active regions as flare-producing or nonflare producing.

The Bayesian version of discriminant analysis is predictive discrimination. Predictive discrimination assigns a probability to membership of a class based on observed properties, and a training sample of class members with corresponding properties. In common with discriminant analysis, it assumes continuous distributions for the properties, so it
is restricted to observed properties which vary continuously (which excludes categorical properties such as sunspot classification). Predictive discrimination is more accurate than discriminant analysis, in particular for small training samples, because it takes into account variability in the training sample.\textsuperscript{28}

Predictive discrimination offers a Bayesian approach to incorporating additional information into the event statistics method of flare prediction. In brief this may be achieved as follows. We consider the relevant classes to be “flare-producing above size $S_1$ within time $T$” and “not flare-producing above size $S_1$ within time $T$”. These classes may be denoted $i = f_1$ and $i = \overline{f}_1$, in an obvious notation. Following the event statistics approach, the observed rate $\lambda_1$ of events above size $S_1$ provides an initial guess $c_1$ for membership of the class $f_1$. The initial guess, which we denote $P(i = f_1)$, plays the role of a prior probability. We assume that the observed data $\mathbf{x}$ used to make the classification is a $d$-dimensional vector, with each element taking a continuous value. A training sample $\{x_{i,jk}\}$, where $j = 1, 2, \ldots, N_i$ labels the number of vector data points and $k = 1, 2, \ldots, d$ labels the components of each vector, is assumed to be available. Each $j$ refers to observations for a particular active region. The components enumerated by $k$ are then relevant properties of the active region, for example, moments of quantities derived from vector magnetic field maps, following Ref. 22. Assuming the data are multinormally distributed, predictive discrimination assigns the probability for membership of the class $i$ as\textsuperscript{27}

\begin{equation}
P(i|\mathbf{x}, \overline{\mathbf{x}}_i, C_i) \propto P(i) P_i(\mathbf{x}|\overline{\mathbf{x}}_i, C_i),
\end{equation}

where $\overline{\mathbf{x}}_i$ and $C_i$ denote the unbiased estimators of the sample means and covariance matrices:

\begin{equation}
\overline{\mathbf{x}}_i = \frac{1}{N_i} \sum_{j=1}^{N_i} x_{i,j}
\end{equation}

and

\begin{equation}
C_{i,jk} = \frac{1}{N_i - 1} \sum_{l=1}^{d} (x_{i,jl} - \overline{x}_{i,l})(x_{i,kl} - \overline{x}_{i,l}).
\end{equation}

For the general case of unequal sample means and covariance matrices for different classes the term $P_i(\mathbf{x}|\overline{\mathbf{x}}_i, C_i)$ on the right-hand side of Eq. (23)
Reconstruction of Nonlinear Force-Free Fields

may be expressed as\textsuperscript{27}

\[ P_i(x|\bar{x}_i, C_i) \propto \left( \frac{N_i}{N_i^2 - 1} \right)^{\frac{d}{2}} \frac{\Gamma \left[ \frac{1}{2} N_i \right]}{\Gamma \left[ \frac{1}{2} (N_i - d) \right]} \left| C_i \right|^\frac{d}{2} \]

\[ \times \left[ 1 + \frac{N_i}{N_i^2 - 1} \delta(x|\bar{x}_i, C_i) \right]^{-\frac{d}{2} N_i}, \quad (26) \]

where

\[ \delta(x|\bar{x}_i, C_i) = (x - \bar{x}_i)' C_i^{-1} (x - \bar{x}_i) \quad (27) \]

is the squared Mahalanobis distance, a measure of the distance of the point \( x \) from the mean of the sample. Equations (23)–(27) provide an estimate for the probability of at least one flare above size \( S_1 \) during time \( T \), based on both observed event statistics and the available additional information. Equations (19) and (20) may then be used to convert this into a prediction for the probability of at least one event above size \( S_2 \) during the time \( T \).

Although the method outlined above gives a general prescription for incorporating additional information into the event statistics method, to date the approach has not been implemented and tested on data. It remains to be seen whether improved predictions result.

4. Summary

This paper presents a brief review of the problem of reconstructing coronal magnetic fields from vector magnetic field boundary data based on the nonlinear force-free model, and of the problem of flare prediction.

A new fast method for calculating nonlinear force-free fields is reviewed.\textsuperscript{5} The time taken by the method scales as \( N^4 \), for calculations on grids with \( N^3 \) points. The motivation for developing fast methods of calculation of nonlinear force-free fields is the reconstruction of coronal magnetic fields based on high-resolution vector magnetic field boundary data, from a new generation of spectro-polarimetric instruments. At present there are unresolved difficulties in applying nonlinear force-free methods to solar data, and these are briefly described.

The event statistics method of flare prediction is also reviewed.\textsuperscript{6,7} This simple procedure uses only the past time history of flaring to make predictions, which have the advantage of being consistent with observed solar flare statistics. A generalization of the method to include additional information, based on Bayesian predictive discrimination, is also
described. The generalization has not appeared in the literature before. The prescription is general, and should permit the incorporation of a variety of additional observational information. It remains to be implemented and tested on solar data.

Acknowledgment

The author acknowledges the support, up to July 2006, of an Australian Research Council QEII Fellowship.

References

This page intentionally left blank
SOLAR RADIO FINE STRUCTURES DETECTED WITH SUPER-HIGH TEMPORAL RESOLUTION IN DECIMETER WAVEBAND

S. J. WANG*, Y. Y. LIU, Y. H. YAN and Q. J. FU
National Astronomical Observatories, CAS
 Beijing 100012, P. R. China
*wsj@bao.ac.cn

Since October 2004, some radio fine structures (FS) were detected with super-high temporal resolution of 1.25 ms, during solar activities in the frequency range of 1.10–1.34 GHz, by spectrometers at Huairou, Beijing. They would reflect energetic particle process on shorter time scale in the low corona. For the three events selected we report the observational characteristics of their FS.

1. Introduction

The solar radio burst emission in dm–m waveband is known to be rather varied. It often reveals spectral fine structures (FS) on dynamic spectra, such as narrowband type III bursts, spikes, fiber or intermediate drift bursts, zebra patterns, pulsations, and others. To some extent, studying such specific signatures can give precise information on physical processes leading to the observed radio bursts, as well as about plasma and magnetic field parameters of the coronal sources and so on. Plasma emission is the emission process for the vast majority of solar radio bursts at decimeter and longer wavelengths. Radio observational characteristics are a narrow band of emission near the plasma frequency and/or its harmonic waves (Chapter 8, Ref. 1). So different frequencies mean different ambient electron densities, and these can lie not only on top of each other in one flux tube, but also side by side in different flux tubes. For an important FS-spike, it was demonstrated by Benz et al. that the spikes are closely related to the electron accelerations region.

A Solar Broadband Radio Spectrometers (SBRS) in microwave band started routine observations after June 2000 at Huairou Station of National Astronomical Observatories (NAOC), Chinese Academy of
Sciences (CAS). After October 2004, a new observation mode went into operation for the radio spectrometer, which is of super-high temporal resolution of 1.25 ms in the range of 1.10–1.34 GHz. The spectral resolution is 4 MHz similar to the former. Some new fine structures have been found with the new mode, for example, fish groups. For the three events selected, we focus on investigating the observational characteristics of those FS in detail.

2. Analysis for Events

Data analysis was performed for the three events which exhibited a complex spectrum. Especially they have very interesting super-FS: fish groups, spikes, and narrowband type III bursts. The main observational characteristics of those super-FS were estimated and the relevant plasma and magnetic field parameters of the burst sources were analyzed. Furthermore, for all of the three events, associated X-ray radiation events were detected by GOES.

2.1. October 31, 2004

In the first place, we see the radio FS of the event on October 31, 2004. As shown in II-(a) of Fig. 1, the radio continua appeared in two phases. The “continuum” emissions are those with broad featureless spectra (here corresponding to two main parts of bright radio emissions). We present here only time profile in single frequency, but not continuum spectra, in order to show the FS superposed on the continua more clearly.

The first phase is much stronger than the second phase. The first flux peak occurred at about 05:27 UT. The radio FS were detected between the two flux peaks in 05:28–05:29 UT at the decay phase of the first radio continuum emission, including many spikes (II-(b) and II-(c) of Fig. 1). Simultaneously, seeing the plot I of Fig. 1, there was only one phase in the X-rays. The X-ray peak was at about 05:30 UT which occurred between the two radio peaks and presented a delay of about 1–2 min after the radio FS. Recently research suggested that radio spikes in microwave band appear not due to the fragmentation of a primary energy release, but due to the “secondary” fragmentation, i.e., due to the presence of the sufficiently strong magnetic field inhomogeneities in the coronal loop. This temporal sequence of the event evolution may suggest that energetic
Fig. 1. The radio emission, spikes, and X-ray radiation on October 31, 2004.
electrons should be accelerated during the first radio phase. The energetic electrons propagated along the inhomogeneous field lines might result in the radio spikes. Then those energetic electrons moved down continuously, causing chromospheric evaporation, and increasing the soft X-ray to its peak.

The polarization degree of the spikes was weak of about 20–30% right-handed polarization (RP). The maximum of frequency drift of the spikes is 2.27 GHz/s, the minimum is 0.91 GHz/s, and the mean is 1.446 GHz/s. Seventy-eight percent of the spikes drifted negative, but twenty-two percent of the spikes drifted positive. The value of frequency drift is a little smaller than the result reported before in lower resolution (10 MHz and 8 ms).\(^6\) Moreover some earlier researches suggested that, at least in metric waveband, spikes observed near the disk center were strongly polarized whereas those near the limb were weakly polarized.\(^7,8\) However this event occurred in the NOAA 10691 AR at position N13W34, but not close to the limb. In decimeter waveband, we have found other spike events which also appeared being weakly polarized but not close to the limb, for example, the spikes in December 27, 2003.\(^9,10\) The polarization of spike is an interesting parameter needed to be interpreted later.

### 2.2. November 3, 2004

In this event, the radio continua appeared in two phases and the first phase is a little stronger than the second phase. The first flux peak occurred at about 03:29–03:30 UT and the second radio peak at about 03:55–04:00 UT. The radio FS, including many fish group bursts (II-(b) of Fig. 2), were detected near 03:26 UT at the beginning of the first radio continuum phase. Simultaneously, seeing the plot I of Fig. 2, there was only one phase in the soft X-ray. The X-ray rises clearly at 03:21, to the peak at about 03:30 UT which occurred almost at the same time as the first radio peak and presented a delay of about 4 min after the radio FS. Radio FS correspond to nonthermal emission and energetic electrons. It was implicated likely that during the rising phase of X-ray, some energetic electrons accelerated before would suffer a complex movement in the magnetic loop to cause the radio FS. After, the energetic electrons poured down continuously, causing chromospheric evaporation, and increasing the soft X-ray to its peak.

The polarization degree of the fish groups was weak of about 10–20% RP. Seeing the plot II-(b) of Fig. 2, we could find that the
Fig. 2. The radio emission, fish group, and X-ray radiation on November 3, 2004.
fish groups are similar to fiber bursts in their shapes. All of the frequency drifts were negative and the mean value was about 400 MHz/s. The mean frequency drift is larger than that of fiber in 2.6–3.8 GHz detected before in lower resolution (10 MHz and 8 ms). The average lifetime of single fish is about 44.0 ms, much shorter than that of single fiber 0.5 s.

2.3. November 6, 2004

In Fig. 3, it is shown that the radio continua appeared in three phases. The radio FS were detected near 00:26 UT, at the beginning of the first radio continuum phase, including many narrowband type III bursts (II-(b) of Fig. 3). Using the data of super-high temporal resolution, the narrowband type III bursts were found appearing a strange structure which seem as a result of fiber bursts modulating spike-like bursts. The FS has a slower frequency drift of about 100 MHz/s (fiber), and a faster drift about 1.8 GHz/s (spike-like). The polarization degree of the narrowband type III burst was very strong near 100% RP. The strange structure was complex and has to be studied later.

Seeing the plot I of Fig. 3, there were three distinct phases in the X-ray that appeared corresponding to the three phases of radio continua. Also three successive flares were observed in 00:11–00:42 UT, 00:44–01:10 UT, and 01:40–02:08 UT which almost come from the same position N10E08, AR 10696. The classifications of them were M9.3/2N, M5.9/2N, and M3.6/2N, respectively. Those observations may indicate that if the relevant coronal activities have three phases, the acceleration processes of energetic electrons have three phases also.

3. Conclusion

The evolution sequence of the radio signatures, including radio fine structures, appeared to be associated with X-ray radiations and even flares. In decimeter waveband, the radio spikes were found appearing weakly polarized but not close to the limb. This is a disagreement with the earlier works which need to be interpreted later. The fish groups are similar to fiber bursts in the shapes, but they have different lifetime. Also the narrowband type III burst appeared a strange structure. All of these need to be studied more later.
Fig. 3. The radio emission, FS, and X-ray radiation on November 6, 2004.
Acknowledgments

This work was supported by the National Nature Science Foundation of China (10225313 and 10333030) and Chinese Academy of Sciences. We wish to thank the anonymous referees for advice on improving this paper.

References

COMMENTS ON THE OBSERVED GALACTIC COSMIC RAY MODULATIONS IN THE HELIOSPHERE

H. S. AHLUWALIA, A. D. ANDREEV, R. C. YGBUHAY
and M. M. FIKANI
Department of Physics and Astronomy
University of New Mexico, Albuquerque, NM 87131-1156, USA

C. LOPATE
Institute for Earth, Oceans, and Space Studies
University of New Hampshire, Durham, NH 03824-3525, USA

This field has rarely been theory driven; instead it has followed a path dictated by exploratory and innovative analyses that provided surprising answers contrary to previously held theoretical notions. It continues to be an exciting field to do research in. The problems pertaining to the energetic particle transport in the turbulent interplanetary magnetic field (IMF), permeating the heliosphere, are challenging to say the least, since the field abounds in controversial issues that remain unresolved for a better part of the last century. We remain committed to developing a more synoptic and an all-encompassing empirical point of view. Our approach is illustrated by analyses of annual mean hourly rates of neutron monitors, muon telescopes (at sea level, at underground sites, and at balloon altitudes), the detectors on board spacecraft, as well as the solar wind velocity and IMF intensity.

1. Introduction

The study of the time variations of cosmic rays has encompassed much of the last century. A diurnal variation of ionization of gases in closed vessels was observed by Wood\(^1\) at the Cavendish Laboratory, long before the discovery of “a penetrating radiation” by Hess.\(^2\) So this year marks the 100th anniversary of the discovery of the diurnal variation of galactic cosmic rays (GCRs)! We take this opportunity to comment on the observed modulations with particular emphasis on diurnal variation, 11-year modulation, and the Forbush decreases. The comments highlight our own research work in these areas, over the past half century, with inferences drawn mainly from the analyses of data obtained with neutron monitors (NMs), muon telescopes
(MTs) located underground and carried aloft on balloons as well as the detectors on board satellites and space probes.

2. Diurnal Modulation

A systematic around-the-clock monitoring of ionization “in closed vessels” (namely, ion chambers) started in the 1930s. Early attempts to establish the heliospheric origin of the daily variation of muon intensity, observed with the shielded ion chambers (ICs) and MTs at global sea level and underground sites, were inconclusive. Main hurdles were: poor counting rate statistics of the detectors; no appreciation of the contributions from the atmospheric and environmental effects; and a lack of knowledge of the coupling functions relating the secondary species (leptons, mesons, baryons) observed on the ground to the GCR spectrum incident at the top of earth’s atmosphere. Also, there was no awareness of the existence of the “heliosphere”, its electrodynamics, and solar activity’s role in causing changes in it. We would be wise to keep in mind that this field has rarely been theory driven; instead it has followed a path dictated by exploratory and innovative analyses that provided surprising answers contrary to previously held theoretical notions.

Using annual mean hourly rate data obtained with a network of NMs, ICs, and MTs at underground sites, properly corrected for atmospheric and environmental effects (Refs. 6–8 and references therein), we established that the steady-state diurnal anisotropy amplitude is independent of GCR rigidity up to a limiting value $R_c$. The heuristic parameter $R_c$ is now found to undergo a solar cycle variation; its physical significance as an important modulation parameter was established later when we showed that it is linearly correlated with the intensity ($B$) of interplanetary magnetic field (IMF) at earth’s orbit; an even better correlation is with the product of the annual mean hourly value of $B$ with the annual mean hourly value of the solar wind speed $V$. An inordinate amount of care is exercised in carrying out these analyses. For example, the annual mean amplitude of diurnal anisotropy is quite small ($<0.5\%$) and it exhibits a remarkable variability. In 1977, it reached a value of $0.04 \pm 0.01\%$ for the underground MTs at Embudo. We devised an extremely sensitive statistical technique to demonstrate the persistence of diurnal variation even when its observed amplitude is small, obviating the argument that diurnal variation is a small effect, akin to finding a needle in the haystack.
We have shown that diurnal anisotropy exhibits two distinct patterns related to the magnetic states of the heliosphere. One state \((A < 0)\) is characterized by the existence of an east–west anisotropy \((A_{\phi})\), aligned tangentially to earth’s orbit \([18\, \text{h local time (LT)}]\). The other \((A > 0)\) is characterized, in addition, by the advent of a radial anisotropy \((A_r)\), aligned along the local noon \((12\, \text{h LT})\). The drift hypothesis also predicts that heliosphere should exist in two states related to the orientation of the solar magnetic dipole moment. When the solar polar field in the northern hemisphere is directed outwards (i.e., \(A > 0\), as in 1970–1980 and 1991–2000) GCR protons drift down from the poles toward the equatorial plane and out along the heliospheric current sheet (HCS). When the solar polar field turns negative (i.e., \(A < 0\), as in 1958–1969 and 1981–1990) the protons drift inwards along the HCS. The numerical simulations indicate that diurnal anisotropy phase change (over a solar magnetic cycle) reported by several colleagues can be understood qualitatively in terms of the drift hypothesis. However, it is not clear yet how much the charged particle drifts actually contribute to the observed phase shift at a given rigidity; few numerical simulations have been performed yet at high rigidities \((R > 10\, \text{GV})\).

3. Transport Parameters

The evidence for the existence of a symmetrical \((G_\theta)\) as well as the asymmetrical \((G_{\theta_a})\) density gradients came from the study of the diurnal anisotropy. In fact, GCR gradients at high rigidities can only be determined from diurnal anisotropy data. A major advance was made in the field when we developed a set of master equations for computing important transport parameters directly from diurnal anisotropy and solar wind data, such as the dimensionless products \(\lambda_\parallel G_r\) and \(r_c G_\theta\), where \(\lambda_\parallel\) is the mean free path parallel to mean \(B\), \(G_r\) and \(G_\theta\) are the radial and transverse components of the gradient and \(r_c (= R/cB)\) is the cyclotron radius of a GCR relativistic proton \((v \sim c)\) of rigidity \(R\). This breakthrough led us to develop a methodology for computing the asymmetric part of the gradient \((G_{\theta_a})\) and the ratio \(\alpha = \lambda_\perp/\lambda_\parallel\) \((\lambda_\perp\) is the mean free path normal to \(B\)) from the observed diurnal anisotropy and solar wind parameters. Following Ahluwalia, we write:

\[
\lambda_\parallel G_r = \left( \frac{3}{v} \right) CV - A_r + A_\phi \tan \psi,
\]
where $C = 1.5$ is the Compton–Getting factor and $\psi$ is the IMF garden hose angle at earth, and

$$r_cG_\theta = \left( \frac{A_\phi}{\cos \psi} \right) - (1 - \alpha) \lambda || G_r \sin \psi . \quad (2)$$

So, annual mean values of the dimensionless product $\lambda || G_r$ may be computed in an independent manner from measured diurnal anisotropy and solar wind parameters; the product is proportional to the diffusive anisotropy.\(^{28}\) Again, following Ahluwalia and Dorman,\(^{26}\) we write:

$$(A_\phi^+) - (A_\phi^-) = (2 \cos \psi) r_c G_\theta^a , \quad (3)$$

where $A_\phi^+$ and $A_\phi^-$ correspond to east–west anisotropy amplitude for away/toward IMF polarities, respectively, and

$$(A_\phi^+) + (A_\phi^-) = (1 - \alpha) \lambda || G_r \sin 2\psi . \quad (4)$$

Since the left-hand side of Eq. (3) represents a difference, the contribution from symmetric gradient, if present, is minimized.\(^{27}\) Also, the value of $G_\theta^a$ is free of seasonal and daily variations in atmospheric temperature and temporal variations of solar activity for muon detectors located on the surface and underground.\(^{29}\)

We computed the value of $\lambda || G_r$ separately for NMs at Deep River, Huancayo/Haleakala and the underground MT at Embudo, for each year and then we took the average of the three values for a year to yield an annual mean value to substitute in Eq. (2) to compute a value for $r_cG_\theta$. Figure 1 depicts a plot of the annual mean value of $\lambda || G_r$, $r_cG_\theta$ and the tilt angle of HCS for 1965–1994; the epochs of solar activity maximum (M), minimum (m), and the solar polar field reversals (hatched areas) are also shown. The following points may be noted:

1. The values of $\lambda || G_r$ are large (>1.0%) in late 1960s and 1980s and small in 1976 and (1992–1994), indicating a Hale cycle dependence. Moreover, its value near the solar minimum for $A < 0$ (1986) is smaller than the value for the solar minimum for $A > 0$ (1976).

2. When $\lambda || G_r$ is large (1968, 1984–1985, 1990–1991), the transverse gradient $G_\theta \approx 0$. Heber et al.\(^{30}\) also find that drift contributions fade with approach to solar maximum. But 1984–1985 is closer to solar minimum. The physical significance of these results is not clear at this time.

3. The value of $r_cG_\theta$ is not constant but varies from one year to next. However, its mean value is negative for $A < 0$ and positive for $A > 0$ epochs; the largest negative value occurs in 1986–1987 period. So, while we
cannot rule out the existence of $G_\theta$, the fact that transverse gradients do not remain steady indicates that we do not understand yet some subtleties of the transport processes in the heliosphere.

(4) Note that $G_\theta \approx 0$ for 1990–1994 when the Ulysses spacecraft embarked on the journey to explore high latitude regions of the sun following solar polar field reversal in 1990. This behavior is in contrast to that observed 22 years earlier (1971–1972) when the annual mean value of $r_c G_\theta$ increased steeply to become positive following the solar polar field reversal for cycle 20. These unexplained results highlight our concern that something is amiss with the drift hypothesis since it is unable to account for the observed long-term changes in $G_\theta$.

(5) At places, the values of $r_c G_\theta$ tend to follow the trend in HCS tilt angle nicely (1972–1977, 1978–1983, 1984–1990) but the overall correlation coefficient is very low (CC = 0.3). Lockwood and Webber also found a strong correlation between the latitudinal gradient and the tilt angle for $>70\text{ MeV/nucleon}$ GCRs for the 1983–1989 period (see their Fig. 4).

(6) Surprisingly, the qualitative trends in our computed values of $r_c G_\theta$ at earth’s orbit are in excellent agreement with those obtained from spacecraft data at irregular intervals of time, at much lower rigidities, in the outer heliosphere. For example, Christon et al. observed a persistent, large, negative latitudinal gradient following abrupt, large decrease of the
tilt angle in early 1985 at locations of Voyager 1, 2 for > 70 MeV/nucleon GCRs. Clearly more work needs to be done to understand these puzzling results and to determine the rigidity dependence of $G_\theta$, if any.

4. Long-Term Modulations

Forbush\textsuperscript{33} established an inverse correlation between sunspot numbers (SSNs) and GCR intensity using IC data. The correlation is valid for a more extensive data set. Figure 2 shows a plot of the 27-day mean hourly CL/NM rate and the 27-day mean smoothed SSNs for 1951–2006; NM rate is normalized to 100% for the month of August 1954. The median rigidity ($R_{m}$) of CL/NM response to GCR spectrum is 11 GV (also see later discussion in Sec. 8). The period covers four complete solar cycles (19–22) and parts of the other two (18, 23) as well as five epochs of solar polar field reversals. There is an inverse correlation between CL/NM rate and the smoothed SSNs as one would expect from Forbush’s work. However, there are significant differences noted below.

1. A repeating pattern is observed in the recovery phase of CL/NM rate; it consists of a broad maximum for $A > 0$ (positive) cycle followed by an inverted “V” recovery for $A < 0$ (negative) cycle, indicating that GCR intensity undergoes a long-term modulation corresponding to solar magnetic (Hale) cycle. The recovery always follows after the polar field has reversed. The reader is referred to Ahluwalia\textsuperscript{34} and references therein for a detailed discussion of this phenomenon.

Fig. 2. Climax neutron monitor 27-day hourly mean rate and 27-day smoothed sunspot numbers are plotted for 1951–2006 (November); see text for details.
(2) As noted by the downward pointing arrows, the recovery during $A > 0$ cycles is to a level lower than those for the $A < 0$ cycles; recovery for (negative) cycle 23 is not complete yet. This effect was noted earlier by Webber and Lockwood\textsuperscript{35} and Ahluwalia.\textsuperscript{34} Potgieter and Moraal\textsuperscript{16} state that they are unable to reproduce this effect in their simulations for NMs that incorporate drifts. Lockwood and Webber\textsuperscript{36} suggest that this difference in intensities on the recovery may result from a modulation in the heliosheath (see their Fig. 7). Earlier, Webber and Lockwood\textsuperscript{37} had argued similarly (see their Fig. 2). We argue that heliosheath-related modulation effects are minimal at earth’s orbit at rigidities greater than 3 GV;\textsuperscript{38} the geomagnetic cut off for CL/NM is 3 GV. A recent numerical simulation supports our view (Potgieter, private communication at COSPAR, Beijing, China, 2006). Reinecke et al.\textsuperscript{39} present an alternate explanation. They interpret the phenomena in terms of “crossovers” of GCR spectra during positive and negative cycles at 6–10 GV, indicated by latitude surveys (see their Fig. 1) with NMs. This effect is also seen at lower rigidities in the spacecraft data for different particle species (protons and helium) but with opposite phase, i.e., one observes a suppression of particles with rigidities below $\sim 1$ GV in negative cycles (1965 and 1987) compared with positive cycles (1977 and 1996); see their Figs. 2 and 3. For spacecraft data, “crossovers” are accounted for in the simulations that include drifts. However, Reinecke et al. are unable to reproduce observed “crossovers” in NM data using the same modulation model. They concede that the two “crossovers” may have different physical causes but do not suggest one. Furthermore, Dorman et al.\textsuperscript{40} dispute the very existence of “crossovers” using their own latitude survey data. So, NM observation remains unexplained.

(3) Unusual variations are seen for the $A > 0$ epoch, for 1973–1975, during recovery from cycle 20. The reader is referred to Ahluwalia\textsuperscript{11} and references therein for a detailed discussion of this observation. Also see discussion below.

(4) For cycle 21, the minimum in GCR intensity occurs in 1982, nearly three years after the solar activity maximum in 1979. A similar situation is observed for cycle 22; SSN maximum is in 1989 and GCR intensity minimum is in 1991. In fact the intensity minimum in 1991 was the lowest ever observed, since continuous monitoring began. This is not expected from the analyses carried out by Forbush. It turns out that the intensity ($B$) of the IMF attains very large values in 1982 and 1991. These observations point to the dominant role played by IMF in causing GCR modulation.\textsuperscript{41}
In the next section we argue that the product $BV$ is a better parameter to study GCR modulation on all time scales.

(5) The shape of modulation profile for cycle 23 is very different from that observed for the prior cycles. Also, it develops a shoulder during its recovery phase. A detailed analysis of this epoch is in progress. However, it is interesting to note that McComas et al. report that Ulysses spacecraft observed a very different heliospheric structure during the declining phase of solar activity cycle 23 compared to previous cycles.

5. Electric Fields in Space

Ahluwalia and Dessler were the first to propose a physical process for convecting GCRs out of the heliosphere. They suggested that GCRs undergo an electric drift ($E \times B$, $E = B \times V$) in Parker’s interplanetary magnetic field spiral ($B$) away from sun, normal to $B$. The drift also contributes to a solar diurnal anisotropy observed on the spinning earth. Parker and Krymsky independently suggested the need for including a diffusive inward flow, in near balance with the convective outward flow, to account for the small amplitude of the diurnal anisotropy. In time, these insights led to the development of the Parker equation, when details of IMF structure and its evolution with time became better understood.

The Parker equation includes contributions by other drifts in an inhomogeneous IMF; an appreciation of their role for causing the heliolatitudinal transport of GCRs came later. Jokipii and Thomas argued that the charged particle drifts lead to “pointy” and “flat” maxima in the recovery phase of the 11-year modulation, at lower GCR rigidities. These patterns are expected to recur during alternate sunspot cycles. We established that an emerging solar polar magnetic field has a significant influence on the recovery profile of 11-year modulation, showing that repetitive patterns exist in the recovery phase of modulation in IC data ($R_m = 67$ GV) obtained at Cheltenham–Fredericksburg. During odd cycles (17, 19, 21) the recovery is completed in 5–8 years whereas for the even cycles (18, 20) the recovery period is reduced to less than half as much. We found no evidence of a “flat” maximum for cycle 18 in IC data. We re-examined this issue using data obtained with NMs at Huancayo (1953–1995) and Climax (1951–1995) as well as IC data obtained at Yakutsk (1957–1995). Again, we found no evidence for a “flat” maximum at higher GCR rigidities for cycle 20 at Huancayo ($R_m = 33$ GV) and Yakutsk.
(\(R_m = 67\) GV), although a broad maximum is present in NM data at Deep River (\(R_m = 16\) GV) and Climax (\(R_m = 11\) GV); see discussion of Fig. 2 above.

6. Modulation Parameter BV

We carried out correlation analyses involving the geomagnetic index \(A_p\), the limiting diurnal anisotropy rigidity \(R_c\), 11-year and 27-day GCR modulations with the solar wind parameters \(B\) and the bulk speed \((V)\) over an extended time period (1963–1998). Interestingly, we found that the time series involving the product of the annual mean values of \(B\) and \(V\) reproduce the observed time variations of the annual mean values of \(A_p\) (CC = 0.92), as well as those of \(R_c\) (CC = 0.8) and GCR variations on time scales of 11-year (CC = 0.84) and 27 days. This points to the importance of the parameter \(BV\) in making short-term forecast for the space weather as well as understanding GCR modulations. The reader is reminded that the parameter \(BV\) is the amplitude of the solar wind electric field.

Figure 3 shows an inverted plot of 27-day hourly average of CL/NM counting rate and the corresponding \(BV\) time series for Bartel’s rotation numbers 1907 (1/1/73) to 1947 (12/17/75), for the period near solar minimum for cycle 20, dubbed “mini-cycle” of GCR modulation by

![Fig. 3. Climax neutron monitor 27-day hourly mean rate and corresponding \(BV\) values are plotted for the Bartel rotation numbers 1907 (1/1/73) through 1947 (12/17/75) for GCR mini-cycle; see text for details.](image)
García-Munoz et al.\textsuperscript{53} These analyses bring out the important role played by solar wind electric field $[\mathbf{E} = \mathbf{B} \times \mathbf{V}]$ in transferring power to the magnetosphere\textsuperscript{54} as well as in the convective removal of GCR from the inner heliosphere.\textsuperscript{11} We have to understand the implications of this line of reasoning on the rigidity dependence of the various forms of GCR modulations, such as 11-year, 27-day, Forbush decreases (FDs), and the solar anisotropies (diurnal and semidiurnal).

Recently, we made some progress to understand the challenge posed by the GCR modulations during cycle 20.\textsuperscript{38} We intend to follow up on this attempt to explore the vital role of the solar wind electric field in magnetospheric and heliospheric phenomena. We now have \textit{in-situ} measurements of the solar wind parameters\textsuperscript{55} and lower rigidity GCRs along Ulysses trajectory. It would be interesting to see if a correspondence exists between the temporal variations of $\mathbf{B}$ and GCR flux at different rigidities along Ulysses's orbit. If so, we may be able to offer an alternate explanation for the observations to those currently available\textsuperscript{56–58} and thereby contribute to an ongoing discussion. We note that progress is being made in understanding the characteristics of GCR transport near the termination shock.\textsuperscript{59–63} This gives us hope that modulation-related problems may be solved in not too distant a future.

More recently, suggestions have been made that time variations of IMF intensity ($\mathbf{B}$) is the basic cause of modulation observed at any rigidity, anywhere in the heliosphere.\textsuperscript{64–67} We disagree with this view,\textsuperscript{68} we show that modulations on all time scales (24 h, 27 days, and 11-year) arise from time variations of the product $\mathbf{B}\mathbf{V}$.\textsuperscript{11} Thereby, we confirm Ahluwalia–Dessler view that convection plays a dominant role in modulating GCR. Since $\mathbf{E}$ depends on $\mathbf{B}$, the observed correlation between $\mathbf{B}$ and modulation parameters is not surprising. Recently, we related the product $\mathbf{B}\mathbf{V}$ to GCR modulation function via the force field approach; see Eq. (3) in Ahluwalia.\textsuperscript{38} We showed heuristically that the diffusion coefficient is proportional to $1/B$ (not $1/B^2$). Also, we argued that heliosheath modulation (at earth’s orbit) is negligible at $R > 3$ GV. Recent simulations have confirmed our inference (Potgieter, private communication at COSPAR in Beijing, China, 2006). Therefore, we question Webber and Lockwood’s\textsuperscript{69} assertion that a heliosheath-related modulation is observable at NM energies (see their Fig. 4). We continue to work on clarifying and resolving the issues from an empirical point of view. The Parker equation can only be solved numerically. Besides, it points to no preferred choice of relevant transport parameters or the configuration of IMF. So an analysis of data must continue in search
of a limited choice for transport parameters, taking account of the new ideas about IMF.\textsuperscript{58} Furthermore, observations show that 3D structure of the heliosphere differs significantly from one solar cycle to the next. In particular, heliospheric structure for cycle 23 seems to be quite different from that for cycles 21 and 22.\textsuperscript{42} So, it is interesting to determine how modulation for cycle 23 is different from previous cycles, over a range of GCR rigidities.

At this point it is interesting to note that Bieber et al.\textsuperscript{70} correlated the annual mean rates of McMurdo (Antarctica) NM ($R_m \sim 16$ GV) with the corresponding values of SSNs, a geomagnetic index (linearly related to $A_p$), $B$, HCS tilt angle, over a magnetic cycle (1965–1988). They were disappointed that, “...none of the several transport theory formulations”, tested by them could produce a correlation significantly better than that between GCR and SSNs. We reached a similar conclusion independently in an analysis involving the modulation parameter $R_c$, applicable to the diurnal anisotropy.\textsuperscript{10} We found that $R_c$ is well correlated to $B$ (CC = 0.7) over an extended time period (1965–1995) but poorly correlated to the IMF turbulence range (CC = 0.56) considered by Bieber et al.\textsuperscript{70} This raises a question whether the scattering processes described by the resonant magnetostatic theory play any role in causing the observed GCR modulations.

7. Median Rigidity of Detector Response

Throughout our discussion in the previous sections, we have emphasized the need to study the rigidity dependence of transport parameters and the observed GCR phenomena in the heliosphere. The rigidity dependence of modulation arises from local as well as global causes. To explore this dependence, one uses data obtained with a variety of detectors at sea level and mountain sites, as well as on balloons, satellites, and space probes. For these studies to be meaningful, it is important to have a clear understanding of the response characteristics of detectors involved. We characterize detectors in terms of their median rigidity of response ($R_m$) to GCR spectrum; 50% of the counting rate lies below it.\textsuperscript{7} Some colleagues define effective rigidity of modulation for NMs in an \textit{ad hoc} manner.\textsuperscript{71} For example, $R_m$ values computed by Lockwood and Webber (L&W) are based on two assumptions. First, modulation ceases above 100 GV. Second, $R_m$ value depends on the modulation function assumed by
them (see their Fig. 2). Our $R_m$ values are computed in a straightforward manner; in particular no assumption is made as to the form of modulation function or the value of $R_c$ for a given epoch. For cycles 21/22, L&W values of $R_m = 5.4/7.0$ GV for Mt. Washington NM may be compared to our value of 10 GV. Later, Lockwood and Webber\cite{Lockwood1984} give $R_m = 10$ GV for Mt. Washington NM in agreement with our value. Later still, Lockwood \textit{et al.}\cite{Lockwood1985} give $R_m = 14$ GV for Mt. Washington NM which exceeds our value; we have shown that solar cycle variation in $R_m$ values is small for the NMs.\cite{Webber1986}

Webber and Quenby,\cite{Webber1987} Lockwood and Webber,\cite{Lockwood1988} Bachelet \textit{et al.},\cite{Bachelet1989} Nagashima \textit{et al.}\cite{Nagashima1990} and Dorman \textit{et al.}\cite{Dorman1991} provide NM response functions. We use the differential response curves of Lockwood and Webber\cite{Lockwood1988} to compute $R_m$ values for several NM sites; these curves are derived from the over-land latitude survey conducted by Carmichael and Berkovitch.\cite{Carmichael1992} Since then data are available from several more surveys undertaken by different research groups over several decades, so $R_m$ values may be computed from these surveys for the global network of NMs to ascertain the range of variations of $R_m$ values for different solar cycles.

8. Forbush Decrease of July 1982

Fillius and Axford\cite{Fillius1983} note that FD of July 14, 1982 reduced GCR intensity to the lowest level for cycle 21, three years away from sunspot maximum, well past the epoch of solar polar field reversal in 1980;\cite{VanAllen1984} see also Fig. 7 in Van Allen and Randall.\cite{VanAllen1985} Tsurutani \textit{et al.}\cite{Tsurutani1986} reported a “Great Magnetic Storm” associated with this event when solar wind velocity reached 1360 km/s with peak $B = 60$ nT and peak Dst = −325 nT (a parameter that indicates the strength of the geomagnetic storm-time ring current). Cliver \textit{et al.}\cite{Cliver1987} discuss solar-terrestrial and heliospheric implications of this event. Lockwood \textit{et al.}\cite{Lockwood1988} studied its rigidity dependence using IMP and NM data only for a limited rigidity range (1–20 GV); $R_m$ values used by them are under-estimated as discussed above (see their Table 1, p. 5449). Fortunately, data are available for this FD from the underground MTs at Embudo, Mawson, and Socorro\cite{Mawson1989} as well as an ensemble of underground MTs at Yakutsk.\cite{Yakutsk1990} So, the rigidity range can be extended for our plot. For MTs we use $R_m$ values computed by Fujimoto \textit{et al.}\cite{Fujimoto1991} based on the theoretical calculations of Murakami \textit{et al.}\cite{Murakami1992} who obtained response functions for muons by solving numerically the equations of hadronic cascades in the
atmosphere; they assumed Feynman scaling for hadronic interactions of GCRs incident at the top of the atmosphere. Their definition of $R_m$ for MTs is the same as ours.

A log–log plot of FD amplitudes and $R_m$ values computed by us (for NMs) and by Fujimoto et al.\(^87\) for MTs is shown in Fig. 4; FD amplitude for Mawson underground MT ($R_m = 164$ GV) may contain a contribution from the isotropic intensity wave.\(^89\) The line represents a power law fit to the data from 1 to 300 GV, the data include those obtained with IMP 8, NMs, and MTs at the underground sites. We believe that $R_m$ value for the IMP data is under-estimated and needs to be revised upwards; we are investigating this issue. An inverse dependence of FD amplitude on $R_m$ is seen very clearly (CC = 0.99). Fenton et al.\(^90\) obtain an exponent of rigidity of $-0.8$ for the Australian network of four NMs; our value ($-0.71$), for a larger range of GCR rigidity (1–300 GV) spectrum, agrees with theirs.

9. Voyager 2 Data

Figure 5 depicts a plot of the data obtained in 1983 onboard Voyager 2 spacecraft at about 11 a.u. At this time a recovery for cycle 22 has set in.
at earth’s orbit (see Fig. 2 above). The data consist of hourly averages of $B$, $V$, product $BV$ and GCR intensity $> 75$ MeV/nucleon, averaged every 6 h. One notes that $BV$ peaks coincide with decreases in GCR intensity, indicating that $BV$ peaks are the signatures of the propagating barriers (MIRs) postulated by Burlaga et al.\textsuperscript{91} So, MIRs are the cause for “steps” noted earlier by Stoker and Carmichael\textsuperscript{92} in NM data. This also implies that enhanced convection at the position of Voyager 2 is responsible for the removal of GCRs in the outer heliosphere. Further out, MIRs form GMIRs.\textsuperscript{93} Burlaga et al.\textsuperscript{91} conclude that the largest modulation occurs under conditions not conducive to large-scale drifts. The recovery at earth should depend upon how and where GMIRs dissolve into the background solar wind. We are beginning to investigate this issue. The results will be reported elsewhere.

In a spherically symmetric approximation, the Parker\textsuperscript{46} equation, for a steady state, yields the so-called convection–diffusion equation,

\begin{equation}
V f - \kappa \frac{\partial f}{\partial r} = 0 ,
\end{equation}
where \( f \) is the GCR distribution function at earth and

\[
\kappa = \beta K_1(r)K_2(R),
\]

(6)

\( \kappa \) is the radial diffusion coefficient at a distance “\( r \)” for a particle of rigidity \( R \).\textsuperscript{94} At high rigidities \(( R > 1 \text{ GV}) \) \( \beta \approx 1 \) and \( \kappa \) is separable into \( K_1 \) and \( K_2 \); also, \( K_2 \) is proportional to \( R \).\textsuperscript{38} The solution for Eq. (5) is given by:

\[
f = f_b \exp(-M),
\]

(7)

where

\[
M = \int V \frac{dr}{\kappa},
\]

where \( M \) (dimensionless) is called the modulation function, its limits are taken from the orbit of the earth \(( r_e )\) to the boundary \(( r_b )\) at the termination shock (TS). Also, \( M \) can be expressed in terms of \( \phi \), the chargeless potential at earth in Volts.\textsuperscript{95}

\[
M = 3 \frac{\phi}{\beta K_2},
\]

(8)

where \( \phi \) is the rigidity loss for a relativistic GCR from TS to the point of observation in the heliosphere. Following Gleeson and Axford\textsuperscript{94} we write,

\[
\frac{J(r, E, t)}{E^2 - \text{(rest mass)}^2} = \frac{J(\infty, E + \Phi)}{(E + \Phi)^2 - \text{(rest mass)}^2},
\]

(9)

where \( J(r, E, t) \) is the GCR differential intensity at the observation point, \( r \), inside the heliosphere, \( J(\infty, E + \Phi) \) is the GCR differential intensity in the local interstellar medium (LIS), the rest mass for proton is taken as 938.8 MeV, \( E \) is GCR total energy, and \( \Phi \) is the energy loss experienced by GCR in coming from infinity; it is related to \( \phi \) as follows:

\[
\Phi = Ze\phi(r, t).
\]

(10)

Figure 6 is a plot of the 27-day average of CL/NM data for 1965–2005 (in the upper and lower panels) on which are superposed the flux of 274.5–398.2 MeV protons from the Goddard Medium Energy instrument, available at the Goddard Space Flight Center website (lower panel) and the modulation energy \( \Phi \) (upper panel) computed using LIS spectrum given by Webber and Highie.\textsuperscript{96} The agreement among the different data sets appears to be satisfactory.
Fig. 6. Climax neutron monitor 27-day hourly mean rate and corresponding values for the particle flux for the Goddard Medium Energy instrument and modulation energy calculated are plotted for 1965–2005; see text for details.

Acknowledgments

HSA thanks the organizers of the Third Annual Asia Oceania Geosciences Society Meeting in Singapore, for the request to prepare an invited review. He is also grateful for the financial support provided to him by the University of New Mexico to attend this meeting.

References

Galactic Cosmic Ray Modulations in the Heliosphere

This page intentionally left blank
In the first part of the paper, neutron monitor data for about four solar cycles are employed in our investigation of the hysteresis effects, convection–diffusion and drift modulations, and by solving the inverse problem determining parameters of the Heliosphere and parameters of high-energy galactic cosmic ray (CR) propagation. In the second part, we solve the inverse problem for small-energy galactic CR employing satellite data; in this case we also take into account the diffusion time-lag.

1. The Inverse Problem for High-Energy Galactic Cosmic Ray Propagation and Modulation in the Heliosphere on the Basis of Neutron Monitor Data

1.1. Hysteresis phenomenon and the inverse problem for galactic cosmic ray propagation and modulation in the Heliosphere

By solving the inverse problem for galactic cosmic ray (CR) propagation and modulation in the interplanetary space on the basis of observation data of CR–solar activity (SA) hysteresis phenomenon important information can be obtained on the main properties of the Heliosphere. Investigation of the hysteresis phenomenon between long-term variations in CR intensity observed at the Earth and SA started about 50 years ago. In the mid 1960s many scientists came to the conclusion that the dimension of modulation region (or Heliosphere) was about 5 AU, and not more than 10–15 AU. It was found that the radius $r_0$ of the CR modulation region
is very small either by analysis of the intensity of coronal green line in some helio-latitude regions (as a controlling SA factor; in this case the result obtained was $r_0 \approx 5$ AU), or by investigation the CR modulation as caused by sudden jumps in SA ($r_0 \approx 10$–$15$ AU). On the other hand in our papers and monographs\textsuperscript{11–17} the hysteresis phenomenon was investigated on the basis of neutron monitor (NM) data for about one solar cycle in the frame of convection–diffusion model of CR global modulation in the Heliosphere, taking into account the time-lag of processes in the interplanetary space relative to processes on the Sun. It was shown that the dimension of the modulation region should be about $100$ AU (much bigger than accepted in those times in literature, 5–15 AU). These investigations were continued on the basis of CR and SA monthly average data for about four solar cycles in Refs. 18–24. Let us note that many authors worked on this problem, used sunspot numbers or other parameters of SA for investigations of CR long-term variations, but they did not take into account the time-lag of processes in interplanetary space relative to processes on the Sun as integral action (see review in Ref. 25). The method, described below, takes into account that CR intensity observed on the Earth at time $t$ is caused by solar processes that started many months before $t$. In our paper\textsuperscript{26} we used this method with CR and SA data for solar cycles 19–22, taking into account drift effects according to Ref. 27. It was shown that it was very important to include drift effects that depend on the sign of solar polar magnetic field (sign of parameter $A$) and determine the difference of total CR modulation at $A > 0$ and $A < 0$, and the effects of the tilt angle between interplanetary neutral current sheet and equatorial plane. It became possible to explain the great difference in time-lags between CR and SA in hysteresis phenomenon for even and odd solar cycles.

1.2. \textit{Hysteresis phenomenon and the model of CR global modulation in the frame of convection–diffusion mechanism}

It was shown in Ref. 11 that the time of propagation through the Heliosphere of particles with rigidity $>10$ GV, to which NM are sensitive, is no longer than one month. This time is at least one order of magnitude smaller than the observed time-lag in the hysteresis phenomenon. This means that the hysteresis phenomenon on the basis of NM data can be considered as a quasi-stationary problem with parameters of CR
propagation changing in time. In this case according to Refs. 28–30:

\[
n(R, r_{\text{obs}}, t) = n_o(R) \exp \left( -a \int_{r_{\text{obs}}}^{r_o} \frac{u(r, t) dr}{\kappa_r(R, r, t)} \right),\]

where \( n(R, r_{\text{obs}}, t) \) is the differential rigidity CR density, \( n_o(R) \) is the differential rigidity density spectrum in the local interstellar medium out of the Heliosphere, \( a \approx 1.5 \), \( u(r, t) \) is the effective solar wind velocity (taking into account shocks and high-speed solar wind streams), and \( \kappa_r(R, r, t) \) is the radial diffusion coefficient that depends on the distance \( r \) from the Sun of particles with rigidity \( R \) at the time \( t \). According to Refs. 12–17 the connection between \( \kappa_r(R, r, t) \) and SA can be described by the relation

\[
\kappa_r(R, r, t) \propto r^\beta \left( W(t - r/u) \right)^{-\alpha},
\]

where \( W(t - r/u) \) is the sunspot number in the time \( t - r/u \). By the comparison with observation data it was determined in Refs. 12–17 that parameter \( 0 \leq \beta \leq 1 \) and \( \alpha \approx 1/3 \) in the period of high SA (\( W(t) \approx W_{\text{max}} \)) and \( \alpha \approx 1 \) near solar minimum (\( W(t) \ll W_{\text{max}} \)). Here we suppose, in accordance with Refs. 18–21, that

\[
\alpha(t) = \frac{1}{3} + \left( \frac{2}{3} \right) \left( 1 - \frac{W(t)}{W_{\text{max}}} \right),
\]

where \( W_{\text{max}} \) is the sunspot number at solar maximum.

According to Eq. (1) the value of the natural logarithm of observed CR intensity global modulation at the Earth’s orbit, taking into account Eqs. (2) and (3), will be

\[
\ln(n(R, r_E, t)_{\text{obs}}) = A(R, X_o, \beta) - B(R, X_o, \beta) F(t, X_o, \beta, W(t - X) | X_E),
\]

where

\[
F(t, X_o, \beta, W(t - X) | X_E) = \int^{X_o}_{X_E} \left( \frac{W(t - X)}{W_{\text{max}}} \right)^{\frac{\beta}{2} + \frac{1}{2}(1-W(t-X)/W_{\text{max})}} X^{-\beta} dX,
\]

\( X = r/u, X_E = 1 \text{ AU}/u, X_o = r_{\text{obs}}/u \) [\( X_E \) and \( X_o \) are in units of average month = (365.25/12) days = 2.628 \times 10^6 \text{s}]}. Solving Eq. (4) based on experimental data will give solutions also for the inverse problem because the regression coefficient \( A(R, X_o, \beta) \) determines the CR intensity out
of the Heliosphere, the regression coefficient $B(R, X_o, \beta)$ characterized the effective diffusion coefficient of CR in the interplanetary space, and $X_o = r_o / u$ characterized the dimension of modulation region. These three coefficients can be determined by correlation between observed values $\ln(n(R, r_E, t)_{obs})$ and the values of $F$, calculated according to Eq. (5) for different values of $X_o$ and $\beta$. In Ref. 18 three values of $\beta = 0, 0.5, 1$ have been considered; it was shown that $\beta = 1$ strongly contradicts CR and SA observation data, and that $\beta = 0$ is the most reliable value. Therefore, we will only consider this value here.

1.3. Even-odd cycle effect in CR and role of drifts for NM energies

To determine $X_{o,\text{max}}$, corresponding to the maximum value of the correlation coefficient for regression equation (4), we compare 11-month moving averages of the Climax NM ($H = 3400$ m, cut-off rigidity $R_c = 2.99$ GV) for solar cycles 19–22 and onset of cycle 23. For each time-lag, $X_o = r_o / u = 1, 2, 3, \ldots, 60$ average months, we determined the correlation between observed and expected CR intensities. The Climax NM data correspond to an effective rigidity of primary CR of about 10–15 GV. For higher energy particles (about 30–40 GV) we used Huancayo ($R_c = 12.92$ GV, $H = 3400$ m)/Haleakala ($R_c = 12.91$ GV, $H = 3030$ m) NM data from January 1953 to August 2000. The results are summarized in Table 1 in columns $A_{\text{diff}} = 0\%$. A big difference can be seen in $X_{o,\text{max}}$ for odd and even solar cycles.

We assume that the observed long-term CR modulation is caused by two processes: the convection–diffusion mechanism (e.g., Refs. 28–31), which is independent of the sign of the solar magnetic field (SMF), and the drift mechanism (e.g., Refs. 27, 32 and 33), what gave opposite effects with changing sign of the SMF. For the convection–diffusion mechanism we use the model described in detail in Ref. 26, shortly given above by Eqs. (1)–(5). For drift effects we use results of Ref. 28, and assume that the drift effect is proportional to the value of the tilt angle $T$ with negative sign at $A > 0$ and positive sign at $A < 0$, and in the period of reversal we again suppose linear transition through 0 from one polarity cycle to other; we assume that average of curves for $A > 0$ and $A < 0$ in these figures characterized convection–diffusion modulation, and difference of these curves double drift modulation. Data on tilt angles for solar cycles 19 and 20 are not available. We used relation between sunspot numbers $W$ and $T$ to make homogeneous
Table 1. Values of $X_{o,max}$ (in average months) for observed data ($A_{dr} = 0\%$) and corrected for drift effects with different amplitudes $A_{dr}$.

<table>
<thead>
<tr>
<th>CY</th>
<th>0%</th>
<th>0.5%</th>
<th>1%</th>
<th>1.5%</th>
<th>2%</th>
<th>2.5%</th>
<th>3%</th>
<th>4%</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>21</td>
<td>18.5</td>
<td>16.5</td>
<td>14.5</td>
<td>12.5</td>
<td>11</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>20</td>
<td>6.5</td>
<td>8</td>
<td>9.5</td>
<td>12</td>
<td>16.5</td>
<td>20</td>
<td>27</td>
<td>34</td>
</tr>
<tr>
<td>21</td>
<td>31</td>
<td>27</td>
<td>23</td>
<td>20</td>
<td>16.5</td>
<td>15</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>22</td>
<td>8</td>
<td>10</td>
<td>11</td>
<td>12</td>
<td>14</td>
<td>16.5</td>
<td>18</td>
<td>24</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CY</th>
<th>0%</th>
<th>0.15%</th>
<th>0.25%</th>
<th>0.35%</th>
<th>0.5%</th>
<th>0.75%</th>
<th>1.0%</th>
</tr>
</thead>
<tbody>
<tr>
<td>19</td>
<td>20</td>
<td>18</td>
<td>16.5</td>
<td>14</td>
<td>12</td>
<td>9</td>
<td>6</td>
</tr>
<tr>
<td>20</td>
<td>10.5</td>
<td>15</td>
<td>18</td>
<td>25</td>
<td>31</td>
<td>39</td>
<td>46</td>
</tr>
<tr>
<td>21</td>
<td>32</td>
<td>23</td>
<td>18</td>
<td>15</td>
<td>12</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>22</td>
<td>9</td>
<td>12</td>
<td>11</td>
<td>12</td>
<td>14</td>
<td>16.5</td>
<td>22</td>
</tr>
</tbody>
</table>

analysis of the period 1953–2000. Based on data for 18 years (May 1976–September 1993), we found that there are very good relation between $\alpha$ and $W$; for 11-month smoothed data:

$$T = 0.349 W + 13.5^\circ$$

with correlation coefficient 0.955. An example for correction of observed CR intensity on the drift effects (to obtain only convection–diffusion modulation) is shown for period January 1953–November 2000 in Fig. 1.

We used 11-month smoothed data of $W$ and determined the amplitude $A_{dr}$ of drift effects as drift modulation at $W_{11M} = 75$ (average value of $W_{11M}$ for 1953–1999). The reversal periods were determined as: August 1949 ± 9 months, December 1958 ± 12 months, December 1969 ± 8 months, March 1981 ± 5 months, and June 1991 ± 7 months. We determined correlation coefficients between the expected integrals $F$ according to Eq. (5) for different values of $X_o = 1, 2, 3, \ldots, 60$ average months with the observed LN(CL11M) and LN(HU/HAL11M), as well as correction for drift effects to $A_{dr}$ from 0.15% up to 4% at $W = 75$.

Table 1 shows the results of the determination of $X_{0,max}$ for solar cycles 19, 20, 21, and 22 without corrections for drift effect, and with corrections to the drift effects on the value of $A_{dr}$ (from 0.5% to 4% for Climax NM and from 0.15% to 1.0% for Huancayo/Haleakala NM).

In Fig. 2 the relationship between $X_{o,max}$ and $A_{dr}$ is shown for Climax NM. From Fig. 2 we can see that the crossover region for odd and even
cycles is $13 \leq X_{o,\text{max}} \leq 16.5$, $1.7\% \leq A_{d\ell} \leq 2.3\%$. For Huancayo/Haleakala NM this region is $13 \leq X_{o,\text{max}} \leq 18$, $0.23\% \leq A_{d\ell} \leq 0.43\%$. Thus we concluded that the amplitude of the drift effect is about $2.0\%$ for Climax NM and about $0.33\%$ for Huancayo/Haleakala NM. Thus, for primary CR with rigidity 10–15 GV the relative contribution of drift effects is about 20–25%. For CR with rigidity 35–40 GV the relative role of drift effects is about two to three times smaller. For $X_{o,\text{max}}$, we obtained about 15 average months for both 10–15 and 35–40 GV ranges, corresponding to $r_o \approx 100$ AU (at average solar wind speed 400 km/s).

1.4. The inverse problem for CR propagation and modulation during solar cycle 22 on the basis of NM data

In the previous section, we considered the relative role of convection–diffusion and drifts in long-term CR modulation on the basis of a
comparison of observations in odd and even cycles of SA: it was shown that the time-lag $X_{o,\text{max}}$ between CR and SA in the odd cycles 19 and 21 decreases with increasing amplitude of the drift effect $A_{dr}$, but in the even cycles 20 and 22, $X_{o,\text{max}}$ increases with increasing $A_{dr}$. To determine $X_{o,\text{max}}$ and $A_{dr}$ separately in the previous section we assumed that as a first approximation $X_{o,\text{max}}$ and $A_{dr}$ are about the same in odd and even solar cycles. In this case the crossover of $X_{o,\text{max}}$ and $A_{dr}$ for odd and even cycles determines the expected values of $X_{o,\text{max}}$ and $A_{dr}$. In this section we try to solve the inverse problem of determining $A_{dr}$ and $X_{o,\text{max}}$ based only on data during solar cycle 22. We will therefore correct the observed CR long-term variation in cycle 22 for drift effects with different values of the amplitude $A_{dr}$; for each $A_{dr}$ we determine the correlation coefficient $R(X_{o},A_{dr})$ of corrected CR long-term variation according to a convection–diffusion model for different values of the time-lag $X_{o}$ (from 0 to 60 average months with one-month steps). Then we determine the value of $X_{o,\text{max}}(A_{dr})$ when $R(X_{o},A_{dr})$ reaches the maximum value $R_{\text{max}}(X_{o,\text{max}}A_{dr})$. For each $A_{dr}$ we will determine $R_{\text{max}}$ and $X_{o,\text{max}}$. It is natural to assume that the most reliable value of $A_{dr}$ will correspond to the biggest $R_{\text{max}}(X_{o,\text{max}}A_{dr})$ value, i.e., when the correction for drift effects is the best (in the frame of the model used for drift effects for long-term CR variations). In this way it will also be possible to
determine the most reliable value for $X_{o,\text{max}}$ characterizing the dimension of the CR modulation region in the Heliosphere. We use the convection–diffusion quasi-stationary model of CR–SA hysteresis phenomenon which was described above (Eqs. (1)–(5)), and the drift model (both were used in the previous section). According to the main idea of the drift mechanism (see Refs. 27, 32, 33 and 36–38) we assume that the CR drift is proportional to the value of tilt angle $T$ and reverses sign with SMF polarity reversal. The important cycle 22 reversal periods was June 1991 ± 7 months.

1.4.1. Results for Climax NM data

According to the procedure described above, we correct the 11-month smoothed data on the drift effect for different values of $A_{dr}$ from 0% (no drift effect) up to 4% at $W = 75$. The dependence of the correlation coefficient on the value of the expected time-lags is shown in Fig. 3. For each value of $A_{dr}$ in Fig. 3 we can easily determined the value of $X_{o,\text{max}}(A_{dr})$ at which the correlation coefficient reaches a maximum value $R_{\text{max}}$.

![Fig. 3. Correlation coefficient $R(X_o, A_{dr})$ for 11-month smoothed data of Climax NM (N39, W106; height 3400 m, 2.99 GV) in cycle 22 for different $A_{dr}$ from 0% up to 4% at $W = 75$.](image)
Galactic Cosmic Ray Propagation in the Heliosphere

The functions $R_{\text{max}}(A_{\text{dr}})$ and $X_{\text{o, max}}(A_{\text{dr}})$ are shown in Fig. 4. Approximating the function $R_{\text{max}}(A_{\text{dr}})$ with a parabola

$$R_{\text{max}}(A_{\text{dr}}) = a A_{\text{dr}}^2 + b A_{\text{dr}} + c,$$

(7)

where $a = 0.004065 \pm 0.000079$, $b = -0.01253 \pm 0.00024$, and $c = -0.9551 \pm 0.0185$ gives a correlation coefficient of $0.9985 \pm 0.0007$.

From Eq. (7) we can determine $A_{\text{dr, max}}$ when $R_{\text{max}}$ reaches its largest value:

$$A_{\text{dr, max}} = -\frac{b}{2a},$$

(8)

which gives $A_{\text{dr, max}} = 1.54 \pm 0.04\%$ at $W = 75$. With this information, we can now correct the Climax NM data of cycle 22 for drifts, with the most reliable amplitude $A_{\text{dr, max}}$ according to Eq. (8) and the function $R(X_o, A_{\text{dr, max}})$ is shown in Fig. 5.

From Fig. 5 we can see that the function $R(X_o, A_{\text{dr, max}})$ can be approximated by a parabola:

$$R(X_o, A_{\text{dr, max}}) = d X_o^2 + e X_o + f,$$

(9)

where $d = 0.000377 \pm 0.000002$, $e = -0.00942 \pm 0.00004$, and $f = -0.906 \pm 0.004$ with a correlation coefficient $0.99994 \pm 0.00003$. 

Fig. 4. Functions $R_{\text{max}}(A_{\text{dr}})$ and $X_{\text{o, max}}(A_{\text{dr}})$ from $A_{\text{dr}}$ at $W = 75$ for Climax NM data in cycle 22.
From Eq. (9) we can determine the most reliable value of $X_{o,\text{max}}$ corresponding to $A_{\text{dr,\text{max}}}$:

$$X_{o,\text{max}} = -\frac{c}{2d}, \quad (10)$$

which gives $X_{o,\text{max}} = 12.5 \pm 0.1 \text{ av. month}$. At the derived values of $A_{\text{dr,\text{max}}}$ and $X_{o,\text{max}}$ the connection between expected and observed CR intensity is characterized by correlation coefficient $R_{\text{max}}(X_{o,\text{max}}, A_{\text{dr,\text{max}}}) = 0.9652$ (see Fig. 5).

1.4.2. Results for Kiel NM data

The function $R_{\text{max}}(A_{\text{dr}})$ for Kiel NM data (sea level; $R_c = 2.32 \text{ GV}$) can be approximated by Eq. (7) with regression coefficients $a = 0.0095 \pm 0.0001$, $b = -0.0250 \pm 0.0004$, $c = -0.960 \pm 0.014$ with a correlation coefficient $0.9992 \pm 0.0004$. Then according to Eq. (8), $A_{\text{dr,\text{max}}} = 1.32 \pm 0.04\%$. Next, we determine $R(X_{o, A_{\text{dr,\text{max}}}})$ using by Eq. (9) with regression coefficients $d = 0.000466 \pm 0.000003$, $e = -0.01191 \pm 0.00007$, $f = -0.897 \pm 0.005$ (correlation coefficient $0.99988 \pm 0.00006$), gives, according to Eq. (10), $X_{o,\text{max}} = 13.4 \pm 0.2 \text{ av. months}$. The derived values for $A_{\text{dr,\text{max}}}$ and $X_{o,\text{max}}$ are about the same as for the Climax NM. In this case the correlation between the predicted and observed CR intensity is characterized by a coefficient of $R_{\text{max}}(X_{o,\text{max}}, A_{\text{dr,\text{max}}}) = 0.977$.

1.4.3. Results for Tyan-Shan NM data

The Tyan-Shan NM (43N, 77E, near Alma-Ata; 3.34 km above sea level, $R_c = 6.72 \text{ GV}$) is sensitive to more energetic particles than the Climax NM.
and the Kiel NM. For the Alma-Ata NM the function $R_{\text{max}}(A_{\text{dr}})$ can be approximated by Eq. (7), with regression coefficients $a = 0.0149 \pm 0.0015$, $b = -0.019 \pm 0.002$, $c = -0.957 \pm 0.009$ (correlation coefficient of 0.9996 ± 0.0002), that gives $A_{\text{dr},\text{max}} = 0.634 \pm 0.012\%$ from Eq. (8). Next, we determined $R(X_o, A_{\text{dr},\text{max}})$ from Eq. (9) with regression coefficients $d = 0.000388 \pm 0.000004$, $e = -0.00845 \pm 0.00005$, $f = -0.917 \pm 0.008$ (correlation coefficient of 0.9997 ± 0.0002), which gives from Eq. (10), $X_{o,\text{max}} = 10.9 \pm 0.2$ av. months. In this case the correlation between the predicted and observed CR intensity is characterized by a coefficient of $R_{\text{max}}(X_{o,\text{max}}, A_{\text{dr},\text{max}}) = 0.963$.

1.4.4. Results for Huancayo/Haleakala NM data

The Huancayo NM (12S, 75W; 3.4 km above sea level, $R_c = 12.92$ GV)/Haleakala NM (20N, 156W; 3.03 km above sea level, $R_c = 12.91$ GV) is sensitive to primary CR particles of 35–40 GV which is about two to three times larger than for the Climax and Kiel NM. For Huancayo/Haleakala NM the function $R_{\text{max}}(A_{\text{dr}})$ from Eq. (7), gives regression coefficients $a = 0.0621 \pm 0.0004$, $b = -0.0165 \pm 0.0001$, $c = -0.978 \pm 0.007$ (correlation coefficient of 0.9998 ± 0.0001), which gives $A_{\text{dr},\text{max}} = 0.133 \pm 0.002\%$ according to Eq. (8). Next, we determined $R(X_o, A_{\text{dr},\text{max}})$ by Eq. (9), we get regression coefficients $d = 0.000406 \pm 0.000001$, $e = -0.00842 \pm 0.00002$, $f = -0.935 \pm 0.002$ (correlation coefficient 0.9998 ± 0.0001), that gives $X_{o,\text{max}} = 10.38 \pm 0.05$ average months from Eq. (10). In this case the correlation between the predicted and observed CR intensity is characterized by $R_{\text{max}}(X_{o,\text{max}}, A_{\text{dr},\text{max}}) = 0.979$.

1.4.5. Main results for the inverse problem for the solar cycle 22 on the basis of NM data

Taking into account drift effects allows us to determine the most reliable amplitude $A_{\text{dr},\text{max}}$ (at $W = 75$) and the time-lag $X_{o,\text{max}}$ (the effective time of the solar wind moving with frozen magnetic fields from the Sun to the boundary of the modulation region at a distance $r_o \approx uX_{o,\text{max}}$), using data only for solar cycle 22. We found that with an increasing effective CR primary particle rigidity from 10–15 GV (Climax NM and Kiel NM) to 35–40 GV (Huancayo/Haleakala NM) decreases in both the amplitude of drift effect $A_{\text{dr},\text{max}}$ (from about 1.5% to about 0.15%) and time-lag $X_{o,\text{max}}$ (from about 13 average months to about 10 average months). So, in cycle 22 the total long-term CR modulation with rigidity 10–15 GV, the relative
role of the drift mechanism was $4 \times 1.5\%/25\% \approx 1/4$ and the convection–diffusion mechanism about $3/4$. This takes into account that observed total 11-year variation in Climax and Kiel NM is $25\%$ and according to Fig. 1 the total change of CR intensity due to drift effects is about four times more than the amplitude $A_{dr}$ at W11M = 75. For rigidities 35–40 GV these values were $4 \times 0.15\%/7\% \approx 1/10$ for the drift mechanism, and about $9/10$ for the convection–diffusion mechanism. If we assume that the average velocity of the solar wind in the modulation region was about the same as the observed average velocity near the Earth’s orbit in 1965–1990, $u = 4.41 \times 10^7 \text{cm/s} = 7.73 \text{AU}/(\text{average month})$, the estimated dimension of modulation region in cycle 22 will be $\sim 100 \text{AU}$ for CR with rigidity of 10–15 GV and about 80 AU for CR with rigidity of 35–40 GV. In other words, at distances more than 80 AU the magnetic field in solar wind and its inhomogeneities were too weak to influence the intensity of 35–40 GV particles.

2. The Inverse Problem for Low-Energy Galactic CR Propagation and Modulation in the Heliosphere on the Basis of Satellite Data

2.1. Diffusion time-lag for low-energy particles

Above, in Sec. 1 we considered how to take into account the time-lag of processes in the interplanetary space relative to processes on the Sun, determined by the value $r/u$. For small-energy particles measured on satellites and balloons, it is necessary to take into account the additional time-lag $T_{dif}(R, r_{obs}, r, r_0)$ caused by the particle diffusion through the Heliosphere from $r$ to $r_{obs}$. This diffusion time-lag can be approximately estimated. It was shown in Refs. 18–20 that as a first approximation the value $u/\kappa_r$ in Eq. (1) can be considered as independent from $r$, and some effective values of solar wind speed $u_{ef}(t)$ and diffusion coefficient $\kappa_{r,ef}(R, t)$ can be used. In this case, instead of Eq. (1), we obtain

$$\frac{n(R, r_{obs}, t)}{n_0(R)} \approx \exp \left(-\frac{a u_{ef}(t)(r_0 - r_{obs})}{\kappa_{r,ef}(R, t)}\right). \tag{11}$$

The diffusion propagation time of CR particles with rigidity $R$ from the distance $r$ to the distance of observations $r_{obs}$ can be approximately
estimated as
\[
T_{\text{diff}}(R, t, r, r_o) \approx \frac{(r_o - r_{\text{obs}})^2 - (r_o - r)^2}{6\kappa_{r,\text{ef}}(R, t)}
\approx C(R, t) \times \frac{(r - r_{\text{obs}})(2r_o - r - r_{\text{obs}})}{u_{\text{ef}}(t)(r_o - r_{\text{obs}})} ,
\]
(12)
where
\[
C(R, t) = -\frac{\ln(n(R, r_{\text{obs}}, t)/n_o(R))}{6a} .
\]
(13)
and \( \kappa_{r,\text{ef}}(R, t) \) is determined by Eq. (11). Instead of the distances from the Sun it is possible to introduce the variables used in Refs. 34 and 35:
\[
X_{\text{obs}} = \frac{r_{\text{obs}}}{u_{\text{ef}}} , \quad X = \frac{r}{u_{\text{ef}}} , \quad X_o = \frac{r_o}{u_{\text{ef}}};
\]
(14)
these variables and \( T_{\text{diff}} \) are in units of average month = \( (365.25/12) \) days = 30.44 days = \( 2.628 \times 10^6 \) s. By combining Eqs. (14) and (12) we obtain
\[
T_{\text{diff}}(R, t, X_{\text{obs}}, X, X_o) \approx C(R, t) \frac{(X - X_{\text{obs}})(2X_o - X - X_{\text{obs}})}{X_o - X_{\text{obs}}} .
\]
(15)
From Eq. (15) it follows that \( T_{\text{diff}}(R, t, X_{\text{obs}}, X, X_o) \) reaches the maximum value at \( X = X_o \), and the coefficient \( C(R, t) \) reaches the maximum value, according to Eq. (13), at the minimum of CR intensity (near the maximum of SA). Then
\[
\frac{T_{\text{diff}}(R, t, X_{\text{obs}}, X, X_o)}{X_o - X_{\text{obs}}} \leq C(R, t) .
\]
(16)
For high and middle latitude NM data (effective particle rigidity 10–15 GV) the amplitude of 11-year modulation is about 25% and according to Eq. (13) we obtain for solar maximum \( C(R, t) \approx 0.028 \). Thus, according to Eq. (16), \( T_{\text{diff}}(R, t, X_{\text{obs}}, X, X_o)/(X_o - X_{\text{obs}}) \leq 0.028 \), i.e., the diffusion time-lag is negligible in comparison with the time propagation of solar wind from the Earth’s orbit to the boundary of Heliosphere. On the basis of Ref. 27 we estimate \( C(R, t) \) for smaller rigidities, observed by satellites. Results are shown in Fig. 6 against tilt angle \( T \), and in Table 2 for maximum and minimum SA.
Fig. 6. The coefficient $C(R,T)$ plotted against tilt angle $T$ for CR particles with rigidities 3, 1, and 0.3 GV.

Table 2. Coefficient $C(R,t)$ for different rigidities, for periods of maximum and minimum solar activity.

<table>
<thead>
<tr>
<th>Solar activity</th>
<th>MAX</th>
<th>MIN</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 GV (protons, 2.2 GeV)</td>
<td>0.107</td>
<td>0.067</td>
</tr>
<tr>
<td>1.0 GV (protons, 430 MeV)</td>
<td>0.30</td>
<td>0.20</td>
</tr>
<tr>
<td>0.3 GV (protons, 43 MeV)</td>
<td>0.55</td>
<td>0.41</td>
</tr>
</tbody>
</table>

2.2. Convection–diffusion modulation for low-energy galactic CR particles

According to Eq. (16), Fig. 6, and Table 2, for low-energy galactic CR particles it is necessary to take into account the additional time-lag caused by the particle diffusion in the interplanetary space. In the first approximation we use the quasi-stationary model of convection–diffusion modulation, described in Sec. 1, but extended to take into account the diffusion time-lag:

$$\ln(n(R,r_{\text{obs}},t)) = A(X_0, \beta, t_1, t_2) - B(X_0, \beta, t_1, t_2)$$
$$\times F(t, X_0, \beta, W(t - X^*)|_{X_{\text{obs}}}).$$  (17)

In Eq. (17) function $F$ is the same determined by Eq. (5), but instead of $X$ we use $X^*$, where

$$X^* = X + C(R, t) \times \frac{(X - X_{\text{obs}})(2X_0 - X - X_{\text{obs}})}{X_0 - X_{\text{obs}}}.  \quad (18)$$
2.3. Low-energy CR long-term variation caused by drifts

According to drift theory (see Refs. 27, 32, 33 and 36–38), we assume that the drifts depend on the value of tilt angle $T$ and the changed sign during periods of the SMF polarity reversal. We assume that the drift effect is proportional to the theoretical value derived from the tilt angle $T$ (or from the sunspot number $W$ through Eq. (6)) with negative sign for general SMF $A > 0$ and positive sign for $A < 0$, and in the period of reversal we suppose linear transition through 0 from one polarity cycle to another.

The theoretical expected values of convection–diffusion modulation $A_{cd}$ and drift modulation $A_{dr}$ have been determined from corresponding figures in Ref. 27; we assume that in these figures the average of curves for $A > 0$ and $A < 0$ characterizes the convection–diffusion modulation (which does not depend on the sign of general SMF), and the difference between these curves represents the double drift modulation (which depends on the sign of general SMF). Figure 7 shows ratios of $A_{dr}/A_{cd}$ for $R = 3$, 1, and 0.3 GV, derived from Ref. 27 based on Eq. (6).

2.4. Satellite proton data and corrections to solar CR increases and the jump in December 1995

We analyze the following data: IMP-8 monthly data of proton fluxes with kinetic energy $E_k \geq 106 \text{MeV}$ ($R \geq 0.458 \text{GV}$) from October 1973 to December 1999 (http://data.ftecs.com/archive/imp_cpm/) and

![Graph showing expected ratios $A_{dr}/A_{cd}$ for $R = 3$, 1, and 0.3 GV in dependence of sunspot number $W$ and derived from Ref. 27 based on Eq. (6).]
GOES daily data of proton fluxes from January 1986 to December 1999 (http://spidr.ngdc.noaa.gov/spidr/) with kinetic energies $E_k \geq 100$ MeV ($R \geq 0.444$ GV), $E_k \geq 60$ MeV ($R \geq 0.341$ GV), $E_k \geq 30$ MeV ($R \geq 0.239$ GV), $E_k \geq 10$ MeV ($R \geq 0.137$ GV), $E_k \geq 5$ MeV ($R \geq 0.097$ GV), as well as fluxes in intervals 60–100, 30–60, 10–30, and 5–10 MeV.

The first problem is that the original GOES data contain many increases caused by SEP events. To exclude these days we sorted daily data for each month and determined the averages from 10 minimal, 10 middle, and 10 maximal daily values. In the present paper, we used averages from 10 minimal daily values for each month. Even by this method the influence of great solar energetic particle events was not totally eliminated (e.g., as in September 1989). These months have been excluded from our analysis. Then, we determined 11-month moving averages.

The second problem is that the original GOES data contain a jump in December 1995. To exclude this jump we compared GOES data for $E_k \geq 100$ MeV with IMP-8 monthly data for $E_k \geq 106$ MeV and estimated the value of jump as 0.006 protons cm$^2$/s/s. For $E_k \geq 60, \geq 30, \geq 10,$ and $\geq 5$ MeV, the value of the jump is 0.012, 0.025, 0.035, 0.040 protons cm$^2$/s/s, respectively. As an example, Fig. 8 shows the corrected IMP-8 data for proton intensities with energy $E_k \geq 106$ MeV.

![Fig. 8. Natural logarithm of monthly and 11-month moving averages IMP-8 data of proton intensities with energy $E_k \geq 106$ MeV, corrected by excluding days with increases mainly caused by SEP events.](image-url)
2.5. Results for proton data obtained by IMP-8 and GOES

In Fig. 9 we show the relationship between the correlation coefficient $R(X_o, A_{dr})$ and the natural logarithm of 11-month moving averages IMP-8 data of proton intensities with energy $E_k \geq 106$ MeV, corrected for SEP and other increases by the method described in Sec. 2.4 above, and corrected for drift effects with different amplitudes as described in Sec. 2.3, and including the diffusion time-lag (important for small-energy particles observed on satellites).

From Fig. 9 it can be seen that $R(X_o, A_{dr})$ reaches the largest values for $A_{dr} \approx 0.1$ (i.e., 10%) with a maximum value of 0.9128 at $X_{o,max} \approx 17$ average months, a little larger than the value obtained for NM data in Sec. 1. A similar result was obtained for monthly data, but with smaller values of correlation coefficient (maximum value 0.8993 at $X_{o,max} \approx 18$ average months).

Figure 10 shows the same results for GOES data. It can be seen that GOES data give similar results to IMP-8 data, but with much larger correlation coefficients. The best correlation is again found at $A_{dr} \approx 0.1$, but with maximum value 0.9793 at $X_{o,max} \approx 15$ average months, and regression equation

$$\ln(I_{cor}) = -3.226 - 0.0525 F,$$

(19)

![Figure 9](image-url)  
Fig. 9. Correlation coefficient $R(X_o, A_{dr})$ for 11-month moving averages of IMP-8 data of proton intensities with energy $E_k \geq 106$ MeV from October 1973 to December 1999, corrected for drift modulation with different amplitudes $A_{dr}$ from 0 to 0.7.
where $F$ is determined by Eq. (5), with $X^*$ from Eq. (18) instead of $X$. From Eq. (19) it follows that the intensity out of Heliosphere $\ln I_o = -3.226$ (for GOES data of protons with energy $E_k \geq 100$ MeV). From these results, using Refs. 23 and 39, we can estimate the dimension of modulation region

$$r_o \approx X_{o,\text{max}}u_{ef} \approx 0.84X_{o,\text{max}}u(1\text{ AU}) \approx 97.4\text{ AU}$$

and the effective radial diffusion coefficient

$$\kappa_r(R_{ef}) = \frac{a_{u_{ef}}^2}{0.0525 \text{ av. month}} \frac{(1\text{ AU})^2}{1.03 \times 10^{23}\text{ cm}^2\text{s}}.$$ (21)

### 2.6. Properties of alpha-particles from satellite data

We used five-min GOES data of low-energy alpha-particle fluxes (in units particles/cm$^2$/s/s/MeV from January 1986 to May 2000 in three energy

<table>
<thead>
<tr>
<th>$E_k$ interval, MeV</th>
<th>$R$ interval, GV</th>
<th>$R_{ef}$, GV</th>
<th>$C_{av}(R_{ef})$</th>
<th>$(v/c)_{ef}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>60–160</td>
<td>0.337–0.554</td>
<td>0.45</td>
<td>0.347</td>
<td>0.23</td>
</tr>
<tr>
<td>160–260</td>
<td>0.554–0.710</td>
<td>0.63</td>
<td>0.257</td>
<td>0.32</td>
</tr>
<tr>
<td>330–500</td>
<td>0.804–1.000</td>
<td>0.90</td>
<td>0.186</td>
<td>0.43</td>
</tr>
</tbody>
</table>
intervals with parameters listed in Table 3 (http://spidr.ngdc.noaa.gov/spidr/).

2.7. Results for alpha-particles in the energy interval 330–500 MeV

The corrected GOES data (by excluding sudden increases caused by solar energetic particle effects) are shown in Fig. 11 (the monthly and 11-month running averages).

We then calculated the expected drift modulation for different \( A_{\text{dr}} \) from 0 (no drift modulation) up to 0.4 (i.e., 40%) using monthly data on tilt angles \( T \) and sunspot numbers \( W \), according to Eq. (6). The values of the correlation coefficients \( R(R_{\text{ref}}, A_{\text{dr}}, X_0) \) are shown in Fig. 12 for different values of \( A_{\text{dr}} \) for solar cycle 22, the whole of which has GOES alpha-particle flux data available.

From Fig. 12 it can be seen that with \( A_{\text{dr}} \) ranging from 0 to 0.40 the maximum correlation coefficient moved from about 8 to 35 average months. In Fig. 13 the relationship between \( A_{\text{dr}} \) and the parameters \( X_{0,\text{max}} \) and \( R_{\text{max}} \) at the highest \( R_{\text{max}} \) values are shown.

![Fig. 11. Natural logarithm of monthly (as LN(1MCOR)) and 11-month moving averages (as LN(11MCOR)) of corrected GOES data of alpha-particle fluxes in the energy interval 330–500 MeV, during January 1986–May 2000.](image)
Fig. 12. Correlation coefficients plotted against $X_0$ at different values of drift modulation amplitude $A_{dr}$ from 0 (no drift correction) to 0.40 for alpha-particle fluxes in energy interval 330–500 MeV during solar cycle 22.

From Fig. 13 we can determine the optimal values of $A_{dr}$ and $X_{o,\text{max}}$ for which the correlation coefficient is highest ($t0.98275$):

$$(A_{dr})_{\text{opt}} = 0.087, \quad (X_{o,\text{max}})_{\text{opt}} = 13.76 \text{ av. months}. \quad (22)$$
3. Conclusion

The information obtained from the solution of inverse problem for satellite alpha-particle data about Heliosphere dimension, CR diffusion coefficient, and CR intensity out of the Heliosphere agrees with those obtained from satellite proton data and from ground-based NM data.

Acknowledgments

This research was partly supported by EU COST-724. The author thanks Prof. Yuval Ne’eman and Dr. Abraham Sternlieb, Prof. Isaac Ben Israel and Dr. Zvi Kaplan, for constant interest and support of Israel Cosmic Ray and Space Weather Center, and Emilio Segre’ Observatory. The author further thanks A. Belov, M. Duldig, E. Eroshenko, N. Iucci, K. G. McCracken, M. Murat, M. Parisi, L. Pustil’nik, M. A. Shea, D. F. Smart, M. Storini, G. Villoresi, V. Yanke, and I. Zukerman for interesting discussions. The work of neutron monitor on Mt. Hermon is supported by Tel Aviv University, UNIRoma-Tre, and IFSI-CNR Collaboration.

References

RADIATION HAZARD FROM LARGE SEP EVENTS FOR AIRCRAFT: MONITORING AND FORECASTING BY USING ON-LINE ONE-MIN COSMIC RAY DATA

LEV DORMAN

Israel Cosmic Ray and Space Weather Center and Emilio Segre’ Observatory
Tel Aviv University, Technion and Israel Space Agency
P. O. Box 2217, Qazrin 12900, Israel
Cosmic Ray Department of IZMIRAN, Russian Academy of Science
Troitsk 142092, Moscow Region, Russia

We show that monitoring and exact forecast of the radiation hazard from great solar energetic particle events for aircrafts can be made by using high-energy particles (few GeV/nucleon and higher) whose transport from the Sun is characterized by much larger diffusion coefficients than for small- and middle-energy particles.

1. The Basis of the Problem

High-energy particles arrive from the Sun much more earlier (8–20 min after acceleration and escaping into solar wind) than the lower energy particles that cause more dangerous situations for people and electronics (more than 30–60 min later). We describe here the principles and experience of the automated program “SEP-Search”. The positive result, which show the exact beginning of solar energetic particle (SEP) event on the Emilio Segre’ Observatory (2025 m above sea level, $R_c = 10.8 \, \text{GV}$), is now determined automatically by simultaneous increase of $2.5\sigma$ in two sections of neutron super-monitor [the same can be made for any neutron monitor (NM) in the World]. With the next one-min data sample, the program “SEP-Search” checks that the observed increase reflects the beginning of real great SEP. If so, the programs “SEP-Research” automatically start working on-line. We determine also the probabilities of false and missed alerts.

The first of the “SEP-Research” programs is “SEP-Research/Spectrum”. We consider two variants: (1) quiet period (no magnetic storm, i.e., no change in cut-off rigidity), and (2) disturbed period in
geomagnetic field (characterized with possible changing of cut-off rigidity). We describe the method of determining the spectrum of the SEP in the first variant (for this we need data for at least two components with different coupling functions). For the second variant we need data for at least three components with different coupling functions. We show that for these purposes the total intensity and some different multiplicities can be used but that it is better to use data from two or three NMs with different cut-off rigidities. We describe in detail the algorithms of the program “SEP-Research/Spectrum”. We show how this program performs with examples of some historical great SEP events.

Then we show that after these two steps it is possible, on the basis of CR data only, to determine important unknown parameters, characterizing SEP generation on the Sun and propagation in the interplanetary space: time of ejection, diffusion coefficient in the interplanetary space and energy spectrum at the source of SEP. To extend the obtained information to very small energies, we also use available satellite one-min data together with NM data. By the method of coupling functions for different altitudes in the atmosphere, we describe some principles of on-line radiation hazard monitoring and forecasting for aircrafts on regular and nonregular lines with dependence of altitude, cut-off rigidity, and shielding. If for some cases the expected radiation hazard will be higher than some definite dangerous level, on-line special alerts will be sent. These programs are tested against some historical great SEP events.

2. The Method to Automatically Search for the Start of Great SEP Events

Let us consider the problem of automatically searching for the start of great SEP events. Of course, the patrol of the Sun and forecast of great solar flares are very important, but not enough: only very small fraction of great solar flares produces dangerous SEP events. As we mentioned in Sec. 1, this forecast can be made by using high-energy particles (few GeV/nucleon and higher) whose transport from the Sun is characterized by much larger diffusion coefficients than for small- and middle-energy particles. Therefore, high-energy particles arrive from the Sun much earlier (8–20 min after acceleration and escaping into solar wind) than the lower energy particles that cause more dangerous situations in space and in the atmosphere. The flux of high-energy particles is very small and is not dangerous for people
and electronics. The problem is that this very small high-energy flux is
not measured with enough accuracy on satellites to use for forecasting (for
very large effective detector areas very large weight is needed). However,
high-energy particles of galactic or solar origin are measured continuously
by ground-based NMs, ionization chambers, and muon telescopes with
very large effective surface areas (many square meters) that provide very
small statistical errors. We show on the basis of data in periods of great
historical SEP events that one-min on-line data of high-energy particles
could be used for forecasting arrival of dangerous fluxes of particles with
much smaller energy. Let us describe the principles and on-line operation of
programs “SEP-Search-1 min”, “SEP-Search-2 min”, and so on, developed
and checked in the Emilio Segre’ Observatory of ICRC. The determination
of increasing flux is made by comparison with intensity averaged from 120 to
61 min before the present \(Z\)th one-min data. For each \(Z\)-min data, start the
program “SEP-Search-1 min”. The program for each \(Z\)th minute determines
the values

\[
D_{A1Z} = \frac{1}{\sigma_1} \left[ \ln (I_{AZ}) - \sum_{k=Z-120}^{k=Z-60} \frac{\ln (I_{Ak})}{60} \right],
\]

(1)

\[
D_{B1Z} = \frac{1}{\sigma_1} \left[ \ln (I_{BZ}) - \sum_{k=Z-120}^{k=Z-60} \frac{\ln (I_{Bk})}{60} \right],
\]

(2)

where \(I_{Ak}\) and \(I_{Ak}\) are one-min total intensities in the sections of neutron
super-monitor A and B.

If simultaneously

\[
D_{A1Z} \geq 2.5, \quad D_{B1Z} \geq 2.5, \]

(3)

the program “SEP-Search-1 min” repeat the calculation for the next \((Z+1)\)th minute and if Eq. (3) is satisfied again, the onset of great SEP is
determined and the program “SEP-Research/Spectrum” starts.

If Eq. (3) is not satisfied, the program “SEP-Search-2 min” searches for
the start of an increase by using two-min data. In this case, the program
“SEP-Search-2 min” will calculate values

\[
D_{A2Z} = \frac{1}{\sigma_2} \left[ \frac{\ln (I_{AZ}) + \ln (I_{AZ, Z-1})}{2} - \sum_{k=Z-120}^{k=Z-60} \frac{\ln (I_{Ak})}{60} \right],
\]

(4)
If the result is negative, then “SEP-Search-3 min” uses the average of three minutes: $Z - 2$, $Z - 1$, and $Z$. If this program also gives a negative result, then the program “SEP-Search-5 min” uses the average of five minutes: $Z - 4$, $Z - 3$, $Z - 2$, $Z - 1$, and $Z$. If this program also gives negative result, i.e., all programs “SEP-Search-K min” (where $K = 1, 2, 3, 5$) give negative result for the $Z$th minute, it means that in the next 30–60-min there will be no radiation hazard from small-energy particles (for this minute in our website shows: Alert — No). After obtaining this negative result, the procedure repeats for the next, $(Z + 1)$th minute, and so on. If any positive result is obtained for some $Z = Z'$, the “SEP-Search” programs check the next $(Z' + 1)$th minute data. If the result is again positive, then the program “SEP-Research/Spectrum” starts. An example of the one-min and 1-h data in real-time scale (updated every minute) with information on the Alert for dangerous cosmic ray (CR) increases on the Mt. Hermon is shown in Fig. 1.

### 3. The Probability of False Alarms

Let us consider that CR intensity fluctuated in which the increase in the NM counting rate was not excluded (though not a real SEP) and could be considered as the start of an SEP event. Let us estimate the probability of this false alarm. Because the probability function $\Phi(2.5) = 0.9876$, then the probability of an accidental increase with amplitude more than 2.5$\sigma$ in one channel will be $(1 - \Phi(2.5))/2 = 0.0062 \text{ min}^{-1}$, or 1 in 161.3 min (in one day we expect 8.93 such increases in one channel). The probability of statistical increases in both channels will be $((1 - \Phi(2.5))/2)^2 = 3.845 \times 10^{-5} \text{ min}^{-1}$ or 1 in 26,007 min $\approx$ 18 days. The probability that the increases of 2.5$\sigma$ will be accidentally in both channels in two successive minutes is $((1 - \Phi(2.5))/2)^4 = 1.478 \times 10^{-9} \text{ min}^{-1}$ or 1 in $6.76 \times 10^8$ min $\approx$ 1286 years. If this false alarm (one in about 1300 years) is sent, it is not critical, because the first alarm is preliminary and can be canceled if in the third successive minute there is no increase in both channels bigger than 2.5$\sigma$ (a third consecutive minute accidental increase probability of this false alarm is negligible: $((1 - \Phi(2.5))/2)^6 = 5.685 \times 10^{-14} \text{ min}^{-1}$ or 1 in $3.34 \times 10^7$ years).
Fig. 1. Graphical presentation of ESO NM-IQSY real-time data, updated each minute: upper circles — one-min data of total neutron intensity from both sections for the last 360 min; lower circles — 1-h data for total neutron intensity for the last 144 h; the numbers shown at the bottom are the CR intensity variations in both sections for one-, two- and three-min data (in units $\sigma$), and the final conclusion for the Alert (for the moment shown on the figure, at 19:38 UT at August 22, 2001; SEP Alert: “No” was obtained after 7 s).

Let us note that the false alarm could be sent in the case of solar neutron event (which really is not dangerous for aircraft), but such events are usually very short (only few minutes) and this alarm will be automatically canceled in the successive minute after the end of a solar neutron event.

4. The Probability of Missed Triggers

Let us now consider CR intensity fluctuations where a real increase in the NM counting rate is largely canceled out by decrease in the measured CR intensity resulting from a statistical fluctuation (not a real CR intensity decrease). In this case the trigger of the real SEP event starting will be missed and corresponding positive Alert will not be initiated. Let us estimate the probability of this mistake. It is necessary to take into account that the probability of missed triggers depends very strongly on the amplitude of increase. Let us suppose, for example, that we have a
real increase of $7\sigma$ (that for ESO corresponds to an increase of about 9.8%). The trigger will be missed if the observed counting rate, as a result of statistical fluctuations, in either channel and in any of the successive minutes remains less than $2.5\sigma$ above the preceding average. For this statistical fluctuation a decrease of more than $4.5\sigma$ is required. The probability of this negative fluctuation in one channel in one-min is $(1 - \Phi(4.5))/2 = 3.39 \times 10^{-6}$ min$^{-1}$, and the probability of missed trigger for two successive minutes of observation simultaneously in two channels is four times larger: $1.36 \times 10^{-5}$. It means that missed trigger is expected only one per about 70 000 events. In Table 1 are listed probabilities $P_{\text{mt}}$ of missed triggers for ESO (where standard deviation for one channel for one-min is $\sigma = 1.4\%$) as a function of the amplitude of increase $A$ (in $\sigma$).

Table 1. Probability of missed trigger versus $A$ (in $\sigma$).

<table>
<thead>
<tr>
<th>$A, \sigma$</th>
<th>6.0</th>
<th>6.5</th>
<th>7.0</th>
<th>7.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P_{\text{mt}}$</td>
<td>$9.3 \times 10^{-4}$</td>
<td>$1.3 \times 10^{-4}$</td>
<td>$1.4 \times 10^{-5}$</td>
<td>$1.1 \times 10^{-6}$</td>
</tr>
</tbody>
</table>

5. On-Line SEP Spectrum Determination from Single Observatory Data

5.1. Coupling functions method

The method of coupling (response) functions$^{1-3}$ allows us to calculate the expected flux above the atmosphere, and out of the Earth’s magnetosphere from the ground-based data.

Based on the latitude survey data$^{4-6}$ and theoretical calculations of meson-nuclear CR cascade in the atmosphere,$^{7,8}$ the polar normalized coupling functions for total counting rate and different multiplicities $m$ can be approximated by the function now referred to in the literature as the Dorman function (introduced in Ref. 9):

$$W_{om}(R) = a_m k_m R^{-(k_m+1)} \exp(-a_m R^{-k_m}),$$

(6)

where index $m$ determined the type of detected CR component. Polar coupling functions for muon telescopes with different zenith angles $\theta$ can be approximated by the same type of functions determined only by two parameters $a_m(\theta)$ and $k_m(\theta)$. Let us note that functions described by Eq. (6) are normalized: $\int_0^{\infty} W_{om}(R) dR = 1$ at any values of $a_m$ and $k_m$. The normalized coupling functions for locations with cut-off rigidity $R_c$,
Radiation Hazard from Large SEP Events for Aircraft

will be

\[
W_m(R_c, R) = \begin{cases} 
  a_m k_m R^{-(k_m+1)} (1 - a_m R_c^{-k_m})^{-1} \exp(-a_m R_c^{-k_m}) & \text{if } R \geq R_c, \\
  0 & \text{if } R < R_c.
\end{cases}
\]

(7)

In the first approximation, the spectrum of primary variation of SEP event can be described by the function

\[
\frac{\Delta D(R)}{D_o(R)} = b R^{-\gamma},
\]

(8)

where \(\Delta D(R) = D(R, t) - D_o(R)\), \(D_o(R)\) is the differential spectrum of galactic CR before the SEP event and \(D(R, t)\) is the total observed spectrum (galactic CR + SEP) at a later time \(t\). In Eq. (8) parameters \(b\) and \(\gamma\) depend on \(t\). Approximation (8) can be used for describing limited intervals of energy in the sensitivity ranges detected by the various components.

5.2. The case of magnetically quiet periods

In this case the observed variation \(\delta I_m(R_c) \equiv \Delta I_m(R_c)/I_{mo}(R_c)\) in some component \(m\) can be described in the first approximation by the function \(F_m(R_c, \gamma)\):

\[
\delta I_m(R_c) = b F_m(R_c, \gamma),
\]

(9)

where

\[
F_m(R_c, \gamma) = a_m k_m (1 - \exp(-a_m R_c^{-k_m}))^{-1} \int_{R_c}^{\infty} R^{-(k_m+1+\gamma)} \exp(-a_m R^{-k_m}) dR
\]

(10)

is a known function. Let us compare data for two components \(m\) and \(n\). From two equations of type Eq. (9) for two indexes \(m\) and \(n\) we obtain

\[
\frac{\delta I_m(R_c)}{\delta I_n(R_c)} = \Psi_{mn}(R_c, \gamma),
\]

(11)

where the special function

\[
\Psi_{mn}(R_c, \gamma) = \frac{F_m(R_c, \gamma)}{F_n(R_c, \gamma)}
\]

(12)
is calculated using Eq. (10). Comparison of experimental results \( \delta I_m(R_c)/\delta I_n(R_c) \) with function \( \Psi_{mn}(R_c, \gamma) \) according to Eq. (11) gives the value of \( \gamma \), and then from Eq. (9) the value of the parameter \( b \). The observed SEP increases for different components allow the determination of parameters \( b \) and \( \gamma \) for the SEP event beyond the Earth’s magnetosphere.

5.3. The case of magnetically disturbed periods

For magnetically disturbed periods the observed CR variation instead of Eq. (9) will be described by

\[
\delta I_k(R_c) = -\Delta R_c W_k(R_c, R_c) + b F_k(R_c, \gamma),
\]

where \( \Delta R_c \) is the change of cut-off rigidity due to change of the Earth’s magnetic field, and \( W_k(R_c, R_c) \) is determined according to Eq. (7) at \( R = R_c \):

\[
W_m(R_c, R_c) = a_m k_m R_c^{-(k_m+1)}(1 - a_m R_c^{-k_m})^{-1}\exp(-a_m R_c^{-k_m}).
\]

Now, for the first approximation of the SEP energy spectrum we have unknown variables \( \gamma, b, \Delta R_c \), and for their determination we need data from at least three different components \( k = l, m, n \) in Eq. (13). In accordance with the spectrographic method\(^\text{10}\) let us introduce the function

\[
\Psi_{lmn}(R_c, \gamma) = \frac{W_l(R_c, R_c) F_m(R_c, \gamma) - W_m(R_c, R_c) F_l(R_c, \gamma)}{W_m(R_c, R_c) F_n(R_c, \gamma) - W_n(R_c, R_c) F_m(R_c, \gamma)}.
\]

Then from equation

\[
\Psi_{lmn}(R_c, \gamma) = \frac{W_l(R_c, R_c) \delta I_m(R_c) - W_m(R_c, R_c) \delta I_l(R_c)}{W_m(R_c, R_c) \delta I_n(R_c) - W_n(R_c, R_c) \delta I_m(R_c)},
\]

the value of \( \gamma \) can be determined. Using this value of \( \gamma \), for each time \( t \), we determine

\[
\Delta R_c = \frac{F_l(R_c, \gamma) \delta I_m(R_c) - F_m(R_c, \gamma) \delta I_l(R_c)}{F_m(R_c, \gamma) \delta I_n(R_c) - F_n(R_c, \gamma) \delta I_m(R_c)},
\]

\[
b = \frac{W_l(R_c, R_c) \delta I_m(R_c) - W_m(R_c, R_c) \delta I_l(R_c)}{W_l(R_c, R_c) F_m(R_c, \gamma) - W_m(R_c, R_c) F_l(R_c, \gamma)}.
\]

So, in magnetically disturbed periods the observed SEP increases for different components also allows to determine parameters \( \gamma \) and \( b \) (for
the SEP spectrum beyond the Earth’s atmosphere), and \( \Delta R_c \), giving information on the ring currents of the Earth’s magnetosphere.

6. On-Line SEP Spectrum Determination from Data of Two or More Observatories During Magnetically Disturbed Periods

Let us consider the case of two CR stations with different cut-off rigidities. In this case, instead of Eq. (13) we will have for stations 1 and 2 the following four equations:

\[
\begin{align*}
\delta I_k(R_{c1}) &= -\Delta R_{c1}W_k(R_{c1}, R_{c1}) + bF_k(R_{c1}, \gamma), \\
\delta I_l(R_{c1}) &= -\Delta R_{c1}W_l(R_{c1}, R_{c1}) + bF_l(R_{c1}, \gamma), \\
\delta I_m(R_{c2}) &= -\Delta R_{c2}W_m(R_{c2}, R_{c2}) + bF_m(R_{c2}, \gamma), \\
\delta I_n(R_{c2}) &= -\Delta R_{c2}W_n(R_{c2}, R_{c2}) + bF_n(R_{c2}, \gamma).
\end{align*}
\]

We also have four unknown variables: \( \gamma \), \( b \), \( \Delta R_{c1} \), \( \Delta R_{c2} \). It is possible to exclude \( b \), \( \Delta R_{c1} \), \( \Delta R_{c2} \) from the system of Eqs. (19)–(22) and finally obtain a nonlinear equation for determining \( \gamma \):

\[
\frac{W_k\delta I_l(R_{c1}) - W_l\delta I_k(R_{c1})}{W_m\delta I_n(R_{c2}) - W_n\delta I_m(R_{c2})} = \Psi_{klmn}(R_{c1}, R_{c2}, \gamma),
\]

where

\[
\Psi_{klmn}(R_{c1}, R_{c2}, \gamma) = \frac{W_kF_l(R_{c1}, \gamma) - W_lF_k(R_{c1}, \gamma)}{W_mF_n(R_{c2}, \gamma) - W_nF_m(R_{c2}, \gamma)}
\]

is a special function that can be calculated for any pair of stations with cut-off rigidities \( R_{c1} \) and \( R_{c2} \), using known functions \( F_k(R_{c1}, \gamma) \), \( F_l(R_{c1}, \gamma) \), \( F_m(R_{c2}, \gamma) \), \( F_n(R_{c2}, \gamma) \) (calculated from Eq. (10)), and values \( W_k \equiv W_k(R_{c1}, R_{c1}) \), \( W_l \equiv W_l(R_{c1}, R_{c1}) \), \( W_m \equiv W_m(R_{c2}, R_{c2}) \), and \( W_n \equiv W_n(R_{c2}, R_{c2}) \) (calculated from Eq. (14)). After determining \( \gamma \) we can solve the system of Eqs. (19)–(22) for the three other unknown variables:

\[
\begin{align*}
\Delta R_{c1} &= \frac{F_k(R_{c1}, \gamma)\delta I_l(R_{c1}) - F_l(R_{c1}, \gamma)\delta I_k(R_{c1})}{W_kF_l(R_{c1}, \gamma) - W_lF_k(R_{c1}, \gamma)}, \\
\Delta R_{c2} &= \frac{F_m(R_{c2}, \gamma)\delta I_n(R_{c2}) - F_n(R_{c2}, \gamma)\delta I_m(R_{c2})}{W_mF_n(R_{c2}, \gamma) - W_nF_m(R_{c2}, \gamma)}.\end{align*}
\]
This method can be generalized for simultaneously using many CR ground detectors for on-line determination of the SEP rigidity spectrum in the interplanetary space.¹¹

7. On-Line Determination of the Time of Ejection, Diffusion Coefficient, and SEP Spectrum at the Source

According to observation data of many events for about 60 years (see reviews in Refs. 3 and 12–21, and on some recent observations in Refs. 22–24) and according to theoretical analysis,²⁵,²⁶ the time variation of SEP flux and energy spectrum for high-energy particles detected on the ground 15–20 min after ejection into solar wind can be described in the first approximation by the solution of isotropic diffusion from the instantaneous source described by the function

\[ Q(R, r', t') = N_o(R) \delta(r') \delta(t'). \]  

(28)

Let us suppose that the time of ejection and the diffusion coefficient are known. In this case the expected SEP rigidity spectrum at the distance \( r \) from the Sun at the time \( t \) after ejection will be

\[ N(R, r, t) = N_o(R) \times \left[ 2\pi^{1/2} (K(R) t)^{3/2} \right]^{-1} \exp \left( -\frac{r^2}{4K(R) t} \right), \]  

(29)

where \( N_o(R) \) is the rigidity spectrum of total number of SEP at the source, \( t \) is the time relative to the time of ejection, and \( K(R) \) is the known diffusion coefficient in the interplanetary space in the period of SEP event. For \( r = r_1 = 1 \) AU and at a time \( t_1 \) the spectrum according to Eq. (8) will be described by the function

\[ N(R, r_1, t_1) = b(t_1) R^{-\gamma(t_1)} D_o(R), \]  

(30)

where \( b(t_1) \) and \( \gamma(t_1) \) are parameters determined from the observed rigidity spectrum at \( t_1 \), and \( D_o(R) \) is the spectrum of galactic CR before the event.

If the time of ejection \( T_e \) and diffusion coefficient \( K(R) \) are unknown, to determine simultaneously on-line \( T_e, K(R) \), and SEP source spectrum \( N_o(R) \), we need information on SEP spectrum at least at three different
times $T_1$, $T_2$, and $T_3$ (in UT). In this case, we obtain for times after SEP ejection into solar wind:

$$t_1 = T_1 - T_e = x, \quad t_2 = T_2 - T_1 + x, \quad t_3 = T_3 - T_1 + x,$$

(31)

where $T_2 - T_1$ and $T_3 - T_1$ are known values and $x$ is unknown value to be determined. From the three equations for $T_1$, $T_2$, and $T_3$ of the type of Eq. (29), and taking into account Eqs. (30) and (31), we obtain

$$\frac{T_2 - T_1}{x(T_2 - T_1 + x)} = - \frac{4K(R)}{r_i^2} \ln \left\{ \frac{b(T_1)}{b(T_2)} \left( \frac{x}{T_2 - T_1 + x} \right)^{3/2} R^{-[\gamma(T_1) - \gamma(T_2)]} \right\},$$

(32)

$$\frac{T_3 - T_1}{x(T_3 - T_1 + x)} = - \frac{4K(R)}{r_i^2} \ln \left\{ \frac{b(T_1)}{b(T_3)} \left( \frac{x}{T_3 - T_1 + x} \right)^{3/2} R^{-[\gamma(T_1) - \gamma(T_3)]} \right\},$$

(33)

After dividing Eq. (32) by Eq. (33) we obtain

$$x = \frac{[(T_2 - T_1)\Psi - (T_3 - T_1)]}{1 - \Psi},$$

(34)

where

$$\Psi = \frac{T_3 - T_1}{T_2 - T_1} \times \frac{\ln \left\{ (b(T_1)/b(T_2))(x/(T_2 - T_1 + x))^{3/2} R^{\gamma(T_2) - \gamma(T_1)} \right\}}{\ln \left\{ (b(T_1)/b(T_3))(x/(T_3 - T_1 + x))^{3/2} R^{\gamma(T_3) - \gamma(T_1)} \right\}}.$$  

(35)

Equation (34) can be solved iteratively: as a first approximation, we can use $x_1 = T_1 - T_e \approx 500$ s which is the minimum time propagation of relativistic particles from the Sun to the Earth’s orbit. Then, by Eq. (35) we determine $\Psi(x_1)$ and by Eq. (34) we determine the second approximation $x_2$. Putting $x_2$ in Eq. (35) we compute $\Psi(x_2)$, and then by Eq. (34) we determine the third approximation $x_3$, and so on. After solving Eq. (34) and determining the time of ejection, we can very easily compute the diffusion coefficient from Eq. (32) or Eq. (33):

$$K(R) = - \frac{r_i^2(T_i - T_1)/4x(T_i - T_1 + x)}{\ln \left\{ (b(T_1)/b(T_i))(x/(T_i - T_1 + x))^{3/2} R^{\gamma(T_i) - \gamma(T_1)} \right\}},$$

(36)

where $i = 2$ or 3.
After determining the time of ejection and diffusion coefficient, it is easy to determine the SEP spectrum at the source:

\[ N_o(R) = 2\pi^{1/2}b(t_i)R^{-\gamma(t_i)}D_o(R)(K(R)t_i)^{3/2}\exp\left(\frac{r^2}{4K(R)t_i}\right), \quad (37) \]

where \( i = 1, 2, \) or \( 3. \)

8. Checking the Model by Calculations of Expected Diffusion Coefficient

To check the model of SEP propagation in the interplanetary space, we first determined the values of \( K(R) \). These calculations have been done according to the procedure described above by assuming that \( K(R) \) does not depend on the distance to the Sun. Results are shown in Fig. 2.

From Fig. 2 it can be seen that at the beginning of the event the derived results are not stable, due to large relative statistical errors. After few minutes the amplitude of CR intensity increase becomes many times larger than \( \sigma \) and we can see a systematical increase of the diffusion coefficient with time, reflecting the increasing of \( K(R) \) with the distance from the Sun.

![Fig. 2. The time behavior of \( K(R) \) for \( R \sim 10 \text{ GV} \).](image)
9. The Case when the Diffusion Coefficient Increases with Distance from the Sun

Let us suppose, according to Ref. 27, that the diffusion coefficient is

\[ K(R, r) = K_1(R) \times \left( \frac{r}{r_1} \right)^\beta. \]  

(38)

In this case

\[
n(R, r, t) = N_o(R) \times r^{3\beta/(2-\beta)}(K_1(R) t)^{-3/(2-\beta)} \exp\left( -\frac{r^{\beta+2-\beta}}{(2-\beta)^2 K_1(R) t} \right) \exp \left( -\frac{r^{\beta+2-\beta}}{(2-\beta)^2 K_1(R) t} \right) \exp \left( -\frac{r^{\beta+2-\beta}}{(2-\beta)^2 K_1(R) t} \right). \]  

(39)

If we know \( n_1, n_2, n_3 \) at times \( t_1, t_2, t_3 \), the final solutions for \( \beta, K_1(R), \) and \( N_o(R) \) will be

\[
\beta = 2 - 3 \left[ \ln \left( \frac{t_2}{t_1} \right) - \frac{t_3(t_2 - t_1)}{t_2(t_3 - t_1)} \ln \left( \frac{t_4}{t_1} \right) \right] \\
\times \left[ \ln \left( \frac{n_1}{n_2} \right) - \frac{t_2(t_2 - t_1)}{t_2(t_3 - t_1)} \ln \left( \frac{n_1}{n_3} \right) \right]^{-1},
\]

(40)

\[
K_1(R) = \frac{r_1^2(t_1^{-1} - t_2^{-1})}{3(2-\beta) \ln(t_2/t_1) - (2-\beta)^2 \ln(n_1/n_2)} \\
= \frac{r_1^2(t_1^{-1} - t_3^{-1})}{3(2-\beta) \ln(t_3/t_1) - (2-\beta)^2 \ln(n_1/n_3)},
\]

(41)

\[
N_o(R) = n_1(2-\beta)^{(4+\beta)/(2-\beta)} \Gamma \left( \frac{3}{2-\beta} \right) \times r_1^{-3\beta/(2-\beta)} \\
\times \left( K_1(R) t_k \right)^{3/(2-\beta)} \times \exp \left( -\frac{r_1^2}{(2-\beta)^2 K_1(R) t_k} \right). \]

(42)

In Eq. (42) index \( k = 1, 2, \) or 3. To check the model, let us again determine the diffusion coefficient from CR observation data. In Fig. 3 the values of parameter \( K_1(R) \) are shown.

From Fig. 3 it can be seen that at the very beginning of event (the first few points) the result is unstable: in this period the amplitude of increase is relatively small, so the relative accuracy is low. After the first points we have a stable result (compare with Fig. 2, where the diffusion coefficient was found as effectively increasing with time).
Fig. 3. Diffusion coefficient $K_1(R)$ near Earth’s orbit (in units $10^{23} \text{cm}^2 \text{s}^{-1}$) in dependence of time (in minutes after 11:40 UT of September 29, 1989).

10. SEP Forecasting by Using only NM Data

By using the first few minutes of the SEP event in NM data we can determine by Eqs. (40)–(42) the effective parameters $\beta$, $K_1(R)$, and $N_0(R)$, corresponding to rigidities 7–10 GV, and then by Eq. (39) we determine the forecasting curve of expected SEP flux behavior for total neutron intensity. We compare this curve with time variation of observed total neutron intensity. In reality, we use data for more than three times by fitting the obtained results in comparison with experimental data to reach the minimal

Fig. 4. Calculation of parameters $\beta$, $K_1(R)$, and $N_0(R)$, and forecasting by Eq. (30) of total neutron intensity (time $t$ is in minutes after 11:40 UT of September 29, 1989; curves — forecasting, circles — observed total neutron intensity).
residual (see Fig. 4 which contains eight panels for times $t = 110$ min up to $t = 220$ min after 11:40 UT on September 29, 1989). From Fig. 4 it can be seen that it is not enough to use only the first few minutes of NM data ($t = 110$ min): the obtained curve forecasts too low intensity. For $t = 115$ min the forecast shows some bigger intensity, but also not enough. Only for $t = 120$ min (15 min of increase after beginning) and later (up to $t = 140$ min) we obtain about stable forecast with good agreement with observed CR intensity. Therefore, we can conclude that the model described by Eqs. (39)–(42) reflects adequately SEP generation on the Sun and propagation in the interplanetary space.

11. SEP Forecasting by the On-Line Use of Both NM and Satellite Data

The results described above, based on NM on-line data, reflect the situation in SEP behavior in the high-energy (more than few GeV) region. For extrapolation of these results to the low-energy interval (more dangerous for aircrafts at high latitudes and altitudes), we use satellite on-line data (now are available through the Internet for public using the GOES satellite one-min data in real-time scale). The problem is how to extrapolate the SEP energy spectrum from high NM energies to very low energies detected by satellite. The main idea of this extrapolation is the following: the source function, time of ejection, and diffusion coefficient in both energy ranges are the same. The source function relative to time is again a $\delta$-function, and relative to energy is determined by Eq. (33). To describe NM and satellite data by one analytical formula we no longer use a power law function with $\gamma = \text{const.}$ according to Eq. (8), but with an energy-dependent index:

$$\frac{\Delta D(R)}{D_o(R)} = bR^{-\gamma}; \quad \gamma = \gamma_o + a \ln \left( \frac{E_k}{E_{ko}} \right)$$

(43)

with maximum at

$$E_{k,\text{max}} = E_{ko} \exp \left( -\frac{\gamma_o}{a} \right).$$

(44)

With parameters $b, a, \gamma_o, E_{ko}$ of primary spectrum beyond the atmosphere in the form of Eq. (34) we determine by the method of coupling functions as described in Secs. 5 and 6 (by fitting to experimental data), but now instead of two components with different coupling functions we need four (additional two components we use from satellite data). Figure 5 shows
results based on the NM and satellite data of forecasting of expected SEP fluxes also in small-energy intervals and comparison with observation satellite data.

From Fig. 5 it can be seen that by using more and more data, the forecast for \( E_k \geq E_0 = 0.1 \) GeV improved and 30–40 min after the beginning of the SEP event the forecasted curve practically coincide with observation data up to about 2500 min.

12. On the Connection of SEP Fluxes with Differential and Integral Radiation Doses Inside Aircraft

In Ref. 28 we introduced a new definition: integral multiplicity \( M_{rd}(R, S, h) \) for radiation dose from one primary CR particle with rigidity \( R \) (in GV) inside the aircraft under shielding \( S \) (in g/cm\(^2\)) at altitude determined by air pressure \( h \) (also in g/cm\(^2\)). In this case the differential radiation dose per unit time \( D_{rd}(S, h, R_c, t) \) will be:

\[
D_{rd}(S, h, R_c, t) = \int_{R_c}^{\infty} D(R, t)M_{rd}(R, S, h)dR ,
\]

where

\[
D(R, t) = D_k(R, t) + D_s(R, t)
\]
Radiation Hazard from Large SEP Events for Aircraft

is the total differential primary rigidity spectrum of CR which contain galactic CR spectrum $D_g(R,t)$ and SEP spectrum $D_s(R,t)$. Note that $M_{rd}(R,S,h)$ is the same for galactic CR and for SEP. As was shown in Ref. 28, $M_{rd}(R,S,h)$ can be very easily determined from latitude surveys of radiation dose measurements at different altitudes and locations with different cut-off rigidities:

$$M_{rd}(R,S,h) = \frac{I_{rd}(S,h,Rc,t_0)}{D_g(R,t_0)}a_{rd}k_{rd}R^{-(k_{rd}+1)} \times (1 - a_{rd}R_c^{-k_{rd}})^{-1} \exp(-a_{rd}R^{-k_{rd}}), \quad (47)$$

where $I_{rd}(S,h,Rc,t_0)$ is the differential radiation dose inside the aircraft from the galactic CR spectrum before SEP event at some time $t_0$. The integral additional radiation dose during the SEP event (additional to obtained from galactic CR) will be

$$I_{rd}^s(S,h,Rc,t_1,t_2) = \int_{t_1}^{t_2} dt \int_{R_c(t)}^{\infty} D_s(R,t)M_{rd}(R,S,h(t))dR, \quad (48)$$

where $D_s(R,t)$ is determined by Eq. (39) with parameters $\beta$, $K_1(R)$, and $N_o(R)$ estimated from Eqs. (40)–(42) on the basis of on-line CR (NM and satellite) one-min data, as was described in Secs. 10 and 11.

13. Alerts in Cases when SEP Events are Expected to be Dangerous

If the predicted fluxes and fluence are expected to be dangerous for aircraft, preliminary "SEP-Alert-1" will be sent a few minutes after the event beginning. As more data become available, better predictions of the expected fluxes will be made and more definitive Alert-2, Alert-3, and so on will automatically be issued. Alerts will give information on the expected time and level of danger for aircraft at different altitudes and at different cut-off rigidities. Experts should decide what to do operationally; for example, to decrease their altitudes to protect crew and passengers as well as electronics and navigation system from radiation hazard.

14. Conclusion

We show that by using on-line CR data from ground NM in the high-energy range and from satellites in the low-energy range during the first
30–40 min after the start of the SEP event, it is possible to predict the expected SEP integral fluxes for different energies up to a few days ahead. The total (event-integrated) fluence of the event, and the expected radiation hazard can also be estimated and corresponding alerts to experts operating different aircrafts at different altitudes and at different cut-off rigidities can be sent automatically. These experts should decide what to do operationally.

Acknowledgments

This research was partly supported by EU COST-724. The author is grateful to Prof. Yuval Ne’eman and Dr. Abraham Sternlieb, Prof. Isaac Ben Israel and Dr. Zvi Kaplan, for constant interest and support of Israel Cosmic Ray and Space Weather Center, and Emilio Segre’ Observatory. The author further thanks A. Belov, M. Duldig, E. Eroshenko, N. Iucci, K. G. McCracken, M. Murat, M. Parisi, L. Pustil’nik, M. A. Shea, D. F. Smart, M. Storini, G. Villoresi, V. Yanke, and I. Zukerman for interesting discussions. The work of neutron monitor on Mt. Hermon is supported by Tel Aviv University, UNIRoma-Tre, and IFSI-CNR Collaboration.

References

Radiation Hazard from Large SEP Events for Aircraft

This page intentionally left blank
MONITORING AND FORECASTING OF RADIATION HAZARD FOR AIRCRAFT FROM GALACTIC COSMIC RAYS

LEV DORMAN
Israel Cosmic Ray and Space Weather Center and Emilio Segre’ Observatory
Tel Aviv University, Technion and Israel Space Agency
P. O. Box 2217, Qazrin 12990, Israel
Cosmic Ray Department of IZMIRAN, Russian Academy of Science
Troitsk 142092, Moscow Region, Russia

Initially we determine the dimension of Heliosphere (modulation region), radial diffusion coefficient and other parameters of convection–diffusion and drift mechanisms of cosmic ray (CR) long-term variation including its dependence on particle energy, level of solar activity (SA), and general solar magnetic field. We obtain this important information on the basis of CR and SA data in the past taking into account the theory of convection–diffusion and drift global modulation of galactic CR in the Heliosphere. By using these results and regularly published predictions of expected SA variation we may make predictions of expected CR intensity variation. We introduce new nominations: integral multiplicity and coupling function for radiation dose inside aircraft caused by galactic CR. By the method of coupling functions we estimate the connection between CR intensity long-term variation and radiation hazard for aircrafts with dependence of altitude, geomagnetic cutoff rigidity, and shielding inside aircraft. We show that in this way we may predict the expected radiation hazard for any aircraft characterized with the dependence on several parameters: altitude, cutoff rigidity, shielding. It thus becomes important to estimate expected long-term changes in the planetary distribution of cutoff rigidities which also influence the galactic CR intensity reaching the atmosphere and thus CR-induced aircraft radiation hazard.

1. The Method

The method, described below, takes into account that the galactic cosmic ray (CR) intensity observed on the Earth is caused by solar processes propagating through the Heliosphere that started many months before. In paper Ref. 1 this method of using CR and solar activity (SA) data for solar cycles 19–22, also taking into account drift effects according to Ref. 2, was
considered. It was shown that it is very important to include drift effects that depend on the sign of solar polar magnetic field and determined by the difference of total CR modulation during $A > 0$ and $A < 0$ polarity cycles and effects of the tilt angles between the interplanetary neutral current sheet and the equatorial plane. It became possible to explain the great difference in time-lags between CR and SA in hysteresis phenomenon for even and odd solar cycles. But, are drift effects important when we consider situations near solar minima? We showed that drift effects became negligible for some short period between two maxima of SA, when the drift effect does not change the sign; moreover, according to Ref. 2 this influence for high-energy particles is expected to be especially small near SA minima. Therefore, in this paper, we will estimate the extent of the modulation region and of the residual CR modulation depending on the primary CR particle rigidity in the last solar minima without taking into account drift effects. This will be done in the context of a model for the global modulation of CR in the Heliosphere, taking into account time-lag processes in the interplanetary space relative to processes on the Sun using results of SA–CR hysteresis effects.

2. Hysteresis Phenomenon and Model of Cosmic Ray Global Modulation

It was shown in Ref. 3 that the time of propagation through the Heliosphere of particles with rigidity $> 10$ GV, to which neutron monitors (NM) are sensitive, is no longer than one month (demonstrated again below on the basis of new data). This time is at least one order of magnitude smaller than the observed time-lag in the hysteresis phenomenon. This means that the hysteresis phenomenon on the basis of NM data can be considered as a quasi-stationary problem with parameters of CR propagation changing in time. In this case according to Refs. 4 and 5

$$\frac{n(R, r, t)}{n_o(R)} \approx \exp \left( - a \int_r^{r_o} \frac{u(r, t)dr}{D_r(R, r, t)} \right), \quad (1)$$

where $n(R, r, t)$ is the differential rigidity CR density; $n_o(R)$ is the differential rigidity density spectrum in the local interstellar medium out of the Heliosphere; $a \approx 1.5$; $u(r, t)$ is the effective solar wind velocity (taking into account shocks and high-speed solar wind streams); and $D_r(R, r, t)$ is the effective radial diffusion coefficient that depends on the distance $r$ from
Radiation Hazard for Aircraft from Galactic CRs

the Sun of particles with rigidity $R$ at the time $t$. According to Refs. 6 and 7 the connection between $D_r(R, r, t)$ and SA can be described by the relation

$$D_r(R, r, t) \propto r^\beta \left( W\left(\frac{t-r}{u}\right)\right)^{-\alpha},$$

(2)

where $W(t - r/u)$ is the sunspot number in the time $t - r/u$. By comparison with observation data it was determined in Refs. 6 and 7 that parameter $0 \leq \beta \leq 1$ and $\alpha \approx 1/3$ in the period of high SA ($W(t) \approx W_{\text{max}}$) and $\alpha \approx 1$ near solar minimum ($W(t) \ll W_{\text{max}}$). Here we suppose, in accordance with Ref. 8, that

$$\alpha(t) = \frac{1}{3} + \left(\frac{2}{3}\right) \left(1 - \frac{W(t)}{W_{\text{max}}}\right),$$

(3)

where $W_{\text{max}}$ is the sunspot number at solar maximum.

According to Eq. (1) the expected value of the natural logarithm of CR intensity global modulation at the Earth’s orbit, taking into account Eqs. (2) and (3), will be

$$\ln(n(R, X_o, \beta, r_E, t)_{\text{exp}}) = A - B \times F(t, X_o, \beta, W(t - X)_{X_E}^{X_o} X^{-\beta} dX),$$

(4)

where

$$F(t, X_o, \beta, W(t - X)_{X_E}^{X_o}) = \int_{X_E}^{X_o} \left(\frac{W(t - X)}{W_{\text{max}}}\right)^{\frac{1}{3} + \frac{2}{3}(1 - W(t - X)/W_{\text{max}})} X^{-\beta} dX,$$

(5)

$X = r/u, X_E = 1 \text{AU}/u, X_o = r_o/u$, and $n(R, X_o, \beta, r_E, t)_{\text{exp}}$ is the expected galactic CR density at the Earth’s orbit dependent on the values of parameters $X_o$ and $\beta$. Regression coefficients $A(R, X_o, \beta, t_1, t_2)$ and $B(R, X_o, \beta, t_1, t_2)$ can be determined by correlation between observed values $\ln(n(R, r_E, t))_{\text{obs}}$ and the values of $F(t, X_o, \beta, W(t - X)_{X_E}^{X_o})$, calculated according to Eq. (5) for different values of $X_o$ and $\beta$ for the period of time between $t_1$ and $t_2$. In Ref. 9 three values of $\beta = 0, 0.5$, and 1 have been considered; it was shown that $\beta = 1$ strongly contradicts CR and SA observation data, and that $\beta = 0$ is the most reliable value. Therefore, we will consider here only this value.

We used monthly data of sunspot numbers and Climax NM data (USA, Colorado, N39, W106, \(H = 3400\) m, \(R_c = 2.99\) GV), as well as Huancayo (Peru, S12, W75, \(R_c = 12.92\) GV, \(H = 3400\) m), or Haleakala (Hawaii, N20, W156, \(R_c = 12.91\) GV, \(H = 303\) m) NM data for the solar minimum (January 1994–January 1997, \(W \leq 40\)). We calculated correlation coefficients \(\rho(X_o)\) between the natural logarithm of observed and expected counting rate according to Eq. (5) with values of \(X_o = r_o/u = 1, 2, 3, \ldots, 60\) av. months (\(X_o\) is measured in units of av. month = 365.25/12 days, \(r_o\) in AU, and \(u\) in AU/av. month). We estimate \(X_{o,\text{max}}\) when correlation coefficients \(\rho(X_o)\) reaches the maximum value. To determine \(X_{o,\text{max}}\) more exactly, we approximated the dependence of \(\rho\) on \(X_o\) in the vicinity of \(X_{o,\text{max}}\) by a parabolic function \(\rho(X_o) = aX_o^2 + bX_o + c\). In this case \(d\rho/dX_o = 2aX_o + b\), and \(X_{0,\text{max}} = -b/2a\). For Climax NM monthly data LN(CL1M) we approximated \(\rho(X_o)\) by

\[
\rho(X_o) = 0.000915X_o^2 - 0.03778X_o - 0.54937 \tag{6}
\]

with correlation coefficient 0.943 ± 0.012, thus obtaining

\[
X_{o,\text{max}} = 20.6 \pm 1.2\; \text{av. months}, \quad \rho_{\text{max}} = -0.939. \tag{7}
\]

For Huancayo/Haleakala NM monthly data LN(HU/HAL1M) we obtained

\[
\rho(X_o) = 0.000859X_o^2 - 0.030173X_o - 0.644715 \tag{8}
\]

with correlation coefficient 0.987 ± 0.003, and

\[
X_{o,\text{max}} = 17.6 \pm 0.5\; \text{av. months}, \quad \rho_{\text{max}} = -0.910. \tag{9}
\]

According to direct measurements on space probes the average solar wind speed for the period 1965–1990 near the Earth’s orbit at \(r = 1\) AU was \(u_1 = 4.41 \times 10^7\; \text{cm/s} = 7.73\; \text{AU/av. month}\). The function \(u(r)\) is determined by solar wind interactions with galactic CR and the anomalous component of CR, with neutral atoms penetrating from interstellar space. According to calculations in Ref. 10 the change of solar wind velocity with the distance
Radiation Hazard for Aircraft from Galactic CRs

\( r \) from the Sun can be described approximately as

\[
u(r) \approx u_1 \left( 1 - b \left( \frac{r}{r_{tsw}} \right) \right), \tag{10}
\]

where \( r_{tsw} \) is the distance to the terminal shock and parameter \( b \approx 0.13/0.45 \) depending on the sub-shock compression ratio and on injection efficiency of pickup protons. On the basis of Eq. (10) we can determine the radius of CR modulation region \( r_{mod} \) from equation:

\[
X_{\alpha,\text{max}} = \int_0^{r_{mod}} \left( u_1 \left( 1 - \frac{br}{r_{tsw}} \right) \right)^{-1} \, dr = - r_{tsw} \ln \left( \frac{-b + r_{mod}/r_{tsw}}{bu_1} \right), \tag{11}
\]

from which follows

\[
r_{mod} = r_{tsw} \left( b + \exp \left( -\frac{X_{\alpha,\text{max}}bu_1}{r_{tsw}} \right) \right). \tag{12}
\]

Let us assume that the radius of modulation region \( r_{mod} \) for Climax NM data (effective rigidity 10–15 GV) is about the same as radius of the Heliosphere \( r_{tsw} \). In this case from Eq. (12) at \( r_{mod} = r_{tsw} \) we obtain

\[
r_{mod} = - \frac{bu_1 X_{\alpha,\text{max}}}{\ln(1 - b)}. \tag{13}
\]

For the average velocity of solar wind at the most reliable value of \( b \approx 0.3 \) we obtain from Eq. (13)

\[
u_{av} = - \frac{u_1 b}{\ln(1 - b)} = 0.84u_1, \tag{14}
\]

which for Climax NM gives \( r_{mod} = 134 \pm 8 \) AU, and for Huancayo/Haleakala NM \( r_{mod} = 114 \pm 4 \) AU.

4. Estimation of Correlation and Regression Coefficients

Determination of regression coefficients \( A \) and \( B \) in Eq. (4) makes it possible to determine the CR intensity outside of the modulation region, and the effective radial diffusion coefficient depending on the effective particle rigidity \( R \). The use of monthly data allows the determination of regression coefficients \( A \) and \( B \) only for integer values of \( X_\alpha \). To increase the accuracy, we also use 11-month-moving averaged data. Therefore, for example, for LN(CL11M) we determined \( A \) and \( B \) for \( X_\alpha = 20 \) (\( A = 8.367430, B = 0.006678 \)) and for \( X_\alpha = 21 \) (\( A = 8.367825, B = 0.006285 \)), and then by
interpolation for $X_{o,\text{max}} = 20.6$. In the same way we determined $A$ and $B$ for $\text{LN(HU/HAL11M)}$.

5. Cosmic Ray Intensity Outside the Heliosphere

The regression coefficient $A$ in Eq. (4) according to Eq. (1) is

$$A = \ln(n_o(R)).$$

(15)

This coefficient determines the galactic CR intensity outside of the modulation region. In our case the coefficient $A$ means the logarithm of the CR intensity outside the modulation region based on the Climax or on Huancayo/Haleakala NM, i.e., corresponds to primary CR particles with effective rigidities about 10–15 and 30–40 GV, respectively. We estimate that the accuracy in determining $\ln(n_o(R))$ for the Climax NM is $\pm 0.004\%$, and for the Huancayo/Haleakala NM it is $\pm 0.008\%$. In January 1994 $\text{LN(CL11M)} = 8.31754$, $\text{LN(HU/HAL11M)} = 7.44217$, that residual modulation (as modulation relative to the CR intensity out of the Heliosphere) was $5.014 \pm 0.004\%$ and $2.144 \pm 0.008\%$ for galactic CR with effective rigidity 10–15 and 30–40 GV. The residual modulation is almost inversely proportional to the effective rigidity of CR particles. Many scientists assume that for minimum SA the NM-detected CR intensity reaches a value very near to the intensity outside the Heliosphere. Let us check this. The maxima of $\text{LN(CL11M)}$ and $\text{LN(HU/HAL11M)}$ were reached in June and July 1997 with values $8.360387$ and $7.46112$ (minimal residual modulations of $0.361 \pm 0.004\%$ and of $0.249 \pm 0.008\%$ for 10–15 and 30–40 GV particles). The results obtained show that even high-energy CR particles (10–15 and 30–40 GV) inside the Heliosphere at the Earth’s orbit never reach the intensity out of the Heliosphere (in interstellar space) even near the minimum of SA: the minimal residual modulations of $0.361 \pm 0.004\%$ and $0.249 \pm 0.008\%$ for 10–15 and 30–40 GV particles were found for June–July 1997, with some time-lag relative to minimum SA (caused by the big dimension of modulation region).

6. Prediction of Cosmic Ray Variations by Integral $F$ Near SA Minimum

Figure 1 shows the time variations of predictions from the integral $F$ [calculated on the basis of monthly sunspot numbers $W$ according to Eq. (5)] compared with the observed natural logarithm of the month’s
average counting for Climax NM LN(CL1M) and for 11-month smoothed LN(CL11M).

In this case we did not take into account the drift effects because according to Ref. 2 for high-energy particles (for protons with energy much more than 1 GeV) near the SA minimum they are negligible in comparison with convection–diffusion modulation which does not depend on the sign of the solar general magnetic field. For Climax NM the correlation coefficient between predicted $F$ and observed values of CR intensity LN(CL11M) was found equal to $0.993 \pm 0.002$. The same analysis for Huancayo/Haleakala NM gave correlation coefficient between predicted $F$ and observed values of CR intensity LN(HU/HAL11M) equal to $0.970 \pm 0.007$.

7. Forecasting CR Intensity During the Periods of Increasing SA

In Sec. 6, we considered forecasting CR intensity near the minimum of SA when the drift effects are negligible. To demonstrate how we can take
into account the drift effects, let us consider, for example, forecasting CR intensity during the period of increasing SA at the onset of the solar cycle during January 1996–August 1999. In this case there is no information on the amplitude of drift modulation $A_{dr}$, which is suggested to be proportional to the theoretically expected value according to Ref. 2 and normalized to sunspot number $W = 75$. If the cycle is only starting, we do not know $A_{dr}$ for this cycle, but we known the type of cycle (odd or even) and we can use published predicted values of sunspot numbers for few years ahead. That lets us use Eqs. (4) and (5) for convection–diffusion modulation and average value of $A_{dr}$ obtained for previous cycles 19–22 in Ref. 1: $A_{dr} \approx 2\%$ and $0.25\%$ at $W = 75$ for Climax NM (effective rigidity of primary particles 10–15 GV) and Huancayo/Haleakala NM (35–45 GV) accordingly. Predicted CR intensity variations (separately expected convection–diffusion modulation and expected convection–diffusion + drift modulations) and observed CR long-term variation during 1996–2000 are shown in Fig. 2 for Climax NM.

It can be seen that in this case taking into account drift effects is sufficient. The correlation coefficient between predicted and observed cosmic radiation is 0.988. For Huancayo/Haleakala NM with $A_{dr} \approx 0.25\%$ at $W = 75$ the coefficient is 0.986.

Fig. 2. Comparison of predicted convection–diffusion modulation $PR_{CD}$ and predicted with taking into account drift effects $PR_{CD} + DR$ with observation OBSLN(CL11M) by Climax NM for period January 1996–August 1999.

Let us consider all periods from 1953 to 2000 by automatically taking into account drift effects during all solar cycles according to the method described above using data of tilt angle and sunspot numbers. Based on data for 18 years (May 1976–September 1993), we found that there was a very good relationship between $T$ and $W$: for 11-month smoothed data $T = 0.349W + 13.5^\circ$ with correlation coefficient 0.955. We used 11-month smoothed data of $W$ (shown in Fig. 3) and determined the amplitude $A_{dr}$ of drift effects as drift modulation at $W_{11M} = 75$ (average value of $W_{11M}$ for 1953–1999). Or the information on reversal periods, we used the following (ftp://ftp.ngdc.noaa.gov/STP/SOLAR_DATA): August 1949 ± 9 months, December 1958 ± 12 months, December 1969 ± 8 months, March 1981 ± 5 months.

![Figure 3](image-url)

Fig. 3. CR data correction for drift effects in 1953–2000 (solar cycles 19–22 and onset of solar cycle 23): LN(CL11M) — observed natural logarithm of Climax neutron monitor counting rate smoothed for 11 months, LN(CLCOR3_DR2%) — corrected Climax data with assumed drift effect according to Ref. 2 with $A_{dr} = 2\%$ at $W_{11M} = 75$. Interval between two horizontal lines corresponds to a 5% variation of CR intensity. Smoothed sunspot numbers $W_{11M}$ are shown for comparison.
months, and June 1991 ± 7 months. The final results for the period January 1953–November 2000 are shown in Fig. 3.

Figure 3 shows five SA minima and in all cases there is very good coincidence of CR intensity corrected for drift effects and observed CR intensity which confirms that in minima of SA drift effects are not important. On the other hand, differences of a few percent in maxima of SA mean that in maxima of SA drift effects are important.

9. On the Connection of Galactic CR Intensity with Differential and Integral Radiation Doses Inside Aircraft

Let us introduce a new definition: integral multiplicity \( M_{\text{rd}}(R, S, h) \) for radiation dose from one primary CR particle with rigidity \( R \) (in GV) inside the aircraft under shielding \( S \) (in g/cm\(^2\)) at altitude determined by air pressure \( h \) (also in g/cm\(^2\)). In this case the differential radiation dose per unit of time \( I_{\text{rd}}(S, h, R_c, t) \) will be:

\[
I_{\text{rd}}(S, h, R_c, t) = \int_{R_c}^\infty D(R, t)M_{\text{rd}}(R, S, h)\,dR , \tag{16}
\]

where \( D(R, t) \) is the differential primary rigidity spectrum of CR. Let us suppose that \( D_{\text{rd}}(S, h, R_c, t) \) is measured in a broad interval of cutoff rigidities, and \( D(R, t) \) is also known. In this case from Eq. (16) we obtain

\[
\frac{\partial I_{\text{rd}}(S, h, R_c, t)}{\partial R_c} = -D(R_c, t)M_{\text{rd}}(R_c, S, h) . \tag{17}
\]

From Eq. (17) follows

\[
M_{\text{rd}}(R, S, h) = -\left( \frac{\partial I_{\text{rd}}(S, h, R_c, t)}{D(R_c, t)\partial R_c} \right)_{R_c \to R} . \tag{18}
\]

The integral radiation dose \( I_{\text{rd}}(S, h, R_c, t_1, t_2) \) during the flight from \( t_1 \) to \( t_2 \) will be determined by

\[
I_{\text{rd}}(S, h, R_c, t_1, t_2) = \int_{t_1}^{t_2} I_{\text{rd}}(S, h(t), R_c(t), t)\,dt \\
= \int_{t_1}^{t_2} dt \int_{R_c(t)}^\infty D(R, t)M_{\text{rd}}(R, S, h(t))\,dR . \tag{19}
\]
10. Main Factors Determining the Variation of Differential Radiation Dose with Time

From Eq. (16) it follows that

\[
\delta D_{rd}(S, h, R_c, t) = \int_{R_c}^{\infty} \delta D(R, t) M_{td}(R, S, h) dR \\
+ \int_{R_c}^{\infty} D(R, t) \delta M_{td}(R, S, h) dR \\
- D(R_c, t) M_{td}(R_c, S, h) \delta R_c .
\]  (20)

Let us consider the relative variations of the differential radiation dose:

\[
\frac{\delta D_{rd}(S, h, R_c, t)}{D_{rd}(S, h, R_c, t_o)} = \int_{R_c}^{\infty} \frac{\delta D(R, t)}{D(R, t_o)} W_{td}(R_c, R, S, h) dR \\
+ \int_{R_c}^{\infty} \frac{\delta M_{td}(R, S, h)}{M_{td}(R, S, h)} W_{td} dR - W_{td}(R_c, R_c, S, h) \delta R_c,
\]  (21)

where

\[
W_{td}(R_c, R, S, h) = \frac{D(R, t_o) M_{td}(R, S, h)}{I_{rd}(S, h, R_c, t_o)}
\]  (22)

is the radiation dose coupling function for aircraft that determines the connection between observed variation of the differential radiation dose inside the aircraft: with variation of primary CR spectrum outside the Earth’s magnetosphere [the first member in the right part of Eq. (21)]; with changing of integral multiplicity [the second member in Eq. (21)]; and with changing of cutoff rigidity [the last member in the right part of Eq. (21)].

11. Analytical Presentation of Radiation Dose Coupling Functions for Aircraft

In many papers it has been shown (see review in Ref. 11) that any coupling function (for NM, for muon telescopes, for balloon and aircraft CR measurements, for CR measurements on satellites) can be presented in analytical form through the polar normalized coupling functions referred in literature as the Dorman function (introduced in Ref. 12)

\[
W_{om}(R) = a_m k_m R^{-(k_m + 1)} \exp(-a_m R^{-k_m}) ,
\]  (23)
where parameters $a_m$ and $k_m$ depend on the air pressure $h$ and the level of SA. Let us note that these functions, described by Eq. (23), are normalized: $\int_0^\infty W_{om}(R)\,dR = 1$ at any values of $a_m$ and $k_m$. The same analytical form will be for the radiation dose coupling function for aircraft $W_{rd}(R_c, R, S, h)$, determined by Eq. (22), but in this case parameters $a_m$ and $k_m$ will depend not only on the air pressure $h$ and the level of SA, but also on the shielding $S$. The normalized coupling functions for a location with cutoff rigidity $R_c$, will be:

$$W_{rd}(R_c, R, S, h) = \begin{cases} a_{rd}k_{rd}R^{-(k_m+1)}(1 - a_{rd}R_{c}^{-k_{rd}})^{-1} \exp(-a_{rd}R^{-k_{rd}}) & \text{if } R \geq R_c, \\ 0 & \text{if } R < R_c. \end{cases}$$

(24)

From Eqs. (22) and (24) the integral multiplicity $M_{rd}(R, S, h)$ for aircraft radiation dose can be determined:

$$M_{rd}(R, S, h) = \frac{I_{rd}(S, h, R_c, t_0)}{D(R, t_0)} a_{rd}k_{rd}R^{-(k_{rd}+1)} \times (1 - a_{rd}R_{c}^{-k_{rd}})^{-1} \exp(-a_{rd}R^{-k_{rd}})|_{R_c \rightarrow R}.$$  

(25)

Equations (18) and (25) show that the integral multiplicity $M_{rd}(R, S, h)$ for radiation dose, an important characteristics for any aircraft, can be determined from aircraft latitude surveys.

12. Monitoring and Forecasting of Radiation Dose Inside the Aircraft

With formulae in Secs. 9–11 and ground-based CR measurements by NM and muon telescopes on-line monitoring of radiation dose inside any aircraft with dependence for shielding $S$, pressure (or altitude) $h(t)$ and cutoff rigidity $R_c(t)$ of the aircraft trajectory according to Eq. (19) can be achieved.

Moreover, on the basis of results obtained in Secs. 1–8, forecasting of expected CR intensity variation can be made, and then by formulae in Secs. 9–11 forecasting of the expected radiation dose for any type of aircraft characterized by some shielding $S$ and any flight trajectory characterized by parameters $h(t)$ and cutoff rigidity $R_c(t)$ are possible.
Acknowledgments

This research was partly supported by EU COST-724. The author thanks to Prof. Yuval Ne’eman and Dr. Abraham Sternlieb, Prof. Isaac Ben Israel and Dr. Zvi Kaplan, for constant interest and support of Israel Cosmic Ray and Space Weather Center, and Emilio Segre’ Observatory. The author further thanks A. Belov, M. Duldig, E. Eroshenko, N. Iucci, K. G. McCracken, M. Murat, M. Parisi, L. Pustil’nik, M. A. Shea, D. F. Smart, M. Storini, G. Villoresi, V. Yanke, and I. Zukerman for interesting discussions. The work of neutron monitor on Mt. Hermon is supported by Tel Aviv University, UNIRoma-Tre, and IFSI-CNR Collaboration.

References

COMPARATIVE MEASUREMENTS OF COSMIC RADIATION MONITORS FOR AIRCREW EXPOSURE ASSESSMENT

I. L. GETLEY
Department of Aviation, University of New South Wales
Sydney, NSW 2052, Australia
getleyian@hotmail.com

L. G. I. BENNETT*, M. L. BOUDREAU, B. J. LEWIS, A. R. GREEN and A. BUTLER
Department of Chemistry and Chemical Engineering
Royal Military College of Canada
P. O. Box 17000, Stn. Forces, Kingston
Ontario, Canada K7K 7B4
*bennett-l@rmc.ca

M. TAKADA† and T. NAKAMURA‡
National Institute of Radiological Sciences, Japan
†mtakada@nirs.go.jp
‡nakamura@cyric.tohoku.ac.jp

Various commercially available electronic personal dosimeters (EPDs) have recently been flown on numerous scheduled airline flights in order to determine their viability as small, convenient monitors to measure cosmic radiation at altitude. Often, frequent flyers or airline crew will acquire such dosimeters and report the readings from their flights, without due regard for the mixed radiation field at altitude, which is different from the intended fields on land. A sampling of EPDs has been compared to two types of spectrometers, which measure the total radiation spectrum. The “HAWK” tissue equivalent proportional counter is considered a reference instrument and measures the total dose equivalent $H^*(10)$. The Liulin-4N and 4SN linear energy transfer spectrometers each have a silicon semiconductor-based PIN diode detector which provides an absorbed dose, $D$, but have been further developed to provide $H^*(10)$. A Thermo Electron FH41B and B-10, and EPD-N2, and several personal dosimeters (Fuji NRY-21 and NRF-20, and RADOS DIS-100) were also flown.
1. Introduction

With the 1990 recommendations of the International Commission on Radiological Protection, in their Publication Number 60, ICRP-60, which included the recognition of aircrew as being occupationally exposed to natural radiation, certain states undertook to implement them. A Directive was issued for European Union (EU) Member States to adhere to by 2000 and Transport Canada issued a Commercial and Business Aviation Advisory Circular to Canadian airlines in 2001, revised in 2006. These documents are consistent in that the intervention level is to be 6 mSv/year and that predictive codes may be used supported by periodic monitoring in lieu of personal or aircraft dosimetry. Subsequently, various research groups have measured the galactic cosmic radiation on jet aircraft and balloons, as well as theoretically or semi-empirically modeled the radiation at altitude. Efforts to determine the additional exposure from solar particle events are also underway.

The tissue equivalent proportional counter (TEPC), which measures both the low-linear energy transfer (LET) and high-LET radiation, is most often utilized at altitude. Supplementary to the TEPC, equipments such as a separate low-LET detector and a high-LET detector, are often flown. For example, the Royal Military College of Canada has an Eberline FHT 191N ionization chamber (IC) and a smart wide-energy neutron detection instrument (SWENDI) or regular Andersson–Braun neutron meter. The TEPC, IC, and SWENDI measure every several minutes throughout the flight. Alternatively, thermoluminescent detectors and neutron-sensitive bubble detectors give an estimate of the total accumulated dose for the flight. Totaling the low- and high-LET contributions of the appropriate combinations of detectors gives consistent total route ambient dose equivalents to those of the TEPC.

This equipment suite has provided the data as a basis for the development of the Predictive Code for Aircrew Radiation Exposure (PCAIRE). However, since the equipment is large, to periodically monitor the performance of a predictive code to be used in the future, compact instruments have been recently evaluated and compared to the existing suite. These include the Liulin spectrometers (models 4N, 4SN, and 4SA), and the Thermo Electron FH41B and B-10, and EPD-N2, Fuji NRF-20, NRY-21 and RADOS Technology DIS-100 electronic personal dosimeters (EPD) (Table 1). They have been contributed by the three authoring groups and have been flown at various times and in various combinations on routes covering the northern and southern hemispheres starting in 2005 and
Cosmic Radiation Monitors for Aircrew Exposure Assessment

Table 1. Available compact spectrometers and detectors for evaluation.

<table>
<thead>
<tr>
<th>Detector</th>
<th>Manufacturer</th>
<th>Dimensions (cm) (L x W x H)</th>
<th>Weight (g)</th>
<th>Battery Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liulin 4N</td>
<td>Solar Terrestrial Influences Laboratory (STIL)</td>
<td>10 x 10 x 5</td>
<td>520</td>
<td>2 Li SAFT cells, ~ 2 months</td>
</tr>
<tr>
<td>Liulin 4SN</td>
<td>STIL</td>
<td>10 x 8 x 2.5</td>
<td>250</td>
<td>12 V packs, ~ 20 h</td>
</tr>
<tr>
<td>Liulin 4SA</td>
<td>STIL</td>
<td>11 x 5.5 x 4.5</td>
<td>550</td>
<td>12 V pack, ~ 15 h</td>
</tr>
<tr>
<td>Display</td>
<td></td>
<td>11.5 x 4 x 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FH41B and B-10</td>
<td>Thermo Electron</td>
<td>5.7 x 3.2 x 11.5</td>
<td>200</td>
<td>19 V, ~ 5 months</td>
</tr>
<tr>
<td>EPD-N2</td>
<td>Thermo Electron</td>
<td>8.6 x 6.3 x 2</td>
<td>110</td>
<td>1 AA, ~ 50 days</td>
</tr>
<tr>
<td>NRY21</td>
<td>Fuji Electronics</td>
<td>10.3 x 5.5 x 1.5</td>
<td>110</td>
<td>1 AAA, ~ 5 days</td>
</tr>
<tr>
<td>NRF20</td>
<td>Fuji Electronics</td>
<td>5.0 x 7.5 x 2.3</td>
<td>77</td>
<td>1 AAA, ~ 2 months</td>
</tr>
<tr>
<td>DIS-100</td>
<td>RADOS Technologies</td>
<td>9.0 x 6.1 x 2.0</td>
<td>132</td>
<td>1 AAA, ~ 2 weeks</td>
</tr>
</tbody>
</table>

ending July 2006. The TEPC is considered the reference instrument and was usually flown with the compact ones under evaluation. When necessary, the IC and SWENDI combination was substituted. Additionally, their results have been compared to those of the predictive codes, PCAIRE, CARI, and EPCARD.

2. Equipment

For most groups, the TEPC, designed by Batelle Pacific Northwest National Laboratories (and subsequently upgraded by Far West Technologies to the HAWK and later HAWK2 TEPC), is considered the reference instrument for in-flight measurements. The detector consists of a propane-filled sphere (12.7 cm in diameter) with a 2.13-mm thick skin comprised of an A150 polymer (to simulate muscle). Together, the propane and the polymer simulate a microscopic volume (2-µm in diameter) of tissue. Both the low-LET, or non-neutron, and high-LET radiation, essentially neutrons, are measured.

The characteristics of the smaller detectors are described below and are summarized in Table 2. The Liulin spectrometers are differentiated from the
Table 2. Characteristics of the small detectors under evaluation.

<table>
<thead>
<tr>
<th>Detector</th>
<th>Capabilities</th>
<th>Method used to obtain $H^*(10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liulin 4N</td>
<td>Spectral information</td>
<td>Download data, spreadsheet</td>
</tr>
<tr>
<td></td>
<td></td>
<td>calculations</td>
</tr>
<tr>
<td>Liulin 4SN</td>
<td>Spectral information</td>
<td>Download data, spreadsheet</td>
</tr>
<tr>
<td>and 4SA</td>
<td></td>
<td>calculations</td>
</tr>
<tr>
<td>FH41B</td>
<td>Gamma</td>
<td>Download data, apply correction</td>
</tr>
<tr>
<td></td>
<td></td>
<td>via spreadsheet</td>
</tr>
<tr>
<td>EPD-N2</td>
<td>Gamma and neutrons</td>
<td>Readout $H_p$ for neutrons and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>gamma</td>
</tr>
<tr>
<td>NRY-21</td>
<td>Gamma and neutrons</td>
<td>Readout $H_p$ for neutrons and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>gamma</td>
</tr>
<tr>
<td>NRF-20</td>
<td>Gamma</td>
<td>Readout $H_p$ for gamma and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>add to bubble detector results</td>
</tr>
<tr>
<td>DIS-100</td>
<td>Gamma</td>
<td>Readout $H_p$ for gamma and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>add to bubble detector results</td>
</tr>
<tr>
<td>Bubble detectors</td>
<td>Neutrons</td>
<td>Count bubbles and multiply by</td>
</tr>
<tr>
<td></td>
<td></td>
<td>scaling factor (neutrons only)</td>
</tr>
</tbody>
</table>

dosimeters and the bubble detectors. The method to obtain all or part of $H^*(10)$ or $H_p(10)$ and $H_p(0.07)$ is also indicated.

The Liulin-4N, 4SN (with GPS), and 4SA (with GPS and alarm) spectrometers, manufactured by Solar Terrestrial Influences Laboratory are compact, lightweight, and portable. All have to be interfaced with a computer using the supplied software to obtain and record the data. The sensing element is a silicon semiconductor-based PIN diode and a multi-channel analyser sorts the inputs into 256 channels. The absorbed dose can then be calculated from the data obtained and, applying an appropriate quality factor, $Q(L)$, to each channel, the ambient dose equivalent $H^*(10)$ is calculated in a similar manner used for the TEPC.

The Eberline FH41B and B-10 are Geiger tube-based meters, designed for measurement of occupational doses, each having a different energy filter that measures photons and charged particles. The FH41B measures $H_x$, photon dose equivalent and the B-10 measures $H^*(10)$, ambient dose equivalent. As the FH41B-10 is not capable of measuring the neutron component directly, only secondary charged particles are detected.

The Thermo Electron EPD-N2 uses three silicon PIN diode detectors, feeding into four counters. It has a direct readout of $H_p(10)$ for both neutron and photon doses. The total doses are calculated using a weighted sum of the four internal counters. Calibration constants determine the weighting of the counts received on each of the four internal counters and are set for
accurate readings in moderated/scattered reactor fields. At these settings, the instrument is known to over-respond to thermal and slow neutrons by approximately a factor of 10.

The Fuji Electronics NRF-20 uses silicon radiation detectors to calculate personal effective dose, $H_p(10)$, and skin dose, $H_p(0.07)$. It is optimized to read gamma effective dose, calibrated with $^{137}$Cs. For these measurements, the unit was set to take dose readings $H_p(10)$ and $H_p(0.07)$ every 10 min in $\mu$Sv/10 min.

The FUJI Electronics NRY-21 uses four semiconductor radiation detectors and is capable of simultaneously measuring the effective dose composed of neutrons and gamma radiation, $H_p(10)$, and local skin dose, $H_p(0.07)$, from beta radiation. For the fast and slow neutron components, two different silicon semiconductor detectors are used. For the fast neutron sensor, a polyethylene radiator inserted in the direction of neutron incidence allows the sensor to act as a recoil proton detector. The slow neutron sensor is doped with natural boron to cause $^{10}$B(n, $\alpha$)Li reactions which are used to detect the low-energy neutrons.

The RADOS Technology DIS-100 measures gamma, beta, and X-radiation for $H_p(10)$ and $H_p(0.07)$, and uses direct ion storage (DIS) technology, which combines three ICs with a nonvolatile electronic charge storage element, allowing it to operate as a real-time personal dosimeter. For the purposes of the at-altitude measurements, the dosimeter was set to store the dose rate every 10 min. RADOS Technologies has developed a DIS-based system that also measures the neutron component of radiation fields, but was not available for these measurements.

3. Comparison of Compact Spectrometers, Dosimeters, and Codes

Since the absorbed dose measurements of the Liulin have been shown to be in excellent agreement with those of the TEPC, a full spectral analysis method has been adopted to determine the ambient dose equivalent $H^*(10)$. Alternatively, a linear and a polynomial method have been suggested. These methods apply linear or polynomial functions of the absorbed dose, which have been derived from calibrations. Since the functions are averaged for all flight-level radiation fields, it is suspected they are limited in their ability to account for field changes. Six flights have been flown with a Liulin and analyzed by these three methods and
compared to the corresponding PCAIRE calculation, as shown in Fig. 1. A maximum of 14% difference was found between all results, well within the measurement and calculation errors.

For the FH41B and B-10 dosimeters, a correction function has been developed based on simultaneous IC + SWENDI measurements on a long haul, return flight route to correct the results for these detectors to a complete $H^*(10)$ value for the total mixed field. This function is linear over the range of vertical cut-off rigidities and requires flight path information in order to be applied.

The dosimeters were flown on various flights throughout 2005, of which 14 are indicated in Table 3, for comparison. For each flight, the total ambient dose from the PCAIRE and EPCARD calculations and the corrected FH41B, Liulin-4N and EPD-N2 are given first, followed by the low-LET results from the EPD-N2, NRF-20, DIS-100, and the EPCARD non-neutron component. It can be seen that the EPD-N2 results are not consistent with the rest, as it is not calibrated for the field at altitude, as noted above. However, the other dosimeters do give comparable numerical values for the total and non-neutron component.

Further comparison between these dosimeters can be seen in Fig. 2 on the flight to AOGS 2005. In the upper half of the figure, the low-LET response is indicated for these dosimeters and compared to the Eberline FHT 191N IC and to an EPCARD non-neutron calculation. All numerical
### Table 3. Comparison of small dosimeter flight ambient dose results.

<table>
<thead>
<tr>
<th>Date (2005)</th>
<th>Flight</th>
<th>PCAIRE (µSv/h)</th>
<th>EPCARD (µSv/h)</th>
<th>FH41B* (µSv/h)</th>
<th>Liulin-4N (µSv/h)</th>
<th>EPD-N2 (µSv/h)</th>
<th>NRF-20 (γ) (µSv/h)</th>
<th>EPD-N2 (γ) (µSv/h)</th>
<th>EPCARD non-neutron (µSv/h)</th>
</tr>
</thead>
<tbody>
<tr>
<td>June 22</td>
<td>SIN–SYD</td>
<td>25.4</td>
<td>19.9</td>
<td>20.53</td>
<td>20.5</td>
<td>220</td>
<td>9†</td>
<td>N/A</td>
<td>10.03</td>
</tr>
<tr>
<td>July 6</td>
<td>MEL–LAX</td>
<td>28.84</td>
<td>27.5</td>
<td>27.94</td>
<td>24.7</td>
<td>250</td>
<td>13‡</td>
<td>20</td>
<td>14.89</td>
</tr>
<tr>
<td>July 8</td>
<td>LAX–AKL</td>
<td>24.8</td>
<td>22.6</td>
<td>25.51</td>
<td>21.8</td>
<td>110</td>
<td>13</td>
<td>10</td>
<td>12.42</td>
</tr>
<tr>
<td>July 15</td>
<td>SYD–LAX</td>
<td>29.1</td>
<td>26.6</td>
<td>29.36</td>
<td>28.5</td>
<td>100</td>
<td>15</td>
<td>10</td>
<td>14.8</td>
</tr>
<tr>
<td>July 17</td>
<td>LAX–SYD</td>
<td>29.4</td>
<td>27.5</td>
<td>27.41</td>
<td>24</td>
<td>140</td>
<td>13</td>
<td>10</td>
<td>15.12</td>
</tr>
<tr>
<td>July 24</td>
<td>PER–SIN</td>
<td>11.2</td>
<td>10.1</td>
<td>12.4</td>
<td>10.7</td>
<td>70</td>
<td>5</td>
<td>0</td>
<td>5.57</td>
</tr>
<tr>
<td>July 25</td>
<td>SIN–LHR</td>
<td>35.2</td>
<td>32.1</td>
<td>29.36</td>
<td>30</td>
<td>250</td>
<td>13</td>
<td>10</td>
<td>15.32</td>
</tr>
<tr>
<td>July 29</td>
<td>LHR–BKK</td>
<td>25.93</td>
<td>30.5</td>
<td>26.76</td>
<td>24.8</td>
<td>50</td>
<td>12</td>
<td>10</td>
<td>14.28</td>
</tr>
<tr>
<td>July 31</td>
<td>BKK–SYD</td>
<td>14.1</td>
<td>15.5</td>
<td>18.06</td>
<td>13.7</td>
<td>50</td>
<td>8</td>
<td>10</td>
<td>8.53</td>
</tr>
<tr>
<td>August 8</td>
<td>SYD–JBG</td>
<td>60.5</td>
<td>67.7</td>
<td>60.05</td>
<td>N/A</td>
<td>380</td>
<td>23</td>
<td>20</td>
<td>27.64</td>
</tr>
<tr>
<td>August 12</td>
<td>JBG–SYD</td>
<td>51.3</td>
<td>54.6</td>
<td>58.77</td>
<td>N/A</td>
<td>250</td>
<td>19</td>
<td>10</td>
<td>22.07</td>
</tr>
<tr>
<td>August 18</td>
<td>SYD–LAX</td>
<td>27.9</td>
<td>26.3</td>
<td>28.39</td>
<td>N/A</td>
<td>240</td>
<td>14</td>
<td>10</td>
<td>14.67</td>
</tr>
<tr>
<td>August 20</td>
<td>LAX–SYD</td>
<td>29.3</td>
<td>28.2</td>
<td>29.36</td>
<td>N/A</td>
<td>100</td>
<td>14</td>
<td>20</td>
<td>15.64</td>
</tr>
<tr>
<td>August 26</td>
<td>SYD–HKG</td>
<td>16.9</td>
<td>15.7</td>
<td>16.32</td>
<td>N/A</td>
<td>60</td>
<td>9</td>
<td>0</td>
<td>9.01</td>
</tr>
</tbody>
</table>

*All FH41B results have been corrected via the function described above.
†The Dis-100(γ) result was 9.
‡The Dis-100(γ) result was 15.
Fig. 2. Compact dosimeter results for Toronto to Singapore flight to AOGS 2005.

values are within the expected experimental and calculation errors of 30%. However, when the neutron component results of two of the dosimeters are compared to bubble detector and EPCARD neutron results, it can be seen that these personal dosimeters indicate values that are much too high.

4. Conclusions

From the evaluation of available compact spectrometers and dosimeters, it has been determined that the Liulin spectrometers provide accurate H*(10) results, within expected experimental error, but entail extensive computations. The FH41B and B-10 dosimeters require position-sensitive calculations, but appear to provide correct results for the low-LET
component only. The various personal dosimeters also gave correct numerical values for the low-LET component, but were unable to measure the high-LET radiation. A need still exists for an accurate, compact, and user-friendly instrument that provides a readout of $H^*(10)$ data during a flight to support the use of a predictive code for career assessment of aircrew exposure as well as detecting solar flares.

Acknowledgments

The support of the Air Canada, Qantas Airways, and Singapore Airlines in allowing the measurements to be made onboard for this study is gratefully acknowledged. The financial support of Transport Canada and the Canadian Air Force is acknowledged.

References

MODELING OF AIRCREW RADIATION EXPOSURE FROM GALACTIC COSMIC RAYS AND SOLAR PARTICLE EVENTS

M. TAKADA
National Institute of Radiological Sciences
4-9-1, Anagawa, Inage-Ku, Chiba 263-8555, Japan

B. J. LEWIS, M. BOUDREAU, H. AL ANID and L. G. I. BENNETT
Department of Chemistry and Chemical Engineering
Royal Military College of Canada
P. O. Box 17000, Stn. Forces, Kingston
Ontario K7K 7B4, Canada

The predictive code for aircrew radiation exposure (PCAIRE) was based on empirical correlations, which were developed from measurement flights during solar cycle 23, for the prediction of the ambient dose equivalent rates. To extend to the extremum conditions of solar modulation and altitude, bounding correlations have been further developed with the LUN transport code and incorporated into the model. For interpolation between the bounding solar-cycle conditions, the new NASA solar modulation model has been used. The conversion ratio of effective dose to ambient dose equivalent, applied to the (measured) PCAIRE estimate for the legal regulation of aircrew exposure, was re-evaluated in this work to take into consideration the new ICRP-92 radiation weighting factors and different possible irradiation geometries of the source cosmic-radiation field. A computational analysis with MCNPX was used to estimate additional aircrew exposure that may result from sporadic solar particle events, considering the geostationary operational environmental satellite data. These predictions were compared to the ambient dose equivalent rates measured with a TEPC onboard an aircraft prior to and during the event, and were further compared to count rate data observed at various neutron monitors on the ground.

1. Introduction
As a result of the recommendations of the ICRP-60 and, in anticipation of possible regulation on exposure of Canadian-based aircrew, an extensive study was carried out to measure galactic cosmic radiation at altitudes over solar cycle 23. Empirical correlations derived from this study were
encapsulated into the predictive code for aircrew radiation exposure (PCAIRE) for the prediction of ambient dose equivalent rates. In order to cover extremum conditions of solar modulation and altitude, theoretical models using the LUIN transport code were developed for incorporation into PCAIRE. Additional updates were evaluated for use in PCAIRE, specifically the revised deceleration parameter model proposed by NASA, the new weighting factors proposed by the ICRP-92, and different possible irradiation geometries.

In addition to the expected galactic cosmic rays (GCR), there exists sporadic solar energetic particles (SEP) resulting from rare solar flares. These highly-variable events occur more frequently during the active phase of the solar cycle and are more challenging to model than the stable GCR component. The MCNPX code was used with the geostationary operational environmental satellite (GOES) data to predict dose rates at altitude and on the ground during solar flares. Once this method was developed, it was compared to neutron monitor and flight measurement data.

2. Model Development

2.1. Ambient dose rate as a function of altitude and solar modulation

The original PCAIRE correlations for the ambient dose equivalent rate were based on experimental data obtained from 1998 to 2003 covering near solar minimum and maximum conditions, which provided two curves $f_1$ and $f_2$, shown respectively, in Fig. 1. They were derived by summing separate measurements obtained with an extended-range neutron rem meter and an ionization chamber.\(^1\) A linear interpolation for intermediate conditions of solar modulation was used;\(^1\) however, there was no extrapolation beyond these (experimental) bounding curves. In order to extend the experimentally-derived model to conditions of low altitude and extremum conditions of solar modulation where measurements proved difficult, ambient dose equivalent rates have been calculated with the LUIN 2000 transport code (Fig. 1).

As shown, since the LUIN analysis does not deviate by more than 10% (which is comparable to the experimental error) at a low heliocentric potential of $U = 390$ MV, the $f_1$ curve is adopted as an extremum curve ($f_{1\text{max}}$) for solar minimum conditions. In contrast, a new curve, $f_{2\text{min}}$, has been utilized to account for extremum conditions during solar maximum
Modeling of Aircrew Radiation Exposure

Fig. 1. Comparison of the ambient dose equivalent rate $\dot{H}_0 (U, \phi$ or $C)$ measured and predicted at different values of solar modulation as a function of the cut-off rigidity.

at $U = 2000$ MV. The dose rate is then linearly interpolated between the given experimental and theoretical curves shown in Fig. 1 employing a solar modulation model, which is characterized by either a heliocentric potential $U$ (in MV), deceleration parameter $\phi$ (in MV), or scaled Climax neutron count rate $C$ (in counts/h/100).

Thus, the normalized ambient dose equivalent rate can be extended to any altitude (up to 20 km) using the correction factor $f_{\text{Alt}}$ (given in Eq. (5) of Ref. 1) using Eq. (1):

$$\dot{H}(R_c, h; U, \phi, C) = \dot{H}_0(U, \phi, C) \cdot f_{\text{Alt}} \cdot f_{\text{Alt}}.$$ (1)

In this way, the effect of solar modulation on the dose rate is based on experimental data where possible, but bounded by theoretical analysis where such data are not yet available. Hence, no extrapolation is now required in the model. However, Eq. (1) cannot be extrapolated to lower altitudes since this relation does not reflect the lengthening of the relaxation length due to the muon component from secondary particle production that dominates the radiation field closer to the Earth’s surface.\(^2\) For a more accurate calculation, $f_{\text{Alt}}$ can be replaced by $f_{\text{Low,Alt}}$ that follows
Fig. 2. Altitude correction function with comparison to LUIN analysis. This function is normalized to unity at 243 g/cm².

From the LUIN theoretical calculations in Fig. 2 for atmospheric depths \( h > 400 \text{ g/cm}^2 \) (\( A < 7.6 \text{ km} \)).

In Fig. 2, the functions of the atmospheric depth (in g/cm²), \( f_{\text{LN}_1} \) and \( f_{\text{LN}_2} \) correspond to a fitted curve of the data at cut-off rigidity values of 7.41 and 0.78 GV, respectively. The functions in Fig. 2 are normalized to unity at 243 g/cm² (10.67 km) and are also plotted with the \( f_{\text{Alt}} \) function for comparison. The \( f_{\text{Alt}} \) equation, currently used in PCAIRE, is in good agreement with the functions \( f_{\text{LN}_1} \) and \( f_{\text{LN}_2} \) for high altitudes (i.e., \( h < 400 \text{ g/cm}^2 \)) and therefore remains in use. As they start to deviate from \( f_{\text{Alt}} \) at \( h > 400 \text{ g/cm}^2 \), the function \( f_{\text{Low,Alt}} \) is applied. The revised low-altitude model therefore provides for an improved ability to calculate route doses of short-haul flights (<2 h of flight time), which typically occur at low altitudes.

### 2.2. Revised solar modulation model

The effect of the solar cycle has been modeled in PCAIRE by correlating the experimental (\( f_1 \) and \( f_2 \)) and theoretical (\( f_{2\text{min}} \)) dose-rate curves to a given value of solar modulation (\( U, \phi, \text{or } C \)). The solar modulation models are also
used in various other aircrew-exposure codes such as CARI and EPCARD to account for the solar-cycle effect.\textsuperscript{3,4} The O’Brien model employed in CARI,\textsuperscript{5} characterized by a heliocentric potential $U$ (in MV), is tabulated by the FAA from daily ground-level counting rates at the Apatity cosmic-ray neutron monitor.\textsuperscript{3} Alternatively, EPCARD is derived from FLUKA Monte Carlo code calculations that employ the primary spectra of Badhwar, where the solar modulation of these spectra is determined via a diffusion-convection model developed by NASA – Johnson Space Centre.\textsuperscript{6} This model implicitly assumes that there is an $\sim$ three-month time lag associated with the time required for the solar wind to carry the solar magnetic field lines out to the solar modulation boundary located at about 100 AU and further depends on the solar magnetic polarity that changes roughly every 11 years.

It has been recently suggested following an analysis of neutron monitoring data that both high-energy and low-energy GCR arrive at Earth’s orbit with small time differences compared to a month.\textsuperscript{7} This study therefore contradicts the original deceleration potential model of NASA which was based on a correlation of the Climax neutron count rate with IMP-8 and ACE satellite data. This three-month delay is not expected since the quasi-steady Fokker–Planck formulation for diffusion, convection, and adiabatic deceleration of particles in the interplanetary magnetic field should account for both the high-energy fluxes for Climax as well as the low-energy fluxes for IMP and ACE. As such, a new correlation has been proposed by O’Neill based on ACE energy spectra data:\textsuperscript{8}

\[
\phi(t) = \begin{cases} 
-1.15674C + 5434.5 & \text{for positive cycles,} \\
-0.9276C + 4534.2 & \text{for negative cycles,} \\
-1.8887C + 8253.75 & \text{for high modulation (}C<3850\text{ count/h/100).}
\end{cases}
\]

Here $\phi$ is the deceleration parameter (in MV) and $C$ is the scaled Climax neutron count rate (in count/h/100). Since the $f_1$ and $f_2$ curves in Fig. 1 are solely empirical, they can be further correlated directly to the Climax neutron monitor count rate (as well as to the heliocentric potential and deceleration parameter).

### 2.3. Effective to ambient dose rate ratios

For radiation protection purposes, a ratio to convert the ambient dose equivalent ($H$) into an effective dose ($E$), $f_{E/H^*(10)}$ is required. The
effective dose rate is finally evaluated as:

\[ \dot{E}(R_c, h; U, \phi, C) = \dot{H} \cdot f_{E/H}^{E/H^*(10)}, \]  

where \( \dot{H} \) is obtained from Eq. (1) with a given choice of solar modulation model.

Fluence-to-effective dose and fluence-to-ambient dose equivalent conversion coefficients have been derived by Pelliccioni\(^9\) using both the recent recommendations of the ICRP-92 for the radiation weighting factors, as well as those based on ICRP-60. These conversion coefficients have been used with particle spectra estimated with the LUIN 2000 transport code to calculate \( E/H^*(10) \) ratios. The calculations have covered ranges of altitudes, geographic positions, and solar modulation values, and are shown in Fig. 3 for an isotropic geometry with ICRP-92 recommendations.

The resulting \( f_{E/H}^1, f_{E/H}^2, \) and \( f_{E/H}^3 \) functions of the pressure altitude \( A \) (km) are taken from averages at 0.7, 12, and 17.6 GV, respectively. Linear interpolations between these functions, which were calculated for an isotropic geometry (ICRP-60 and 92), give the overall function, \( f_{E/H}^{E/H^*(10)} \), shown in Fig. 3. The ICRP-92 coefficients were also evaluated for a hemispherical geometry (not shown). Equation (3) therefore provides a function that relates the effective dose rate to the ambient dose equivalent rate.

![Fig. 3. \( E/H^*(10) \) Ratio as a function of atmospheric depth for various cut-off rigidities and heliocentric potentials (for an isotropic geometry and ICRP-92 recommendation).](image)
prediction of the effective dose rate at any geographic position (or cut-off rigidity), altitude (up to $\sim 20$ km), and period in the solar cycle. This quantity can then be suitably integrated over a given flight route for a route dose prediction using either a great circle route or waypoints for the flight.$^2$

3. Aircrew Radiation Exposure During Solar Particle Events

A computational methodology to enable an estimate of additional aircrew exposure that may result from SEP events follows. In this work, primary source spectra from both GCR and SEPs have been propagated through the atmosphere using the MCNPX code and tested against measured data.

3.1. GCR and SEP primary source spectra

For the GCR analysis, the local interstellar spectrum models for the cosmic ray protons and helium$^{10}$ have been used in the LUIN 2000 code and attenuated to the top of the atmosphere. These models show excellent agreement when compared with data obtained from balloon experiments and the space shuttle using various spectrometers.$^{10}$ Only the lower-energy particles are affected by solar modulation.

Since the SEP spectra are highly variable, real-time satellite measurements have been used for source input data. The space environment monitor instruments on the NOAA GOES provide measurements of solar and galactic energetic particles, solar X-rays, energetic particles in the Earth’s magnetosphere and geomagnetic-field variations. An energetic particle sensor and high-energy proton and alpha detector monitor the incident flux density of protons, alpha particles, and electrons over a large range of energy.$^{11}$ The five-min proton and helium flux data are distributed at the NOAA web site: http://spidr2.ngdc.noaa.gov/spidr/. For analysis of SEP exposure, the particle fluxes must be extrapolated to a high energy of 10 GeV. This extrapolation is done via a fitting of the GOES data to a power-law expression for the differential flux as a function of the particle energy $E$ as shown in Fig. 4.$^{12}$

3.2. MCNPX

The MCNPX code (version 2.5) was used to determine the particle production and transport in the atmosphere. Although secondary particles
are produced by interaction of primary cosmic ray particles with atmospheric nuclei, only the production of neutrons and protons were considered. The atmosphere was divided into 36 concentric shells using an average air density for a given shell thickness. Secondary particle energy spectra produced from an incident mono-energetic source particle was tracked in the analysis. A combined particle spectra (at a given altitude) was therefore obtained by summing the secondary particle spectra derived from each mono-energetic primary particle based on the initial proton spectrum and helium spectrum. The geometry used was spherical \((4\pi)\) for GCR and planar for SEP events, although the latter simplification is worthy of further study.

As a preliminary test, the interstellar GCR spectrum was used to predict neutron and proton spectrum on the ground and at 17 km. These results were compared to those measured by Goldhagen and Gordon and were determined to be in reasonable agreement. As a next step, a GCR and an SEP spectrum from the GOES-11 satellite was used to predict the ambient dose rates at altitude before and during GLE-65. For the latter, the GCR spectrum was first subtracted from the GOES data and PCAIRE calculations of the GCR dose component were added to the resulting the ambient doses from the solar flare. These results were compared to ambient doses measured with a TEPC at altitude on a flight during this event (Fig. 5).

\[
\phi(E)dE = \frac{C}{R} \left( \frac{R}{R_c} \right)^{\gamma} dE
\]
As shown, predicted doses remain within 20% of the measured doses. In order to account for rigidity effects, $R_c$ values were estimated for the associated SPE. Only those particles with energy greater than the corresponding energy for these $R_c$ values were summed. Since the model was found to be extremely sensitive to $R_c$, detailed information of the flight path and storm effects were considered. This approach was further tested against neutron monitor data on the ground.

This model has been generated in a manner that would allow for near real-time calculation of solar flare dose rates in the event of an SEP. Incorporation of the models into a computer code, and eventually PCAIRE, is now underway.

4. Conclusions
Correlations developed with the LUIN transport code are in agreement with experimentally-derived models and are further used to provide extensions of the models to bounding conditions of solar modulation and altitude. The deceleration parameter used in PCAIRE, to account for effects of solar modulation, has been updated to include the more recent NASA model for consistency between all three solar modulation models. In addition, calculation of new $E/H^*(10)$ ratios, considering the revised ICRP-92 radiation weighting factors, shows that these ratios are mostly affected
by altitude and cut-off rigidity. Since they are close to unity, the ambient
dose is therefore now a reasonable estimate of effective dose. Solar flare
models currently under development have shown promising agreement with
available flight and neutron monitor data, encouraging continuation with
its development.

Acknowledgments
The financial support of Transport Canada, the National Science and
Engineering Research Council and the Canadian Air Force is acknowledged.

References
1. B. J. Lewis, M. Desormeaux, A. R. Green, L. G. I. Bennett, A. Butler, M.
293.
M. A. Shea and D. F. Smart, CARI-6M, Civil Aerospace Medical Institute,
Federal Aviation Administration, Oklahoma City, OK, USA (2001).
4. H. Schraube, G. Leuthold, W. Heinrich, S. Roesler, V. Mares and
G. Schraube, EPCARD (European Program Package for the Calculation of
Aviation Route Doses), GSF National Research Centre for Environment and
Health (2002).
System, Federal Aviation Administration Report DOT/FAA/AM-05/14, July
2005.
12. International Organization for Standardization, Probabilistic Model for
Particle Fluences and Peak Fluxes of Solar Energetic Particles, ISO TS 15391,