S C I E N C E & T E C H N O L O G Y P O L I C Y I N S T I T U T E

Next Step Space Weather Benchmarks

December 2019

Approved for public release;
distribution is unlimited.

IDA Group Report
NS GR-10982

Log: H 19-000615

IDA SCIENCE & TECHNOLOGY
POLICY INSTITUTE
1701 Pennsylvania Ave., Suite 500
Washington, DC 20006-3602

About This Publication

This work was conducted by the IDA Science and Technology Policy Institute under under contract NSFOIA-0408601, Project NF-20-4516, “SWORM Goal 1 Next Step Benchmarking.” The views, opinions, and findings should not be construed as representing the official positions of the National Science Foundation or the sponsoring office.

Acknowledgments

The contributors to the Next Step Space Weather Benchmarks Report wish to express their gratitude for the support from the National Science Foundation (NSF) and the National Aeronautics and Space Administration (NASA) in sponsoring the effort.

For More Information

Thomas J. Colvin, Project Leader
tcolvin@ida.org, 202-419-5461

Mark S. Taylor, Acting Director, Science and Technology Policy Institute
mtaylor@ida.org, 202-419-5491

Copyright Notice

SCIENCE & TECHNOLOGY POLICY IN STITUTE

IDA Group Report NS GR-10982

Next Step Space Weather Benchmarks

Geoffrey Reeves, Chair
Thomas J. Colvin, Executive Secretary
Jericho Locke, Executive Secretary

Induced Geo-Electric Fields
Pete Riley, Leader
Jeff Love
Antti Pulkkinen
Adam Schultz
Emanuel Bernabeu
Alan Thomson

Ionizing Radiation
Christina Cohen, Leader
Joe Giacalone
Therese Moretto Jorgensen
Juan Rodriguez
Tim Guild
Delores Knipp

Ionospheric Disturbances
Susan Skone, Leader
Anthea Coster
Keith Groves
Jonathan Makela
Ethan Miller
Roger Varney

Solar Radio Bursts
Dale Gary, Leader
Tim Bastian
Gregory Fleishman
Stephen White
Angelos Vourlidas
Jade Morton
Jasmina Magdalenic

Upper Atmosphere Expansion
David Jackson, Leader
Sean Bruinsma
Yue Deng
Eric Sutton
Tzu-Wei Fang
John Emmert

Executive Summary

In June 2018, the Space Weather Operations, Research, and Mitigation (SWORM) Subcommittee of the National Science and Technology Council (NSTC) released the *Space Weather Phase 1 Benchmarks*. The benchmarks specify the nature and intensity of extreme space-weather events and provide a point of reference from which to improve understanding of their effects. The purpose of developing benchmarks is to inform engineering standards, vulnerability assessments, risk estimates, decision points and thresholds for action, more effective mitigation procedures and practices, and response and recovery planning.

The refinement of the Space Weather Phase 1 Benchmarks is an important step toward improving national resilience to the effects of space weather—a point that is reinforced in the NSTC's new *National Space Weather Strategy and Action Plan*, released in March 2019. The strategy calls for a Phase 2 benchmarking effort that will produce more rigorous benchmarks than were possible during Phase 1. Refined space weather benchmarks may also play a role in the implementation of Executive Order 13865, *Coordinating National Resilience to Electromagnetic Pulses*, issued by the White House in March 2019, as the physics and potential threat of extreme space weather share commonalities with human-caused electromagnetic pulses.

As a first step toward refining the Phase 1 benchmarks, the National Science Foundation (NSF) and the National Aeronautics and Space Administration (NASA) asked the IDA Science and Technology Policy Institute (STPI) to facilitate the Next Step Space Weather Benchmarks effort. The purpose of the effort was to engage the expertise of the U.S. and international space weather scientific community to make recommendations that would improve the Phase 1 benchmarks, including identifying any outstanding gaps. This report, written by an expert panel of 32 international space weather scientists, is the culmination of the Next Step Space Weather Benchmarks effort. Inputs to the analysis include: white papers on space weather benchmarks solicited through a request for community input; a workshop that convened the expert panel and broader members of the space weather community to identify gaps and recommendations; and a subsequent Town Hall to receive community input on the panel's draft conclusions.

The expert panel is organized into five working groups, one for each space weather phenomenon. Each working group was asked to assess the Phase 1 benchmarks according to the following three criteria. First, are the chosen *quantities* the best way to measure extreme events associated with each space weather phenomenon? For example, geomagnetic disturbances may be measured by nanotesla per minute, volts per kilometer, etc., and the quantity chosen affects the usability, accuracy, and extensibility of the benchmarks. Second, are the calculated *values* for each quantity reasonable? Third, is the methodology used to calculate the values sound? Based primarily on the answers to these three questions, each working group makes recommendations to the Phase 2 benchmark effort and provides near- and long-term research recommendations that will improve the ability to understand and benchmark extreme space weather events. Selected elements of each working group's analysis and recommendations are summarized below.

Space weather can produce *induced geo-electric fields* on the surface of Earth that interfere with the operation of high-voltage power transmission systems and other critical infrastructure. Extreme electric fields are highly variable across the country so Phase 1 provided 100-year benchmark values as a geographic map; however, benchmark values were not computed for the entire map due to a lack of magnetotelluric (MT) data. The panel noted several other gaps. First, information is needed on the time evolution of the electromagnetic field (i.e., a waveform), which is required for system operators and engineers to increase system resilience. Second, observational data has insufficient spatial, temporal, and frequency resolution to obtain maximum values. Third, theoretical maximum values were not estimated.

The highest priority recommendation of the working group for induced geo-electric fields is the completion of the MT survey for the continental United States, so that the 100-year benchmarks can be completed. Further, the panel proposes follow-on observational activities that will address data resolution issues. They also recommend modeling activities that could lead to benchmark waveforms and theoretical maximum values.

Ionizing radiation is hazardous for satellites, airline communication, airline crews, and astronauts. These hazards can be divided into three major types of radiation: solar energetic particles (SEPs), galactic cosmic rays (GCRs), and radiation belt particles. The Phase 1 benchmarks, which were accompanied by a supplemental report that expands upon the Phase 1 work, extensively cover these hazards; however, gaps remain in important areas: all ionizing radiation areas—SEP, GCR, and radiation belt particles—are missing benchmarks in important energy ranges; some types of radiation are missing benchmarks for important particle species; and some of the models used to generate the benchmarks are not validated for extreme space weather. Generally, the benchmarks require more data to improve their energy range coverage and reduce uncertainties.

To address gaps in the species and energy ranges for which data is available, the working group for ionizing radiation recommends new space-based observations provided by instruments at Earth-Sun Lagrange point 1 that measure SEPs and instruments in Earth orbit to measure SEPs and GCRs. They also recommend that the government facilitate the public release and analysis of existing data. For instance, the Alpha Magnetic Spectrometer, which is installed on the International Space Station, has data critical to SEP benchmarks that should be made available for analysis.

Ionospheric disturbances can affect the propagation of communication, navigation, and timing signals. These disturbances are driven primarily by three types of space weather phenomena: solar flares, solar energetic particle events, and geomagnetic storms. The Phase 1 benchmarks did not estimate values for either the 100-year or theoretical maximum events for any of the three disturbance drivers, instead providing qualitative descriptions of extreme events and a few extreme values based on the 2003 Halloween storm. The working group noted the following gaps: several of the chosen quantities for ionospheric benchmarks should be revised; the Phase 1 analysis made insufficient use of published literature and public databases; and no methodologies were proposed for calculating benchmark values.

The working group for ionospheric disturbances recommends new benchmark quantities to better characterize ionospheric disturbances, suggest methodologies for estimating these quantities, and provide rough calculations of benchmark values where possible. The panel finds that available data and models are sufficient to produce initial estimates of all benchmark values; their recommendations focus on studies that will improve and validate the proposed methodologies to produce benchmark values. Until the limits of the currently available data and models are assessed through benchmarking, it would be premature to recommend large new observational or research efforts.

Potentially hazardous *solar radio bursts* are emitted routinely—on average, every 3.5 days during solar maximum and every 18.5 days at solar minimum—and can disrupt the radio spectrum on which the United States relies extensively. One of the most important systems that may be disrupted is the global positioning system (GPS). The Phase 1 methodology did not make a distinction between incoherent and coherent emissions, which is noteworthy because they occur with different intensities and only coherent emissions are likely to interfere with communication and navigation systems. Many of the gaps impeding higher fidelity benchmarks are due to limitations in observational data, including, for example, lack of data on the polarization of the bursts, important because GPS is only affected by right-hand circularly polarized emissions and coherent emissions tend to be strongly circularly polarized; monitored frequencies are too limited to reliably measure extreme flux events; and current solar observation instruments lack the resolution to resolve the source area and hence determine the intrinsic brightness of emissions, which impedes the ability to understand the causes of radio bursts and makes any theoretical maximum estimate highly uncertain.

The working group for solar radio bursts recommends two new observational systems that will address the previously mentioned gaps. The first system will produce solar radio data with continuous coverage of the Sun, with the necessary frequency coverage, frequency range, polarization, and dynamic range to estimate high fidelity benchmarks. The second system will provide the measurements necessary to understand the cause and evolution of extreme radio bursts, which in turn will enable the creation of a credible theoretical maximum benchmark. The working group also proposes a methodology for benchmarking coherent emissions based on currently available data, which tend to lack polarization data. They provide an estimated benchmark value and recommend future research to improve and validate the benchmark value.

Upper atmosphere expansion benchmarks characterize the ability of extreme space weather to increase neutral mass density in the upper atmosphere. Increased neutral mass density enhances drag on satellites and space debris. This enhanced drag can interfere with the ability to monitor debris trajectories and reduce the operational lifetime of a satellite. The Phase 1 report provided benchmarks for three drivers of neutral density change: solar extreme ultraviolet (EUV) and far ultraviolet (FUV) radiation; enhancement of an extreme solar flare by moderate intensities of solar EUV; and geomagnetic disturbances (GMDs). The working group for upper atmosphere expansion identifies gaps related to physics-based models of the thermosphere, including the use of inputs (e.g., Kp, F10.7) that are poor proxies for the true drivers of atmospheric expansion and deficiencies in the observational data required to improve the models. Such modeling and observational gaps present a challenge for estimation of benchmark values. They also note that the Phase 1 benchmarks may be missing important contributors to satellite drag (e.g., ionospheric plasma) and operational lifetime (e.g., atomic oxygen density).

This working group recommends a new approach to EUV measurements that ensures all future EUV observational missions have sufficient overlap to enable cross-calibration of their data. Future instruments should also be designed to include at least the three spectral bands most relevant to thermospheric modeling and be designed for resilience to degradation by incorporating in-flight calibration capabilities. The working group makes multiple recommendations to update physics-based models so that magnetospheric responses to extreme events and the magnetosphere's interaction with the thermosphere can be used to improve 100-year benchmark values and reduce their uncertainty. Improved understanding and models are also necessary for creation of credible theoretical maximum benchmarks. Finally, the working group also recommends that the effects of ions and atomic oxygen be investigated for potential inclusion in future benchmarks.

There are also issues and recommendations that cut across more than one of the five benchmarked phenomena and should be addressed as part of the Phase 2 effort. Neither Phase 1 nor the Next Step Benchmarks effort provides benchmarks for the duration of an extreme event. A benchmark for event duration is critical for improving national resilience to the effects of space weather and should be prioritized. Across all benchmark areas, the calculation of extreme values and their credible uncertainties were also inhibited by limited data and limited physical understanding of the phenomena. We, the full panel, recommend the development of a dedicated data collection plan applicable to the benchmarks, releasing and merging data sets needed for improved benchmarks, and establishing a regular cadence for updating benchmarks as new data and updated models emerge. Further, we recommend calculating benchmark values for more frequent events than just 1-in-100 year events to provide estimates with operationally useful statistical uncertainties.

Enabling our cross-cutting and phenomenon-specific recommendations may require science funding agencies to increase the priority of applied research relative to basic physical understanding. Data cleaning, cross-calibration of heterogeneous data, the application of statistical analysis techniques, and publicly releasing data sets for community analysis are critically important activities for benchmarks, but have been difficult to fund under traditional research programs. Funding agencies may also need to broaden the scope of model development considered for funding. Traditional model development has focused on increased fidelity and forecast accuracy of relatively common space weather events, while models developed specifically to capture extreme distributions or ensemble modeling of extreme conditions are uncommon, though immensely valuable. We recommend that funding agencies support these types of applied data analysis and modeling efforts to improve space weather benchmarks.

Contents

1. Introduction

A. Background

In June 2018, the Space Weather Operations, Research, and Mitigation (SWORM) Subcommittee of the National Science and Technology Council (NSTC) released the *Space Weather Phase 1 Benchmarks*. The benchmarks specify the nature and intensity of extreme space-weather events and provide a point of reference from which to improve understanding of their effects. They characterize the space weather environment and are intentionally technology agnostic. The purpose of developing benchmarks is to inform engineering standards, vulnerability assessments, risk estimates, decision points and thresholds for action, more effective mitigation procedures and practices, and response and recovery planning.

Benchmarks characterize extreme space weather environmental parameters and are technology agnostic

Space weather benchmarks are being developed in a phased approach. The Phase 1 benchmarks effort, called for by the 2015 *National Space Weather Strategy* and the *National Space Weather Action Plan*, was intended to be an initial, quick-turnaround analysis based on existing data sets and studies, conducted by teams of subject matter experts across nine Federal departments and agencies. The *Space Weather Phase 1 Benchmarks* document defines benchmarks for five phenomena associated with space weather events: induced geo-electric fields, ionizing radiation, ionospheric disturbances, solar radio bursts, and upper atmospheric expansion. For Phase 2, SWORM will more rigorously analyze and refine the benchmarks to improve their precision and utility. Refining the space weather benchmarks contributes to improving national resilience to the effects of space weather—a priority reinforced in the NSTC's new *National Space Weather Strategy and Action Plan*, released in March 2019.

Phase 1 benchmarks were the result of a quick-turn analysis

Phase 2 benchmarks will be more rigorous

The physics and potential threat of extreme space weather share some commonalities with a human-caused electromagnetic pulse (EMP). Specifically, efforts taken to improve grid resilience with respect to extreme induced geo-electric fields in turn improve grid resilience relative to the effects of the slow-rise-time portion of an EMP blast. Similarly, efforts to improve infrastructure resilience with respect to the effects of benchmarked ionospheric disturbances or ionizing radiation may also overlap with the ability to be resilient to the

Extreme space weather has some overlap with EMP effects of an EMP. In March 2019, the White House issued Executive Order 13865, *Coordinating National Resilience to Electromagnetic Pulses*, which specifically calls out the importance of geomagnetic disturbances—effectively induced geo-electric fields—to improved resilience from the effects of EMP. The Next Step Benchmarks and eventual Phase 2 Benchmarks will be valuable inputs informing both space weather and EMP resilience efforts.

Refined benchmarks may benefit implementation of space weather and EMP policies

B. Purpose

This report represents the next step in an ongoing effort to prepare the Nation for the consequences of an extreme space weather event. To inform the refinement of the Phase 1 Benchmarks, the National Science Foundation (NSF) and the National Aeronautics and Space Administration (NASA) have sponsored the Next Step Benchmarks (NSB) task. The goal of the task was to gather input from the space weather research and operations/user communities on how to refine and improve the Phase 1 Benchmarks. This includes: identifying new research or data sets that may be used to improve the benchmark values; identifying gaps in existing data sets and methodologies that hinder the ability to produce high-confidence benchmark values; and suggesting future research activities that may improve the accuracy of and confidence in the benchmark values by closing the gaps. The NSB task is not producing the Phase 2 benchmarks, nor does it serve as an interim update to the Phase 1 values; instead, the community identifies gaps and recommendations produced by the NSB task identify near- and long-term research priorities and will be a valuable input for the Phase 2 Benchmark team.

Goal of this task is to inform the refinement of the Phase 1 Benchmarks

This task is not Phase 2

C. Approach

To accomplish this goal, STPI helped to assemble 32 of the world's leading space weather scientists. Geoffrey Reeves was chosen to chair the entire effort. For each of the five phenomena, a prominent scientist was chosen to lead a working group on that topic. The leaders of the working groups then choose their team members to balance the expertise needed for a thorough analysis of the Phase 1 document.

This report was produced by a panel of 32 top space weather scientists from around the world

The panel first sought feedback on the Phase 1 benchmark report from the broader space weather scientific and user communities. A request for community input was published online and broadly advertised through the major electronic newsletters used by the space weather community, such as AGU SPA, GEM, CEDAR, SHINE, etc.

The broader community submitted white papers on improving benchmarks

In April 2019, the IDA Science and Technology Policy Institute (STPI) organized a 3-day workshop in Denver, Colorado to solicit further input. This

The panel engaged with the broader community at a workshop in April, 2019 workshop was broadly advertised, including on the National Oceanic and Atmospheric Administration (NOAA) Space Weather Prediction Center (SWPC) website, and participation was solicited from key stakeholder groups in both the research and user communities. At the workshop, plenary talks by several leaders of the Phase 1 benchmark effort familiarized the group with the previous work and outstanding gaps. Members of the space weather user community participated in a panel discussion, introducing the scientists to the concerns of the user community and generating ideas for how to make benchmarks more operationally useful. For the remainder of the first day, and the following two days of closed-door sessions, the working groups formed breakout sessions and began their analysis. To conclude the workshop, each working group presented its preliminary results and a plan for subsequent writing and research.

Analysis continued through the summer of 2019 with panel members working remotely but coordinating through e-mail and teleconferences. Working group leaders guided the analysis on each of the five space weather phenomena. The executive committee consisting of the working group leads, the chair, and executive secretaries met by telecon at least once per month to coordinate the overall activity.

A Town Hall meeting was held in Washington, DC in September 2019 to present a draft version of the panel's findings and to collect additional feedback from the Federal space weather community. A draft report of the panel's findings was distributed to town hall attendees and other space weather experts for review and comment. The responses from the request for community input, the workshop, and the town hall informed the panel's effort to improve the accuracy, rigor, and utility of the benchmarks, as reported in this document.

An initial draft of this report was distributed and discussed at a Town Hall in September 2019

Throughout these activities, each working group was asked to answer the five questions quoted below.

1. **Quantity Assessment: Are the Phase 1 benchmark quantities well-aligned with the objectives and use cases stated in the Phase 1 Document?**

   The objectives and use-cases of the benchmarks were described in Phase 1 as "to provide input for creating engineering standards, developing vulnerability assessments and risk estimates, establishing decision points and thresholds for action, understanding risk, developing more effective mitigation procedures and practices, and enhancing response and recovery planning." The panel assessed whether the benchmark quantities (e.g., the variables and their units) identified in Phase 1 were appropriate for representing space hazards posed by extreme space environments.

To make the assessment, the panel was asked to specifically assess the usability of the benchmarks. For instance, to enhance resilience of the bulk power system to the effects of a geomagnetic disturbance, grid operators can more readily use a benchmark for induced electric fields measured in volts/km than a benchmark for deviations in the Earth's magnetic field measured in nanotesla. Thus, induced geo-electric fields measured in volts/km would be well-aligned with the objectives of the Phase 1 Benchmark effort.

2. **Value Assessment: Are the benchmark values reasonable and up-to-date based on current understanding?**

   The panel assessed whether the values expressed by the benchmarks for 1-in-100 year events and theoretical maxima are likely to be the best estimates possible. For example, a theoretical maximum benchmark value would be assessed as not reasonable if higher values had been found in the literature or observed in databases.

3. **Methodology Assessment: Is the methodology used to derive the benchmark values up-to-date, rigorous, and compelling?**

   The panel assessed the methods used to calculate the Phase 1 values and suggest methodological improvements. An up-to-date, rigorous, and compelling methodology will increase the confidence in the benchmark values, and thus their value and usability. As a representative example, the panel determined that separating solar radio bursts by type (coherent versus incoherent emissions) would more accurately specify the distribution of the extreme events that have the greatest impact on vulnerable systems.

4. **Near-term Recommendations: What are recommendations and priorities for updates that could be done now or in the near term?**

   The panel provided recommendations for activities that could quickly address gaps identified in the previous three assessments. These near-term recommendations could be used to inform an interim update of the Phase 1 values or for the Phase 2 analysis; e.g., by leveraging new literature or data, merging multiple data sets, or expanding the use of models.

5. **Long-term Recommendations: What are recommendations for longer-term studies or research that would improve benchmark values, reduce their uncertainties, or improve their usability?**

The panel also identified opportunities for research that could have dramatic impacts on improving the benchmark values and advancing general understanding of extreme space weather events. These recommendations, such as new observational data sets or platforms, could be used to inform a research roadmap aimed at continuously improving both understanding and the capability to define and prepare for the effects of extreme space weather events.

D. Structure of Document

Chapters 2–6 individually address each of the five space weather phenomena that were defined by SWORM and benchmarked in Phase 1. The five questions outlined above appear as sections titles within each chapter. Chapter 7 concludes with gaps and recommendations common to multiple—if not all—benchmarked phenomena.

Throughout the report, the sidebar is used to call out important summary information. Gaps that are identified within each chapter are labeled as such using italics, followed by a brief description of the gap. Similarly, recommendations for Phase 2 will be identified in the sidebar, again with an italicized title and accompanying brief description. In the discussions of near- and long-term recommendations, the sidebar is used to clearly connect these recommendations back to one or more previously identified gaps. This will be done with italics to indicate which gap is being addressed, followed by a verbatim restatement of the gap.

Gap X
The report calls out gaps in the benchmarks that require additional research

Phase 2 recommendation
The report calls out recommendations to the Phase 2 team

Addresses Gap X
The report calls out gaps in the benchmarks that require additional research

E. Acknowledgments

The panel wishes to acknowledge and thank the individuals and organizations that contributed to the development and writing of this report. We thank Mike Wiltberger (NSF) for initiating and sponsoring this effort to engage the research community in the development of space weather benchmarks; Jim Spann (NASA) for sponsoring the Workshop in Colorado and the Town Hall in DC; Bill Murtagh and Jinni Meehan (NOAA) for supporting the workshop and providing the panel with the collaboration tools used to develop this document; Seth Jonas (STPI) for coordinating with NSF and NASA to help launch this effort; and Robin Dorsey, Lisa Wallace, and Zimika Stewart for their invaluable logistic support for the workshop and town hall. Finally, we thank all of the research and user community members who contributed to the next step benchmark effort, whether by responding to the request for community input, participating at the workshop and town hall, or by providing feedback on the draft final report. Your time and effort was critical to the results of this report.

2. Induced Geo-Electric Fields

Space weather can generate geomagnetic storms that induce geo-electric fields in the Earth's electrically conducting interior; these geo-electric fields can interfere with the operation of high-voltage power transmission systems and other critical infrastructure. Quantifying these fields requires that we evaluate both geomagnetic field variation realized during intense magnetic storms and the Earth's surface impedance (Pirjola 2002; Thomson et al. 2009). Surface impedance describes the electromagnetic response of the Earth to geomagnetic field variations driven by electric currents in the ionosphere and magnetosphere. The Phase 1 team used the relatively limited magnetotelluric (MT) survey data available at the time, in conjunction with a sinusoidal magnetic field statistically estimated from ground-based magnetic observatories, to establish a 1-in-100-year geo-electric benchmark (Love et al. 2016). The team found benchmark values to be sensitive to geography—at some places differing by more than two orders of magnitude across distances of a few hundred kilometers.

A. Quantity Assessment

The 1-in-100-year benchmarks are well-aligned with the objectives and use cases as stated in the Phase 1 document. The amplitude of the electric field is quantified in volts per kilometer, which are the units desired by operators of critical infrastructure, such as the bulk power system and natural gas pipelines. Additionally, the electric field is represented as a map that captures the geographic variability of extreme values.

The Phase 1 induced geo-electric field benchmarks did contain some gaps. Theoretical (or statistical) maximum values were not estimated. An additional gap is that extremes in duration of the induced field were not calculated; instead, the benchmark calculations assumed a 600-second duration for the geomagnetic disturbance. Finally, the Phase 1 benchmarks do not provide time series information—i.e., a reference waveform—for the amplitude, duration, and frequency of the induced geo-electric field. Operators of the electric power grid require a waveform to assess the thermal effects on power transformers (NERC 2017), reactive power loss, power line phase, and voltage and frequency stability. Lack of a waveform is a significant gap.

Gap A
Theoretical maximum values were not estimated

Gap B
Geomagnetic disturbance duration was not estimated

Gap C
No time-series information

B. Value Assessment

The 1-in-100-year benchmarks field amplitudes calculated by the Phase 1 team are still representative for individual MT impedance responses; however, the amplitudes were not estimated for much of the country due to limited MT data. At the time, the MT survey had only gathered impedance tensors for approximately half of the contiguous United States. At the time of this writing, more MT survey data is available that Phase 2 should use to expand the geographic extent of the benchmark values. Additional benchmark-related studies have also been performed, for example, using new MT data in the Northeast United States (Love et al. 2019; Lucas et al. 2018).

Phase 2 recommendation
Use existing MT survey data to expand the benchmark values

The largest gap impeding adding empirically-based values for geo-electric hazards across the continental United States was—and continues to be—the incomplete national-scale MT survey. By the time it ended in 2018, the NSF EarthScope program completed the survey on a nominally 70-km station grid across approximately two-thirds of the continental United States. NASA is currently sponsoring an extension of the MT survey to cover parts of the southwestern United States; but large areas remain to be surveyed under EarthScope-like protocols. Additionally, there is no impedance data in Alaska, Southern Canada, and Northern Mexico. Power grids and pipeline networks cross borders over much of North America, and thus the United States grid is susceptible to geomagnetically induced currents in bordering territory.

Gap D
Incomplete surveys of surface impedance for the continental United States

Gap E
No MT survey data for Alaska and U.S. border countries

C. Methodology Assessment

The Phase 1 benchmarks represented a significant advance over previous practice. The benchmarking efforts produced maps that show the geographic granularity of instantaneous 1-in-100-year geo-electric amplitudes over a specified window of time and at the highest frequency resolvable by the method and data used. However, the methodology suffered from multiple gaps.

The main gap is that the benchmarks do not fully capture the temporal or spatial complexity of the induced geo-electric field. The temporal-spatial variability in the vector geomagnetic field at ground level, within the frequency band of concern for geomagnetically induced currents (GIC) response due to space weather (approximately 10^{-4} Hz – 1 Hz), has great complexity. The complex geomagnetic field interacts with the MT ground impedance to induce a geo-electric field at ground level with significant temporal-spatial variability.

The available MT survey data may not sufficiently determine the spatial and frequency complexity of the induced geomagnetic fields. The current frequency bandwidth and 70km grid spacing does not capture the highly

Gap F
Insufficient spatial and frequency resolution of MT survey data

localized variations in Earth's impedance, potentially affecting benchmark values in regions of high complexity by 1–2 orders of magnitude. A related issue is that a single measured MT impedance reflects only local geological structures; thus, applying single-site impedance information to GIC calculations for over 100 km transmission line scales may not accurately represent the actual field imposed on the line. Consequently, while the benchmark quantifies the geo-electric field and empirical MT impedances from arrays of MT observation sites—which is arguably the best way to characterize the field—care must be taken in applying MT impedance-based geo-electric fields in GIC calculations to account for local effects.

Gap G
Insufficient spatial and frequency resolution of geomagnetic observational data

The other major gap was the sparse long-term geomagnetic monitoring data from across the continental United States and reported at an appropriate sample rate. Available geomagnetic observatory data used in the Phase 1 analysis have a one-minute sampling rate; the Phase 1 team believed that greater amplitudes of induced electric fields might be induced by higher frequencies of geomagnetic variation than the data can support due to its low sampling rate. For this reason, the limited sampling rate of the geomagnetic field used in Phase 1 limits the accuracy of maximum values for induced geo-electric fields. This limit, combined with the short timeline set by the Space Weather Action Plan, motivated the Phase 1 team to develop simple latitude-dependent mapping of geomagnetic activity, limiting the fidelity of the benchmark values.

Improved long-term geomagnetic monitoring is also needed across the conterminous United States. This would sensibly be concentrated between about 45° and 65° north magnetic latitude, a zone that experiences intense and spatially-complex storm-time disturbances. Improved monitoring might also be focused on areas where we now know that geo-electric hazards are high, such as in the Northeast and Northern Midwest United States, as well as other areas that may be identified in future as the MT survey is completed across North America. Given reports (personal communication) from Mexican colleagues of large GICs on the Mexican power grid, improved geomagnetic monitoring in Mexico is also desirable.

Gap H
Need more research to create benchmark waveforms

A final methodological gap is that statistical benchmark quantities (e.g., 100-year amplitudes) are not easily harmonized with the time-series qualities of storm-time geo-electric vector variation. Markovian statistical modeling of the signal and generation of waveform ensembles with the model is a possible route for further exploration that could be used to build models for the storm-time geo-electric field variations (Pulkkinen et al. 2006). One could consider extending the Markovian model parameters to extreme conditions and then simulating ensembles of representative waveforms. Characterization of the time variation necessitates improved (statistical and theoretical) understanding of the external electric current sources in the ionosphere and magnetosphere that drive the electromagnetic induction process in the ground.

D. Near-term Recommendations

Near-term 1: National scale MT survey

Addresses Gap D
Incomplete surveys of surface impedance for the continental United States

The highest priority for establishing meaningful geo-electric benchmarks across the continental United States is completing the national scale MT survey. Once this is completed, credible benchmark values can be computed for the Nation. Without this data, however, benchmark values cannot be reliably estimated for large portions of the country.

Near-term 2: Construct Earth conductivity models

Addresses Gap D
Incomplete surveys of surface impedance for the continental United States

In addition to direct application of empirical MT impedances, Earth conductivity models can be constructed (Kelbert et al. 2019) and then used to support spatially continuous estimation of surface impedance. We recommend analysis of the full observed vector magnetic and electric field waveforms, and—to the extent possible—using 1-sec and 10-sec data sets that capture sharp changes in the geomagnetic field associated with more intense induced geo-electric fields (Ngwira et al. 2015).

Near-term 3: Validate 70km MT survey data in GIC applications

Addresses Gap F
Insufficient spatial and frequency resolution of MT survey data

Due to the locality of the MT impedances, 70 km spacing data needs further validation in GIC applications. There needs to be a rigorous comparison between GIC modeled based on MT survey values and observed GIC in locations where the network topology and Direct Current (DC) characteristics are well known. The validation work will require close collaboration with industry partners.

Near-term 4: Validate and use geo-space models to calculate benchmarks

Addresses Gap A
Theoretical maximum values were not estimated

Research on magnetic storms remains a priority. Geo-space models need to be validated against ground magnetic field measurements both from permanent magnetic observatories and from denser networks of magnetic variometer stations. Of particular interest is empirical analysis of the spatiotemporal correlational relationship of localized geomagnetic disturbance across the conterminous United States with global magnetic indices (such as Dst). Also of interest is improved quantification of historical magnetic storm intensities, which are especially poorly determined pre-1957.

Additionally, we recommend that geo-space models be applied to support generation of higher fidelity induced geo-electric field benchmarks. Modern geo-space models have reached a level of maturity that might make their use possible for extreme storm studies (Ngwira et al. 2013). Using solar wind drivers representative of extreme events (to be developed by solar/heliospheric physicists), one could drive ensembles of geo-space models to generate collections of geomagnetic and geo-electric field waveforms. Further, with careful consideration of model limitations and uncertainties, modern geo-space models can also be applied in studies of theoretical geo-electric field maxima.

E. Long-term Recommendations

Long-term 1: Conduct localized MT surveys with greater resolution

Addresses Gap F
Insufficient spatial and frequency resolution of MT survey data

Once the initial MT survey has been completed on its nominally 70-km station grid (the "Level 1" survey), we recommend the commencement of a series of more localized surveys at higher spatial scale and high frequencies. This "Level 2" survey would use instrumentation with greater frequency bandwidth, capable of obtaining conductivity structure information on finer spatial scales and in areas with greater cultural electromagnetic noise. This may be required in areas with geographically complex surface impedance, such as the northeastern United States. Additionally, we recommend that regions of known high hazard, such as those adjacent to New York, be prioritized for survey.

Long-term 2: Extend the MT survey across United States' borders

Addresses Gap E
No MT survey data for Alaska and U.S. border countries

We also recommend extending the MT survey into Alaska, southern Canada, and Northern Mexico. Expanding the survey would permit improved modeling of North American-wide subsurface conductivity structure and, from that, spatially continuous mapping of surface impedance. This expanded modelling would enable a more comprehensive calculation of the hazards to U.S. critical infrastructure, improving U.S. resilience to the effects of space weather.

Long-term 3: Implement a program to investigate the needed spatial coverage of magnetic field variations

Addresses Gap G
Insufficient spatial and frequency resolution of geomagnetic observational data

We recommend that the United States implement a program using mobile variometer stations to demonstrate new geomagnetic observation posts. These observations are needed to provide the inputs for electric field hazard calculations by adequately capturing magnetic field variations. This small-scale study would inform NSF on the optimum placing for future permanent magnetic observatories, as well as long-duration, multi-model variometer/MT stations. These stations could form the core of a facility that would support a number of applications including GIC, ionospheric, and the solid Earth. By relying on the lessons learned from the EarthScope MT array and follow-on efforts, we would again recommend initially focusing on regions where the hazards are known to be high, such as northern Minnesota and/or regions within the Northeast and Appalachia/Piedmont.

This recommendation should be considered within the context of other international efforts, which would likely lead to fruitful collaboration through the sharing of datasets, cross-validation of the results, and/or exchange of analysis techniques. For example, both Australia and China currently have continental MT surveys underway. Additionally, both Ireland and the UK have completed partial surveys and have expressed interest in collaborating with the United States on issues with spatial and temporal complexity and how these influence optimum MT surveying and magnetic monitoring.

Long-term 4: Deploy a geomagnetic monitoring campaign

Addresses Gap G
Insufficient spatial and frequency resolution of geomagnetic observational data

Based on the results of the prototype station placements, we recommend that the United States deploy campaign-style magnetometers to better quantify spatiotemporal complexity of storm-time geomagnetic field variation. This might include deployment of magnetometers in arrays of different sizes and different station spacing, possibly placed temporarily at different latitudes.

The development of a campaign event program using both electric and magnetic field (i.e., MT) sensors, ideally in conjunction with the deployment of prototype variometer stations, would allow researchers to test the fidelity of induced geo-electric field estimates based on previously calculated impedances. If possible, these should occur at the time of a geomagnetic storm to maximize usefulness of the data.

Long-term 5: Support efforts that model time-dependent magnetic fields

Addresses Gap B
Geomagnetic disturbance duration was not estimated

And Gap C
No time-series information

And Gap H
Need more research to create benchmark waveforms

Global geo-space models have now reached a level of maturity such that they may be able to provide ensemble estimates of the magnetic field during a large geomagnetic storm. Modelling would allow researchers to produce an independent prediction of ground electric fields for every realization of the ensemble and at every location, providing an independent examination of the extreme value statistics. Thus, we recommend that funding agencies support proposals that address these types of modeling efforts. Further work is required to assess how well current and near-future models can produce the time-dependent variations in the magnetic field, as well as over what spatial resolution. Other efforts, such as a NASA, Living With a Star Program (LWS)-funded extreme space weather event "Focused Science Team" are underway, and we recommend that the panel solicit input from them.

Long-term 6: Support research to generate waveform benchmarks

Addresses Gap B
Geomagnetic disturbance duration was not estimated

And Gap C
No time-series information

And Gap H
Need more research to create benchmark waveforms

It is important to fund research aimed at creating meaningful geo-electric field waveforms for 1-in-100 year or theoretical maximum contexts. While current techniques for estimating electric field benchmarks have proven useful, we suggest that other promising approaches should be considered and potentially invested in, including eigenvalue analysis and machine learning. Although speculative at the moment, these may lead to better estimates of the spatial and temporal variations, as well as more rigorous constraints on the uncertainties in these estimates.

Long-term 7: Improve modelling methods for MT impedances

Addresses Gap D
Incomplete understating of surface impedance for the continental United States

Finally, methods for modeling MT impedances need to be improved to better accommodate spherical-Earth geometries and to better accommodate non-plane-wave effects associated with localized ionospheric sources. This could incorporate magnetospheric-ionospheric modeling with the use of Earth-surface impedance derived from MT surveys, assisting with the development of theoretical upper-limits on geo-electric hazards.

References

Kelbert, A., Bedrosian, P. A. & Murphy, B. S. 2019. "The first 3D conductivity model of the continental US: Reflections on geologic structure and application to induction hazards." In *Geomagnetically Induced Currents from the Sun to the Power Grid* (eds J. Gannon, Z. Xu, A. Swidinski), pp. 127-151, American Geophysical Union.

Love, J. J., Pulkkinen, A., Bedrosian, P. A., Jonas, S., Kelbert, A., Rigler, E. J., Finn, C. A., Balch, C. C., Rutledge, R., Waggel, R. M., Sabata, A. T., Kozyra, J. U. & Black, C. E. 2016. "Geo-electric hazard maps for the continental United States." *Geophys. Res. Lett*., 43(18), 9415-9424, doi:10.1002/2016GL070469.

Love, J. J., Lucas, G., Bedrosian, P. A. & Kelbert, A. 2019. "Extreme-value geo-electric amplitude and polarization across the Northeast United States." *Space Weather*, 17(3), 379-395, doi:10.1029/2018SW002068.

Lucas, G. M., Love, J. J. & Kelbert, A. 2018. "Calculation of voltages in electric power transmission lines during historic geomagnetic storms: An investigation using realistic Earth impedances." *Space Weather*, 16(2), 185-195, doi:10.1002/2017SW001779.

North American Electric Reliability Corporation. 2017. *Transformer Thermal Impact Assessment White Paper*. October 2017.

Ngwira, C. M. et al. 2013. "Simulation of the 23 July 2012 extreme space weather event: What if this extremely rare CME was Earth directed?" *Space Weather*, 11(12), 671-679.

Ngwira, C., A. Pulkkinen, E. Bernabeu, J. Eichner, A. Viljanen, and G. Crowley. 2015. "Characteristics of extreme geo-electric fields and their possible causes: Localized peak enhancements." *Geophysical Research Letters*, 42(17), 6916-6921, doi:10.1002/2015GL065061.

Pirjola, R. 2002. "Review on the calculation of surface electric and magnetic fields and of geomagnetically induced currents in ground-based technological systems." *Surv. Geophys*., 23 (1), 71-90, doi:10.1023/A:10148160, 2002.

Pulkkinen, A. et al. 2006. "Role of stochastic fluctuations in the magnetosphere-ionosphere system: A stochastic model for the AE index variations." *Journal of Geophysical Research: Space Physics*, 111(A10).

Pulkkinen, A., E. Bernabeu, A. Thomson, A. Viljanen, R. Pirjola, D. Boteler, J. Eichner, P. J. Cilliers, D. Welling, N. P. Savani, R. S. Weigel, J. J. Love, C. Balch, C. M. Ngwira, G. Crowley, A. Schultz, R. Kataoka, B. Anderson, D. Fugate, J. J. Simpson and M. MacAlester. 2017 "Geomagnetically induced currents: Science, engineering and applications readiness." *Space Weather*, 15(7), 828–856, doi:10.1002/2016SW001501.

Schultz, A. 2010. "A continental scale magnetotelluric observatory and data discovery resource." *Data Sci*. J., 8, IGY6-IGY20, doi:10.2481/dsj.SS_IGY-009.

Thomson, A. W. P., A. J. McKay and A. Viljanen. 2009. "A review of progress in modeling induced geo-electric and geomagnetic fields with special regard to induced currents." *Acta Geophysica*, 57(1), 209-219, doi:10.2478/s11600-008-0061-7.

3. Ionizing Radiation

Ionizing radiation poses a number of hazards for satellites, astronauts, airline crews, and airline communication. Phase 1 divided these hazards into three major types of radiation: solar energetic particles (SEPs), galactic cosmic rays (GCRs), and radiation belt particles. For each of these types of radiation, the Phase 1 team selected energy ranges, types of particles, and locations that they expected to pose the greatest hazard to critical infrastructure, providing particle fluxes and fluences as benchmark values.

The Phase 1 Benchmarks estimated SEP, GCR, and radiation belt benchmarks

Characterizing hazards of ionizing radiation is a difficult task due to the variety of sources, forms, and locations involved. This is particularly true for the radiation belt hazards. We commend the Phase 1 team on their efforts to be as complete as possible. It should be noted that a longer, supplemental report (SR) provides more detail regarding the methodology and, in some cases, additional benchmark values. We strongly recommend that the Phase 2 team review the SR as well as what is contained in the Phase 1 report (herein referred to as "primary report" or PR). In this chapter, we note where additional details were found in the SR material without labeling them as gaps in the PR.

Phase 2 Recommendation
Review the supplementary report as well as the Phase 1 report

As was done in the primary report, our evaluations are divided into three main categories. The term SEPs, rather than solar proton event (SPE), will be used throughout this document to avoid confusion. In the radiation belt category, we include co-located populations, such as hot plasmas and SEPs with access to the inner magnetosphere, that are not radiation belt particles but nonetheless pose a hazard to technological systems or humans in space.

A. Quantity Assessment

For all benchmarks, it is recommended that values for "maximum observed to date" be given, as this is often more useful (and available) than theoretical maxima.

Phase 2 Recommendation
Provide maximum observed values for all phenomena

1. SEPs

In general the benchmarks for SEPs are well-aligned with the objectives. To better compare with the hazards due to GCRs, the SEP benchmarks should be given in intensity units.

Phase 2 Recommendation
Represent GCR and SEP benchmarks in intensity units

We note that PR Table 3 (SEP benchmarks) only has values for Geosynchronous Equatorial Orbit (GEO), while other altitudes are more useful for particular users (e.g., aviation operators). The SR does contain some material on additional orbits, identifying Low Earth Orbit (LEO) and Medium Earth Orbit (MEO) values as a gap in the benchmarks. We recommend the Phase 2 team calculate benchmarks for the most commonly used launch trajectories, including electric orbit-raising transfer orbits, and provide a methodology for users to adjust values to their specific needs. NOAA SWPC will be providing aviation radiation advisories for effective dose for airlines, based on International Civil Aviation Organization (ICAO) requirements, specifically the thresholds in ICAO Document 10100 (2018). We recommend the Phase 2 team determine whether the Phase I benchmark quantities are appropriate and consistent with values needed to support the aviation advisory thresholds.

Phase 2 Recommendation
Calculate SEP benchmarks at common orbits other than GEO

Phase 2 Recommendation
Confirm consistency of SEP benchmarks with aviation needs

SEP intensities change by orders of magnitude throughout a given event and their time profiles (i.e., how long intensities are greater than a certain threshold) also exhibit substantial event-to-event variability. Given the customary engineering use of fluence for SEP events, we recommend that event durations also be given along with proton intensities to simplify conversion. We further recommend that benchmarks for smaller event durations be given—e.g., the worst 5-minute intensity, the worst 1-day intensity, the worst week intensity.

Phase 2 Recommendation
Provide benchmarks for a variety of event durations

The PR identified particle intensities at energies > 500 MeV as a known and important gap. Particles at these energy levels penetrate through the atmosphere and create a shower of secondary radiation that reaches aviation altitudes (Matthia et al. 2014) and the ground. Since the writing of the report, detailed observations from the PAMELA mission have been published, including SEP spectra extending to over a GeV in energy for large SEP events (Bruno et al. 2018). Although these observations span a rather limited time period (2006–2014), they have been used to further calibrate the high energy observations available from the GOES spacecraft, providing additional information useful to determining benchmarks for the high energies (Bruno 2017).

Gap A
Missing benchmark for SEP intensities >500 MeV

A benchmark for heavy ion intensities is missing. Heavy ion intensities are important for satellites as they can cause single event upsets as well as present additional radiation hazards for planned human missions. There is some treatment of this in the SR and we recommend that the Phase 2 benchmarks include a benchmark or several (e.g., one per element) for heavy ions.

Phase 2 Recommendation
Create benchmarks for heavy ions

Although the PR identified electrons as a specific hazard in the overview section, there is no further discussion of electron-related hazards. The Phase 2 team should examine the hazards related to SEP electrons and whether electron

Phase 2 Recommendation
Consider creating a benchmark for SEP electrons

benchmark values should be developed. If users are not concerned with energetic electrons, then the overview material should be modified appropriately.

The Phase 2 team should consider the inclusion of a benchmark for linear energy transfer (LET, MeV cm^2 mg^{-1}) spectra. Radiation effects engineers work with GCR and SEP ion spectra that have been transformed from number flux versus energy to number flux versus LET. Typically, these spectra are integral rather than differential in LET and combine the contributions of all elements from hydrogen through uranium (Xapsos et al. 2007). Moreover, worst-case SEP event-integrated fluence spectra (particles cm^{-2}) are needed so that radiation effects engineers can predict the number of single event effects during the worst-case event. Because the derivation of these LET curves is complicated—requiring multiple empirical LET curves and singularities in the transformation—it is not practical that users calculate the values from the benchmark heavy-ion number-flux spectra, hence the recommendation for inclusion.

Phase 2 Recommendation
Consider creating a benchmark for linear energy transfer

2. GCRs

Galactic cosmic rays are the easiest to characterize as they are well measured and generally slowly varying. The benchmarks given in the PR are well-aligned and sufficient for the objectives and identified use cases. Our only recommendation is for the Phase 2 team to consider whether having a benchmark associated with the Pfotzer maximum (Reitz 1993) would be useful. This quantity corresponds to the region of the atmosphere where the radiation level due to GCRs reaches a maximum.

Phase 2 Recommendation
Consider creating a benchmark for the region of atmosphere where GCRs are a maximum

3. Radiation Belts

Benchmarks associated with radiation belt hazards are particularly difficult to capture in a simple table as the hazard varies significantly depending on the orbit/location. It may be more useful to users to organize the benchmarks by common orbits, e.g., GPS, LEO, GEO, and electric orbit raising/low-thrust orbits (Horne and Pitchford 2015; Lozinski et al. 2019; Messenger et al. 2014) and then provide a clear prescription on how to calculate the hazards for other orbits. Additionally, the current benchmarks give both integral and differential fluxes. In an effort to reduce the number of values presented, we recommend including only differential fluxes, from which integral values can generally be easily calculated.

Phase 2 Recommendation
Consider organizing radiation belt benchmarks by orbits

Phase 2 Recommendation
Only provide values for differential fluxes

Spacecraft surface charging can be a significant cause of spacecraft anomalies due to electrostatic discharges. Benchmarks related to this hazard were not determined and were identified as a gap in the PR. We suggest the Phase 2 team consider developing new benchmarks that are relevant to the hazards of surface charging and refer to the following sources when developing such benchmark values: Mateo-Velez et al. 2018, the NASA guideline for mitigating in-space charging effects (NASA-HDBK-4002A), and the ISO standard, *Plasma Environments for Generation of Worst-Case Electrical Potential Differences for Spacecraft* (ISO 19923:2017).

Phase 2 Recommendation
Consider developing benchmarks relevant to surface charging

Conditions relating to high-speed solar wind streams are insufficiently addressed. A recent paper by Meredith et al. 2016 found that the radiation belts respond significantly differently to the impact of high-speed streams. The effect of these streams on the radiation belts should be examined and benchmarks developed accordingly.

Phase 2 Recommendation
Better address the effect of high-speed solar wind streams

Proton hazards at the radiation belts are insufficiently addressed. The PR primarily focuses on electrons. Only PR Table 7 (proton benchmarks) provides any benchmarks on protons and these are for low energy protons in the slot region appropriate for solar array degradation. The work should be expanded to include energies > 10 MeV in the inner Van Allen belt, where the primary hazards to spacecraft are single event effects and the increased accumulation of total ionizing dose and displacement damage. An ideal dataset for specifying this hazard is the Van Allen Probes Relativistic Proton Spectrometer instrument (Mazur et al. 2013).

Phase 2 Recommendation
Expand proton benchmarks to be relevant for satellite hazards beyond solar array degradation

B. Value Assessment

1. SEPs

The values given for the SEP benchmarks are reasonable based on observations of recent extreme events (e.g., the benchmarks are approximately one order of magnitude higher than these values).

2. GCRs

The benchmark values for GCRs are generally reasonable. The spectrum given in PR Table 4 (GCR 1-in-100 year benchmarks) has an unexpected bump near 30 GeV; the intensities should be reviewed and corrected if need be. Values could be compared to recent Voyager measurements. It should be noted that the value of phi (for the degree of modulation) chosen is only explained in the supplemental report. It is recommended that at least a brief statement be derived from this discussion and included in the Phase 2 report for better understanding.

Phase 2 Recommendation
Review the phase 1 GCR spectrum for possible errors

Phase 2 Recommendation
Provide a justification for the chosen degree of modulation

3. Radiation Belts

Most of the values for the radiation belt benchmarks are reasonable. Values for the inner belt should be revisited using data from the Van Allen Probes/Relativistic Proton Spectrometer as these new data have substantially improved our understanding of the inner radiation belts. Additionally there may be a typo in PR Table 7 (1-in-100 year proton fluxes) as it is odd that the 5 MeV flux is less than the 3 MeV and the 10 MeV value by two orders of magnitude. These values should be verified.

Phase 2 Recommendation
An important new data set is available and should be used

Phase 2 Recommendation
Review the phase 1 table of 100-year proton fluxes for errors

C. Methodology Assessment

Before discussing each of the three phenomena, we note a methodological gap, concerning the use of geo-magnetic cutoff models, that cross-cuts the SEPs, GCRs, and radiation belt benchmarks. Extreme value theory is used to estimate the 1-in-100 year benchmark values, which requires an estimate of the probability distribution relating the intensity of the space weather phenomenon with its frequency. The ionizing radiation benchmarks use a geomagnetic cutoff model to estimate the needed distribution; however, there has been recent work discussing the errors in common cutoff models (O'Brien et al. 2018; Kress et al. 2013, 2015).

Gap B
Geomagnetic cutoff models used in Phase 1 may have errors

1. SEPs

Generally the methodology for establishing the benchmark values is rigorous and compelling. In some cases, more recent data or newer models are now available and should be exploited to update the values.

One methodological challenge is that, even with the SR, the methodology for determining the stated uncertainties was not sufficiently discussed (e.g., the upper limit + 1 sigma values in PR Table 3 [SEP benchmarks]). Towards this end, we recommend the Phase 2 team examine Jiggens et al. (2018) who use the recently developed SAPPHIRE model to calculate multiple 1-in-N-year proton-event values, which can provide an indication of the uncertainties. SAPPHIRE is similar to the PSYCHIC model referenced in the Phase 1 report but takes advantage of the SEPEM database, which provides cross-calibrated GOES solar proton fluxes since 1974.

Phase 2 Recommendation
Review new literature on 1-in-N year proton events

The SR includes analysis of heavy ion benchmarks—the methodology for determining them is not sufficiently explained. Ion spectra are shown in SR Figure 5 for an event that is clearly enriched in Fe; however, this characteristic

Phase 2 Recommendation
If heavy ion benchmarks are included, the methodology must be sufficiently explained

is not typical for large or extreme events and thus the resulting benchmarks may not be suitable.

2. GCRs

The methodology for determining the GCR benchmarks is up-to-date, rigorous, and compelling. The methodology of calculating the theoretical maximum by assuming no modulation is correct; however, the PR or SR did not clarify how the uncertainties were determined. More clarity is needed on how the quoted uncertainties of 10–25% were determined as well as what the source was for the assumed composition values. Furthermore, GCR values outside of the heliosphere have not been empirically verified.

Phase 2 Recommendation
Provide more clarity on how the theoretical maximum and its uncertainties were calculated

Gap C
Theoretical maximum values for GCRs could not be empirically verified or made practical

Generally, GCRs are well measured; however, in the region of 0.5–2 GeV, where modulation begins to be substantial, there are not routine, continual measurements (with the exception of Alpha Magnetic Spectrometer data that are not made public).

Gap D
GCRs are not well measured in the low energy, 0.5–2 GeV range

3. Radiation Belts

Some of the methodology for determining the radiation belt benchmarks is rigorous and compelling. There are several new data sources that are potentially relevant to updating the benchmark values:

- Recent public release of GPS electron fluxes derived from CXD dosimeters on multiple satellites. (Morley et al. 2016) Released in 2017. https://www.ngdc.noaa.gov/stp/space-weather/satellite-data/satellite-systems/gps/;
- Recently GPS integral solar proton fluxes were cross-calibrated with GOES data and recalculated using SEPEM effective energies (Carver et al. 2018); and
- Van Allen Probes data; specifically the RPS instrument.

The scaling methodology using the AE9 and AP9 models is not appropriate and should be revisited. Extreme events are outside the range of validity for these models; however, it is possible to run the models in a manner more suitable for such events. To improve the benchmarks for the electrons in the radiation belts, it is recommended that Phase 2 rerun the AE9 and AP9 models in a manner more appropriate to extreme conditions.

Phase 2 Recommendation
Revisit the methodology for running the AE9 and AP9 models

Additionally, the recently released GPS data can be used to inform the extreme values of electron flux above L ~ 4.The British Antarctic Survey has a model, which they have run for 30 years (Horne et al. 2018), that may be used instead. This model should be examined and utilized (if possible) for improved values for the outer radiation belt.

Much of the research previously referenced, as well as in the PR and SR, involved cross-calibrations of two or more data sets. We recommend that this be applied as a best practice to all of the data sets used to create benchmarks in the next phase because observations from different instruments for the same period of time could be used to estimate observational uncertainties. Similarly, for those benchmark values that require modeling, it is recommended that ensembles of models be run and the variability in the results be used to estimate the uncertainties. Several datasets critical to improving benchmark values should be reprocessed and calibrated.

Phase 2 Recommendation
Cross-calibrate datasets and models to calculate better benchmark values and estimate uncertainty

Gap E
Need to reprocess some key datasets

Finally, we note the critical importance of datasets to setting benchmark values of ionizing radiation. While calibrating datasets and supplementing observations with modelling can improve benchmarks, accurately characterizing the intensity, duration, and frequency of extreme events will require long-term and continuing observations. Datasets can be improved by new instruments at new locations, but are threatened by aging infrastructure.

Gap F
Need to continue and expand long-lasting, science quality data streams

D. Near-term Recommendations

Near-term 1: Publicly release the Alpha Magnetic Spectrometer data

Addresses Gap A
Missing benchmark for SEP intensities >500 MeV

Determining benchmarks for SEP energies above a few hundred MeV can be done utilizing two specific datasets. The PAMELA data (Bruno et al. 2018) has been published for 30 large SEP events from cycles 23 and 24. The second data set is from the Alpha Magnetic Spectrometer which is installed on the International Space Station and has measured high energy spectra for a number of SEP events. Unfortunately, these data are not available publicly; it is recommended that action be taken to release these data. The instrument studies an energy range not routinely monitored by spacecraft.

Near-term 2: Validate GCR theoretical maximum

Addresses Gap C
Theoretical maximum values for GCRs could not be empirically verified or made practical

As a check on the reported theoretical maximum benchmark values (e.g., local interstellar medium values), it is recommended that the values be compared to recent measurements from the Voyager spacecraft, which are now both in interstellar space.

E. Long-term Recommendations

We recommend that data obtained from paleo measurement methods (e.g., analysis of ice cores) be used to validate and/or test some of the current benchmarks. This would provide a longer time base over which to evaluate the variability of extreme events. In a similar vein, there are a number of revised archival databases that should be exploited in addition to new data streams as they become available.

Phase 2 Recommendation
Use paleo measurement methods to improve current benchmarks

Long-term 1: Combine data sets to produce >500 MeV SEP benchmarks

Two specific studies are recommended that would provide significant improvements/expansion of the benchmark values. A study combining data from GOES, PAMELA and neutron monitors to fully characterize the >500 MeV portion of the SEP spectrum and its variability is considered a high priority—with limited investment it can be performed now. This study would fill an identified benchmark gap and provide significant improvements/expansion of the benchmark values.

Addresses Gap A
Missing benchmark for SEP intensities >500 MeV

Long-term 2: Support neutron monitors

Neutron monitor observations are currently one of our best measurements of the high-energy SEP population and provide the most enduring data repository. These data are particularly valuable for extreme space weather benchmarks and can provide early warning during fast-rising events. Funding to maintain neutron monitors has declined in the recent past and jeopardizes this valuable asset. We recommend that the funding for these installations/ instruments be sustained at a level needed for regular maintenance, data processing, and the training of early career scientists required to assure the longevity of this data resource.

Addresses Gap F
Need to continue and expand long-lasting, science quality data streams

Neutron monitors provide high-quality long-term SEP data, but their funding is in jeopardy

Long-term 3: Support new space-based SEP monitoring instruments

Regarding SEP measurements, it is critical to have continual SEP monitoring at the Earth-Sun Lagrange point 1, which is outside the influence of Earth’s magnetic field, preferably over a range of elements (not just protons) and energies. These measurements need to be of science-level quality, rather than lower quality measurements used for monitoring and now-casting radiation levels, in order to further the understanding of the system and the related radiation hazards.

Addresses Gap F
Need to continue and expand long-lasting, science quality data streams

Science-quality SEP measurements at L1 are critical

Further, measurements of the SEP spectrum between 500 MeV and 2 GeV are needed on a regular basis for an extended period of time. This could be

Also addresses Gap A
Missing benchmark for SEP intensities >500 MeV

obtained from a new spacecraft, an instrument added to another spacecraft, or a future iteration to the GOES satellites. The same instrument could also examine the GCR protons in the energy range at which modulation begins to have strong effects.

The Air Force has a goal of including radiation detectors (i.e., energetic charged particle sensors) on all Air Force satellites. This effort should be supported and expanded; a standard package could be created that is flown on every mission. The package could also include sensors that help address our suggested surface charging benchmark.

Long-term 4: Update magnetic cutoff models

Addresses Gap B
Geomagnetic cutoff models used in Phase 1 may have errors

It is recommended that the magnetic cutoff models utilized for the benchmarks be updated and/or a new model be built based on recent work. In particular, there has been recent work discussing the errors in common cutoff models (O'Brien et al. 2018; Kress et al. 2013, 2015). REACH, a constellation of 32 dosimeters in LEO (Mazur et al. 2017), could possibly be a useful validation tool for any new model or model improvements. GPS data may also be used to study and validate the dynamics of the geomagnetic cutoff and its impact on SEP and radiation belt intensities.

Long-term 5: Enable GCR measurements in the 0.5–2 GeV range

Addresses Gap D
GCRs are not well measured in the low energy 0.6–2 GeV range

We recommend enabling GCR measurements in the 0.5–2 GeV to improve the benchmarks. Focusing accuracy at these lower energies would be particularly valuable if the upcoming solar cycles are significantly weaker than has been previously observed.

Long-term 6: Reprocess the POES and DMSP datasets

Addresses Gap E
Need to reprocess some key datasets

Of highest priority for improved radiation belt benchmarks is the reprocessing of the POES and DMSP datasets. This will provide an important long-term dataset that will provide needed information regarding variability and observed worst cases.

Long-term 7: Study theoretical and observed electron flux values

Addresses Gap E
Need to reprocess some key datasets

A systematic study comparing the theoretical flux maxima derived for nonrelativistic (Kennel and Petschek 1966) and relativistic (Summers and Shi 2014) electrons in the radiation belts to the 1-in-100 year values and most extreme values observed should be done.

Long-term 8: Investigate particle tracing for GCR benchmarks

The current state of the art particle tracing through the heliosphere should be examined to determine whether there is room for improvement over the older models currently in use. This is of lower priority largely because, while useful, it is unclear how much of an improvement could be made over the current theoretical maximum (which is already of limited practical use).

Addresses Gap C
Theoretical maximum values for GCRs could not be empirically verified or made practical

References

Bruno, A., Bazilevskaya, G. A., Boezio, M., Christian, E. R., de Nolfo, G. A., Martucci, M., and Ryan, J. M. 2018. “Solar energetic particle events observed by the PAMELA mission.” *The Astrophysical Journal*, 862(2), 97.

Bruno, A. 2017. “Calibration of the GOES 13/15 high-energy proton detectors based on the PAMELA solar energetic particle observations.” *Space Weather*, 15(9), 1191-1202.

Carver, M. R., Sullivan, J. P., Morley, S. K., & Rodriguez, J. V. 2018. “Cross calibration of the GPS constellation CXD proton data with GOES EPS.” *Space Weather*, 16, 273– 288. https://doi.org/10.1002/2017SW001750

Horne, R. B., Phillips, M. W., Glauert, S. A., Meredith, N. P., Hands, A. D. P., Ryden, K., & Li, W. 2018. “Realistic worst case for a severe space weather event driven by a fast solar wind stream.” *Space Weather*, 16, 1202– 1215. https://doi.org/10.1029/2018SW001948

Horne, R. B., and Pitchford, D. 2015. “Space Weather Concerns for All‐Electric Propulsion Satellites.” *Space Weather*, 13, 430– 433, doi:10.1002/2015SW001198.

International Civil Aviation Organization. 2018. *Manual on Space Weather Information in Support of International Air Navigation*. ICAO Document 10100, https://www.icao.int/airnavigation/METP/Panel%20Documents/Doc.10100.Space%20Weather%20Manual%20FINAL%20DRAFT%20Version.pdf

Jiggens, Piers, Daniel Heynderickx, Ingmar Sandberg, Pete Truscott, Osku Raukunen, Rami Vainio. 2018. “Updated Model of the Solar Energetic Proton Environment in Space.” *J. Space Weather Space Clim*. 8 A31. DOI: 10.1051/swsc/2018010

Kennel, C. F., and Petschek, H. E. 1966. “Limit on stably trapped particle fluxes.” *J. Geophys. Res*., 71(1), 1– 28, doi:10.1029/JZ071i001p00001.

Kress, B.T., J. V. Rodriguez, J.E. Mazur and M. Engel. 2013. “Modeling solar proton access to geostationary spacecraft with geomagnetic cutoffs.” *Adv. Space Res*., 52, 1939-1948.

Kress, B. T., M. K. Hudson, R. S. Selesnick, C. J. Mertens, and M. Engel. 2015. “Modeling geomagnetic cutoffs for space weather applications.” *J. Geophys. Res. Space Physics*, 120, 5694–5702, doi:10.1002/2014JA020899.

Lozinski, A. R., Horne, R. B., Glauert, S. A., Del Zanna, G., Heynderickx, D., & Evans, H. D. R. 2019. “Solar cell degradation due to proton belt enhancements during electric orbit raising to GEO.” *Space Weather*, 17, 1059– 1072. https://doi.org/10.1029/2019SW002213

Matéo-Vélez, J.-C., Sicard, A., Payan, D., Ganushkina, N., Meredith, N. P., & Sillanpäa, I. 2018. “Spacecraft surface charging induced by severe environments at geosynchronous orbit.” *Space Weather*, 16, 89–106. https://doi. org/10.1002/2017SW001689

Matthia et al. 2014. “Numerical Calculation of the Radiation Exposure from Galactic Cosmic Rays at Aviation Altitudes with the PANDOCA Core Model.” *Space Weather*, 12, 161–171, doi:10.1002/ 2013SW001022

Mazur, J.E. 2017. “SA21C-05: Monitoring Space Radiation Hazards with the Responsive Environmental Assessment Commercially Hosted (REACH) Project.” Fall AGU Meeting. New Orleans, LA, December 2017. https://agu.confex.com/agu/fm17/meetingapp.cgi/Paper/212711

Mazur, J.E., et al. 2013. “The Relativistic Proton Spectrometer for the Radiation Belt Storm Probes Mission.” *Space Science Reviews* Volume 179, Issue 1–4, pp 221–261 https://rd.springer.com/article/10.1007/s11214-012-9926-9

Meredith, N. P., Horne, R. B., Isles, J. D., Ryden, K. A., Hands, A. D. P., and Heynderickx, D. 2016. "Extreme internal charging currents in medium Earth orbit: Analysis of SURF plate currents on Giove-A." *Space Weather*, 14, 578– 591, doi:10.1002/2016SW001404.

Morley, S. K., Sullivan, J. P., Henderson, M. G., Blake, J. B., & Baker, D. N. 2016. "The Global Positioning System constellation as a space weather monitor: Comparison of electron measurements with Van Allen Probes data." *Space Weather*, 14, 76–92. https://doi.org/10.1002/ 2015SW001339

O'Brien, T. P., Mazur, J. E., & Looper, M. D. 2018. "Solar energetic proton access to the magnetosphere during the 10–14 September 2017 particle event." *Space Weather*, 16, 2022– 2037. https://doi.org/10.1029/2018SW001960

Reitz, G. 1993. "Radiation environment in the stratosphere." *Radiat. Prot. Dosim*., 48(1), 5–20.

Scott R. Messenger, Frankie Wong, Bao Hoang, Cory D. Cress, Robert J. Walters, Craig A. Kluever, and Glenn Jones, "Low-Thrust Geostationary Transfer Orbit (LT2GEO) Radiation Environment and Associated Solar Array Degradation Modeling and Ground Testing," *IEEE Transactions On Nuclear Science*, Vol. 61, No. 6, December 2014, 3348-3355

Summers, D., and Shi, R. 2014. "Limiting energy spectrum of an electron radiation belt." *J. Geophys. Res*. Space Physics, 119, 6313–6326, doi:10.1002/2014JA020250.

Xapsos, M. A., Stauffer, C., Jordan, T., Barth, J. L., & Mewaldt, R. A. 2007. "Model for cumulative solar heavy ion energy and linear energy transfer spectra." *IEEE Transactions on Nuclear Science*, 54(6), 1985–1989.

4. Ionospheric Disturbances

The Phase 1 ionospheric disturbances benchmarks focus on describing properties that can affect radio propagation for communication, navigation and timing. The ionospheric disturbances are driven primarily by three types of space weather phenomena: solar flares, solar energetic particle events, and geomagnetic storms. The ionospheric characteristics of interest were defined by:

- Ionospheric radio absorption as a function of frequency;
- Total electron content (slant, vertical, and rate of change);
- Ionospheric turbulence that alters the phase and amplitude of transmitted signals (i.e., scintillation); and
- Peak ionospheric densities and the height of the peak.

This document classifies the Phase 1 benchmark quantities into three categories: F-region, turbulence, and absorption. The F-region is of practical importance as a dominant influence on radio propagation. Total electron content (TEC)—i.e., the number of electrons between the transmitter and receiver—is an F-region benchmark that is directly related to signal delays in trans-ionospheric communications. Other F-region benchmarks include the peak ionospheric density in the F2 layer (NmF2) and the height of the F2 layer (hmF2). Turbulence benchmarks are those associated with phase (σ_ϕ) and amplitude scintillation (S4) of transmitted signals. Absorption is characterized in Phase 1 by the user quantities of highest affected frequency (HAF) and maximum usable frequency (MUF).

A. Quantity Assessment

1. F-region: TEC, NmF2 and hmF2

The F-region benchmark quantities selected in the Phase 1 document are generally well-aligned with objectives. Specifically, maximum TEC as well as temporal and spatial gradients of TEC are useful benchmark quantities.

Phase 2 recommendation
TEC values should be expressed as VTEC, independent of look angle

Phase 2 recommendation
TEC gradient values should be provided in a technology-independent unit

However, TEC values should be expressed as vertical TEC values (VTEC), which are independent of look angle. VTEC describes the vertically integrated ionospheric electron density and can be scaled to different slant paths as appropriate for impact analysis. Additionally, the TEC gradient values should be provided in a unit independent of any technological system, such as

TECu/km, that readily describe physical characteristics of the propagation medium; the current benchmark quantities are only provided in units specific to the GPS L1 range.

Quantities NmF2 (peak ionospheric density) and hmF2 (height of the peak of the layer) describe key physical parameters in the propagation medium that translate readily into user impact. The 2015 Space Weather Action Plan also called for an ionospheric refractive index benchmark; however, a refractive index benchmark is not recommended as it is of limited value in the absence of complete ionospheric profiles.

2. Turbulence: phase and amplitude scintillation

The benchmark quantities suggested by Phase 1 are not the most effective parameters for capturing physical characteristics of ionospheric irregularities. While the proposed phase (σ_ϕ) and amplitude scintillation (S4) benchmarks are popular and often used in ionosphere studies, largely due to their ease of derivation from raw observables collected by ground-based instruments, both σ_ϕ and S4 scintillation indices are frequency and geometry dependent and σ_ϕ depends on scan velocity. Both indices are dependent on the processing interval and de-trending filter used (especially σ_ϕ). S4 is a normalized quantity with a theoretical maximum value of approximately 1.

These benchmark indices reflect cumulative ionospheric effects (refraction and diffraction due to turbulence) on measured RF signals and are limited by inherent data collection and processing assumptions**.** It is recommended to consider using a system independent ionospheric turbulence measure such as C_kL, where C_k is the 1 km cross-section of the power-spectral density of the ionospheric irregularity and L is the thickness of the irregularity layer.

Phase 2 recommendation
Consider using a system independent turbulence measure such as CkL

C_kL is a measure of the total power in the electron density irregularities along a path passing through the entire ionosphere. It is expressed as log of the height-integrated irregularity strength, either simulated or observed, calculated on line-of-sight paths from ground to an overhead satellite. This would be more representative of the actual environment than σ_ϕ and S4 and, while perhaps less intuitive for some users, it fully captures the effects on both amplitude and phase. Vertical C_kL benchmarks would provide flexibility for computing propagation parameters based on frequency, geometry, dynamics, etc. Additionally, the scintillation indices σ_ϕ and S4 could be derived from C_kL.

Gap A
Proposed benchmark quantities may not be intuitive for users

3. Absorption: HAF and MUF

The absorption benchmark quantity is not well-aligned with the Phase 1 report. The Phase 1 analysis proposed a benchmark parameter of HAF which is generated in space weather products such as NOAA's D-RAP model and is dependent on absorption thresholds (e.g., 1 dB globally or 10 dB in polar regions). The Phase 1 report did not define the attenuation threshold to be applied for the HAF benchmark. Additionally, despite its use by system operators, HAF is not a widely accepted community standard; for example, it is not in the International Telecommunication Union recommendations (ITU-R). It is recommended that a modified benchmark for absorption (in dB units) for a given frequency at vertical incidence be considered instead. For example, riometers directly measure absorption at 30 MHz and ITU-R P.531-13 quotes values in this quantity. Phase 2 could solicit community input on usefulness of this type of benchmark versus the HAF.

Phase 2 recommendation
Consider a modified benchmark for absorption for a given frequency at vertical incidence

It is recommended that two different benchmark values be provided for extreme solar flare and extreme solar energetic particle events. It is also suggested that a third benchmark be investigated and estimated for auroral absorption.

Phase 2 recommendation
Provide separate benchmarks for solar flare and solar energetic particle events

Phase 2 recommendation
Consider a new benchmark for auroral absorption

B. Value Assessment

Due to challenges in modeling ionospheric disturbances, the Phase 1 report only listed some values of the 2003 Halloween event as a temporary surrogate for benchmark values.

1. F-region: TEC, NmF2 and hmF2

The Phase 1 surrogate value of 250 TECu for maximum TEC is not appropriate. Higher values have been published (e.g., almost 350 TECu in Figure 3 of Mannuci et al. 2005) and observed (e.g., greater than 350 TECu found in the public Madrigal database). These larger values are associated with the equatorial anomaly and storm enhanced density, both of which can occur over the contiguous United States. A more appropriate benchmark value would be 350 (and perhaps 400) TECu with duration of several hours.

Phase 2 recommendation
Consider a TECu benchmark of 350–400 with a duration of several hours

The Phase 1 value of 40 cm/km for the TEC spatial gradient is specific to GPS L1 (derived from Figure 13 of Datta-Barua et al. 2010) and should be translated to a system independent value of 2.5 TECu/km. The Phase 1 value of 15 cm/s for the TEC temporal variation is similarly specific to GPS L1 (related to Figure 4 of Datta-Barua 2004) and should be translated to a system independent value of 1 TECu/S. Existing literature (Pullen et al. 2009) suggests

Phase 2 recommendation
TEC spatial gradient value should be 2.5 TECu/km

Phase 2 recommendation
TEC temporal variation value should be 1 TECu/S

that larger gradients have been observed than the stated Phase 1 values. For both TECu and its gradients, the Phase 1 analysis made insufficient use of published literature and public databases.

Gap B
Insufficient use of TEC Data

Phase 1 did not quantify NmF2 and hmF2 benchmarks. However, observational evidence from the past two solar cycles exists and can inform approximate estimates. Extreme vertical TEC benchmarks translate directly into NmF2 values (following Gerzen et al. 2013):

Phase 2 recommendation
Can use VTEC benchmark to calculate NmF2 benchmarks

$$N_{mF_2} = \frac{\mathrm{vTEC}}{\tau} = 1.24 \times 10^{10}\left(f_{oF_2}\right)^2$$

where values of τ represent ionospheric slab thickness and are in the range of 270–420.

2. Turbulence: phase and amplitude scintillation

Phase 1 studies did not provide numerical estimates.

3. Absorption: HAF and MUF

Phase 1 studies did not provide numerical estimates.

C. Methodology Assessment

As previously noted, the Phase 1 approach for estimating the F-region benchmarks was to provide values observed in the Halloween storm. Other benchmark values were not quantitatively estimated. As such, the ionospheric benchmarks are effectively lacking a methodology to assess. Instead of an assessment, this section will provide a suite of methodological suggestions for the Phase 2 benchmark team to consider.

1. F-region: TEC, NmF2 and hmF2

The elevation-angle dependence of the TEC spatial gradient parameter needs to be described and/or the benchmark values normalized to vertical, per the discussion in the section on quantity assessment. For example, literature provides examples of spatial gradients of 42.5 cm/km (exceeding the Phase 1 benchmark; Pullen et al. 2009), but these are for lower elevation angles and therefore longer slant paths (and higher values of along-path TEC). The benchmark should specify a vertical look angle, with an equation to translate to different elevation angles, for mapping to slant paths as appropriate for a given technological system.

Phase 2 recommendation
If not using VTEC, the TEC spatial gradient needs to provide more information to be useful

Ionosphere-plasmasphere models with high upper boundaries (above 600 km) exist and could be used to complement F-region observation studies for both profiled and integrated electron density values; model extreme inputs of F10.7 (400 sfu) and vertical drift (200 m/s) are appropriate drivers. Machine learning methods may provide additional insight about ionospheric response to solar drivers, taking advantage of long-term solar and TEC data sets. Combining physics-based modeling with such methods could provide new understanding of extreme events. This approach can be investigated for both 100-year and max benchmarks.

Phase 1 assumed that peak values of NmF2 and hmF2 would occur during geomagnetic storms, but not all extreme ionospheric variability is directly driven by storms; for example, extended intense solar maxima yield peak densities similar to those found in the biggest storms but over wider spatial regions. Likewise, very low solar activity can be equally variable. Driver cases other than geomagnetic storms should be considered for future refinement of the methodology. Low values of the MUF are also of critical interest to system users and operators due to associated limitations in channel capacity and availability. Methods should be explored to determine low value extremes.

Phase 2 recommendation
Consider driver cases other than geomagnetic storms

The NmF2 and hmF2 analyses should be coordinated with investigations of the TEC 100-year and maxima/minima—such that there is consistency in the ionosphere characterization—in addition to considering extreme value analyses for NmF2. Physics-based models can be run with high and low solar flux (e.g., EUV and F10.7) inputs for NmF2 analysis and with prompt penetration fields consistent with 200 m/s vertical drift for hmF2. It is also useful to coordinate with the working groups for solar radio bursts, ionizing radiation, and upper atmosphere expansion to ensure consistency in extreme value quantification when specifying model inputs. This is a priority for informing Phase 2 research efforts.

Phase 2 recommendation
Coordinate NmF2 and hmF2 with investigations of TEC extrema

Phase 2 recommendation
Coordinate with other working groups to ensure consistency in modelling inputs

Based on the evaluation of Phase 1 values, it is recommended to consider spatial and temporal evolution of future benchmark values. Also, to the extent practical and valid, benchmarks should be harmonized with the existing engineering tools for HF propagation:

Phase 2 recommendation
Consider spatial and temporal evolution of future benchmarks, and harmonize them with existing HF propagation tools

- *VOACAP (Voice of America Coverage Area Prediction) parameters;*
- *Recommendation ITU-R P.533-13 Method for the prediction of the performance of HF circuits (2015);*
- *Recommendation ITU-R P.534-5 Method for calculating sporadic-E field strength (2012); and*

- *Recommendation ITU-R P.581-2 The concept of "worst month" (1990)*

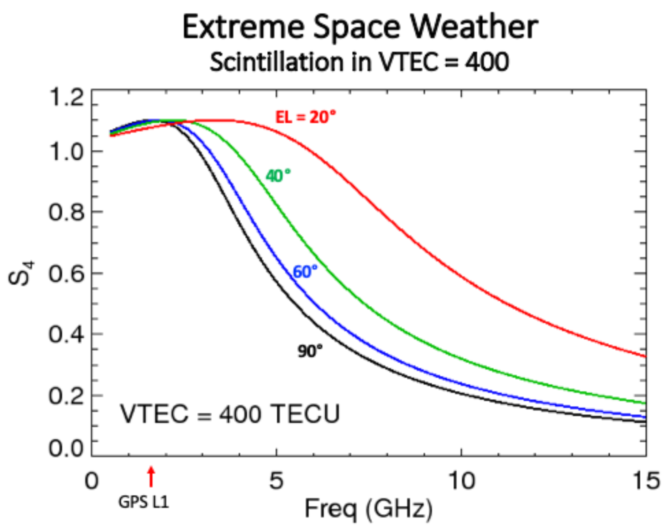
Some of the ITU-R recommendations include complex and detailed descriptions of the ionosphere and its variability. It is encouraged that development of future benchmarks reflect the attention to detail already inherent in the engineering standards.

2. Turbulence: phase and amplitude scintillation

TEC benchmarking can be used to inform turbulence benchmarks 100-year and max (expressed as C_kL or other physical parameters). Scintillation strength depends on the integrated absolute change in electron density along a ray path. It would be feasible to derive a measure of scintillation by starting with the benchmark extreme TEC value, assuming a vertical column in the ionosphere populated with irregularities throughout, and assuming characteristics such as relative variations from background TEC consistent with observed spectra (e.g., WBMOD default). From this approach scintillation parameters (phase and amplitude indices and/or C_kL) may be calculated for extreme events. This can be accomplished now using existing knowledge. A rough estimate of such calculations is provided below; however, a systematic investigation of this approach would be necessary to ensure a high fidelity estimate.

Phase 2 recommendation
We provide a methodology for benchmarking turbulence

Gap C
New methodology for turbulence benchmark is not yet validated



Extreme scintillation levels calculated by imposing irregularities in a background ionosphere with an extreme vertical TEC value of 400 TECu. The amplitude scintillation index, S4, is given as a function of frequency for radio propagation paths at four different elevation angles. The red arrow indicates the GPS L1 frequency (1.575 GHz) where S4 is saturated at all propagation angles.

Figure 1. Extreme scintillation levels calculated by imposing irregularities in a background ionosphere with an extreme vertical TEC value of 400 TECu

As an example, a calculation was performed by scaling up measured integrated turbulence strength (C_kL) from an ionosphere with vTEC = 100 TECu. Since C_kL is proportional to electron density variance, and relative variations are assumed constant, then C_kL should scale with the square of density (or total electron content). This approach results in C_kL = 1.6e37 which is then used to calculate corresponding values of the amplitude scintillation index, S4, shown above in Figure 1.

3. Absorption: HAF and MUF

It is recommended that the Ionizing Radiation Working Group provide a benchmark for integrated proton flux @ GOES (in units of $cm^{-2}s^{-1}sr^{-1}$) for use with existing empirical formulas. For example, Sauer and Wilkinson (2008) provide simple formulas for absorption at 30 MHz (based on the Thule Riometer):

Phase 2 recommendation
Ionizing radiation working group should provide a benchmark for integrated proton flux @ GOES

$$A_d=0.115[J(E>5.2\text{ MeV})]^{1/2}\text{ dB (daytime)}$$

$$A_n=0.020[J(E>2.2\text{ MeV})]^{1/2}\text{ dB (nighttime)}$$

The Ionizing Radiation Working Group evaluated the integrated fluxes for the October 1989 extreme event. Inserting those fluxes into the formulas above yields a maximum absorption of 34.66 dB at 30 MHz; however, this event is more extreme than any event in the database used to develop these empirical formulas. First-principles modeling should be conducted to test validity limits of these empirical relationships for extreme flux levels.

Gap D
Empirical absorption models are not validated for extreme inputs

For flares the spatial extent is the whole dayside and a typical X-ray flare duration model, similar to that which exists in D-RAP, could be used with maximum flare duration provided by solar experts. For SEP the spatial extent is the polar cap. The Phase 1 report suggests this region is 25 degrees around the geomagnetic pole; however, the justification for this estimate is weak. It is recommended to confirm or re-investigate such values for extent of the polar cap. The typical SEP duration is days and ITU-R P.531-13 (Figure 14) suggests 5 days. Benchmarks for extreme values of SEP duration could be provided by the Ionizing Radiation Working Group. Knowledge of extreme solar drivers is currently insufficient and is needed to quantify extreme ionospheric response.

Phase 2 recommendation
Confirm or re-investigate the spatial extent for SEP events at the polar cap

Gap E
Insufficiently characterized ionospheric response to extreme solar drivers

D. Near-term Recommendations

The panel focused on recommending methodologies for establishing quantitative benchmark values. The panel assesses that available data and models are sufficient to estimate benchmark values. The following near-term recommendations reflect the priority of gathering and analyzing existing data sources and investigating the applicability of proposed methods for estimating benchmarks.

Near-Term 1: Literature and Data Review for Extreme TEC values

Addresses Gap B
Insufficient use of TEC data

Further literature review is recommended for TEC values, both 100-year and max, including TEC gradients. A comprehensive study of not only the magnitude of observed extreme values but also their duration would provide critical information for operators' continuity and availability metrics and is also recommended.

For example, Jakowski and Hoque (2019) define and study the Gradient Ionosphere indeX (GIX) and the Sudden Ionospheric Disturbance index, which may suggest new approaches to capturing extreme event characteristics. A more exhaustive search through existing databases of TEC values and a subsequent statistical analysis would yield more robust values. Extensive historic ionosonde and GPS/GNSS data exist and can be exploited for this purpose. Notably the IGY 1957–1958 data correspond to the highest solar activity levels in approximately 400 years while 2008–2009 corresponds to lowest solar activity levels since systematic ionospheric soundings. Historic data should be made accessible through online tools, such as those implemented previously for the Space Physics Interactive Data Resource.

Gathered data should be made publicly available

Near-term 2: Turbulence Fidelity

Addresses Gap C
New methodology for turbulence benchmark is not validated

TEC benchmarking can be used to inform turbulence benchmarks, 100-year and max (expressed as C_kL or other physical parameters). In assuming plasma density irregularities to be a fraction of the background ionospheric TEC, absolute changes in density can be estimated along the ray path and integrated for new turbulence measures using the methodology recommended in Section C.2. This can be accomplished using existing knowledge.

Near-term 3: Investigate and Extend the Validity of Absorption Models

Addresses Gap D
Empirical absorption models are not validated for extreme inputs

As previously noted, HAF is a less satisfactory benchmark quantity than absorption at 30MHz; however, models for absorption tend to be empirically derived and potentially invalid for extreme inputs. First-principles modeling should be conducted to test and enhance the validity limits of these empirical relationships for extreme proton flux levels.

E. Long-term Recommendations

Until the limits of the currently available data and models are assessed, it would be premature to recommend large new observational or research efforts. The long-term recommendations focus instead on improving the utility of the newly proposed benchmark quantities and increasing knowledge of ionospheric response to extreme solar flares.

Long-term 1: Improve the Utility of Proposed New Benchmark Quantities

Addresses Gap A
Proposed benchmark quantities may not be intuitive for users

If the Phase 2 benchmarks use a physics-based unit for measuring TEC, such as TECu, it is recommended that guidelines be developed for engineers and operators to translate from the physics-based units to impact on systems. For example, a single frequency-dependent scale factor translates TECu into range observation errors for Global Navigation Satellite Systems. An ionospheric range delay is defined as being approximately equal to $(40.3/f^2)*TEC$, where f is the radio-wave frequency. The range delay is due to the radio wave traveling through the ionosphere at a speed different from the speed of light in a vacuum. The ionospheric range delay for 50 TECu (or $50*10^{16}$ el/m^2) is approximately 10 m at L-band, 100 m at UHF, and 800 m at VHF. A simple rule-of-thumb (range delay multiplied by a satellite geometry scale factor) further translates such range errors into positioning errors for real-world global navigation satellite systems (GNSS) applications. Guidelines that translate benchmark values into system impacts will improve the utility of the benchmarks.

Similarly, if a new system-independent benchmark such as C_kL is pursued, approaches or guidelines for translating into scintillation indices and/or system impacts should be developed for user communities. The physical dependence of such impacts on ionospheric irregularities is known but not necessarily intuitive for users and operators.

Long-Term 2: Study Extreme Solar Drivers for Ionospheric Absorption

Addresses Gap E
Insufficiently characterized ionospheric response to extreme solar drivers

Knowledge of extreme solar drivers is a priority for future research efforts in quantifying extreme ionospheric response. Estimates of X-ray and EUV benchmarks for an extreme solar flare would be useful for this purpose. It is recommended to investigate empirical and first-principles modeling of response to that extreme solar flare, and further studies could be conducted for extent and duration of impact. Duration is a particularly important consideration for absorption benchmarks.

References

Datta-Barua, S. 2004. “Ionospheric threats to space-based augmentation system development.” In *Proc. ION GNSS 2004*, pp. 21-24. 2004. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.298.7056&rep=rep1&type=pdf

Datta-Barua, S., J. Lee, S. Pullen, M. Luo, A. Ene, D. Qiu, G. Zhang, and P. Enge. 2010. “Ionospheric threat parameterization for local area global-positioning-system-based aircraft landing systems.” *Journal of Aircraft* 47, no. 4 (2010): 1141–1151. https://arc.aiaa.org/doi/pdf/10.2514/1.46719

Gerzen, T., Jakowski, N., Wilken, V., and Hoque, M. M. 2013. “Reconstruction of F2 layer peak electron density based on operational vertical total electron content maps.” *Ann. Geophys*., 31, 1241-1249, 2013. https://doi.org/10.5194/angeo-31-1241-2013

ITU-R Recommendation ITU-R P.531-13. 2016. “Ionospheric propagation data and prediction methods required for the design of satellite services and systems.” International Telecommunication Union.

ITU-R Recommendation ITU-R P.533-13. 2015. “Method for the prediction of the performance of HF circuits.” International Telecommunication Union.

ITU-R Recommendation ITU-R P.534-5. 2012. “Method for calculating sporadic-E field strength.” International Telecommunication Union.

ITU-R Recommendation ITU-R P.581-2. 1990. “The concept of ‘worst month.’” International Telecommunication Union.

Jakowski, N. and Hoque, M. M. 2019. “Estimation of spatial gradients and temporal variations of the total electron content using ground-based GNSS measurements.” *Space Weather*, 17, 339–356, 2019. (https://doi.org/10.1029/2018SW002119)

Mannucci, A. J., B. T. Tsurutani, B. A. Iijima, A. Komjathy, A. Saito, W. D. Gonzalez, F. L. Guarnieri, J. U. Kozyra, and R. Skoug. 2005. “Dayside global ionospheric response to the major interplanetary events of October 29–30, 2003 ‘Halloween Storms.’” *Geophysical Research Letters* 32, no. 12, 2005. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004GL021467

Pullen, S., Y. S. Park, and P. Enge. 2009. “Impact and mitigation of ionospheric anomalies on ground-based augmentation of GNSS.” *Radio Science* 44, no. 1, 2009. https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2008RS004084

Sauer, H.H. and Wilkinson, D.C. 2008. “Global mapping of ionospheric HF/VHF radio wave absorption due to solar energetic protons.” *Space Weather* 6: doi: 10.1029/2008SW000399. issn: 1542-7390, 2008.

5. Solar Radio Bursts

Solar flares emit potentially hazardous solar radio bursts, on average occurring every 3.5 days during solar maximum and every 18.5 days at solar minimum (Nita et al. 2002). These bursts disrupt the radio spectrum on which the United States relies extensively. The purpose of this benchmark is to enhance awareness of the threat of extreme solar radio bursts to infrastructure owners/operators, provide input for engineering standards for wireless communication and navigation systems, aid in the assessment of vulnerability and risk to those systems, and to help guide development of mitigation procedures and establish thresholds for action. One of the most important types of these systems, GNSS, is unique in using right-hand circular polarization (RCP) for transmission and reception. Although solar radio bursts occur equally in both senses of circular polarization, only RCP bursts affect GNSS.

A. Quantity Assessment

The benchmark quantities are mostly, but not fully, aligned with the Phase 1 objectives. In Phase 1, the solar radio burst benchmark was divided into three frequency ranges—VHF, UHF, and microwave—based on users' operating frequencies and theoretical understanding of solar radio emission mechanisms. In addition, separate benchmarks were given for two specific bands—the GNSS frequency range and the frequency used for the F10.7 index. The Phase 1 is not fully aligned with its objectives, in part, because it did not estimate values for the theoretical maximum benchmark. For the quantities that are estimated, we find that they are well-aligned with the objectives and use cases in the Phase 1 document, except that we do not see the need for a separate F10.7 benchmark. The F10.7 index is an important, long-running index of the general level of solar activity that is used as a proxy for ionizing radiation affecting the Earth's upper atmosphere. Thus, the quiet time F10.7 flux is important for other benchmarks, but only as a proxy. The radio emission itself is not ionizing and hence when additional radio flux density is generated during solar bursts the F10.7 proxy is no longer reliable and cannot be used as a measure of ionization of the Earth's atmosphere. Because the 10.7 cm (2800 MHz) band has no special role in wireless communication and navigation systems, we recommend that it be dropped as a separate solar radio burst benchmark.

Gap A
No theoretical maximum benchmark

Phase 2 Recommendation
The F10.7 benchmark has no identifiable purpose and should be dropped

An additional benchmark should be added to characterize the duration of an extreme event**,** as the consequences of communication and navigation outages increase with time. For instance, a burst only momentarily reaching threshold and causing outages will have little practical impact, while a similar burst above threshold for 10s of minutes can seriously impact GNSS navigation and air traffic control over the entire sunlit hemisphere (Cerruti et al. 2008; Marqué et al. 2018). Phase 2 should conduct research to add a benchmark for event duration above various thresholds. This new benchmark can possibly be derived from the existing database of light curve information. A preliminary search of the data suggests that lifetimes for >10,000 sfu events range from 10 min–400 min.

Phase 2 Recommendation
Add a benchmark for event duration

Extreme bursts may last from 10 to 400 minutes

B. Value Assessment

The Phase 1 team arrived at benchmarks using data published in the early 2000s, mainly the 40-year NOAA database used by Nita et al. (2002). Another 20 years of additional data exist that should be used to derive the Phase 2 benchmarks. The NOAA database should be checked against the smaller but independent Nobeyama Radio Polarimeter database. We believe that the Phase 1 approach has already yielded the best-available benchmark value in the "microwave" category, but this should be verified.

Phase 2 Recommendation
Additional 20 years of data exists and should be used

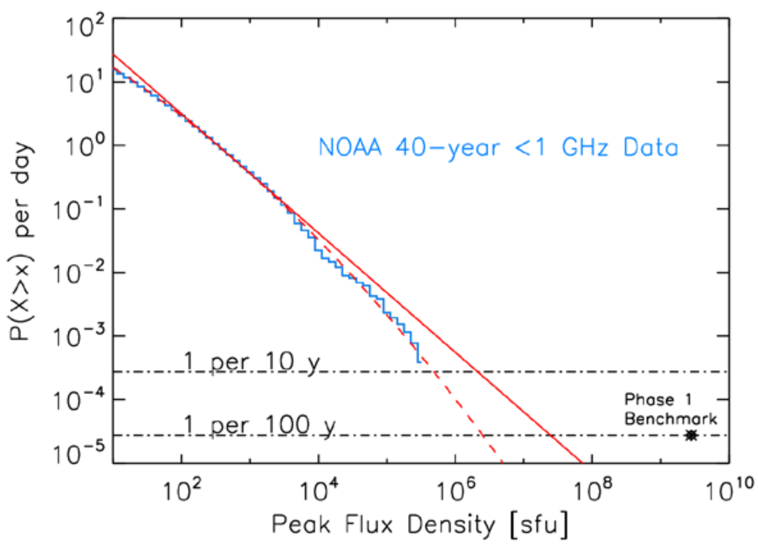


Figure 2. Distribution of bursts below 1 GHz, from the 40-yr (1960–1999) NOAA database used by Nita et al. (2002).

The extrapolated VHF benchmark value seems unreasonably high based on the observed population of events in that frequency range. As shown in Figure 2 (based on the NOAA 40-yr database used by Nita et al. 2002), of all bursts below 1 GHz, no bursts observed from 1960–1999 exceed 3 x 10^5 sfu.

Extending the distribution with a powerlaw fit to 1 event per 100 y only reaches 2×10^7 sfu. The log-normal (dashed curve) fit yields an even smaller 100-y flux density. The Phase 1 benchmark of 2.8×10^9 sfu is shown for comparison. In any case, this benchmark should be reevaluated based on new methodology as described in the next section.

Phase 2 Recommendation
The VHF benchmark should be reevaluated based on methodology presented in this report

C. Methodology Assessment

The Phase 1 process was limited to existing published research, but that research, while a useful guide, was not intended specifically to address the extreme solar radio burst hazard. We find that the 1-in-100 year benchmark methodologies and values can be improved.

Based on our current understanding of the cause of extreme solar radio burst flux densities in the VHF and UHF frequency bands (the latter includes the GNSS frequencies), the extreme events are dominated by coherent emission mechanisms. The published research makes no distinction between events dominated by weaker, incoherent mechanisms and stronger, coherent mechanisms, so the data from which the benchmarks were extrapolated mix these two possibly quite different populations. This methodological gap may contribute to an underestimate of the solar radio bursts that affect critical infrastructure. A new analysis of existing data is called for that attempts to separate solar bursts into two populations—those dominated by incoherent emission, which are not likely to be a threat to communication and navigation systems, and those dominated by coherent emission, which are a threat because of their ability to attain extremely high (nonthermal) brightness temperatures.

Gap B
Benchmarks made no distinction between coherent and incoherent bursts in VHF and UHF bands

The gap may underestimate the threat to critical infrastructure from coherent bursts

The panel discussed ways to separate the incoherent and coherent burst populations in the existing data. Based on our understanding of incoherent bursts, which are due to the well-understood gyrosynchrotron emission mechanism, we expect the flux densities to fall monotonically to lower frequencies. Coherent processes, in contrast, tend to grow stronger at low frequencies. This creates the classic Castelli U-burst spectral shape (Guidice and Castelli 1975) in which the spectrum falls toward lower frequencies over some range and then rises sharply again at still lower frequencies. Coherent processes also tend to produce strongly circularly polarized emission, in contrast to the typically low to modest degree of polarization for incoherent bursts.

The two chief sources of radio burst flux density measurements are the U.S. Air Force worldwide Radio Solar Telescope Network (RSTN), which does not measure polarization, and the Nobeyama (Japan) Radio Polarimeters

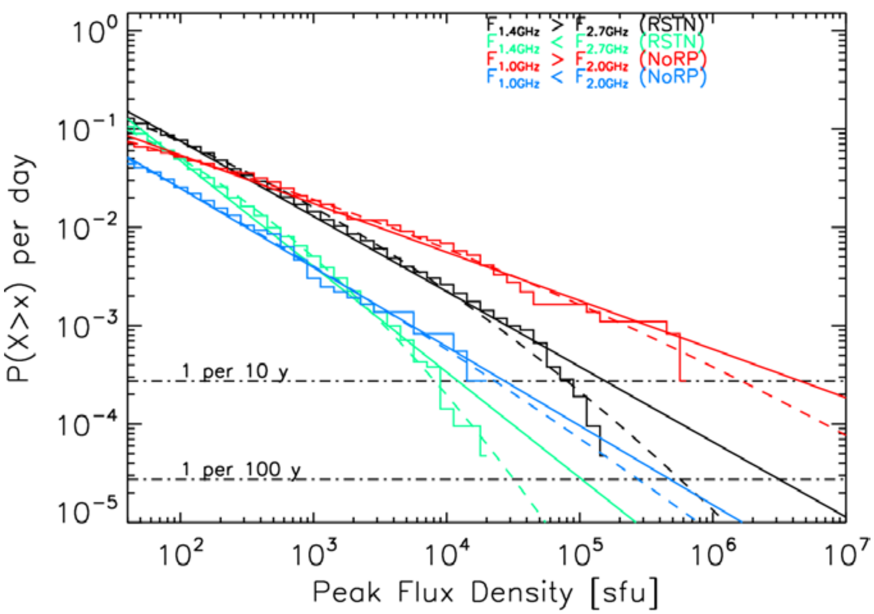
(NoRP), which does measure polarization. The panel has done a preliminary test for selection of coherent bursts as follows:

Phase 2 Recommendation
The following methodology may be used to separate coherent from incoherent bursts

1. Select NoRP bursts above 1000 sfu at 1 GHz;
2. Check that they are stronger at 1 GHz than at 2 GHz (i.e., Castelli U-bursts); and
3. Check that such bursts have other characteristics of coherent emission (high degree of polarization, low temporal correlation with the higher-frequency, incoherent emission).

We suggest that this is promising as a relatively robust way of verifying the coherent nature of the emission. Steps 1 and 2, as well as the temporal correlation of step 3, will also work with the more voluminous RSTN data, so we have performed a *preliminary* selection and have split the population into incoherent and coherent bursts based only on steps 1 and 2 (for now). This is shown in Figure 3, which we emphasize is only illustrative and must be done in a more rigorous way, folding in step 3.

Proposed methodology is used to roughly estimate 100-year values



The RSTN and NoRP databases have been tentatively split into coherent bursts (those for which the spectrum rises to lower frequencies—black for RSTN, red for NoRP) and incoherent bursts (those for which the spectrum continues to fall to lower frequencies—green for RSTN and blue for NoRP). Solid lines are power law fits, while dashed lines are log-normal fits to the data.

Figure 3. Preliminary results of extreme coherent and incoherent bursts

Figure 3 shows several trends relevant to the Solar Radio Burst benchmarks for extreme events. First, the distribution of bursts identified as incoherent have significantly steeper slopes than those identified as coherent, so as expected, an extrapolation of the coherent population predicts significantly

Preliminary estimate of 100-year coherent burst value is less than the Phase 1 value (> 1E9 sfu) stronger 1 in 100 year maximum flux densities. Second, the NoRP distributions, although based on fewer events, show trends closer to a power law than the RSTN distributions. The fall-off of larger events in the RSTN database may be a result of saturation of the RSTN receivers. Third, the distributions are well fit by both log-normal (dashed curves) and power law (solid lines) fits, but a statistical approach such as done by Riley and Love (2016) should be performed to both establish errors in the fits and possibly to determine whether one type of fit is statistically better. While this preliminary methodology is promising, additional research is required to establish the benchmark values and to reduce attendant uncertainty.

Gap C
Additional research needed on coherent burst methodology

The above approach is dramatically limited by the lack of available measurements, which are incomplete in the following ways:

Gap D
Lack of available measurement data relevant for benchmarks

1. The data are heavily weighted to RSTN measurements, which lack polarization information and also saturate for bursts greater than 100,000 sfu;

Gap D.1
Lack polarization and suffer saturation

2. The data for both NoRP and RSTN are limited to a small number of discrete frequencies, whereas the extreme flux densities are known to occur in narrow (and unpredictable) frequency ranges demonstrated to be missed in the monitored frequencies; and

Gap D.2
Monitored frequencies are too limited to reliably measure extreme flux events

3. In certain cases the reported flux densities in the database are lower than reported by other well-calibrated instruments, partly due to the saturation effects.

Gap D.3
Discrepancies exist between flux density data and measurements from other instruments

The authors of the Phase 1 document identified another methodological gap: a simple extrapolation of the power law distribution of burst flux densities may not properly capture the extreme tail of the distribution. This method may not be appropriate to estimate 1-in-100 year benchmark values and provides no insight into theoretical maximum values. Other techniques of extreme value analysis based on the latest research (Riley and Love 2016) may provide more robust values as well as better uncertainties.

Gap E
Simple power law distribution may be an incorrect representation of extreme burst densities

Finally, although we understand that extreme solar radio burst events, at least in the VHF and UHF bands, are due to coherent emission processes, the specific reason for the extreme flux densities reached by these bursts is largely unknown, due mainly to lack of imaging information about the sources and spatial evolution of these events. This knowledge gap makes the establishment of theoretical maximum benchmarks highly uncertain.

Gap F
Lack the solar imaging data necessary to understand the cause of extreme flux densities

Quite generally, extreme flux densities are the product of two key parameters: (1) the brightness temperature (intensity) that is characteristic of the

Extreme flux densities depend on the size of the emission area and the intensity of the emission

emission mechanism and (2) the source area over which the emission is simultaneously occurring. A simplified expression for the flux density in solar flux units is

$$F_I = F_{RCP} + F_{LCP} \simeq 12\, f^{\,2}_{GHz}\, T_7\, A_{20} \text{ [sfu]}$$

where f_{GHz} is the frequency in GHz, T_7 is the effective temperature T_{eff} in units of 10^7 K, A_{20} is the source area in units of 10^{20} cm^2, and we have explicitly noted that the flux density in Stokes I is the sum of the flux density in right and left circular polarization. The source area is bounded by natural limits of the Sun (e.g. the linear scale of an emitting active region, of order 3 arcmin, gives $A_{20} \simeq 170$), so that the brightness temperature can be seen as the dominant uncertainty. However, the only way to ascertain the brightness temperature is to spatially resolve the source of the emission.

Need instruments to spatially resolve the emission's source area and brightness

It may be possible to bound the brightness temperature through a better theoretical understanding of the emission mechanism(s) involved, which may be different between bursts, or in different frequencies within a given burst. For example, the dominant contribution to the gyrosynchrotron (incoherent) optical depth comes from electrons with a kinetic energy not exceeding their rest energy—thus, $T_{\mathrm{eff}} < 5\ 10^9$ K. This value sets a reliable upper limit on the theoretical brightness of incoherent emission processes on the Sun. For coherent bursts, on the other hand, $T_b = 10^{17}$ K appears to be a general upper bound. When these limits in brightness temperature are accepted, that leaves only the upper bound of the source area as the major unknown. It should be recognized, however, that for extreme brightness temperatures like $T_b = 10^{17}$ K, even a small source area yields extreme flux densities, and a new question arises asking for the likely limits of source area for coherent emission mechanisms.

D. Near-term Recommendations

Based on the above tentative results, we have several suggestions for additional research that can be done in the near term to better establish both the benchmark values and the uncertainties.

Near-term 1: Repeat separation of bursts for higher fidelity benchmarks

Addresses Gap B
Benchmarks made no distinction between coherent and incoherent bursts in VHF and UHF bands

And Gap C
Additional research needed on coherent burst methodology

More carefully repeat the above experiment to separate the burst populations into coherent and incoherent bursts. In particular, incorporate step 3 above, based on the temporal and polarization properties of the bursts. This means going beyond the NOAA database, which lists only peak flux density, and examining the light-curve data for each individual burst.

While looking at individual burst light curves, check for possible saturation effects, especially for bursts seen with both NoRP and RSTN. Some quantitative assessment of the effects of saturation will be important to include, if possible.

After separate populations of coherent and incoherent bursts are obtained, apply the statistical fits of both linear and log-normal functions to subsets of the data derived from the original populations, and do the statistical tests following Riley and Love (2016) to better establish uncertainties; possibly allow a choice between power law and log-normal functions based on the statistical properties of the fits.

E. Long-term Recommendations

The panel has two recommendations for the longer-term data acquisition and research that is essential to improve the benchmark values and their uncertainties, both for the 1-in-100 year benchmarks and for the theoretical maximum, which was not attempted by the Phase 1 study.

Long-term 1: Improve operational infrastructure for solar radio burst monitoring

Addresses Gap D
Lack of available measurement data relevant for benchmarks

To build an adequate database for extrapolation, an urgent national need is to enhance the monitoring capabilities for solar radio bursts with the following key features:

- Worldwide (24/7) coverage of the Sun (3 or 4 stations around the globe) with rapid (near-real-time) availability of the data, (the latter not specifically for the purpose of setting benchmarks, but for meeting the goals of risk mitigation and thresholds for action);
- Broad frequency coverage (20 MHz–20 GHz) is needed to address the full range of vulnerabilities of current wireless communication and navigation systems. Future systems may require extending the high-frequency limit to 35 or even 50 GHz;

Addresses Gap D.2
Monitored frequencies are too limited to reliably measure extreme flux events

- Continuous frequency coverage over the above range with at least 5% frequency resolution ($\Delta\nu/\nu = 0.05$);
- Dual circular polarization, for assessment of impact on GNSS, whose signals are right-hand circularly polarized;

Addresses Gap D.1
Lack polarization and saturate too quickly

- Large dynamic range (>1,000,000 sfu) to minimize saturation effects; and

- Implementation of a regularly executed absolute and cross-calibration plan.

Addresses Gap D.3
Discrepancies exist between flux density data and measurements from other instruments

This added capability will, over one or two solar cycles, create the database needed to properly determine the true distribution of events and enable the more robust extrapolation to firmly establish the benchmarks and uncertainties. It will also provide the real-time monitoring to warn system operators and users of events in progress. With current technology, a system with these capabilities need not be expensive. Development of a prototype would likely cost a few million dollars, while cloning the system for additional stations would be much cheaper. The cost of operating the worldwide system should not exceed the current cost of operating RSTN.

Long-term 2: Fund solar observational platforms necessary for basic research into extreme solar radio burst causes and characteristics

Addresses Gap F
Lack the solar imaging data necessary to understand the cause of extreme flux densities

And Gap A
No theoretical maximum benchmark

As previously mentioned, the reasons for extreme flux densities during solar radio bursts are unknown, mainly due to a lack of imaging information about the sources and spatial evolution of these events. To address the need for additional understanding of solar radio burst causes and characteristics, additional research and associated observation platforms are needed. For instance, to make progress on theoretical maximum benchmark values requires bounding the brightness temperature and source area; to generate these bounds, researchers need simultaneous high spatial, spectral, and temporal resolution observations (broadband imaging spectropolarimetry).

Need high spatial, spectral, and temporal resolution of the Sun

Investments in research with existing instruments such as the solar-dedicated Expanded Owens Valley Solar Array (EOVSA), and non-solar-dedicated instruments such as the Jansky Very Large Array (VLA) and Low Frequency Array (LOFAR) can begin to provide the needed knowledge on topics such as:

Existing instruments can provide some of the needed knowledge

- Identification of emission mechanism(s);
- Source location, bounds on source size, and source structure;
- Magnetic topology; and
- Estimates of brightness temperature and polarization of bursts.

However, none of the existing instruments has the combination of frequency range, frequency resolution, time resolution, and image quality needed to place an upper bound on source area. This is because individual bursts due to the most likely coherent mechanism, the Electron-Cyclotron Maser (ECM), have a characteristic lifetime of 5–10 ms in the UHF band and frequency

No existing instrument can produce data needed to estimate a theoretical maximum flux density value bandwidths of only a few percent, in a typical source size of only 100 km. Observing the mechanism requires a national, solar-dedicated facility, such as the Frequency Agile Solar Radiotelescope (FASR), which has received strong recommendations in several decadal surveys.

The Frequency Agile Solar Radiotelescope would provide the needed capability

References

Bruno, A., Bazilevskaya, G. A., Boezio, M., Christian, E. R., de Nolfo, G. A., Martucci, M., and Ryan, J. M. 2018. "Solar energetic particle events observed by the PAMELA mission." *The Astrophysical Journal*, 862(2), 97.

Cerruti, A. P., Kintner, P. M., Gary, D. E., Mannucci, A. J., Meyer, R. F., Doherty, P. and Coster, A. J. 2008. "Effect of intense December 2006 solar radio bursts on GPS receivers." *Space Weather*, 6, S10D07. doi:10.1029/2007SW000375.

Guidice, D. A. and Castelli, J. P. 1975. "Spectral Distributions of Microwave Bursts." *Solar Physics* 44, 155. doi:10.1007/BF00156853.

Marqué, C., Klein, K.-L., Monstein, C., Opgenoorth, H., Pulkkinen, A., Buchert, S., Krucker, S., Van Hoof, R. and Thulesen, P. 2018. "Solar radio emission as a disturbance of aeronautical radionavigation." *Journal of Space Weather and Space Climate*, 8, A42. doi: 10.1051/swsc/2018029.

Nita, G. M., Gary, D. E., Lanzerotti, L. J. and Thomson, D. J. 2002. "The Peak Flux Distribution of Solar Radio Bursts." *Astrophysical Journal*, 570, 423. doi:10.1086/339577.

Riley, P., and J. J. Love. 2016. "Extreme geomagnetic storms: Probabilistic forecasts and their uncertainties." *Space Weather*, 15, doi:10.1002/2016SW001470.

6. Upper Atmosphere Expansion

The Phase 1 Upper Atmosphere Expansion benchmarks describe changes in atmospheric neutral mass density from extreme space weather events. Such changes in neutral density in turn affect the drag on LEO satellites, serving as the primary means by which space weather influences satellite dynamics. This poses two distinct hazards to operational spacecraft: (1) the direct effect of enhanced drag on the spacecraft, changing its orbit, increasing the uncertainty of its position, and reducing its orbital lifetime; and (2) the indirect effect of atmospheric expansion on the ability to monitor the trajectories of debris, including objects with high area-to-mass ratios, for collision avoidance. The first hazard primarily affects small satellites at lower altitudes (below ~400 km), while the second hazard is pronounced in the heavily populated 600–1000 km region. These hazards may become more acute with emerging commercial initiatives to establish large constellations (thousands) of small communications satellites, as well as the growing use of CubeSats for commercial and research purposes. The benchmarks characterize the range of neutral density variation that can affect satellite operations, critical for both orbital prediction/tracking and collision avoidance.

Increases in upper atmosphere density can reduce the lifetime of satellites and complicate space situational awareness

The Phase 1 team investigated three drivers for neutral density changes: solar extreme ultraviolet (EUV) and far ultraviolet radiation (FUV); solar EUV radiation enhancement during solar flares; and coronal mass ejections (CMEs) driving geomagnetic storms (GMDs). For each of these three drivers, the Phase 1 benchmarks provided estimates for 1-in-100 year events and theoretical maximum values at a few representative altitudes.

Neutral density is driven by three factors: extreme radiation, radiation enhancements of solar flares, and GMDs

A. Quantity Assessment

The current benchmarks in general are well-aligned with the objectives of the Phase 1 document. While all three types of space weather events for which benchmarks are calculated are important, we note that the impact of solar flares and enhanced EUV/FUV is less than for geomagnetic storms. A similar benchmarking effort in the UK also focuses more strongly on geomagnetic storm impacts. Although we do not currently recommend refining the list of drivers, we suggest further discussion with users to determine if any can be prioritized.

Phase 2 recommendation
GMDs appear to be the dominant driver of neutral density

Solar flare and geomagnetic storm impacts are only shown at 400 km, and this should be extended to other altitudes. This could be easily expanded for the geomagnetic storm benchmarks by mining the existing simulations for the 1-in-100-year event, with the caveat that the upper boundary of the physics-based model will be below the upper region of interest (i.e., 850 km) during quiet times. Le et al. (2016) used a physics-based model to simulate the effect of an extreme (X40) flare and showed that the change in neutral density was around 100% at 400 km, but smaller below (around 20% at 200 km) and greater above (around 200% at 600 km). This indicates the different impact of flares on density at different altitudes and could be used to guide the production of a revised benchmark.

Gap A
Benchmarks of solar flare and GMD impacts only at 400km

The metrics used for relative density changes are inconsistent and confusing. Specifically, the relative density changes associated with an extreme geomagnetic storm are with respect to a *baseline storm* (Halloween or Bastille Day storm), but relative density changes for flares and EUV/FUV changes are with respect to a *quiet time baseline*. Therefore, the Phase-1 CME benchmark of 400% should be read as "the density during the 1-in-100-year event at 400 km is expected to be 400% larger than the density during the 2003 Halloween storm." Given that the 2003 Halloween storm is itself an extreme event where densities rose by a factor of more than 3, the 1-in-100-year storm increases from pre-storm conditions by a factor of 15 or more. Similarly, reported observed storm-time increase relative to quiet time background is up to 750% (Sutton et al. 2005; Krauss et al. 2015; 2018). The revised report should modify the geomagnetic storm-based density changes to also be with respect to quiet time conditions.

Gap B
Relative density changes are measured with respect to inconsistent baselines

Phase 2 Recommendation
GMD driven benchmark should use a quiet time baseline

User feedback indicated that density changes should be expressed in absolute as well as in relative terms. This is important since quiet-time density varies by around an order of magnitude between solar minimum and solar maximum (at 400 km). Table 1 (a) modifies the EUV benchmarks shown in Table 1 of the Phase 1 report by including absolute density values. Absolute densities should also be reported for the solar flare and CME benchmarks.

Phase 2 Recommendation
Density changes should also be expressed in absolute terms

Table 1. Benchmarks for global mean neutral density response to solar EUV on > 1-day timescales

Altitude	**Baseline global average density** (daily F10.7 of 240 sfu with an 81-day mean of 200 sfu)	**100-Year Benchmark** (absolute density / percent increase in response to a daily F10.7 of 390 sfu with an 81-day mean of 280 sfu)	**Theoretical Maximum** (absolute density / percent increase in response to a daily F10.7 of 500 sfu with an 81-day mean of 390 sfu)
250 km	1.06E-10 kg/m^3	1.59E-10 kg/m^3 / 50%	2.17E-10 kg/m^3 / 105%
400 km	7.79E-12 kg/m^3	1.50E-11 kg/m^3 / 93%	2.06E-11 kg/m^3 / 165%
850 km	2.69E-14 kg/m^3	8.09E-14 kg/m^3 / 200%	1.06E-13 kg/m^3 / 296%

The composition (in particular, atomic oxygen) is not included as a benchmark. Fluences of atomic oxygen can lead to degradation of satellite materials (Avcu and Celik 2003; Hooshangi et al. 2016), including sensitive instruments (Green et al. 2012). At solar maximum, atomic oxygen is the dominant atmospheric species between ~200 and 750 km (Emmert 2015). Expansion caused by a large EUV enhancement or an extreme CME could lift large amounts of atomic oxygen to higher altitude. Although the adverse effects of atomic oxygen exposure are generally seen as cumulative, a prolonged extreme event could produce a fluence that is a significant fraction of what would normally be expected over a satellite's lifetime. Besides potentially shortening operational lifetimes, this could also complicate the calibration of affected instruments. We recommend that Phase 2 consider the development of a separate atomic oxygen benchmark.

Gap C
Missing composition as a benchmark

Phase 2 Recommendation
Consider a new benchmark for atomic oxygen density

We also identify the following relatively minor gaps with the benchmark quantities. First, benchmark values calculated with respect to quiet conditions may not be sufficient. A potential gap is that benchmarks are presented only for one level of solar activity, rather than a range typical of a solar cycle. We recommend that solar minimum and solar moderate baselines should also be considered for Phase 2. Next, as expanded on in the methodology section below, it was also not possible to provide density changes for a theoretical maximum geomagnetic storm. Finally, while details on the event duration appear to some degree within the Phase 1 text, duration should be made clearer in any revised report to aid the utility of the report for satellite operators and others.

Phase 2 Recommendation
Include benchmarks with respect to solar minimum and solar moderate baselines as well

Phase 2 Recommendation
Include more detail on duration to air the usability of the benchmarks

B. Value Assessment

The benchmark values are generally reasonable and up-to-date. The reported changes in relative density appear to be fairly reasonable. Estimates of uncertainty are also reasonable, although for some cases the estimates are solely based on the expertise of the modelers involved; including other models (e.g., DTM in addition to NRLMSISE-00) would help better characterize uncertainties.

We note that the Phase 1 study does not include new work on extreme solar flares, which may inform the benchmark value for neutral density enhancement caused by solar flares. A recent study by Tschernitz et al. (2018) has suggested X80 is the theoretical maximum flare, while Tsiftsi and De la Luz (2018) used extreme value analysis to indicate a 1-in-100 year maximum of X30.8, with confidence limits between X21.3 and X72.1. The latter study also reported a 1 in 150 year maximum flare of X33.6, with confidence limits of X22.7 and X87.5.

Phase 2 Recommendation
Include new work on extreme solar flares

C. Methodology Assessment

The methodology is generally up-to-date, rigorous, and compelling. Phase 1 places less emphasis on observations compared to models. Many of the benchmarks were calculated using empirical models (NRLMSISE-00) and physics-based models (CTIPe). By contrast, observed changes in neutral density were only reported with respect to extreme solar flare impacts. This is because of issues with the relatively short duration of the observational dataset, biases between different observations, and the ability of the models to better provide a self-consistent representation of the state of the thermosphere. The revised report should contain an extended discussion on the issues with and usefulness of the observations, as well as a discussion of the relative merits of empirical and physics-based models. It should also be made clear that physics-based models can fail for extreme (or theoretical maximum) scenarios.

Phase 2 Recommendation
Include an extended discussion on issues with observations and their usefulness

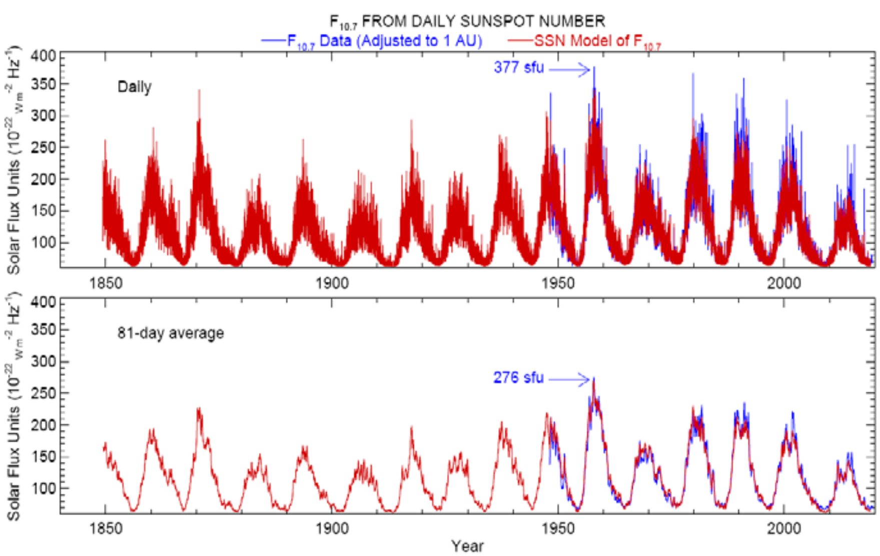
The Weimer model (for driving the high-latitude electric field) and auroral precipitation patterns (e.g., TIROS/NOAA) used in physics-based models are based on observations and therefore are not likely to be credible theoretical maximum events. In addition, the Phase 1 team used an assumption of how Joule heating saturates in the model simulations to inform 1-in-100 year events. These assumptions were made based on one type of saturation and could be improved.

The theoretical maximum F10.7 values used in the Phase 1 study (daily 500 sfu, 81-day average 390 sfu) were assumed in the absence of any literature on the topic. Refinement of this assumed theoretical maximum (in terms of both

F10.7 and the estimated EUV/FUV irradiance) will require input from the solar physics community.

Phase 2 Recommendation
Work with the solar physics community to refine the F10.7 theoretical maximum values

For the Phase 1 EUV/FUV benchmarks, the maximum daily and 81-day average F10.7 values (at 1 AU) were retrieved from the historical record (which extends back to 1947). These values are respectively 377 sfu (23 Dec 1957) and 276 sfu (21 Nov 1957). Adjusted to the nearest Earth-Sun distance, the historical maxima are 390 sfu/285 sfu. Figure 4 shows an extension of the F10.7 record back to 1850 by regressing F10.7 against the sunspot number record. Although there is some uncertainty in this approach, it seems plausible that the extreme values that occurred in 1957 represent a maximum over the past 170 years and therefore provide an upper-bound estimate of a 100-year event. The historical reconstruction of EUV irradiance by Lean et al. (2011, Figure 10) suggests that the 1957 event was a 400-year maximum. However, a more rigorous statistical analysis is needed to better define a 100-year event. Additionally, there is considerable uncertainty in the reliability of F10.7 as an EUV proxy at these high F10.7 levels.



The red curves show a model of F10.7 in terms of the sunspot number, using the formulation $F_{10.7}^{Fit} = a + b \cdot \langle SSN \rangle + c \cdot (SSN - \langle SSN \rangle)$, where SSN is the daily sunspot number and <SSN> is the running annual average sunspot number. The arrows highlight the maximum values in each panel. The F10.7 values (Tapping 2013) were obtained from https://spaceweather.gc.ca/solarflux/sx-en.php. The sunspot numbers are version 2.0 (Clette et al. 2016) from the World Data Center SILSO, ROB, Brussels (http://www.sidc.be/silso/datafiles).

Figure 4. Daily F10.7 (top, blue) and 81-day average F10.7 (bottom, blue).

The correlation between F10.7 and EUV is not linear and not well understood at very high values of EUV due to the small number of concurrent observations. For very high activity, EUV appears to increase at a slower rate

Gap D
F10.7 is not a good proxy for EUV

than F10.7 (Dudok de Wit and Bruinsma 2011). Because the relationship between F10.7 and EUV is not linear, using the F10.7 index as an input argument to describe the response to changes in solar EUV/FUV irradiance, as is done in most empirical models, results in potential errors (Viereck et al. 2001). Like empirical thermosphere models, empirical models of solar spectral irradiance that use F10.7 as an input argument combine daily and 81-day time scales (Richards et al. 1994; Lean et al. 2011); using EUV irradiance data to fit thermospheric density data essentially eliminates the need for the 81-day average term (Emmert 2015, section 4.2).

The advantage of using F10.7 and other solar radio flux indices (e.g., DTM uses F30 [Bruinsma et al. 2015]) is that the radio flux measurements are well calibrated, continuously available, and span the entire historical record of thermospheric measurements. For the purpose of benchmarking extreme EUV events, however, F10.7 presents two disadvantages. First, it is unlikely that the combinations of the daily and 81-day average F10.7 derived to fit the historical thermospheric data are accurate for more extreme conditions, and they may not even be accurate for the very high solar activity conditions in the existing record. Second, the need to estimate two benchmark input parameters (daily and 81-day average values) adds a second, non-physical, degree of freedom and complicates the definition of the upper atmospheric expansion benchmarks.

Ideally, the upper atmospheric expansion EUV benchmark would be defined in terms of the total irradiance in that spectral band, on a time scale of ~1 day. However, measurements of EUV irradiance have a limited temporal span (~33 years) and suffer from instrument drift and biases among instruments. To address that challenge, a robust EUV composite or a longer-term reconstruction would provide a direct basis for estimating a solar EUV benchmark that could be used to derive atmospheric expansion benchmarks. Some EUV composites and reconstructions have already been developed (Haberreiter et al. 2017; Lean et al. 2011), but not for the purpose of estimating EUV 100-year events or the EUV theoretical maximum. With an accurate EUV composite or reconstruction, the empirical approach to estimating upper atmospheric expansion benchmarks could be applied and extrapolated more robustly. The empirical models may need to be re-fit using the new solar inputs.

Gap E
EUV measurements are currently limited

Phase 2 Recommendation
Develop a EUV composite or reconstruction to better estimate a EUV benchmark

For the phase 1 EUV Flare benchmarks, the 1-in-100-year and theoretical maximum flares were respectively assumed to be X30 and X40 in the absence of a strong consensus from the literature. While the continuous record of X-ray flux monitoring in the 1-8 A wavelength band from GOES and SMS satellites dates back to 1974, saturation effects constrain the record to flares of around X17 or less (e.g., the 4 November 2003 flare saturated the GOES detector at

Gap F
The maximum flares for the EUV benchmarks were assumed at a high degree of uncertainty

X17.4, but was estimated to be X28). Several subsequent authors have argued that the actual flare magnitude could have been as large as X35–45 (Thomson et al. 2004; Woods et al. 2004; Brodrick et al. 2005). An additional complication is caused by the unknown correlation between X-ray and EUV/FUV enhancements during a flare, the latter being more directly related to the resultant thermospheric expansion. Considering the inherent unknowns in this process, the values of X30 and X40 seem reasonable at this time. However, assuming these values incurs large uncertainties for the benchmark.

For the phase 1 CME benchmarks, only the 1-in-100-year benchmark was defined and only at 400 km altitude. Solar wind conditions as estimated for the 1859 Carrington event by Li et al. (2006) along with the Weimer convection and expected saturation models were used to arrive at a peak global Joule heating rate of 10,000 GW. These values were used as input to the CTIPe model, with baseline conditions equivalent to those just prior to the December 2006 "AGU" storm. Given the difficulty in constructing the 1-in-100-year case, the theoretical maximum CME event is unfeasible without making a series of unfounded assumptions, and it was not calculated. There is no more information now on how to represent this than when the Phase 1 document was written. Most theoretical models use storm-time drivers that are parameterized using existing observations. Without knowing how these drivers respond to the theoretical worst event, it is extremely challenging to estimate thermosphere change. NASA has sponsored a focused science topic (FST) to improve magnetosphere-ionosphere-thermosphere coupled models to better simulate extreme events, the outcomes of which should be directly relevant to future benchmarks.

Gap G
It is currently unfeasible to calculate a theoretical maximum CME event

The Kp index was used to evaluate the phase 1 benchmarks, but Kp (ap) is limited to 9 (400 2nT) and cannot accurately characterize extreme events. Ideally, the representation of extreme geomagnetic forcing of the thermosphere would be done using coupled Sun-to-Earth physics-based models. However, such a coupled model will take a long time to develop. Instead, the assessment used empirical or standalone ionosphere-thermosphere physics-based models that characterize geomagnetic forcing using an index, very often Kp or the related ap. For 1-in-100 year extreme events, an advantage of the Kp record is that it extends longer than most other indices (it extends back to the 1930s). In addition, the Kp index is endorsed by the International Association of Geomagnetism and Aeronomy (IAGA). At our request, the provider of the Kp index, GFZ, has started research on extending the range of the index. First results indicate that a few major past storms, such as the 2003 Halloween storm, would be upgraded to a Kp=10 storm. Updating its definition in order to reach a higher maximum will require preparatory work followed by a test period, and in the

Gap H
Kp index is limited and cannot accurately characterize extreme events

case of success, adoption by IAGA. Some have argued that Kp is not the most appropriate way of representing the geomagnetic forcing of the thermosphere and have proposed alternative indices (Chambodut et al. 2015).

The phase 1 study focuses on the response of the neutral mass density. Above ~500 km altitude, the plasma density of the ionosphere can be a significant fraction of the neutral mass density. NRLMSISE-00 includes an anomalous oxygen component that nominally represents the contribution of atomic oxygen ions and hot neutral atomic oxygen to observed mass densities derived from satellite drag. However, Capon et al. (2019) found that the MSIS anomalous oxygen densities are much smaller than ion densities in IRI-2016 (Bilitza et al. 2017). Furthermore, the physics of the drag interaction between spacecraft (and debris) and ions is different from the interaction between spacecraft and neutrals, due to spacecraft charging. Capon et al. (2019) reviewed existing work and conducted new simulations that showed that the electrodynamic contribution can significantly increase ionospheric drag. A better understanding of the separate responses of the ionosphere and neutral thermosphere to extreme events and their contribution to satellite drag would help reduce uncertainty in the benchmark values.

Gap I
Insufficient understanding of separate responses of the ionosphere and neutral thermosphere

All upper atmospheric expansion benchmarks require better characterization of uncertainty. A 100% uncertainty was reported for all the theoretical maximum benchmarks. For the benchmarks that are based on empirical models, this uncertainty estimate was based on the expertise of the model developers, indicating that the theoretical maximum events are not included in the training data used to develop these models. We may want to consider how to better quantify this uncertainty, possibly via use of extreme event analysis. The 1-in-100 year extreme events are closer to the observed range of the training data, and a better assessment of the uncertainty in these benchmarks may come from running benchmark estimates with different empirical models. Such models produce a range of results (Bruinsma et al. 2017; 2018), so a multi-model estimate of the benchmarks will make a useful contribution to the uncertainty estimates. We recommend that the benchmark simulations should be rerun with multiple models. This is easy for DTM, but harder for JB. We should also examine flux for extreme cases.

Gap J
All benchmarks require better characterization of uncertainty

Phase 2 Recommendation
Rerun benchmark simulations with multiple models

We need to ensure that the extreme case parameter values (e.g., maximum solar flares) are consistent with those chosen by other groups.

Phase 2 Recommendation
Ensure that parameter values are consistent with those of the other benchmarks

D. Near-term Recommendations

Near-term 1: Expand scope of solar flare and geomagnetic storm benchmarks

Addresses Gap A
Benchmarks of solar flare and GMD impacts only at 400km

There is a need to expand the benchmarks for flare and geomagnetic storm response at 250 and 850 km and at very low and high solar activity. The Phase 1 report only reported thermosphere expansion caused by large CMEs at 400 km altitude. The response at 250 km is readily available as part of the simulations carried out for the Phase 1 report. However, a difficulty arises at 850 km since the physical models needed to assess the response to CMEs have upper boundaries near the exobase. While the exobase will be above 850 km in extreme storms, the quiet-time simulation will have to be extrapolated in order to establish a baseline for the event. Accurate extrapolation will require the inclusion, either self-consistently (in the long-term) or in an ad hoc manner (in the short-term), of neutral helium. Proper treatment of this lighter species will further improve the percentage increase metric accuracy, but will not affect the reported absolute values.

New index simulations for flare responses should be run to expand altitudinal coverage (i.e., 250 and 850 km) and solar activity baselines of the benchmarks. The Phase 1 benchmark relied on empirical evidence from satellite observations made near an altitude of 400 km during the 2003 Halloween flares, and was therefore limited to the sampling and prevailing conditions during that event. Estimating the response at different altitudes and over varying phases of the solar cycle will require a physics-based modeling approach. Le et al. (2016) explore this and can be used as a starting point.

Near-term 2: Conduct a research review to standardize benchmarks

Addresses Gap B
Relative density changes are measured with respect to inconsistent baselines

Conduct a literature review to standardize tables, baselines (e.g., Halloween storms), timing (e.g., daytime), and type of value (e.g., global average). Provide both percentage increase relative to baseline values and absolute values of the neutral density to contextualize table data. For future CME benchmarks, impacts should be reported with a baseline from quiet conditions instead of from a Bastille/Halloween storm.

Near-term 3: Analyze neutral density response to past storms

Addresses Gap G
It is currently unfeasible to calculate a theoretical maximum CME event

Conduct an analysis of the response of the neutral density to significant past storms. Some analysis has already been done. For example, Sutton et al. (2005) used CHAMP observations during the October 2003 geomagnetic storm, and Krauss et al. (2015; 2018) used GRACE and CHAMP observations from 2003–2015, and reported largest density enhancements (at 490 km) of up to 750% (relative) and up to 4 x 10–12 kg m-3 (absolute). The latter also indicated that the impact of co-rotating interaction regions (CIRs) on density is similar to that from weaker CMEs, which may be of relevance to satellite operators during solar minimum, when CMEs are rare but CIRs are common.

And Gap J
All benchmarks require better characterization of uncertainty

There are limited data available, but reprocessing of accelerometer data to make the inferred densities from the various satellites self-consistent will enable a more comprehensive analysis of observed storm-time effects. In the longer-term, we need to conduct more observations in order to get a base for a statistical event.

Near-term 4: Examine forcing and saturation in physics-based models

Addresses Gap G
It is currently unfeasible to calculate a theoretical maximum CME event

Examine forcing and responding saturation in physics-based models. The cross-polar cap potential (CPCP) associated with ionospheric convection varies roughly linearly with the solar wind electric field for nominal conditions but saturates or asymptotes to a constant value of ~ 200 kV for large electric field during magnetic storms, which is related to the response of the magnetosphere to extreme conditions in the solar wind. Various interpretations of the saturation phenomenon have been offered and are fully discussed in Siscoe et al. (2004), Ridley et al. (2005), Shepherd et al. (2007), and Kivelson et al. (2008). Most mechanisms entail a role for the region 1 current system and reflection of Alfven waves. Since CPCP is strongly related to the total geomagnetic energy deposited into the upper atmosphere, the saturation of CPCP indicates the saturation of geomagnetic energy deposition.

E. Long-term Recommendations

Long-term 1: Develop an EUV observation system

Addresses Gap D
F10.7 is not a good proxy for EUV

And Gap E
EUV measurements are currently limited

Develop a robust EUV observation system. EUV observations must be calibrated accurately for each mission and across different missions. This is needed but presently not achieved (Vourlidas and Bruinsma 2018). This would provide higher quality inputs to models.

Long-term 2: Quantify impacts to the magnetosphere from extreme events

Addresses Gap I
Insufficient understanding of separate responses of the ionosphere and neutral thermosphere

Quantify the impact of extreme events on the magnetosphere through physics-based model simulations and analysis of magnetospheric saturation trends that occurred in past storms. The response of the magnetosphere to an extreme event is poorly known and may modulate the energy flow from the solar wind into the upper atmosphere and the consequent magnitude of the Joule and auroral heating rates. Specifically we should (1) exploit understanding of the magnetosphere (saturation, mass loading), and thermosphere cooling to improve modeling of the neutral density (and wind) response to extreme events; and (2) leverage activities of NASA FST to provide additional information about extreme events.

And Gap J
All benchmarks require better characterization of uncertainty

Long-term 3: Update the Kp index to accommodate more extreme values

Addresses Gap H
Kp index is limited and cannot accurately characterize extreme events

Geomagnetic Kp Index definition should be updated to accommodate more extreme values. GFZ has recently started investigating how to implement an updated calculation. The index update must be proposed to IAGA at the earliest opportunity.

Long-term 4: Account for contributions of ion density

Addresses Gap I
Insufficient understanding of separate responses of the ionosphere and neutral thermosphere

Account for contributions of ion density to satellite drag at higher altitudes. This contributes to the general aim of improving benchmarks and reducing uncertainty.

Long-term 5: Conduct research on change in atomic oxygen density

Addresses Gap C
Missing composition as a benchmark

Engage the satellite user community to assess the hazard of atomic oxygen fluence during extreme events. If it is determined to be a credible hazard, conduct research on the change in atomic oxygen density in response to extreme space weather events and consider the development of a separate atomic oxygen benchmark.

References

Avcu, S., and Celik, B. 2003. "Structural material selection and processing for low Earth orbit spacecraft regarding atomic oxygen effects." *International Conference on Recent Advances in Space Technologies*, 2003. RAST '03.

Bilitza, D., D. Altadill, V. Truhlik, V. Shubin, I. Galkin, B. Reinisch, and X. Huang. 2017. "International Reference Ionosphere 2016: From ionospheric climate to real-time weather predictions." *Space Weather*, 15. doi:10.1002/2016SW001593

Brodrick, D., S. Tingay, and M. Wieringa. 2005. "X-ray magnitude of the 4 November 2003 solar flare inferred from the ionospheric attenuation of the galactic radio background." *J. Geophys. Res*.,110, A09S36, doi:10.1029/2004JA010960.

Bruinsma, S. L. 2015. "The DTM-2013 thermosphere model." *J. Space Weather Space Clim.*, 5. A1, doi:10.1051/swsc/2015001

Bruinsma, S., E. Sutton, S.C. Solomon, T. Fuller-Rowell, M. Fedrizzi. 2018. "Space Weather Modeling Capabilities Assessment: Neutral Density and Orbit Determination at LEO." *Space Weather*, doi.org/10.1029/2018SW002027

Bruinsma, S.L., D. Arnold, A. Jäggi, N. Sánchez-Ortiz. 2017. "Semi-empirical thermosphere model evaluation at low altitude with GOCE densities." *Journal of Space Weather and Space Climate*, http://dx.doi.org/10.1051/swsc/2017003

Capon, C. J., B. Smith, M. Brown, R. Abay, and R. R. Boyce. 2019. "Effect of ionospheric drag on atmospheric density estimation and orbit prediction." *Adv. Space Res.*, 63. 2495-2505, https://doi.org/10.1016/j.asr.2019.01.013

Chambodut A, Marchaudon A, Lathuillère C, Menvielle M, Foucault E. 2015. "New hemispheric geomagnetic indices with 15 min time resolution." *J Geophys Res Space Phys*, 120, 9943–9958, DOI: 10.1002/2015JA021479

Clette, F., Cliver, E.W., Lefèvre, L, Svalgaard, L., Vaquero, J.M., Leibacher, J.W. 2016. "Preface to Topical Issue: Recalibration of the Sunspot Number." *Solar Phys*., 291, DOI: 10.1007/s11207-016-1017-8

Dudok de Wit, T., S. Bruinsma. 2011. "Determination of the most pertinent EUV proxy for use in thermosphere modeling." *Geophys. Res. Lett*., doi:10.1029/2011GL049028

Emmert, J. T. 2015. "Thermospheric mass density: A review." *Adv. Space Res*., 56. 773–824, doi:10.1016/j.asr.2015.05.038

Green, J. C., et al. 2012. "The Cosmic Origins Spectrograph." *ApJ* 744 60, https://doi.org/10.1088/0004-637X/744/1/60

Haberreiter, M., M. Schöll, T. Dudok de Wit, M. Kretzschmar, S. Misios, K. Tourpali, and W. Schmutz. 2017. "A new observational solar irradiance composite." *Journal of Geophysical Research: Space Physics*, 122. doi:10.1002/2016JA023492

Hooshangi, Z., Feghhi, S. A. H., and Saeeedzadeh, R. 2016. "The effects of low earth orbit atomic oxygen on the properties of Polytetrafluoroethylene." *Acta Astronica*, 119, 233-240, doi:10.1016/j.actaastro.2015.11.031

Le, H., L. Liu, Z. Ren, Y. Chen, H. Zhang,and W. Wan. 2016. "A modeling study of global ionospheric and thermospheric responses to extreme solarflare." *J. Geophys. Res. Space Physics*, 121,832–840, doi:10.1002/2015JA021930.

Lean, J. L., T. N. Woods, F. G. Eparvier, R. R. Meier, D. J. Strickland, J. T. Correira, and J. S. Evans. 2011. “Solar extreme ultraviolet irradiance: Present, past, and future.” *J. Geophys Res*., 116. A01102, doi:10.1029/2010JA015901

Li, X., Temerin, M., Tsurutani, B. T., Alex, S. 2006. “Modeling of 1–2 September 1859 super magnetic storm.” *Adv. Space Res*., 38, 273-279, doi:10.1016/j.asr.2005.06.070

Richards, P. G., J. A. Fennelly, and D. G. Torr. 1994. “EUVAC: A solar EUV flux model for aeronomic calculations.” *J. Geophys Res*., 99. 8981–8992

Tapping, K. F. 2013. “The 10.7 cm solar radio flux (F10.7).” *Space Weather*, 11. doi:10.1002/swe.20064

Thomson, N. R., C. J. Rodger, and R. L.Dowden. 2004. “Ionosphere gives size of greatest solar flare.” *Geophys. Res. Lett*.,31, L06803, doi:10.1029/2003GL019345

Viereck, R., L. Puga, D. McMullin, D. Judge, M. Weber, and W. K. Tobiska. 2001. “The Mg II Index: A Proxy for Solar EUV.” *Geophys. Res. Lett*., 28. 1343–1346

Vourlidas, A., S. Bruinsma. 2018. “EUV Irradiance Inputs to Thermospheric Density Models: Open Issues and Path Forward.” *Space Weather*, doi:2017SW001725

Woods, T. N., F. G. Eparvier,J. Fontenla, J. Harder, G. Kopp, W. E. McClintock, G. Rottman, B. Smiley, and M. Snow. 2004. “Solar irradiance variability during the October 2003 solar storm period.” *Geophys. Res. Lett*.,31,L10802, doi:10.1029/2004GL019571.

7. Cross-Cutting Issues and Recommendations

Each of the previous chapters has highlighted issues specific to determining benchmarks for their phenomena. Across the five categories, the panel also identified issues and recommendations relevant to many or even all of the phenomena. In this concluding chapter, we summarize cross-cutting issues and identify how addressing them may improve the space weather benchmarking process as a whole.

A. Issues and Recommendations for Phase 2

1. Updated Data and Models

Across all benchmarks, the panel identified datasets and models that were not included in the Phase 1 effort but that would be useful for developing extreme space weather benchmarks. These resources were either not referenced by the Phase 1 team or have been developed since the benchmarks were released. Similarly, new data and models will continue to be published before and after the Phase 2 benchmarks, even after implementation of the research recommendations in this report. To ensure that benchmark values are useful and up-to-date, the panel recommends that the benchmarks be updated on a regular cadence as appropriate to take advantage of new models and data streams.

Recommendation
Establish a regular cadence for new updating benchmarks with new models and data

2. Extreme Values and Uncertainties

The Phase 1 team attempted to specify the 1-in-100 year and theoretical maximum values for space weather phenomena. Across all benchmark quantities, identifying these extreme values was problematic—if not currently impossible. The limitations on specifying benchmarks are driven by two challenges. The first is limited data. Only a few data sets span more than 50 years, and most less than several decades; extrapolations from such limited data inherently introduce large uncertainties. The second limitation is inadequate physical understanding of the phenomena. Physics-based models might be used to extrapolate beyond observations, but we generally lack sufficient physical understanding of the processes that produce the extremes and, in many cases, existing models cannot yet handle extreme input conditions.

Challenge
Limited data and understanding of phenomena

While specifying the 1-in-100 year and theoretical maximum values is a valid goal, achieving that goal can be problematic. Large uncertainties in the 1-in-100 year or theoretical maximum benchmark values reduce the utility of the values for real world planning and preparedness. Furthermore, some stakeholders expressed the desire for intervals shorter than 100 years.

Challenge
Large benchmark uncertainties reduce their utility

Recommendation
Develop a 1-in-N year approach

Therefore, the panel recommends that future efforts to develop extreme space weather benchmarks consider a 1-in-N year approach. Appropriate values for N might be 10, 20, 30, 50, and 100 years. Several sectors of the user community suggested that a variety of time spans would enable greater use of the benchmarks in a wider variety of applications. Additionally, decreasing the time over which the benchmarks specify maxima decreases uncertainty for specific quantities; however, the panel notes that the only way to truly reduce uncertainties in benchmark values is through new research and observations.

3. Critical Data Sets

Any quantification of extreme values will be limited by the data used to estimate those extremes—whether the data are the benchmark quantities themselves or data used as input to models. Better data sets and longer spans of data collection are critically important for improving the space weather benchmarks in the future. The panel recognizes tradeoffs when determining which data collection methods to prioritize. For example, new data sets do not provide the ease of analysis that is afforded by continued collection of current observations; however, where older data streams are limited, new observations can provide better information and greater utility for the benchmarks. In addressing and coordinating these tradeoffs, the panel recommends that the benchmarks and national preparedness as a whole would benefit from a dedicated data collection plan directed specifically at improving the space weather benchmarks.

Recommendation
Develop a dedicated data collection plan for the benchmarks

In some cases, additional analysis could allow multiple datasets to be combined and utilized to extend the observational time-horizon or increase richness. This approach allows new data streams to be combined with older observations or proxy data sets that are relevant to benchmark quantities. The panel recognizes the value to reducing benchmark uncertainty through combining data sets and recommends that future benchmarking efforts attempt to exploit older and less well-validated data sources to the extent possible.

Recommendation
Use and merge multiple available data sets when updating benchmarks

4. Duration of Events

In most use cases, the hazards posed by extreme space weather events are a combination of intensity of the conditions and how long they persist. For example, a communication/navigation outage lasting 5 minutes has drastically different consequences than one lasting more than a day. Similarly, for ionizing radiation it is often the fluence (intensity integrated over time) that determines the severity of effects. The panel determined that, in many cases, extreme values are less useful if unaccompanied by duration, and thus recommends that further benchmarking captures duration along with intensity of the extreme events.

Recommendation
Capture duration along with intensity in the benchmarks

5. User-Community Engagement

Since the Benchmarks report focuses chiefly on the environment, there is a risk that the end user requirements are not adequately addressed. It is critically important that the benchmarks are actionable to end-users. As an example, user feedback at the NSB Workshop indicated that 1-in-100 year events or theoretical worst cases are not always appropriate to user needs, prompting the recommendation from this panel to introduce a 1-in-N year approach. The panel recognizes that much deeper involvement of the user/stakeholder communities in future benchmark studies could improve existing benchmarks and inform where new benchmarks might be needed. Similarly, a better understanding of how individual benchmarks are used for planning and preparedness would help guide future research priorities. The panel recommends that the Phase 2 benchmarks are refined in close collaboration with organizations that represent and understand their respective users' needs.

Recommendation
Refine Phase 2 benchmarks in close collaboration with users

Engagement with user communities could be facilitated by adopting a more holistic approach, where worst case space weather environments are assessed by a wide range of relevant experts, including scientists, policy-makers, emergency planners, and other end users. The panel notes that all of these communities (including research scientists) need to more fully understand the objectives of the benchmarks, how they should be used, and how they differ from other existing space weather activities such as forecasting or basic research. A series of focused, user-to-scientist/scientist-to-user workshops would help to accelerate future benchmark development. Finally, it is important to identify, interface with, and learn from other parallel benchmark efforts; examples include the geomagnetic disturbance benchmark created by the North American Electric Reliability Corporation and the *Space Weather Worst-Case Environments* produced by the UK Space Environment Impacts Expert Group.

6. Translation of Benchmark Values into System Effects

Some of the environmental parameters benchmarked in the Phase 1 document relate closely to user needs, e.g., induced geo-electric field values measured in volts per kilometer across the continental United States, which are directly useable by operators of the electric power grid. Other environmental parameters are more abstract or require a non-trivial calculation to translate benchmark values into effects on technology systems; examples include suggestions by this panel to use a system independent ionospheric turbulence measure such as C_kL or to organize radiation belt ionizing radiation benchmarks by common orbits. The panel recognizes that the fundamental physical quantities cannot always be used to directly specify engineering requirements. Therefore, the panel recommends that, where needed, the benchmark documentation provide guidelines for engineers and operators to translate the benchmarks into values relevant to their needs.

Recommendation
Consider adding guidelines for translating benchmarks to usable quantities

7. Approaches in Other Fields

Space weather is not the only nor the first field to develop benchmark values for use by operators, engineers, and responders. Other fields have well developed methodologies for categorizing and communicating extreme events, such as the 1-to-5 categorization Saffir-Simpson Hurricane Wind Scale or the Richter Scale for earthquakes. A review of these and other fields could improve the space weather benchmarking process as well as provide a metric to which the space weather hazards can be compared. The panel recommends that benchmarking efforts adopt holistic comparisons to other fields, including a review of applicable methodologies. This approach would be similar to and could build on the work of the United Kingdom Space Environment Impacts Expert Group (SEIEG).

Recommendation
Adopt holistic comparisons to other extreme event fields

B. Recommendations for Future Research

Each focus area in this report makes specific recommendations for research that should be conducted in both the near and long term to improve the space weather benchmarks. The panel has also identified some general, cross-cutting issues.

Firstly, the panel recognizes that improving the space weather benchmarks represents a new direction for the research community and for research funding agencies. The goals are sometimes more closely associated with applied research aimed more at quantification and less at basic physical understanding (although the two are not mutually exclusive). One example is applying advanced statistical analysis techniques to space weather data, including the fields of extreme value analysis and uncertainty quantification. Similar critically important activities have had trouble obtaining funding under traditional research programs. Those include cleaning data sets to remove artifacts such as measurement saturation that can significantly degrade analysis of extremes;

cross-calibration of heterogeneous data sets to increase the span of data available for analysis of extremes; and making data sets publicly available to increase opportunities for analysis. Traditional research programs evaluate proposals based on advancing understanding of fundamental physical processes and phenomena. Proposals directed at developing and quantifying space weather benchmarks may not be a good fit to the current research program objectives. The panel recommends that research funding agencies, such as NASA and NSF, implement new research programs that directly address the unique applied research demands of improving space weather benchmarks.

Recommendation
Funding agencies should implement new programs for benchmark-related research

We also suggest that there are new opportunities to apply physical models to improve the space weather benchmarks. Traditionally, model development has been more focused on greater physical fidelity and/or better forecast accuracy. Models developed specifically for comparison against long-term observations of extreme distributions are not common but could be immensely valuable. Similarly, ensemble modeling specifically directed at extreme conditions could provide important new insights as well as helping extrapolate conditions that have not been observed in the historical data sets; the variability in results from multiple models could also help to estimate uncertainties. The panel suggests that research funding agencies consider research priorities in modeling that advance physical models with the goal of understanding the origin and consequences of extreme space weather conditions.

Recommendation
Funding agencies consider supporting research in modeling extreme space weather

We lack understanding of space climatology, which refers to the long-term trends of solar activity. We know that space weather conditions and extreme events vary from one solar cycle to another, but our understanding of longer-term trends is poor. The unusually low level of solar activity in solar cycle 24 only adds to our uncertainty. Since much of the data used to estimate statistical benchmarks come from the previous few solar cycles, better estimates could be produced by understanding how those data fit into solar activity trends. The panel recommends that the government fund new studies of space climatology that can be leveraged to improve benchmark value estimates.

Recommendation
Government should fund new studies of space climatology

Finally, we note that scientific understanding is lacking in the area of emergent phenomena, which refers to systems that have a tipping point where extreme conditions are produced by fundamentally different processes than those that produce merely intense conditions. In general, we do know yet know which systems contain emergent phenomena that need to be understood before accurate benchmarks can be specified. The panel recommends research to explore how disparate physical processes may interact to produce unexpectedly extreme space weather events.

Recommendation
Research unexpected emergent phenomena in extreme space weather events

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	XX-12-2019
2. REPORT TYPE	FINAL
3. DATES COVERED *(From - To)*	
4. TITLE AND SUBTITLE	Next Step Space Weather Benchmarks
5a. CONTRACT NUMBER	NSFOIA-0408601
5b. GRANT NUMBER	
5c. PROGRAM ELEMENT NUMBER	
5d. PROJECT NUMBER	NF-20-4516
5e. TASK NUMBER	
5f. WORK UNIT NUMBER	
6. AUTHOR(S)	Reeves, Geoffrey Colvin, Thomas J. Locke, Jericho W. et al.
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)	IDA Science and Technology Policy Institute 1899 Pennsylvania Avenue, NW, Suite 520 Washington, DC 20006-3602
8. PERFORMING ORGANIZATION REPORT NUMBER	IDA Group Report NS GR-10982
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	National Science Foundation 2415 Eisenhower Avenue Alexandria, Virginia 22314
10. SPONSOR/MONITOR'S ACRONYM(S)	NSF
11. SPONSOR/MONITOR'S REPORT NUMBER(S)	

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution is unlimited (9 December 2019).

13. SUPPLEMENTARY NOTES

14. ABSTRACT

In June 2018, the Space Weather Operations, Research, and Mitigation (SWORM) Subcommittee of the National Science and Technology Council (NSTC) released the Space Weather Phase 1 Benchmarks. The benchmarks specify the nature and intensity of extreme space-weather events and provide a point of reference from which to improve understanding of their effects. The National Science Foundation (NSF) and the National Aeronautics and Space Administration (NASA) asked the IDA Science and Technology Policy Institute (STPI) to lead the Next Step Space Weather Benchmarks effort to engage the expertise of the U.S. and international space weather scientific community to identify gaps and make recommendations to improve the benchmarks. To produce this report, STPI helped to assemble a panel of 32 space weather experts, draft and release a request for white papers from the international space weather community on space weather benchmarks, organize a workshop that convened the panel and members of the space weather community to identify tentative gaps and recommendations, and organize a Town Hall to share the panel's conclusions. This report contains the final analysis and recommendations of the expert panel.

15. SUBJECT TERMS

Geomagnetic Disturbance (GMD); Induced Geo-Electric Fields; Ionizing Radiation; Ionospheric Disturbances; National Space Weather Strategy and Action Plan (NSWSAP); Resilience; Solar Radio Bursts; Space Weather; Space Weather Operations, Research, and Mitigation (SWORM); Upper Atmosphere Expansion

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON	19b. TELEPHONE NUMBER *(Include area code)*
Unclassified	Unclassified	Unclassified	Same as Report	73	Wiltberger, Michael	703-292-4690

Reset