Role Of the Sun and the Middle atmosphere/thermosphere/ionosphere In Climate (ROSMIC): a retrospective and prospective view

William Ward1,4†, Annika Seppälä2†, Erdal Yiğit3, Takuji Nakamura4, Claudia Stolle5, Jan Laštovička6, Thomas N. Woods7, Yoshihiro Tomikawa4, Franz-Josef Lübken8, Stanley C. Solomon9, Daniel R. Marsh9,10, Bernd Funke11 and Duggirala Pallamraju12

Abstract

While knowledge of the energy inputs from the Sun (as it is the primary energy source) is important for understanding the solar-terrestrial system, of equal importance is the manner in which the terrestrial part of the system organizes itself in a quasi-equilibrium state to accommodate and re-emit this energy. The ROSMIC project (2014–2018 inclusive) was the component of SCOSTEP’s Variability of the Sun and Its Terrestrial Impact (VarSITI) program which supported research into the terrestrial component of this system. The four themes supported under ROSMIC are solar influence on climate, coupling by dynamics, trends in the mesosphere lower thermosphere, and trends and solar influence in the thermosphere. Over the course of the VarSITI program, scientific advances were made in all four themes. This included improvements in understanding (1) the transport of photochemically produced species from the thermosphere into the lower atmosphere; (2) the manner in which waves produced in the lower atmosphere propagate upward and influence the winds, dynamical variability, and transport of constituents in the mesosphere, ionosphere, and thermosphere; (3) the character of the long-term trends in the mesosphere and lower thermosphere; and (4) the trends and structural changes taking place in the thermosphere. This paper reviews the progress made in these four areas over the past 5 years and summarizes the anticipated research directions in these areas in the future. It also provides a physical context of the elements which maintain the structure of the terrestrial component of this system. The effects that changes to the atmosphere (such as those currently occurring as a result of anthropogenic influences) as well as plausible variations in solar activity may have on the solar terrestrial system need to be understood to support and guide future human activities on Earth.

Keywords: Climate, Atmosphere, Middle atmosphere, Thermosphere, Ionosphere, Solar irradiance, Energetic particles, Atmospheric coupling, Trends
1 Introduction

The primary goal of solar terrestrial physics is to understand the physical foundations of the structure and variability associated with the coupled Sun-Earth system. The challenges are significant, since the important physics of the various components of this system (Sun, Solar wind, magnetosphere, ionosphere, and atmosphere) differ and dominate in different spatial regions. Furthermore, the important time scales of the atmospheric responses to the forcings become longer closer to the surface of the Earth. Significant progress has been made in understanding the physics within each region. However, to facilitate progress, the boundary conditions linking the components have tended to be simplified. As a result, interactions between the different regions have not been realistic, and a full Sun to Earth system description has not been developed.

Recently, progress on establishing a full physically self-consistent understanding of the solar-terrestrial system has been made. This is due to the development of new observational techniques, advances in the sophistication and computing power of models, and the generation of tools allowing processes from the Sun to the Earth to be visualized and analyzed. This evolution is reflected in the nature of recent Scientific Committee on Solar-Terrestrial Physics (SCOSTEP) programs where investigations of the coupling between regions has been encouraged (see Shiokawa and Georgieva (2021)).

The recently completed VarSITI program (Variability of the Sun and Its Terrestrial Impact, 2014–2019) supported four projects: SEE (Solar Evolution and Extrema), MiniMax24/ISEST (International Study of Earth-affecting Solar Transients), SPECIMEN (Specification and Prediction of the Coupled Inner-Magnetospheric Environment), and ROSMIC (Role Of the Sun and the Middle atmosphere/thermosphere/ ionosphere In Climate). Interactions between the various projects were encouraged. In PRESTO (PREdictability of variable Solar-Terrestrial cOupling), the 2020–2024 SCOSTEP science program, the system aspects of solar-terrestrial relations is further emphasized by defining projects in terms of their temporal and spatial relationships through the full system.

The ROSMIC project was devoted to understanding the impact of the Sun on the Earth’s middle atmosphere/lower thermosphere/ionosphere (MALT1) and its tropospheric climate. It sought to estimate the importance of this natural forcing over time scales from minutes to centuries, in a world where anthropogenic forcing is driving large-scale changes across all atmospheric regions. It was a continuation of the science initiated during the CAWSES (Climate And Weather of the Sun-Earth System) programs under which many of the intricacies and subtleties of the solar-terrestrial system started to be recognized. The essential insight was that the terrestrial component of the system plays an active role in determining the character of the system and its response to variations in solar input. On average, the terrestrial component of the system is in equilibrium with respect to solar inputs, i.e., the incoming and outgoing energy fluxes are equal. However, the physical properties of the terrestrial system and the non-local and non-linear processes that are involved determine the form that this equilibrium takes and its response to variations in solar input.

Two scientific issues stimulated work to develop insight into the terrestrial response. The first was that while significant sensitivity to solar variability in the stratosphere and troposphere is observed, the mechanisms causing these dependencies are difficult to determine (Gray et al. 2010). The second was the recognition that, in addition to radiation and energetic particles, much of the communication between layers of the atmosphere is due to waves (most of which are excited in the troposphere) and that these waves propagate to significant heights and cause ionospheric and thermospheric variability (Oberheide et al. 2015).

Over the course of the ROSMIC project, research into these issues has continued, and although they have not been settled, our understanding has advanced. The inconsistency between different observations and methods to determine the solar spectral irradiance variations is being resolved (Woods et al. 2015, 2018, Yeo et al. 2017). Together with total solar irradiance, solar spectral irradiance is now being included along with energetic particle forcing as input to model runs for the Coupled Model Intercomparison Project Phase 6 (CMIP6) of the World Climate Research Program (Matthes et al. 2017). Investigations to identify the subtle processes leading to the solar influence of the lower atmosphere are progressing, and new challenges have been identified (Gordon 2020; Dhomse et al. 2016; Thibblemont et al. 2015; Chiodo et al. 2019). New complexities in the upward coupling from the troposphere and stratosphere to the mesosphere, thermosphere, and ionosphere associated with waves are now being addressed. These include recognition of the importance of secondary gravity wave generation to wave penetration into the thermosphere (Becker and Vadas 2018) and pole-to-pole coupling of dynamical signatures (Karlsson and Becker 2016; Smith et al. 2020) and that the presence of wave signatures in the thermosphere and ionosphere need not be due to direct propagation (Miyoshi and Yamazaki 2020). Sudden stratospheric warmings and their global influences are being extensively studied (Butler et al. 2015; Pedatella et al. 2018a). They are the largest perturbations to the middle atmosphere on record and provide insights into the mechanisms of the coupling between atmospheric layers.

The need for coordinated ground observations and campaigns and associated modeling started to be
addressed during ROSMIC. Examples include the Interhemispheric Coupling Study by Observations and Modeling (ICSONM) project, the Antarctic Gravity Wave Instrument Network (ANGWIN) (Moffat-Griffin et al. 2019), and The Deep Propagating Gravity Wave Experiment (DeepWave) (Fritts et al. 2016). Modeling advanced significantly with whole atmosphere models increasing in sophistication and including the ionosphere (Liu et al. 2018; Miyoshi et al. 2018; Solomon et al. 2019), and data assimilation efforts start to include mesospheric observations (Eckermann et al. 2018).

The scientific questions which guided ROSMIC activi-
ties were as follows:

- What is the impact of solar forcing on the entire atmosphere? What is the relative importance of variability in solar irradiance versus energetic particles?
- How are the solar signals transferred between the thermosphere and the troposphere?
- How does coupling within the terrestrial atmosphere function (e.g., what are the roles of gravity waves and turbulence?),
- What is the impact of anthropogenic activities on the MALTI?
- Why does the MALTI show varying forms of long-term variations?
- What are the characteristics of reconstructions and predictions of TSI and SSI?
- What are the implications of trends in the ionosphere/thermosphere for technical systems such as satellites?

During VarSITI, ROSMIC supported the activity in the community through existing activities (i.e., Workshop on Coupling in the Atmosphere-Ionosphere System, ANtarctic Gravity Wave Instruments Network (ANGWIN), High Energy Particle Precipitation in the Atmosphere/SOLAR Influences for SPARC (Stratospheric Processes and their Role in Climate) (HEPPA-SOLARIS) Workshop, 13th International Workshop on Layered Phenomena in the Mesopause Region (LPMR), and 9th IAGA - ICMA/IAMAS - ROSMIC/VarSITI/SCOSTEP workshop on Long-Term Changes and Trends in the Atmosphere) as well as through the organization of sessions at international meeting such as COSPAR, IUGG, and the VarSITI symposia.

This paper is organized as follows. The topic is introduced in an overview section which summarizes the physical context underlying the atmosphere/ionosphere components of the Sun/Earth system. This is followed by two sections which address the progress in the two most significant pathways for modulation and propagation of atmospheric/ionospheric responses to solar variability to solar variability namely coupling from above and coupling from below. The fourth section consists of a summary of progress in observing and understanding long-term trends in the mesosphere and thermosphere (this includes progress associated with themes 3 and 4). The paper concludes with a summary and thoughts on future progress in this area.

2 Overview: general physical principles of the energetics and organization of the atmosphere/ionosphere

Identifying the mechanisms which underlie the mean global atmospheric and ionospheric structures and their variability is especially important at this time. First, we do not fully understand the processes which couple the various regions of the atmosphere and ionosphere. Second, uncertainties remain in establishing the extent to which these structures and their variability might change in response to changes to solar inputs and/or to anthropogenically driven changes to atmospheric composition (especially those involved in radiative effects in the atmosphere). Either of these directly changes the thermal structure and circulation of the atmosphere and initiates a cascade of secondary effects associated with atmospheric waves and their filtering and the transport of radiatively active constituents. As noted in more detail below, many of these effects are non-linear and non-local and point to the complex and active character of the terrestrial part of the solar-terrestrial system.

The zeroth-order energy balance between solar radiation incident on the Earth and outgoing radiation from the Earth is well known and covered in most atmospheric texts. More complex considerations of the nature of this balance for the lower atmosphere can be found in review papers (Trenberth et al. 2009; von Schuckmann et al. 2016) where the effects of wind and current systems and eddies are also quantified. Thermodynamic considerations, involving the analysis of the total energy of the atmosphere (the sum of the internal and potential) relative to the energy of an isothermal reference atmosphere, provide some constraints on the partitioning of the energy. Recent analysis along these lines (Bannon 2013) shows that the percentage of the total energy of the atmosphere (2.5 Giga Joules) available for atmospheric motion (termed available potential energy) is a small percentage (<1%) of the total energy although these motions are responsible for much of the meridional transport of energy toward the poles. This implies that much of the absorbed solar energy is devoted to maintaining the internal and potential energy of the atmosphere with a small amount being available for motion. In his analysis, Bannon (2013) suggests that the rate of conversion of available energy to kinetic energy is ∼ 5-7 W m⁻². While this is small relative to the solar constant, 1361 W/m², it plays an essential role in the terrestrial response to solar
radiation, as it is associated with the poleward transport of energy and the global distribution of radiatively active constituents. An analysis along these lines, which identifies the vertical variation of the partitioning of the incident energy, appears not to have been undertaken to this point, although Kwak and Richmond (2017) have described the relative contributions of momentum forcing and heating in the lower high-latitude ionosphere, albeit without complete consideration of upward coupling.

Figure 1 summarizes the relationships between the components involved in establishing the large-scale structures of the middle atmosphere. The large-scale structural elements are in green, and the processes influencing their form are in blue. The complexity of these relationships is illustrated by considering the effects a shift in the thermal structure might have. Through thermal wind balance, the zonal wind structure would alter. These shifts in the thermal structure and zonal wind structure would modify the wave propagation conditions and location where the waves dissipate. This modification in dissipation location would induce secondary circulations and feedback into the thermal and wind structures. Changes to the winds would also affect the transport of constituents and possibly their distribution. Any effects that this transport might have on the distribution of radiatively active species would alter the heating, further feeding back into the thermal structure. The full impact of the initial change in thermal structure requires consideration of all these potential adjustments simultaneously. Changes to any other element would result in a similar cycle.

Figures 2 and 3 are overview figures which provide a context for the processes and phenomena discussed later in this paper. Figure 2 is a schematic of the drivers of the coupling processes noted above. The solstice zonal mean temperature and wind structures are background fields in the figure with the temperatures being the color background (blue being cooler and yellow warmer) and the winds, the contouring (solid green, eastward; dashed blue, westward). The three regions of most significant solar radiative heating are indicated through the white boxes, the wave processes by orange boxes and black arrows, and the energetic particle precipitation by green boxes and arrows. The induced residual circulation is indicated by the white arrows. This figure provides an indication of the spatial relationships between the various processes.

Figure 3 presents the phenomena important to coupling processes in the MALT. It includes phenomena relevant to the observation of these processes as well as drivers of variability which provide the means to empirically investigate coupling processes. These are all mentioned later in the paper. Also included is a summary of various ground-based and satellite observation types with the height ranges over which observations are provided. The background to the figure again is the solstice zonal mean temperature. Drivers of atmospheric variability are identified by ovals and include sudden stratospheric warmings (SSW), the quasi-biennial oscillation (QBO), and El Niño-Southern Oscillation (ENSO). The height range of their influence is indicated by the vertical white arrows. These phenomena cause variations in the large-scale wind
and/or temperature fields and their variability which allow the strength and nature of the coupling processes to be investigated. Polar mesospheric clouds (PMC) appear in the summer upper mesosphere and are the result of upwelling over the summer pole associated with the induced residual circulation. The photochemistry involves the cycling of constituents from the well-mixed lower atmosphere into the thermosphere where they are dissociated and then diffuse downward to the mesosphere where they recombine to form the original molecular species. The cycling between molecular and atomic oxygen is particularly important for the formation of the airglow which provides an important means of probing the dynamics of the mesopause region. Ozone is a constituent whose distribution depends on photochemistry and transport. The downward transport of NO\textsubscript{x} which affects ozone chemistry is one of the significant mechanisms of downward coupling. The aurora is directly associated with downward fluxes of electrons along field lines and hence a signature of downward influence. The ionospheric dynamo and plasma bubbles are ionospheric phenomena whose form is influenced by upward propagating waves.

The main terrestrial-associated solar-terrestrial coupling components, namely energetic particle precipitation, solar irradiance absorption, and coupling associated with waves and constituent transport, are summarized briefly below. This serves as an introduction to the more detailed review of recent progress which follows. Of importance during the ROSMIC project was facilitating work which helped clarify the role and importance of these various processes in the atmosphere/ionosphere response to solar variability. Also of importance was work being undertaken to model and assimilate space weather effects on the ionosphere and thermosphere. This topic is
not reviewed in this paper, but the interested reader can find a review of the current capabilities in Scherliess et al. (2019).

2.1 Energetic particles
Energetic charged particles, mainly electrons and protons from the Sun and the Earth’s magnetosphere, deposit energy into the atmosphere, particularly in the polar regions (see Fig. 2), where the particles are guided by the Earth’s magnetic field. Solar protons, ejected from the Sun, typically have energies between about 1 MeV to a few hundred MeV and have direct access to the high polar latitude middle atmosphere during solar proton events (SPEs) as shown in Fig. 4 (Seppälä et al. 2014). This range of energies means that they can directly impact the altitudes from the stratosphere to mesosphere. Energetic electrons, with auroral electrons with energies less than 10 keV, and medium energy electrons with energies from tens of keV up to a few MeV, originate in the magnetosphere and as a result of magnetospheric processes, enter the atmosphere near the auroral ovals (as shown in Fig. 4). Note that we will not explicitly cover galactic cosmic rays (GCR), which mainly impact the troposphere; here, the reader is directed to the recent comprehensive review article by Mironova et al. (2015).

The process of all these particles interacting with the atmosphere is known as energetic particle precipitation, or EPP. As a result of EPP, whether of protons or electrons, ionization is enhanced, leading to production of odd-hydrogen ($\text{HO}_x = \text{H} + \text{OH} + \text{HO}_2$) and odd-nitrogen ($\text{NO}_x = \text{NO} + \text{NO}_2$) in the middle atmosphere. As these are both known catalysts in ozone loss reactions, solar influence via variations in EPP levels has implications to both atmospheric chemical and radiative balance hence potentially affecting surface climate via downward coupling of introduced wind anomalies (Seppälä et al. 2013). $\text{HO}_x$ is chemically short lived and thus so are its atmospheric impacts. In contrast, $\text{NO}_x$ produced by EPP (known as “EPP-NO$_x$”) in the middle atmosphere is mainly destroyed by photolysis and so has a long chemical lifetime during polar winter. It is therefore subject to transport by the downward circulation over the winter pole. One of the challenges with estimating EPP effects is that, while good observations on proton fluxes and energy
spectra during the large but sporadic SPEs exist, the limitations in electron observations have long hindered good characterization of both flux and energy spectrum of precipitating electrons. Unlike solar radiation, there is not a clear relation to the 11-year solar cycle, and the role of magnetospheric processes (such as those addressed in the VarSITI SpeCIMEN project) in electron precipitation provides another challenge in their quantification.

2.2 Solar irradiance variability
Solar radiation is the dominant energy input to the Earth system with about 70% absorbed by the atmosphere, land, and ocean, and the remainder scattered and reflected back to space (L’Ecuyer et al. 2015). This energy partially determines the temperature and structure of the atmosphere and warms the Earth surface. Globally, a delicate balance is maintained between incoming solar radiation, the Earth’s albedo (fraction of radiation reflected back to space), and outgoing long-wave infrared radiation that can be altered by greenhouse gases, clouds, and aerosols. Changes in solar irradiance have both direct and indirect effects on the Earth climate system, and the roles of solar irradiance are evident in many climate records (e.g., Solanki et al. (2013); Ermolli et al. (2013); Lean et al. (2005)). There are two dominant direct solar heating effects in the lower atmosphere: the "bottom-up" mechanism is the solar irradiance warming the Earth’s surface and upward coupling, and the “top-down” mechanism whereby solar ultraviolet radiation absorbed in the stratosphere couples downward. The winds and circulations in the atmosphere and oceans invoke complicated feedbacks that introduce non-linear and non-local responses (e.g., Meehl et al. (2008)). The solar radiation is also critically important in the Earth’s upper atmosphere.
vacuum ultraviolet (VUV: shorter than 200 nm) photons originate in the Sun’s chromosphere, transition region, and corona and deposit their energy in the Earth’s ionosphere, mesosphere, and thermosphere, being regions strongly influenced by space weather events like solar flares.

A well-known example of the non-linear response of the atmosphere to solar forcing (and an example of how the atmosphere modulates solar forcing) is what has been termed the natural “thermostat” effect (Mlynczak et al. 2003). During periods of enhanced geomagnetic effects (coronal mass ejections and particle precipitation) and variations in solar irradiance, the production of NO and the associated near-infra-red cooling is enhanced. As these events are also associated with enhanced heating of the atmosphere, the enhanced cooling serves to reduce the temperature response of the lower thermosphere to these inputs. The extent to which this process influences the cooling was described using 12 years of SABER observations by Mlynczak et al. (2014). Recent refinements to our understanding of this cooling process appeared in studies by Knipp et al. (2017) and Nischal et al. (2019).

The solar irradiance varies on all time scales with the key variations being the 11-year solar activity cycle (22-year magnetic cycle), the 27-day solar rotation variability, and short-term (seconds to hours) variability during solar flares. The solar variability is highly dependent on wavelength, which is related to the source regions in the solar atmosphere. The total solar irradiance (TSI: integrated over all wavelengths) and the solar spectral irradiance (SSI) in the near ultraviolet (NUV, 300–400 nm), visible (VIS, 400–800 nm), and near infrared (NIR, 800–3000 nm) comprises the bulk of the solar radiative energy and vary the least amount at about 0.1% over the 11-year solar cycle and typically have less variability for shorter time scales (e.g., Woods et al. (2018); Lean et al. (2005)). These NUV-VIS-NIR wavelengths are most important for climate studies for assessing the influence of the Sun on Earth and for comparison with other natural processes (such as volcanic eruptions and the El Niño Southern Oscillation) and human activity (such as greenhouse gas production from fossil fuel combustion). The SSI is much more variable in the ultraviolet (UV) ranges: middle ultraviolet (MUV; 200–300 nm), far ultraviolet (FUV; 120–200 nm), extreme ultraviolet (EUV; 10–120 nm), and soft X-rays (SXR, 0.1–10 nm) but involves only 1.5% of the solar radiative energy. The solar cycle variability in the UV is less than 15% for MUV wavelengths, about a factor of two for many EUV and FUV wavelengths, and even to factors of more than a 1000 for most SXR wavelengths (e.g., Woods et al. (2018); Woods and Rottman (2002)). These UV ranges are most important for atmospheric research and for space weather studies and applications involving satellite operations, communications, and navigation.

### 2.3 Transport and upward coupling through waves

The details of the chemistry and dynamics underlying the zonal mean structure of the stratosphere and mesosphere and its variability have been recently reviewed by Baldwin et al. (2018). Much of the structure of the middle atmosphere is due to a combination of radiative forcing and transport of momentum and constituents by mean flows (i.e., the residual circulation) and mixing which are driven in part by a broad spectrum of atmospheric waves. In the middle atmosphere, the combined effect of the residual circulation and mixing is now starting to be termed the Brewer-Dobson circulation (Baldwin et al. 2018). The current consensus is that the Brewer-Dobson circulation will be enhanced as a result of climate change. However, because of the complexities of the linkages between the different components identified in Fig. 1, exact drivers of this change remain difficult to determine (Butchart 2014).

The situation in the upper mesosphere and lower thermosphere remains less studied and understood. Recent advances in our knowledge of sudden stratospheric warmings reveal that the various regions of the atmosphere are closely coupled. The associated wave dissipation in the polar stratosphere not only influences the state of the middle atmosphere but drives global changes to the state of the thermosphere and ionosphere (Pedatella et al. 2018a). Nevertheless, details of the mechanisms which maintain the average conditions above the stratopause remain poorly understood. Wave dissipation in the mesosphere is considered important for the cycling of air through the stratosphere and mesosphere and associated age of air calculations (see for example Kovács et al. (2017)). The cycling of constituents and associated photochemistry across the mesopause region is still being examined in terms of global temporal means (Garcia et al. 2014; Gardner et al. 2019; Swenson et al. 2018) as opposed to zonal means or regional averages. There is indirect evidence of a wave dissipation-driven, lower thermosphere to stratosphere circulation (opposite in direction to the better known residual circulation in the middle atmosphere) (Qian et al. 2017; Qian and Yue 2017) which provides some indication of zonal mean structures at these heights. Apart from this, there is little information on the seasonal variation and latitudinal structures of constituents and dynamics at these heights. This remains an important topic for future research.

The source of the wave activity is predominantly in the troposphere. As the waves propagate upwards, they interact with the mean wind and temperature fields, which influence their propagation and dissipation, and in turn, the stress associated with wave dissipation modifies the global fields. These wave processes are therefore non-linear and non-local. The nature of the wave field as a function of height was explored in detail in general circulation models of sufficient resolution and height.
(Shepherd et al. 2000; Koshyk and Hamilton 2001; Hamilton et al. 2008). They noted that the rotational components dominated the kinetic energy spectrum lower in the atmosphere with the divergent component becoming stronger with height and in the mesosphere becoming as strong as the rotational component. The divergent component is associated with atmospheric tides and gravity waves. Their increased relative importance above the stratopause is reflected in the extensive literature devoted to exploring gravity wave effects throughout the middle and upper atmosphere (and references therein Liu et al. (2014b); Yi˘git and Medvedev (2015); Liu et al. (2017); Becker and Vadas (2018); Miyoshi and Yi˘git (2019)). Of particular interest in these investigations is the analysis of SSWs as they provide the means to investigate structural differences in wave forcing and propagation for conditions which deviate significantly from normal (Manney et al. 2009; Goncharenko et al. 2013; Ern et al. 2016; Siddiqui et al. 2017; Pedatella et al. 2018b). The character of the waves penetrating and affecting the ionosphere and thermosphere is a topic that is of significant current interest and continues to evolve. The waves of interest and some of their characteristics are summarized by Liu (2016), and the density variations associated with various atmosphere and ionospheric phenomena are summarized in Liu et al. (2017).

3 Coupling from above

This section will focus on the aspects of coupling from above via solar forcing on the atmosphere and climate system. We will address EPP and solar irradiance separately. As there have been significant advances on the modeling capability that impact both, we will first address the latest progress in modeling during the VarSITI period and briefly discuss some remaining challenges. In this context, the earlier works of Gray et al. (2010) and Seppälä et al. (2014) outline the progress on the topic of solar influence on climate, as a result of previous SCOSTEP science programs. Many of the data sets important to this area are from satellite missions that are no longer active. The lack of planned future missions, important for validating models and new ideas on the nature of this coupling, is of considerable concern to the community.

The key questions initially identified by the Working Group on Solar Influence on Climate were as follows:

1. Drivers of solar variability: How well do we know their magnitude and variability?
2. Mechanisms and coupling processes: How is the solar signal transferred down to the troposphere and surface?
3. Solar influence on climate: What are the uncertainties in establishing the long-term effect?

3.1 Progress and challenges in modeling

Various advances have been made in both improving modeling capability and identifying current limitations and future requirements. Reviews that contain details and insights to atmospheric modeling have recently been published by, e.g., Baldwin et al. (2018); Maher et al. (2019).

One issue of importance to modeling is the influence of solar UV irradiance on climate variability. Chiodo and Polvani (2016) addressed the limitations arising from the assessments of this influence when interactive (coupled) stratospheric ozone chemistry is not included. Model integrations with coupled chemistry are computationally expensive. For example, the WACCM (Whole Atmosphere Community Climate Model) model has a throughput of 7.5 simulated years/day on the US–based Yellowstone supercomputer (Smith et al. 2014) with interactive stratospheric chemistry. This is increased to 14.8 simulated years/day with specified (non-interactive) chemistry with the same number of CPUs (central processing unit), nearly a doubling of performance. Thus, for better computational efficiency, interactive chemistry is often omitted. Chiodo and Polvani (2016) showed that inclusion of interactive stratospheric ozone chemistry reduces the surface warming signal from solar irradiance significantly, by a third, when contrasted to predictions from non-interactive (specified) chemistry. They conclude that models that do not take into account the responses in stratospheric chemistry are likely overestimating the surface level response to solar variability.

Another issue of importance is the credibility of model representations of the downward coupling of EPP-induced chemical responses. The multi-model, multi-satellite intercomparison work of Funke et al. (2017) examined how well various medium-top (model top lid at about ~ 80 km) and high-top (model top lid above 120 km) models performed in this capacity. They examined the ability of these models to reproduce polar downward transport of high-altitude carbon monoxide and odd nitrogen (NOx = NO + NO2) as well as observed middle atmosphere temperatures during a dynamically perturbed Northern Hemispheric (NH) winter. The models performed reasonably well until the polar vortex was disrupted by a SSW event. These events frequently occur in the Northern Hemisphere causing major disturbances to the dynamical state of the atmosphere. They lead to the mixing of air masses and, sometimes, in the recovery phase, a reformation of the stratopause at typically mesospheric altitudes. These latter events are known as elevated stratopause (ES) events and can result in effective transport from the mesosphere and lower thermosphere (MLT) into the stratosphere (see also discussion in Section 4).

The descent of air masses as the newly formed stratopause moves downwards to 50 km is challenging for
models to reproduce, with most simulating a too rapid return of the stratopause to its typical height. Funke et al. (2017) determined that model improvements are needed to address the lack of dynamical representation of ES events. To date, the closest representation of the downward descent following an ES event has required either relaxing (“nudging”) the model toward assimilated meteorological fields up to about the altitude of 90 km (Siskind et al. 2015), or the more self-consistent approach of using data assimilation in a chemistry climate model such as WACCM-DART (Pedatella et al. 2018b). This approach is only possible for times when meteorological observations (and thus assimilated products) are available and the models can be constrained to observed dynamics. The model improvements called on by Funke et al. (2017) have not been fully implemented and remain necessary to improve the dynamical capabilities of atmospheric and climate models.

Meraner et al. (2016) found that the parameterization of non-orographic gravity waves played an important role in improved representation of the MLT to stratosphere descent following ES events in the Hamburg Model of the Neutral and Ionized Atmosphere (HAMMONIA). Their results suggest that improvements to gravity wave parameterizations may need to be made to simulate realistic downward transport brought on by ES events. As the ES events can bring MLT air, rich in EPP-NO\textsubscript{x}, into the stratosphere, Randall et al. (2015) further highlighted the need to have a realistic representation of energetic particle precipitation (EPP) to bring down sufficient amount of NO\textsubscript{x} into the stratosphere.

The missing EPP contributions and the associated underestimation of EPP-NO\textsubscript{x} in the polar mesosphere have lead to two significant developments in climate modeling during VarSITI. Firstly, we now have the first long-term energetic electron precipitation (EEP) dataset for use in climate modeling. This is now incorporated into the recommended solar irradiance and energetic particle forcing for CMIP6 simulations (Matthes et al. 2017) and will be addressed further later in the Section 3.2. The second major improvement is the development of the first fully coupled climate model with comprehensive lower ionosphere (D-region) ion chemistry (307 reactions of 20 positive ions and 21 negative ions) (Verronen et al. 2016). This version of the WACCM model, the whole atmosphere component of the Community Earth System Model (CESM), called WACCM-D enables detailed studies of the global lower ionosphere and its response to external (e.g., solar) and internal (e.g., dynamical) forcing. WACCM-D includes fully interactive chemistry, radiation, and dynamics. For the purposes of EPP studies, the inclusion of detailed ion chemistry leads to improved representation of production of NO\textsubscript{x} and HO\textsubscript{x} gases, by removing the need for parameterizations, and the re-partitioning of nitrogen compounds (Andersson et al. 2016; Funke et al. 2017; Orsolini et al. 2018).

There has been an increased interest in the role of the stratosphere in tropospheric variability and the potential for improved predictions. While it still remains to be determined if stratospheric influence is relevant for daily weather forecasts (Baldwin et al. 2018), growing evidence is supporting inclusions of stratospheric information to produce skillful seasonal forecast from a few months to up to a year ahead (Thiéblemont et al. 2015; Dunstone et al. 2016). There are further indications that the ability to simulate stratospheric variability, whether in ozone, SSW occurrence or large-scale modes such as the quasi-biennial oscillation (QBO) leads to improvements elsewhere in the atmosphere/climate system (Baldwin et al. 2018). For the top-down coupling, this would suggest that mesospheric and thermospheric chemical and dynamical processes contributing to stratospheric variability also need to be investigated—one of the questions to be addressed in PRESTO (Shiokawa and Georgieva 2021) is the predictability in sub-seasonal to decadal variability for the atmosphere and climate.

### 3.2 Energetic particle precipitation

The widespread shift from SPE focused, event type studies (e.g., Funke et al. 2011), toward more general consideration of EPP impacts on the atmosphere that we saw during the CAWSES-II program (Seppälä et al. 2014) has continued during VarSITI. The current consensus is that EPP, in the form of EEP, modulated by magnetospheric activity, is an important driver for the chemical variability of the polar middle atmosphere. This needs to be captured in chemistry-climate models in order to correctly represent natural polar ozone variability. Figure 4 illustrates some of the key differences in the sources and impact areas of SPEs and different types of EEP.

In particular, Stone et al. (2018) highlighted the need to account for SPEs when evaluating the recovery of stratospheric ozone due to chlorofluorocarbons (CFCs). SPEs could have both direct (in situ) and indirect (via transported EPP-NO\textsubscript{x}) influences on stratospheric ozone. Denton et al. (2017; 2018) analyzed ozonesonde data from the NH polar region and found that, once seasonal background variability was taken into account, lower stratospheric ozone (altitudes below 35 km) was reduced in excess of 30 days from the start of the SPEs, with an average depletion of 5—10%. However, these results were recently disputed by Jia et al. (2020).

As SPEs are only a fraction of the total EPP, it is likely that further improvements on simulated ozone variability on decadal scales would arise from inclusion of EEP. This has been partially addressed by the development of the first EEP proxy model for inclusion of ionization from the so called “medium energy electrons” (with energies up
to 1 MeV) in climate models (van de Kamp et al. 2016; Matthes et al. 2017; van de Kamp et al. 2018). As mentioned earlier, this is now the official input to CMIP6 model simulations going into IPCC AR6. Overall, the first EEP proxy is an underestimation of the total electron precipitation, and recent improvements to EEP observations (Peck et al. 2015; Nesse Tyssøy et al. 2016; Clilverd et al. 2017; Oyama et al. 2017; Nesse Tyssøy et al. 2019) will likely further aid our understanding of variability in this type of solar forcing into the atmosphere and lead to inclusion of more accurate levels of EEP in model studies (Smith-Johnsen et al. 2018).

To test the van de Kamp et al. (2016) EEP model, Andersson et al. (2018) ran simulations with the WACCM model to investigate the impact of inclusion of electron precipitation on middle atmosphere chemistry. They found that, on average, mesospheric ozone was reduced by up to 20% by inclusion of the new EEP forcing, while in the stratosphere, there was an additional 7% ozone loss when contrasted to simulations without EEP. They further noted that on solar cycle scales, the inclusion of EEP in WACCM doubled the stratospheric ozone response. These results can be contrasted to the multi-satellite observations presented by Andersson et al. (2014), who found that on solar cycle timescales, EEP can drive mesospheric ozone variations of up to 34%.

In their multi-satellite observational study, Damiani et al. (2016) showed evidence that EPP in general has influenced Antarctic upper stratospheric ozone since 1979, at a level of 10–15% O₃ depletion on a monthly basis. This is slightly more than found by Fytterer et al. (2015), who estimated Antarctic ozone depletion of 5–10% using observations from a different set of satellite instruments. Using two independent chemistry-climate models (CCMs), Arsenovic et al. (2019) and Pettit et al. (2019) showed that only by including EEP were models able to reproduce stratospheric ozone anomalies found in satellite observations. A number of further studies (e.g., Zawedde et al. 2016; Smith-Johnsen et al. 2017; Smith-Johnsen et al. 2018; Newnham et al. 2018) have demonstrated EEP impacts on both NOₓ and NO₃ gases.

Together, these suggest that there is need to improve the EEP forcing (the current EEP proxies are known to underestimate precipitating levels, see Nesse Tyssøy et al. 2019)) to ensure better representation of ozone variability in the middle atmosphere. A wider source of EEP are also likely to influence ozone variability in the mesosphere. Seppälä et al. (2015), Turunen et al. (2016), and Seppälä et al. (2018) investigated the potential impacts on mesospheric ozone levels from auroral substorms (which are known to occur frequently), pulsating aurora, and the more energetic relativistic electron microbursts, respectively. They found all to influence ozone levels by up to tens of percent, but thus far, this has not been verified from observations, nor is it clear to what extent these types of electron precipitation are included in the existing EEP proxy (Matthes et al. 2017).

To capture the descent of MLT NOₓ into the stratosphere, Funke et al. (2014b) analyzed 10 years of satellite observations from the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) instrument onboard the Envisat satellite. The instrument simultaneously measured several different middle atmosphere constituents, enabling the extraction of purely EPP produced reactive nitrogen, NOₓ, from the record. By the end of the polar winter/start of spring, the descent of NOₓ into the stratosphere was observed to reach altitudes of 30 km and below in the NH and 25 km and below in the SH during the time period studies. The results highlighted the asymmetries between the two poles arising from the very different dynamical conditions controlling the atmosphere in the north and the south: In the SH, Funke et al. (2014b) show a steady annual descent of EPP-NOₓ from the mesosphere to the stratosphere inside the stable SH polar vortex, with the overall amount of NOₓ depending on the solar cycle and level of geomagnetic activity. In the NH, not only did the overall levels of NOₓ depend on the solar and geomagnetic activity, but also the large dynamical variability present in the polar atmosphere. SSW/ES events had a significant impact on the descent rates as well as the amount and timing of the NOₓ reaching the stratosphere. Funke et al. (2014a) use these EPP-NOₓ observations to derive the total amount of NOₓ molecules in the polar atmosphere reaching altitudes below 70 km. They found a nearly linear correlation between the amount of NOₓ and the geomagnetic Ap index for the winter period extending to early spring, confirming earlier works by Randall et al. (2007); Seppälä et al. (2007), with the NH again showing large responses to dynamical perturbations.

Based on these findings from the MIPAS data set, Funke et al. (2016) presented a semi-empirical, a p-driven, proxy model for EPP-NOₓ descending through mesosphere and upper-stratosphere. The model provides a seasonal dependent flux of EPP-NOₓ descending through given vertical levels. One of the main uses of this semi-empirical model is as an upper boundary condition (UBC) for chemistry-climate models to emulate the descent of NOₓ rich air from the MLT region. For CMIP6, the recommended way of taking into account production of EPP-NOₓ above the upper boundary of medium-top models is by employing the Funke et al. (2016) UBC (Matthes et al. 2017). Figure 5 outlines the different approaches of inclusion of sources of EPP in model simulations.

Following from the work of Funke et al. (2014b), Gordon (2020) recently showed evidence that not only can EPP-NOₓ reach altitudes of 25 km and below by the end of the Antarctic polar winter, but once in the stratosphere,
it remains in the stratospheric NO\textsubscript{x} column throughout the spring until breakup of the SH polar vortex. They found that the role of EPP in the springtime Antarctic stratospheric NO\textsubscript{x} column is modulated by the QBO. They proposed that as CFCs are reduced, EPP-NO\textsubscript{x} will likely take a larger role in Antarctic springtime ozone loss processes.

As we see from above, there is a growing body of evidence that EPP plays an important role in atmospheric chemical balance, extending to decadal scales. Including this source of variability in medium-top and high-top models is now possible for the first time due to development of the CMIP6 EPP forcing input and the EPP-NO\textsubscript{x} UBC (Matthes et al. 2017), but further improvements are still needed. Questions remain unanswered at the end of the VarSITI program: What is the exact size of the EEP-NO\textsubscript{x} source and how important is overall EPP-induced chemistry to climate variability?

While the latest meteorological re-analysis investigations (e.g., Maliniemi et al. (2018); Salminen et al. (2019)) continue to suggest that there are important implications to atmospheric dynamical variability on decadal scales, the results from simulations with a range of CCMs are somewhat inconclusive (Arsenovic et al. 2016; Meraner and Schmidt 2018; Sinnhuber et al. 2018), particularly when it comes to relevance of EPP forcing to surface level climate variability. However, importantly, results from the latest model studies by Arsenovic et al. (2016), and the multi-model study by Sinnhuber et al. (2018) are in agreement with earlier model investigations (Baumgaertner et al. 2011), and also with earlier meteorological reanalysis studies (Seppälä et al. 2013). They found that inclusion of EPP resulted in changes in polar winter stratospheric temperatures (Arsenovic et al. 2016) or radiative heating patterns (Sinnhuber et al. 2018). The Sinnhuber et al. (2018) study further found that state-of-the-art models...
accounting for EPP are able to reproduce observed chemical effects. The introduced radiative forcing changes in the polar stratosphere are of a similar order to those caused by UV variability in the tropics, however, with alternating sign in mid winter (polar night) and spring.

Recently, Maliniemi et al. (2020) addressed the benefit of accounting for the full description of the CMIP6 solar forcing (both irradiance and EPP) in future projections. By looking at the amount of NO\(_x\) descending into the stratosphere, they were able to conclude that climate change will likely increase the EEP-related atmospheric effects toward the end of the century.

3.3 Solar irradiance
This section will outline the recent progress in the understanding of variability in solar irradiance, in both long-term (solar cycle scales) and short-term, models of solar variability, and, finally, the atmospheric and climate impacts.

3.3.1 Long-term variability—solar cycle
The nearly periodic 11-year solar cycle is manifested by the regular presence of numerous large sunspots during solar maximum and few, if any, sunspots during solar minimum. These sunspots are the result of magnetic activity rising up through the Sun's photosphere. Three decades of space-based research document dramatic increases in solar photon, particle, and plasma energy output accompanying the increase in sunspot numbers. Figure 6 shows that in the recent solar cycles 23 and 24, the TSI increased during the maximum by about 0.14% relative to solar minimum and that the SSI HI Lyman-\(\alpha\) emission at 121.6 nm

![Figure 6](image)

**Fig. 6** Solar variability over solar cycles 23 and 24. Examples of solar variability over solar cycles (SC) 23 and 24 for a the total solar irradiance (TSI) and b the bright ultraviolet HI 121.6 nm Lyman-\(\alpha\) emission. The SORCE TSI observations shown here are being extended with TSIS-1 observations. Important solar UV irradiance records over solar cycle (SC) 23 and 24 have been established with TIMED SEE observations, and the TIMED SEE and SDO EVE observations will be extended into the next solar minimum and cycle 25. These extensions are particularly important for overlapping with the new GOLD and ICON missions to observe the ionosphere-thermosphere during cycle 25.
increased by about a factor of 1.7 at maximum relative to minimum. It is also important to note the difference in cycle maximums. In particular, the ratio of the solar variability, defined as maximum minus minimum, at cycle 24 maximum to the variability in cycle 23 is 0.84 for TSI and 0.65 for Lyman-$\alpha$. Moreover, cycle 24 maximum has proven to be the weakest during the past 90 years.

As described earlier, the SSI variability is very wavelength dependent. The solar cycle variability for the SSI as a function of wavelength is shown in Fig. 7. These observations are from NASA's Solar Radiation and Climate Experiment (SORCE (Rottman et al. 2005)), launched in 2003 and also the Thermosphere Ionosphere Mesosphere Energetics and Dynamics (TIMED (Woods et al. 2005)) spacecraft, launched in 2001. The exact amount of SSI variability from the SORCE mission is under debate primarily in the visible where the amount of variability is smaller than the measurement uncertainty and in the near infrared (NIR) where many wavelengths have out-of-phase (negative) variability over the solar cycle (e.g., Ermolli et al. (2013)). The NASA Total and Solar Spectral Irradiance Sensor (TSIS-1) observations that started in 2018 are anticipated to address the SSI variability more accurately for the visible and NIR ranges.

The TSI at the 2008–2009 minimum also appears to be lower by 0.2 W/m$^2$ (−140 ppm) than the previous minimum in 1996 (Fröhlich 2009); however, the 100 ppm uncertainty of this TSI change makes this finding less certain (e.g., Kopp and Lean (2011)). Observations taken during the next solar minimum in 2019–2020 will be very intriguing in terms of determining whether the secular trend of lower solar activity is continuing or not. There are already some indications that the solar activity could continue to decline as indicated from studying

![Fig. 7 SORCE solar spectral irradiance variability. The SORCE solar spectral irradiance variability results from Woods et al. (2015), shown in black, are compared to the SATIRE-S (blue) and NRLSSI (red) model estimates for solar variability between February 2002 (max) to December 2008 (min). The top panel shows the relative variability in percent, and the bottom panel shows the absolute variability in irradiance units. The dashed lines are out-of-phase (negative) solar cycle variability results. The irradiance variability (max-min) in broadbands is provided, and those numbers are in units of mW/m$^2$. The total variability is for the 115 to 1600 nm range. This figure is adapted from Woods et al. (2015).](image-url)
the overlapping solar activity bands of the 22-year solar magnetic cycle (McIntosh et al. 2015). Furthermore, this next cycle minimum in 2019–2020 will be characterized even better than the 2008–2009 minimum (e.g., Chamberlin et al. (2009); Woods et al. (2009)) with the continuation of the SORCE, TIMED, and SDO missions and the addition of new ionosphere-thermosphere measurements by NASA’s Global-scale Observations of the Limb and Disk (GOLD (Eastes et al. 2017)) and Ionospheric Connection Explorer (ICON (Immel et al. 2018)).

3.3.2 Short-term variability—flares
Solar flares have long been an interest for sudden ionosphere disturbances and their effect on radio communication (e.g., Dellinger (1937)). Flare observations have been made for decades in the visible, primarily in H-α (e.g., Ellison (1946)), and also in the SXR and EUV ranges from sounding rocket and satellite experiments (e.g., Friedman (1963)). Hudson (2010, 2011), Doschek and Feldman (2010), Lang (2009), and Aschwanden et al. (2009) provide reviews of recent progress in understanding flares from observations that involve the Solar and Heliospheric Observatory (SOHO), Transition Region and Coronal Explorer (TRACE), Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI), Hinode (Solar-B), and Solar TErrestrial RElations Observatory (STEREO) missions. These satellites include imagers in X-ray and EUV broadbands and imaging spectrographs with high spectral resolution, but over a limited EUV range.

The new and exciting aspects of solar irradiance observations for flare studies are with the Solar Dynamics Observatory (SDO) EUV Variability Experiment (EVE (Woods et al. 2012)). They include spectral coverage over the full EUV range from 0.1 to 106 nm with 0.1 nm resolution and the continuous monitoring of the solar activity at a high cadence of 10 s. There is also a continuous monitoring of the solar EUV images with cadence of 12 s with the SDO Atmospheric Imaging Assembly (AIA (Lemen et al. 2012)). The AIA image time series permit physical interpretation of the EVE (full-disk) irradiance flare variability. The EVE flare observations have revealed that many EUV emissions do not behave like the X-ray variations that are often used for classifying the flare magnitude and as a proxy for EUV emissions in models such as the Flare Irradiance Spectral Model (FISM (Chamberlin et al. 2007; Chamberlin et al. 2008)). Therefore, the EVE flare observations are important to improve the understanding of flare energetics and their impacts on Earth’s space environment.

Flare events can be studied in detail with EVE’s full-disk EUV spectral irradiance as long as only one major flare is happening at a time, which happens frequently. Woods et al. (2011) provide examples of the four major phases seen during flares with the EVE data. These phases include (1) the impulsive phase best seen in transition region emissions such as He II 304 Å, (2) gradual phase seen in hot coronal emissions such as the Fe XX/Fe XXIII 133 Å, (3) coronal dimming seen in cool corona emissions such as Fe IX 171 Å, and (4) EUV late phase (ELP), which has a second, broad peak one to 5 h after the main flare phases and seen best in the Fe XVI 335 Å emission. The X-ray flare classification by the Geostationary Operational Environmental Satellite (GOES) X-Ray Sensor (XRS) is based on the peak SXR variations during the gradual phase, and the derivative of the GOES XRS 1-8 Å emission can be a proxy for the impulsive phase, as related to the Neupert effect (Neupert 1968). Prior to EVE observations, flare observations in the SXR and radio wavelengths were usually decomposed into an impulsive phase with significant non-thermal signatures and a gradual (slow) mostly thermal phase that follows the impulsive phase (Donnelly 1976; Hudson 2010; Hudson 2011). The coronal dimming and EUV late phase effects are observable primarily in the EUV emissions. Each flare can have its own unique behavior; some flares have all four of these phases, and some flares only have the gradual phase (by definition from the X-ray flare identification by GOES/XRS). For more detailed information, Hudson (2010) reviews flare processes and phases, and Hock (2012) identifies different categories of flares based on the new SDO/EVE and SDO/AIA observations of hundreds of flares. Notably, the eruptive flares tend to have impulsive phase, gradual phase, and coronal dimming, and some eruptive flares also have the EUV late phase. The C8.8 flare on 2010 May 5 is a good example when all four components clearly exist as shown in Fig. 8. The various EUV emissions have one or more of these aspects in their time series, and the four emissions that best highlight each component are included in this figure. The EUV spectral variations during this C8.8 flare are shown in Fig. 9. [Prior to the SDO mission, the flare irradiance models have been using the GOES X-ray signal as a proxy for the gradual phase and the derivative of the X-ray signal as a proxy for the impulsive phase emissions. It is clear now with the SDO EVE and AIA measurements that at least two additional flare components—(a) coronal dimming for cool coronal emissions and (b) an EUV late phase for warm coronal emissions—are required for improved modeling of the EUV irradiance. [While coronal dimming and long-duration events like post-flare giant arches have been known for some time, their impact on EUV irradiance is now being clarified with the new SDO observations. The new EVE results are also very important for many space weather applications as deposition of the solar EUV irradiance into the Earth’s atmosphere depends on the spectral variability and on the timing that determines the local (regional) effects on Earth. For example, the ionospheric F layer is expected to have an
additional increase one to 5 h after the GOES X-ray peaks for EUV late phase flares. These late-phase flares can be significant because they can enhance the total EUV irradiance flare variation by a factor of 40% or more when the EUV late-phase contribution is included. Furthermore, the occurrence of late-phase flares are clustered before and after each solar cycle minimum and has a minimum frequency of occurrence during cycle maximum (Woods 2014). Another advance is that the coronal dimming observations can serve as proxies for the mass and velocity of Earth-directed coronal mass ejections (CMEs) (Mason et al. 2014; Mason et al. 2016).

3.3.3 Models of solar variability
There are several models of the solar spectral irradiance variability that provide temporal coverage when there are gaps in the observations and also for extending back to times prior to the advent of satellite solar measurements. The physics-based solar irradiance models, such as the Solar Radiation Physical Model (SRPM (Fontenla et al. 2011; Fontenla et al. 2014)), can also provide insight into the causes for the solar variability. A few of the more commonly used models for climate studies (e.g., CMIP6) are the Naval Research Laboratory TSI model (NRLTSI (Lean et al. 2005)), Naval Research Laboratory SSI model (NRLSSI (Lean et al. 2005; Lean et al. 2011), NRLSSI2 (Coddington et al. 2016)), and Spectral and Total Irradiance Reconstruction (SATIRE (Ball et al. 2014; Krivova et al. 2011; Wenzler et al. 2004; Fligge et al. 2000)). These models provide daily estimates for the solar variability and also extend back to the Maunder Minimum period, a time of low solar activity in the 1600s. It should be noted that, as pointed out by Matthes et al. (2017), some of the models show a discrepancy over the last three solar cycles: while SATIRE predicts a significant downward trend of the baseline SSI at solar minimum, this is not as apparent in NRLSSI2 (a combination of these models is used for CMIP6 recommended SSI and TSI forcing datasets).

For space weather studies, the models need to be at high cadence of at least once a minute in order to describe flare variations. The Flare Irradiance Spectral Model (FISM (Chamberlin et al. 2007; Chamberlin et al. 2008)) is one of the solar irradiance models used for studying flare effects in the ionosphere and thermosphere (e.g., Qian et al. (2019b)). FISM has also been extended to be used at Mars (FISM-M (Thiemann et al. 2017)), and similar algorithms as used by FISM have been adopted for
Fig. 9 Flare spectral variations from the EVE MEGS channel (6–37 nm) for the C8.8 flare on 2010 May 5 as adapted from Woods et al. (2011). A The pre-flare spectrum. B–D The variability between the pre-flare irradiance and the main phase, coronal dimming, and EUV late phase, respectively. These results used 5-min averages taken at the times indicated in Fig. 8 as vertical dashed lines.
NOAA’s operational solar irradiance model (Thiemann et al. 2019).

3.3.4 Influence of solar variability on the atmosphere and climate system

Representation of variations in solar irradiance, both in TSI and SSI, as outlined in the above sections, is of high importance for capturing the influence the Sun has on the atmosphere and climate system. Not only does it link directly to radiative heating, but solar radiation (in the UV) also modulates stratospheric ozone budget (for detailed descriptions of the processes, see, e.g., (Gray et al. 2010; Matthes et al. 2017)).

Kodera et al. (2016) provide a recent update on understanding of the physical mechanisms for solar radiative forcing of the Earth’s surface, through coupling of the stratosphere and troposphere, including interactions with the ocean and sea surface temperatures (SST). They write that “The role of the ocean as heat storage can be seen as persistent surface temperature anomalies from winter to spring. In addition, ocean currents advect SST anomalies to higher latitudes, which may introduce further delayed response.” Kodera et al. (2016) highlight how zonal mean zonal wind anomalies extend from the stratosphere into the troposphere during winter, but these disappear in the spring as the polar-night jet dies out. This is found to coincide with temperature anomalies diminishing over continents, while those over ocean basins (east of continents) are still developing. They also point to the role of oceanic frontal zones (regions of strong temperature gradients in SSTs and surface air temperatures forming in regions of confluent cool and warm ocean currents) as one of the key regions where solar cycle related surface level warming takes place.

The solar minimum between cycles 23/24 was anomalous with regard to its length as well as its depth and thus could provide new insights in the longer term (secular) variations that underlie the 11-year activity cycle. Satellite drag data indicate that the thermosphere was lower in density, and therefore cooler, during the protracted solar minimum period of 2008–2009, than at any other time in the past 47 years (Emmert 2009). Satellite measurements indicate that solar EUV irradiance was also lower in 2008 than during the previous solar minimum. However, secular change due to increasing levels of CO₂ and other greenhouse gases, which cool the upper atmosphere, also plays a role in thermospheric climate, and changes in geomagnetic activity could also contribute to the lower density. Solomon et al. (2010); Solomon et al. (2011); Solomon et al. (2013) conclude that CO₂ and geomagnetic activity play small but significant roles and that the primary cause of the low temperatures and densities is the unusually low levels of solar EUV irradiance.

The upper atmosphere and ionosphere are influenced by forcing from above through solar cycle variations, short term solar flares, or dissipation of magnetospheric energy. Within the ROSMIC period, ESAs Swarm satellite mission (Olsen et al. (2013)) started orbiting the Earth in 2013. Due to its nature as a geomagnetic mission, its data were mainly explored to investigate ionospheric current systems in the E- and F-region. These were often successfully combined with data from the CHAMP mission (2000–2010) to achieve a multi-year data set.

Many of the results associated with these missions are relevant to coupling from below (see the Section 4.3 below). However, solar and geomagnetic effects were also addressed. These have been reviewed in several papers. Laundal et al. (2018) developed an empirical model of the high-latitude E-region and auroral field-aligned currents and their responses to the direction and strength of the solar wind for different seasons, during geomagnetically quiet and active times Alken et al. (2017) and Yamazaki and Maute (2017) reviewed the current knowledge mid- and low-latitude currents in the F-region and E-region (the latter Sq and EEJ), respectively, derived climatological behaviors from these long-term observations, and compared and assessed recent model results to simulate these currents.

Maycock et al. (2016) revisited our understanding of the solar cycle signal in stratospheric ozone, using updated satellite datasets. They found that on annual scales, the signal in stratospheric ozone is much smaller than previously estimated, but note potential, substantial, monthly scale variations. Ball et al. (2019) used new ozone composite datasets to also estimate the solar signal in the upper stratosphere. They determined a statistically significant “U-shaped” annual mean response (peaking at approximately 3% near the equator) that is likely the imprint of seasonal response variations. Good agreement was found with model simulations using the SOCOL and WACCM models.

Dhomse et al. (2016) called into question how well we understand the response of upper stratospheric ozone to the 11-year solar cycle. They suggest that there may be large uncertainties in observations, the impact from changing GHG emissions, recent changes in the amplitude of the 11-year solar cycle, and potential challenges arising from meteorological reanalysis data sets. While, on annual scales, the solar cycle signal in stratospheric ozone may be smaller than previously estimated, accounting for this solar cycle ozone response in global models is important. This was highlighted by Maycock et al. (2018) where they pointed to the necessity of accounting for the ozone response to adequately capture solar cycle impacts on the atmosphere. How this is done must be carefully considered, particularly for models that do not have interactive chemistry.
The influence of the 11-year solar cycle on North Atlantic via the North Atlantic Oscillation (NAO) has been a topic of high interest, particularly in the past two decades. This is not least due to the potential increased predictability of large scale climate patterns influencing the large metropolitan areas around the North Atlantic and extending into central Asia. Recently Thieblemont et al. (2015), Gray et al. (2016), and Roy (2018), found further evidence of the link between the solar cycle and these large scale climate patterns, pointing to the role of downward propagation of the solar signal from the stratosphere to the troposphere. However, Chiodo et al. (2019) proposed that the signal in the North Atlantic climate that several previous studies have attributed to the 11-year solar cycle is, in fact, a chance occurrence, resulting from internal variability in the climate system. If this is the case, it would limit the usefulness of the 11-year solar cycle as a predictor for large scale circulation. The authors note that there are several caveats that may affect their conclusions, but nonetheless, call for much needed caution when interpreting quasi-decadal signals, such as those from the solar cycle, in the existing, relatively short, observational records. In this context, studies providing insights into the detailed physical mechanisms behind the so called “top-down” coupling are important. investigations such as that by Lu et al. (2017b); Lu et al. (2017a) who investigate roles of planetary wave breaking, reflection and resonance in the propagation of the solar signal from the stratosphere to the troposphere, may be able to provide some further understanding of the details of the solar cycle signals in climate.

As discussed earlier, there have been some indications that solar activity could continue to decline in the next decades. Changes in the solar input warrant further investigation to assess the potential impact of a potential future “grand solar minimum.” Following initial work by Meehl et al. (2013), Maycock et al. (2015) and Ineson et al. (2015) investigated the combined effects of lower solar activity (0.12% reduction in TSI, 0.85% reduction in UV), reducing ozone depleting substances and increasing GHG concentrations. They confirmed that there was negligible impact on global mean surface temperatures (around −0.1 K), which could not offset the projected impacts of anthropogenic climate change. However, the predicted regional-scale changes were much larger, particularly for mid- to high latitudes. Both the polar vortex and the tropospheric midlatitude jet were found to respond to the reduced solar activity (indicating links to top-down coupling) resulting in changes in large scale dynamical patterns in both hemispheres. These in turn could potentially drive “larger regional scale surface climate effects” (Ineson et al. 2015), highlighting the need to carefully consider changes in both anthropogenic and natural drivers together when assessing future projections.

3.4 Future challenges
What will the Sun do in the future? Variations in the 11-year solar cycle as well as longer cycles are important for any future projections of the state of the atmosphere and climate. Continued observations of solar activity both for radiation and particles is vital for our understanding. At the same time from the atmosphere side, we are relying on aging missions to detect changes in atmospheric composition in response to forcing from above and below. The growing unease in the scientific community has not diminished during VarSITI: we are heading into the future relying more and more on models that continue to need a “ground truth” from satellite observations to enable increased accuracy.

What will be the role of solar forcing for climate in the future? One of the big challenges we are now facing is understanding how the atmosphere and climate response will adjust due to changes brought on by enhanced GHGs and climate change, recovery of the Antarctic ozone hole (which will influence both the chemical and dynamical state of the SH atmosphere), and even the potential future grand solar minimum—together. Most studies thus far have focused on the solar irradiance changes but with the developing EPP proxies we are now seeing the first evidence that there is benefit in accounting for the full description of the CMIP6 solar forcing in future projections (Maliniemi et al. 2020). On a smaller scale, the question on whether there is improved seasonal prediction skill from inclusion of solar activity is showing some promise.

4 Coupling from below
ROSMIC’s working group “Coupling by Dynamics” focused on the dynamical and thermodynamical processes originating in the lower and middle atmosphere, influencing the upper atmosphere. Modeling, observations, and theoretical studies have all been relevant to the interests of this working group. In addition to the physical processes impacting the atmosphere-ionosphere system from above discussed in the Section 3, upward propagating lower atmospheric processes can shape the structure and evolution of the upper atmosphere. This is generally referred to as “coupling from below.” The key science questions of the Coupling by Dynamics Working Group have been as follows:

1. Wave coupling: What are the influences of lower atmospheric waves on the state and evolution of the thermosphere-ionosphere?
2. Electrodynamic coupling: How does atmospheric dynamics constrain electrodynamics in the ionosphere?
3. Small-scale dynamics: How can we characterize significance of small-scale structures for the large-scale features in the upper atmosphere?
These science questions have been actively studied by the solar-terrestrial community and fostered collaborations among atmospheric and space scientists. The vast majority of the recent studies have indicated that the primary mechanism of coupling from below is via internal atmospheric waves. As noted in the Overview (see Figs. 1 and 2 for a schematic context), they play a role in the transfer of energy and momentum, the mixing of constituents, and development of large-scale flows where they dissipate and atmospheric and ionospheric variability wherever they are present. Their role in the downward propagation of NO\textsubscript{x} and the formation of elevated stratopauses has already been mentioned in the section on downward coupling.

Overall, the lower atmosphere is the primary source of internal waves. At higher altitudes in the atmosphere higher-order waves can be found (Satomura and Sato 1999; Yasui et al. 2018; Becker and Vadas 2018), which are produced via either primary waves or other dynamical and thermal processes, such as instabilities and localized heating taking place in situ. These waves span a broad spectrum of periods (or frequencies) and horizontal and vertical scales. The key components of the spectrum of atmospheric waves are gravity waves, tides, Kelvin waves, and Rossby-planetary waves, covering a range from a few tens of kilometers to several thousand kilometers. The influence of the upward propagating waves can be substantially modulated by lower and middle atmospheric transient processes, such as sudden stratospheric warmings (Pancheva et al. 2008; Pancheva et al. 2009; Tomikawa et al. 2012; Nayak and Yiğit 2019; Yiğit and Medvedev 2012; Yiğit et al. 2014; Yiğit and Medvedev 2016; Pedatella et al. 2018a; Yamazaki et al. 2020). These modulations occur primarily due to the underlying changes in the background atmospheric circulation and temperature, which directly effect wave propagation and dissipation.

Comprehensive reviews of how upward propagating internal waves influence the upper atmosphere and ionosphere have been recently provided in the work of Yiğit and Medvedev (2015) and Liu (2016). Figure 10 adopted from Yiğit and Medvedev (2015) gives a good overview of the different neutral atmospheric and ionospheric layers and coupling from below. They discuss various theoretical, numerical modeling, and general circulation modeling approaches to internal waves, as well as observation of internal wave activity in the middle and upper atmosphere. In particular, the importance of SSWs has been highlighted in their review of wave-induced processes.

![Fig. 10](image-url) Vertical coupling in the atmosphere-ionosphere system, from the troposphere to the thermosphere-ionosphere. The atmosphere is influenced from below by internal waves (gravity waves, tides, planetary waves, Kelvin waves) and from above by solar and geomagnetic processes, referred to as space weather processes. These mechanisms produced wave-like disturbances in the upper atmosphere, known as traveling ionosphere/atmospheric disturbances (TIDs/TADs). The thermosphere and ionosphere coexist, interacting with each other in a two-way coupled manner. Taken from Yiğit and Medvedev (2015)
The review by Liu (2016) concentrates on the variability in the ionosphere and thermosphere associated with waves. The review by Yiğit and Medvedev (2016) specifically capitalized on the dynamical and thermal influence of atmospheric GWs during SSWs. In this context, they turned their attention to the assumption and limitations to GW parameterizations in atmospheric models. Observed changes in the upper atmosphere during SSWs were discussed in detail, emphasizing that more work needs to be conducted in determining the effects of GWs on the circulation and thermal budget of the upper atmosphere during SSWs. The importance of this aspect of gravity waves was discussed earlier (in the Section 3) in reference to the elevated stratosphere and the downward transport of NO\textsubscript{x}. Modeling efforts with HAMMONIA and other GCMs have also indicated that physics of nonorographic gravity waves must be taken into account for a more realistic representation of the wave-mean flow interactions and mixing necessary for this downward transport.

### 4.1 Modeling of vertical coupling by waves

Numerical models are very powerful in the investigation of basic physical processes, i.e., generation, propagation, and dissipation, of waves and in studying how dynamical processes associated with waves and other transient processes influence different atmospheric layers. In essence, vertical coupling is the exchange of energy, momentum, and mass within the atmosphere in the vertical direction. In a stratified atmosphere with nearly exponentially decreasing background density, waves grow in amplitude with increasing altitude in order to conserve energy, with the growth rate modulated by background atmospheric dissipation. Therefore, models can serve as a tool to study how waves are generated, how they propagate from their source, the altitude variation of their dissipation, and as a result, their influence on the dynamical and thermal structure of the atmosphere. In addition, they facilitate the identification of the altitudes where the various types of wave perturbations dominate and what their characteristics are. Wave propagation and dissipation are the primary mechanisms for the coupling between the lower and upper atmosphere.

While there are a number of different genres of atmospheric models, all focus on providing deeper insight into the physical mechanisms of either already observed or yet to be observed phenomena. In doing so, they utilize different analytical and numerical techniques. Model predictions often motivate observers to organize targeted campaigns and measurements and observations stimulate model development and simulations.

One flavor of the modeling of coupling processes is based on idealized numerical models. These generally involve simplified background conditions and specified wave characteristics and provide conceptual frameworks useful for the interpretation of observations. The majority of these numerical models focus on small-scale GWs (e.g., Hickey et al. (2009); Hickey et al. (1993); Kshevetskii and Gavrilov (2005); Gavrilov and Kshevetskii (2013)). Such models are often either one-dimensional or two-dimensional and facilitate the examination of the propagation and dissipation of GWs under various circumstances and at very high-resolution. Earlier studies of the propagation and dissipation of acoustic-gravity waves in the thermosphere (e.g., Hickey and Cole (1988)) have provided significant motivation for recent numerical efforts. Heale et al. (2014b) have used a nonlinear, fully compressible, two-dimensional numerical model to study the effects of dissipation on GW packets in the thermosphere. Such models can also be used to investigate wave propagation over large horizontal distances as a function of direction of propagation relative to the average modeled background winds (Heale et al. 2014a). Control simulations conducted in these studies revealed that waves can undergo varying degrees of reflection, ducting, and trapping, which ultimately influence the extent of their horizontal propagation.

Another type of numerical model is the finite-volume anelastic model that discretizes the anelastic Navier-Stokes equations (Lund and Fritts 2012). One advantage of using finite volume numerical formulation is that they include exact numerical conservation laws. Fritts et al. (2018) have used this anelastic model to explore GW dynamics in the vicinity of mesospheric inversion layers (MILs). They found that MILs can cause GWs to yield instabilities below the altitude where they would usually occur.

Another type of numerical wave model was used in the work by Gavrilov and Kshevetskii (2015) to better investigate the details of dynamical and thermal effects on the upper atmosphere of waves originating in Earth's troposphere. They focused on the nonlinear propagation of GWs, generated with a fixed forcing, to the middle and upper atmosphere. They demonstrated that GWs can directly propagate to the thermosphere and also found out that shorter GWs with slower horizontal phase speeds generate relatively smaller mean heating and wave-induced velocity fluctuations in the upper atmosphere. Shpynev et al. (2015) emphasized that during strong SSWs, the main stratospheric jet stream can amplify GWs to large amplitudes, which facilitate their breakdown. Hickey et al. (2015) have used a two-gas full-wave model to examine the role of various thermospheric dissipation mechanisms on the wave-induced fluctuations of atomic oxygen and molecular nitrogen. They found that viscosity and thermal conduction can alter the oscillation of one species relative to another and can impact the relative phases.
While in a realistic atmosphere, GWs are generated primarily by weather related processes and lower atmospheric features, such as convection, orography, jets, frontal systems, and various localized heating events, idealized numerical models usually represent GWs by a specified initial wave packet (e.g., Gavrilov and Kshevetskii (2013); Gavrilov et al. (2015); Fritts et al. (2018)). On the other hand three-dimensional general circulation models (GCMs) with coarse grid resolutions utilize GW parameterizations, which are analytical representations of GW physics. Typically, larger scale waves are resolved in the model provided that their generation mechanisms are accounted for.

There are various types of gravity wave parameterizations. Some are nonlinear; others are linear schemes (for a discussion see Medvedev and Yiğit (2019)). Some are designed solely for the middle atmosphere; some cover the entire atmosphere. The common purpose of GW schemes is to represent subgrid-scale GWs, which are not sufficiently resolved in coarse-grid GCMs. Yiğit et al. (2014)’s simulation study using the whole atmosphere GW parameterization (Yiğit et al. 2008) implemented in the Coupled Middle Atmosphere Thermosphere-2 (CMAT2) GCM, extending from the lower to the upper atmosphere, demonstrating that small-scale GWs produce up to 50% of the wind variability during a minor sudden stratospheric warming. This latter study was motivated by earlier GCM simulation studies demonstrating direct GW propagation from the lower atmosphere to the middle thermosphere (Yiğit et al. 2009; Yiğit et al. 2012; Yiğit and Medvedev 2012) and by idealized numerical simulations (Hickey and Cole 1988; Gavrilov and Kshevetskii 2013). The whole atmosphere scheme takes into account the dissipation of upward propagating GWs due to eddy diffusion, radiative damping, nonlinear interactions (Medvedev and Klaassen 2000), molecular diffusion and thermal conduction, and ion-neutral friction, known as ion-drag. Since gravity wave processes are ubiquitous features of all terrestrial and planetary atmospheres, GW parameterizations developed for Earth models can also be tested in other planetary models. For example, this scheme has also been used in a Martian GCM to study coupling between GWs and ice clouds (Yiğit et al. 2018), which highlights the broader impact of terrestrial GW studies. Meanwhile, it is well-established that GW processes play a significant dynamical and thermodynamical role also in other planetary atmospheres (Yiğit and Medvedev 2019).

Medvedev et al. (2017)’s theoretical work implemented the influence of the geomagnetic field in the physics of GW dissipation in the whole atmosphere scheme. In addition to the effect of ion drag on the large-scale thermospheric circulation, the effect of ion friction on the propagation and dissipation of small-scale GWs has been taken into account in their formulation. Their numerical experiments have shown that taking into account the geometry of the geomagnetic field can increase GW drag by more than a hundred m s$^{-1}$ day$^{-1}$, GW cooling/heating rate by ~15 K day$^{-1}$, and GW-induced temperature variance by several K in the ionosphere-thermosphere region. Although GCMs, extending into the upper atmosphere, often include the influence of the magnetic field on the large-scale atmosphere, they do not include direct magnetic field effects on GW propagation and dissipation. The impact of this limitation is yet to be explored.

While GW parameterizations provide a practical tool in studying GWs, high-resolution GCMs are increasingly resolving GW-induced perturbations in Earth’s (e.g., Miyoshi and Fujiwara (2008); Watanabe and Miyahara (2009); Gardner and Schunk (2011); Miyoshi et al. (2014); Liu et al. (2014b); Liu et al. (2014a); Becker and Vadas (2018)) atmosphere. However, these models can be computationally challenging to run for longer time scales, especially if they extend from the lower atmosphere to the thermosphere. Another issue is that high-resolution models often damp small-scale GWs above the mesopause via an artificial hyperdiffusivity, which can potentially lead to an underestimation of GW activity at higher altitudes. Miyoshi et al. (2014)’s high-resolution GCM simulations as well as those described in Liu (2016) confirmed that the GW drag due to upward propagating GWs from below plays an important role in the general circulation of the thermosphere. This confirms predictions from previous coarse-grid GCM studies with whole atmosphere GW parameterizations (e.g., Yiğit and Medvedev (2009); Yiğit et al. (2009); Yiğit and Medvedev (2010)). Their results also provide robust evidence that mean thermospheric effects of GWs can be successfully parameterized in lower-resolution GCMs.

Using the CMAT2-GCM coupled with the whole atmosphere GW scheme, Yiğit and Medvedev (2017) analyzed the interaction of the small-scale GWs with the migrating diurnal tide. Their GCM simulations have suggested the GWs significantly modulate the diurnal tide in the MLT as well as in the thermosphere. Their work suggested that GW parameterization should consider a broad spectrum of wave harmonics and describe wave propagation and dissipation properly in order to capture GW-tidal interactions in the atmosphere. Greer et al. (2018) have recently used the Kühlungsborn mechanistic circulation model (KMCM) to select GW characteristics that helped to better interpret thermospheric airglow emissions. Recently, Miyoshi and Yiğit (2019) have implemented the whole atmosphere GW parameterization (Yiğit et al. 2008) into a whole atmosphere GCM, which covers the region from the ground to the exobase. They showed that GWs attenuate the semidiurnal solar tide (SW2) in the lower
thermosphere and modify its latitudinal structure above 150 km.

The vertical coupling induced by large-scale waves can be studied with GCMs as well. Karpov et al. (2018) studied with the Global Self-Consistent Model of the Thermosphere, Ionosphere, and Protonosphere (GSCM TIP, (Namgaladze et al. 1988)) the response of the upper atmosphere, Ionosphere, and Protonosphere (GSCM TIP, (Namgaladze et al. 1988)) to an SSW. They have demonstrated significant changes in thermospheric and ionospheric characteristics, such as the a shift in the maximum atomic oxygen concentration to lower latitudes in the Southern Hemisphere. Lilienthal and Jacobi (2019) studied the generation mechanism of the terdiurnal tide with the Middle and Upper Atmosphere Model (MUAM) model. Their simulations have indicated that the primary excitation of the terdiurnal tide is due to the terdiurnal component of solar radiation absorption in the troposphere and stratosphere. Secondary sources, such as nonlinear tide–tide interactions and gravity wave–tide interactions, can also play a role. Geißler et al. (2020) have used the same GCM to study the forcing mechanisms of the migrating quarterdiurnal tide (QDT). Among the studied mechanisms were the absorption of solar radiation by ozone and water vapor, nonlinear tidal interactions, and gravity wave–tide interactions. Nonlinear processes and GW forcing were found to be important for the QDT besides the primary mechanism of solar radiation forcing.

The nature of short-term tidal variability was identified as a significant issue for our understanding of tides by Oberheide et al. (2015) in their summary of CAWSES II activities for Task Group 4. Advances in this area have been made both in the refinement of data analysis techniques which permit shorter term variability to be diagnosed from satellite observations and the use of general circulation models to explore the nature of this variability. The interesting issues being explored include resolving the stark difference in gravity wave forcing between parametrizations which do not include propagation times and horizontal transport and those which do (Ribstein and Achatz 2016), the lack of correlation between the short-term variability of tropospheric sources and the amplitude of tidal components in the mesopause region (Du et al. 2014) and why tidal amplitudes exhibit short-term variability (of the order of days) (Baumgarten et al. 2018). Recent studies addressing these questions include the statistical approach of Vitharana et al. (2019) and the Hough mode extension approach of Orland (2017). These questions regarding tidal variability and interactions with gravity waves remain open and illustrate the complexity of resolving interactions between waves and determining the impact of their dissipation.

Overall, models provide a powerful methodology of studying all types of waves. However, observations are needed in order to complement, validate, or challenge the numerical findings.

4.2 Observation of coupling processes

In recent years, the global distribution of gravity wave activity in the middle atmosphere has been obtained by satellite temperature measurements with high vertical or horizontal resolution such as Microwave Limb Sounder (MLS), Sounding of the Atmosphere using Broadband Emission Radiometry (SABER), Atmospheric Infrared Sounder (AIRS), and High-Resolution Dynamics Limb Sounder (HIRDLS). The area from the Andes Mountains to the Antarctic Peninsula has been realized as a “hot spot” of gravity wave activity through these satellite observations (Ern et al. 2018a; Hocke et al. 2016; Meyer et al. 2018). Nonetheless, observations of gravity waves around the Antarctic continent are limited. Therefore, the zonal wind acceleration/deceleration induced by gravity waves is uncertain and so are climate change projections (Geller et al. 2013). In particular, the “missing gravity wave drag” near 60° S latitude causes a strengthening of the southern hemisphere polar vortex and a delay of its breakdown, which prevents the quantitative reproduction of the Antarctic ozone hole (McLandress et al. 2012).

Since satellites observe only a part of the spectrum of gravity waves (i.e., “observational filter”, see Alexander (1998)), ground-based and in situ measurements of GWs, despite their limited coverage in time and space, have become important to obtain the entire spectrum of gravity waves. The Mesosphere-Stratosphere-Troposphere (MST)/Incoherent Scatter (IS) radar is able to obtain the time-altitude section of gravity wave activity and its momentum flux in all frequency bands through observations at a high time cadence and vertical resolution. Sato et al. (2017) utilized the PANSY radar, the only MST/IS radar in the Antarctic, and found that the momentum transfer to the Antarctic summer mesosphere is mainly due to long-period gravity waves (i.e., inertia gravity waves). On the other hand, in situ observations with a super-pressure balloon can provide a horizontal distribution of the GW activity and its momentum flux for the entire (intrinsic) frequency band at a single altitude (Vincent and Hertzog 2014; Jewtoukoff et al. 2015). Lidar observations of gravity waves are now available (both day and night) of suitable quality and resolution for mean wave characteristics to be derived and compared with models (Baumgarten et al. 2017; Strelnikova et al. 2020). A recent gravity wave climatology at 54° N and 69° N based on nearly 15,000 h of lidar measurements has some similarity to reanalysis models but also significant discrepancies (Strelnikova et al. 2020). It is now generally accepted that simple gravity wave parametrization scenarios such as those being used in several GCMs, cannot cover all aspects of the physical processes.
related to gravity waves, i.e., their observed spatial and temporal intermittency and their interaction with tides (Baumgarten et al. 2018). Still, satellite, ground-based, and in situ observations of global gravity wave activity are expected to provide the means to improve representation of the middle atmosphere general circulation in climate models. These measurements provide information on the spatial and temporal distribution of the amplitudes and spectra of gravity waves and the associated energy and momentum transfer (see for example Ern et al. (2016); Ern et al. (2018b)). These distributions along with the observed background zonal mean meridional and zonal winds provide a valuable framework for validating the assumptions underlying GW parameterizations and their impact in models.

The horizontal propagation and intermittency of gravity waves must be taken into account to improve the parameterization of gravity wave drag in climate models (Plougonven et al. 2020; Senf and Achatz 2011), because they control the horizontal and vertical distribution of gravity wave drag. The horizontal propagation is mainly detected by satellite observations (Yamashita et al. 2013; Fritts et al. 2016), and the intermittency is characterized through balloon, radar, and lidar observations (Hertzog et al. 2012; Minamihara et al. 2020; Baumgarten et al. 2018). Observational evidence for secondary GWs is also emerging through lidar and satellite observations (Bossert et al. 2017; Song et al. 2017; Kogure et al. 2020).

Airglow observation in the mesopause region is a useful means to identify GWs propagating from below and determine dynamical vertical coupling. To date, ground-based airglow imagers have been installed and are in operation at various locations. The Optical Mesosphere Thermosphere Imagers (OMTIs) are located at more than twenty sites from high to low latitudes, mainly in the northern hemisphere. The recent extension during the VarSITI period was eight new OMTIs at around 60°N in magnetic latitude which covers all longitudinal regions (Shiokawa et al. 2017). In the high latitude Southern Hemisphere, or in the Antarctic region, airglow imaging at more than eight stations form an observation network called ANGWIN (Antarctic Gravity Wave Imager/Instrument Network) (Moffat-Griffin 2019). A dense network of airglow imagers covering a wide area without gaps is in operation in the midlatitudes of China (Xu et al. 2015). The advantage of such ground-based imagers is the opportunity for long-term monitoring of GWs (Tsuchiya et al. 2018) and simultaneous observations with co-located instruments such as wind measurement radars and/or temperature measurement lidars, e.g., Taylor et al. (2019); Criddle et al. (2020). Variabilities with local time, day-to-day, seasonal and inter-annual scales have been studied, as well as vertical propagation from the ground using background wind/temperature profiles.

Airglow imaging from satellites is also a means to identify coupling processes associated with GWs. Concentric GWs observed by satellite imagers with large fields of view (FOV) indicate the impact of meteorological disturbances such as typhoons/cyclones. They excite gravity waves which propagate up into mesosphere, thermosphere, and ionosphere. IMAP/VISI on board International Space Station has shown global maps of these concentric waves and their seasonal variations (Perwitasari et al. 2016). Observations from the Visible Infrared Imaging Radiometer Suite (VIIRS) on board the Suomi/NPP satellite have demonstrated that a hurricane excites concentric GWs when the intensity is maximum (Xu et al. 2019). Effort has been exerted on the development of analysis methods for airglow images. Matsuda et al. (2014) obtained power spectra of airglow perturbation in the horizontal phase velocity domain (M-transform) using 3-D Fourier transforms. This new technique can quantitatively indicate GW interaction with the background mean wind and temperature during their vertical propagation from below (Matsuda et al. 2017; Perwitasari et al. 2018). On the other hand, various approaches, such as machine-learning (Lai et al. 2019), for extracting GW events have also been tried.

Meteor radars have also been used to study the dynamics in the mesopause region. Recently, a regional network for meteor radar observations with a common scattering volume has been developed as a powerful tool for this purpose (Stober and Chau 2015; Chau et al. 2017; Chau et al. 2019; Stober et al. 2018; Vierinen et al. 2019). Wind vectors can be accurately derived for a smaller area with this new network observation technique. The three dimensional structure of wind fields over a wide horizontal range is observed (Stober and Chau 2015). New dynamical parameters of a wind field such as the horizontal divergence and relative vorticity can be determined, an example of which is described by Chau et al. (2017) for observations over northern Scandinavia. The possibility of vertical wind velocity measurement is also indicated. These new techniques are expected to be highly beneficial for studying mesosphere-thermosphere coupling in near future when combined with the EISCAT3D radar being constructed in northern Scandinavia (https://eiscat. se/eiscat3d-information/). Vierinen et al. (2019) even stepped forward to develop a method to estimate the kinetic energy of turbulence using the anticipated observations.

A SSW is the most drastic event in the middle atmosphere, and its impact is widespread. In recent years, “elevated stratopause” events where the stratopause is reformed at an altitude of about 80 km following a SSW has attracted strong interest (see also Section 3.1 for the role of these events for downward coupling). Global temperature observations in the middle atmosphere by satellites such as SABER and MLS capture the overall structure of
these events and established their occurrence after major SSWs in the Northern Hemisphere (Manney et al. 2008; Manney et al. 2009). Prolonged periods of westward wind in the stratosphere following SSWs, enhance westward gravity wave drag above 80 km altitude through the filtering of upward-propagating GWs. This causes a strong downwelling over the pole and the associated adiabatic heating results in a reformation of a stratopause (temperature maximum) around 80 km (Tomikawa et al. 2012; Chandran et al. 2013). The circulation changes during elevated stratopause events are observed to be associated with the enhanced downward transport of NOx, which, as noted earlier, affects the ozone budget in the stratosphere and mesosphere (Randall et al. 2006; Randall et al. 2009). In addition, radar and satellite observations have shown that changes in the wind system and the distribution of trace constituents in the stratosphere and mesosphere during SSW have a significant impact on the ionosphere through the modulation of solar and lunar atmospheric tides in the thermosphere (Chau et al. 2015). Recently, using total electron content (TEC) data from five GPS stations combined with TIMED/SABER data, Nayak and Yiğit (2019) analyzed GW activity variations in the ionosphere during the major SSW of December 2008/January 2009, demonstrating that there is significant variability in GW activity and associated TEC variability during the SSW.

On the other hand, interhemispheric coupling is also attracting attention as a pattern of meridional teleconnection due to SSWs. The occurrence of polar mesospheric clouds in the Antarctic mesopause region is observed to decrease following SSWs in the Northern Hemisphere (Karlsson et al. 2007). Kornich and Becker (2010) proposed a coupling mechanism between the SSW in the winter hemisphere and the warming of the summer mesosphere due to modulation of the global gravity wave forcing. However, another mechanism has also been suggested (Smith et al. 2020), and observational validation is required. An international collaborative observation campaign, ICSOM (Interhemispheric Coupling Study by Observations and Modeling; see also http://pansy.eps.s.u-tokyo.ac.jp/icsom/), has been carried out as an official activity of SCOSTEP/VarSITI/ROSMIC since 2016. ICSOM is composed of the global MST/IS radar network from the Arctic to the Antarctic with many complementary radio and optical observations.

The resonance scattering lidar which involves scattering from the meteoric metal layers (e.g., Fe, Na, K, Ca, and Ni) can measure temperature, wind, and the density of the metallic atoms in the MLT region with a cadence of minutes and high vertical resolution. Recent notable results include direct measurements of eddy heat fluxes and thermal diffusion coefficients in the mesopause region (e.g., Guo et al. 2017) and the discovery of neutral metal layers up to ~200 km altitude in the thermosphere (e.g., Chu et al. 2011). The eddy heat fluxes and thermal diffusion coefficients in the mesopause region were derived from the turbulence-scale perturbations in temperature and wind obtained with an Na wind/temperature lidar. The observation suggests that vertical constituent transport induced by gravity waves can be estimated by lidar measurement (Gardner et al. 2019). Liu et al. (2016) derived the temperature and horizontal wind profiles up to 140 km with an Na lidar at Andes Lidar Observatory in Cerro Pachon, Chile, during a thermospheric Na enhancement. Chu and Yu (2017) used a thermosphere-ionosphere Fe/Fe+ (TIFe) model to infer that a major source of the thermospheric Fe layers observed in Antarctica between 120 and 200 km is the neutralization of metallic ions Fe+. The ions are transported upward by the polar electric field from their main deposition region (~80–110 km) into the ionosphere. Observations of Ca+, the only metallic ion in the MLT region observable with a ground-based lidar, provide insights into the behavior of ions in the E and F region ionosphere. Recent high-resolution Ca+ density measurements revealed various structures and the evolution of the Ca+ layer between 80 km and 180 km altitudes. These results provide information on the coupling of the E and F regions (Ejiri et al. 2019b; Ejiri et al. 2019a; Raizada et al. 2020; Wu et al. 2020). Recently, mesospheric Ni layers were also detected by lidar over Alaska and Germany (Collins et al. 2015; Gerdig et al. 2019). Large differences, however, in Ni density between the two sites were observed.

Thermospheric dayglow emissions obtained using a novel daytime imaging spectograph (Pallamraju et al. 2013) from Hyderabad (a low-latitude location in India) and radio (digisonde) measurements from Ahmedabad (typically a region under the crest of the equatorial ionization anomaly in Indian longitudes) of several years’ duration have been used to assess gravity wave activity, both in terms of their occurrence numbers and their vertical propagation speeds. From their study of the occurrence statistics, it was found that the percentage of days when waves in the spectral periods of gravity wave range occurred was greater for times of relatively higher solar activity duration (in 2013) compared to those of low solar activity (in 2011) (Laskar et al. 2015). Furthermore, by using a new method to extract neutral gravity wave information from digisonde measurements (Mandal et al. 2019), vertical wave propagation speeds and vertical scale sizes of waves were diagnosed. These revealed the presence of a systematic enhancement in these parameters with solar flux (Mandal et al. 2020; Mandal and Pallamraju 2020).

Overall, the community’s ability to explore the nature of coupling processes in the MALTI has advanced significantly during the period of the ROSMIC project.
New observation techniques, data analysis methods and approaches to instrument networks have been developed. These include advances to ground-based radar, lidar and airglow imagers, as well as satellite instruments. While the associated data sets and their analysis have clarified some aspects of coupling from the ground to the thermosphere/ionosphere, they have also generated many new questions.

4.3 Ionospheric processes and the equatorial electrojet

Forcing from the middle and lower atmosphere also affects the ionosphere and ionospheric currents. This connection is believed to occur due to alteration of the thermospheric winds though atmospheric waves, such as, planetary waves and tides. Winds in turn drive the daytime ionospheric dynamo, which means that especially the E-region electric field is affected. At mid and low latitudes, where the forcing from above is less direct, the forcing from below can cause alterations in ionospheric structures of a similar magnitudes to those occurring during geomagnetic storms.

The E-region dynamo mainly affects the daytime E-region currents, most prominently the Equatorial electrojet (EEJ) and the mid-latitude solar quiet (Sq) currents (e.g., Yamazaki and Maute (2017)), and the low-latitude plasma distribution. This occurs through the direct driving of the equatorial ionization anomaly by the equatorial E-region electric field (Stolle et al. 2008). Due to high conductivities along the magnetic equator at the dayside, the EEJ, being an additional signal to the low latitude Sq system, is especially sensitive to these variabilities. Its signals have recently been used extensively to investigate atmosphere/ionosphere interactions. Though the EEJ is typically eastward around local noon, it was found that there is a longitudinal dependence of the occurrence and strength of the westward electrojet (often called counter electrojet (CEJ)) in the morning and afternoon hours. This is mainly related to the well-known longitudinal wave-4 structure in the background equatorial electrojet strength resulting from nonmigrating solar tides and stationary planetary waves (Singh et al. 2018; Soares et al. 2018). Soares et al. (2018) further suggested that additional longitudinal variability in CEJ can be attributed to the effect of migrating tides and regional geomagnetic field gradients. In particular during SSW events, modulation of the ionosphere (i.e. the EEJ) has often been identified. Forbes and Zhang (2012); Yamazaki et al. (2012); Fejer et al. (2010), suggested that enhanced lunar semidiurnal tides play a major role in altering the equatorial electrodynamics. Later, the effects of both the semidiurnal lunar and solar tides were suggested as contributing factors (Siddiqui et al. 2018).

Most recently, bursts coherent with the quasi-6 day wave have been identified in the EEJ during the SH SSW of September 2019 (Yamazaki et al. 2020).

Post-sunset equatorial plasma irregularities are a severe threat to trans-ionospheric radio waves used for navigation, and can lead to their interruption (e.g., Xiong et al. (2016)). The formation of these irregularities is generally thought to be favored during the strong eastward electric field and upward plasma density gradients present in the lower F region after sunset when conditions are suitable for Rayleigh-Taylor instabilities to develop. Different scenarios have been proposed regarding the initiation of the growth of irregularities. A currently favored scenario proposes upward propagating gravity waves as the source of the initiating perturbations in the bottom side F region (see Kelley et al. (2003); Woodman (2009)). Recent work, based on multi-instrumental ground-based observation and novel regional analyses from satellite-based optical data, has provided further evidence of the connection between GWs and plasma irregularities (e.g., Sivakandan et al. (2019); Eastes et al. (2019); Das et al. (2020)).

4.4 Future challenges

While the study of wave associated upward coupling is a relatively new topic for ionospheric and thermospheric investigations, it has been under way for several decades in the middle atmosphere. The dissipation of gravity waves in the mesosphere is now generally accepted as the cause for the solsticial pole to pole circulation. However, the particular waves causing this circulation are yet to be identified observationally and models do not yet have sufficient resolution to faithfully simulate sources and propagation conditions so that these waves can be identified. Secondary wave generation has been identified as an additional mechanism for wave generation which further complicates our understanding of the height variation of the character of these waves. Thus, one of the major challenges for the future remains the identification of the sources, propagation conditions, interactions and dissipation mechanisms for these waves. A related challenge is determining the causes of constituent distributions in the mesosphere and lower thermosphere since they are dependent on the induced large scale flows and mixing associated with waves of all types. This requires the further development of a detailed understanding of the wave breaking process and any turbulence that results.

Solutions to these questions require advances in modeling capabilities (from whole atmosphere GCMs to high resolution simulations of wave propagation), the development of data assimilation techniques for the middle and upper atmosphere, instruments capable of higher spatial and temporal resolution and more extensive international observational networks. Local time variations in winds and temperatures associated with tides mean that short-term diagnoses (i.e., daily) of the dynamical conditions in the upper atmosphere using observations from single
satellites cannot be cleanly achieved. In principle, a multiple satellite mission or ground based chains of suitably selected instruments can address this problem but the issues of funding and international coordination make this difficult in practice. In addition, the lower mesosphere and lower thermosphere are difficult regions to observe because of the lack of suitable natural phenomena which can be easily and inexpensively observed. To observe these regions, new instrument techniques need to be developed.

The upward coupling of waves into the ionosphere and lower thermosphere is now an active area of research. Whole atmosphere GCMs which include ionospheric physics and ionospheric models which include realistic lower boundary conditions are currently available to explore this coupling and are being refined. As this research is at the interface between the neutral atmosphere and ionosphere, progress can best be made when researchers familiarize themselves with the complementary concepts and analysis context which have been developed to understand each region.

The lack of observations in the lower thermosphere (noted above) remains a significant concern. Without them, the models cannot be validated, and data assimilation for these heights remains a tentative venture. The establishment of suitable observing capabilities which allow the dynamical conditions in the middle and upper atmosphere to be monitored is the greatest challenge for future advances in this field.

5 Trends
Working group 3 of ROSMIC has been dealing with long-term trends in the mesosphere, lower thermosphere (MLT region) and related part of the ionosphere (below about 120 km). Working group 4 treated long-term trends in the thermosphere and ionosphere (above about 120 km). The key question for both working groups was as follows:

- What is the impact of anthropogenic activities on the mesosphere-thermosphere-ionosphere (MTI) system?

This is a long-term strategic question and ROSMIC contributed to developing answers to this question. Some important findings reached during the ROSMIC project, concerning trends in the MLT region, including the related part of the ionosphere, are briefly outlined in this section. In addition, a few results concerning higher levels are also mentioned, specifically those from studies dealing with effects throughout the MTI. It is important to realize that while long-term change in the atmospheric concentration of CO₂ is the primary driver of trends, some secondary trend drivers must also be taken into account, e.g., long-term changes of the stratospheric ozone concentration for trends in the MLT region.

The main driver of long-term trends in the MLT region is CO₂. The apparent discrepancy between the observed (ACE/FTS and TIMED/SABER) and modeled long-term trend of the CO₂ volume mixing ratio has been resolved (Qian et al. 2017; Rezac et al. 2018). The reason for the apparent difference was found to be the neglect of some influences on observational data during the derivation of concentrations from observations. The observed trend of CO₂ in the MLT region now does not differ significantly from the modeled trend of about 5%/decade. This is approximately equal to the trend at Earth's surface.

With the issue of CO₂ mixing ratio trend resolved, attempts to model the atmospheric response to trace gas trends gain additional credibility. Global simulations of temperature change due to anthropogenic trace gas emissions, that extended from the surface, through the thermosphere and ionosphere, to the exobase were performed by Solomon et al. (2018); Solomon et al. (2019) using the Whole Atmosphere Community Climate Model – eXtended (WACCM-X). Warming in the lower atmosphere switched to cooling above the lower stratosphere. The cooling is dramatic in the thermosphere, 2–3 K per decade for the global mean annual mean, with the higher rates corresponding to solar minimum conditions, and the lower rates for solar maximum. This dependence of global change on solar activity is due to solar-driven increases in radiatively active gases other than carbon dioxide, such as nitric oxide. The dependence is fairly small in comparison with previous observational and modeling studies, but is commensurate with updated satellite drag analysis by Emmert (2015). They found that any trend magnitude dependence on solar cycle is weak, relative to the trend uncertainties. The reason for this study was that since solar and geomagnetic activity have tended to decline over the past several cycles, it has become difficult to disentangle the relative effects of increasing CO₂ over the same time period.

An ancillary result of these simulations is an estimate of the solar cycle effect on temperatures as a function of altitude. These simulations used modest, five-member ensembles. Observed sea surface temperatures rather than a fully coupled ocean model was used, so any solar cycle effects were not statistically significant in the lower troposphere. Temperature change from solar minimum to maximum increased from near-zero at the tropopause, to about 1 K at the stratopause, to approximately 500 K in the upper thermosphere, commensurate with the empirical evidence, and previous numerical models.

Changes in stratospheric ozone concentration play an important role in trends in the MLT region. Observational data reveal a change of temperature trends in the MLT region that coincide with a change of ozone trends in the second half of the 1990s. The trends in the mesopause region (∼90 km) temperature change from no trends...
to negative trends, whereas trends in mesospheric (~75 km) temperatures become less negative. WACCM4 model simulations confirm that temperature trends in the mesosphere are clearly different in 1975–1995 versus 1995–2015. There is also a tendency that the trends of both zonal and meridional summer prevailing winds are changing after 1995. Using a numerical circulation model with realistic ozone and carbon dioxide fields and with tropospheric/stratospheric circulation taken from ERA reanalyses, the observed trend changes can be reproduced qualitatively (Jacob et al. 2015).

Despite several objections (von Zahn 2003; Thomas et al. 2003), polar mesospheric clouds (PMC) are still considered to play a major role in the evaluation of trends in the summer mesosphere at mid- to polar latitudes. Observations of PMC by ground-based observers (visual and lidar) and by satellites now cover several decades which allows the study of the dependence of the morphology of PMC on various processes such as the solar cycle, volcanoes, and greenhouse gases such as CO$_2$ and H$_2$O. For example, the ice water content (IWC) in bright clouds increased rapidly in the period 1979–1990 but became almost constant in the period 1990–2013. This pattern correlates with the behavior of stratospheric ozone. Sensitivity analysis indicates comparable roles for the influence of temperature and H$_2$O on boreal trends in IWC but mainly a temperature influence on austral trends in IWC (Hervig et al. 2016).

Recent studies using LIMA/MIMAS (Leibniz-Institute Middle Atmosphere model; Mesospheric Ice Microphysics And tranSport model) for a period of 138 years (1871–2008) suggest that long-term trends of PMC parameters are most prominent after 1960 when the increase of CO$_2$ and H$_2$O accelerates (Lübken et al. 2018; Lübken et al. 2020). Trends at polar latitudes are rather different from those at middle latitudes. Some tendencies are caused by the increase of water vapor (due to increasing methane abundance): for example brightness, albedo, backscatter coefficient (relevant for lidar), and occurrence rates. On the other hand, the weak decrease of (geometric) PMC altitude is primarily due to a shrinking of the atmosphere below PMC heights. LIMA/MIMAS model results are consistent with long-term observations by ground based observers (visual or lidar) and by satellite measurements (Bremer and Berger 2002; Kirkwood et al. 2008; Pertzsev et al. 2014; Fiedler et al. 2017; Russell III et al. 2015). It should be noted that the observational data base is still limited and does not yet allow for a clear cut statement regarding the role of PMC for detecting climate change effects in the middle atmosphere (Lübken et al. 2018; Lübken et al. 2020).

The solar cycle has a direct impact on PMC as has been demonstrated in many experimental and simulation studies. They show that there is a clear anti-correlation between solar activity and PMC characteristics (brightness, occurrence, etc.) which is caused by the effects of solar heating and the photo-dissociation of water vapor (Lübken et al. 2009). Surprisingly, a solar cycle response is missing in satellite observations after 2002 which is presumably caused by a not yet fully understood suppression of the solar cycle response of water vapor (Thomas et al. 1991; Hervig et al. 2019).

A long (1979–2016) series of winds, combining low frequency and meteor radar wind measurements at Collm (51° N, 13° E), Germany reveals trends in zonal and meridional winds. The trend in zonal wind is a gradual change from strong eastward to weak eastward. The trend in meridional wind is a weakening of its magnitude over time (both northward and southward) in all months of the year. Figure 11 illustrates this result, with fewer regions exceeding ±8 m/s (orange and red in winter, blue and green in summer) in recent years.

Qian et al. (2019a) compared model simulations with WACCM-X with the wind measurements from Collm and found that dynamics have a significant impact on trends in mesospheric temperatures and, in particular, in mesospheric winds, where the effects of dynamics appear to be of primary importance. Global average mesospheric temperature trends peaked in the lower and middle mesosphere in qualitative agreement with observations. The observed trends in winds were stronger and more statistically significant than the simulated trends.

Liu et al. (2017) analyzed latitude-time cross-sections to determine the linear trend term of the relative GW potential energy as a function of height during each month for the altitude range of 30–100 km based on SABER data over 2002–2015. They found the monthly trends to be both positive and negative, changing remarkably from month to month. This is an example of temporal variability of trends in atmospheric wave activity. Examples of significant regional variability are also known. This variability makes investigation of trends in wave activity difficult.

Hall et al. (2016) reported some increase in the turbopause height in summer at Tromsø, (70° N) and a statistically significant change of the turbopause height in Saskatoon (52° N) and in Tromsø in winter. From those results they inferred a decrease of atomic oxygen concentration in the lower thermosphere.

Calculation of ionospheric trends requires removal of solar cycle effects from the data. Analysis of trends in foE (heights around 110 km) shows that the solar correction (using both F10.7 and solar Lyman-α flux) changes with time (Laštovička et al. 2016), perhaps due to changes in the relation between solar ionizing radiation and solar proxies (there are indications of such changes on the Sun in cycles 23 and 24, e.g., Balogh et al. (2014)). Danilov and Konstantinova (2019) found a strong seasonal and diurnal
variation of trends in foE, the trends being remarkably stronger in winter than in summer.

5.1 Future challenges
Progress in investigating long-term trends in the MLT region, which largely reflect the global change in the atmosphere, during VarSITI was significant but many questions remain open. Future investigations of trends in MLT should deal with filling gaps in scenario of trends in the MLT region and removal/explanation of controversies, particularly with respect to the following:

1. Investigations of trends in atmospheric wave activity, which is the key problem of trends in the MLT region
2. Further development and improvement of complex models
3. Monitoring and investigations of changes of secondary (= non-CO$_2$) trend drivers and of their impacts on trends

6 Conclusions
The ROSMIC science topics covered large parts of the middle atmosphere, thermosphere, ionosphere, and climate research. As we have seen above, significant progress on various aspects of these was made during VarSITI, with access to improved models and continued high-quality observations proving critical for these advances. One of the remaining substantial challenges is considering the picture, presented in Fig. 1, as a whole. While this is not a straightforward task, it will be critical for the next SCOSTEP science program, PRESTO, **PREdictability of variable Solar-Terrestrial cOupling**, in order to assess how different parts of the system interact, with the goal of improving predictability in different scales, from temporal to spatial.

Evident in all topics, is the importance of the interactions between various components of the terrestrial part of the Sun/Earth system. While the specification of particular parameters have been identified as targets for future work (i.e., flux and energy spectra of electrons and constituent transport—downward coupling; the gravity wave spectrum and its dissipation—upward coupling; wave activity—trends), they are important in terms of allowing the full system and its interactions to be better defined. A prime example of this is the SSW which was identified as significant in several contexts. For downward coupling, it was notable for its role in the downward transport of constituents and changes in the large-scale thermal structure. In upward coupling, it was implicated in the development of global meridional teleconnections, as a modifier of wave signatures and middle atmosphere circulation and as the driver of significant changes in ionospheric current systems. A significant development during ROSMIC is the recognition that although different
aspects of SSWs are important depending on the context, inclusion of all aspects is necessary to fully understand them and their impact.

Mention was made of the successful incorporation of satellite and ground-based observations to provide a global perspective on various phenomena during ROSMIC. Also highlighted was the successful development of whole atmosphere models and associated parameterizations to provide the means to explore the interactions between the various components of the atmosphere/ionosphere. These two scientific thrusts need to continue being supported in the future. Although the recently launched GOLD (Eastes et al. 2017) and ICON (Immel et al. 2018) missions will substantially enhance our understanding of the ionosphere and thermosphere, it is concerning that plans to ensure continuation and enhancement of the satellite measurements that led to our current understanding of the MLT are not in place. They would provide an extremely valuable complement to these new missions.

As has been emphasized throughout this paper, we are moving past the time in solar terrestrial system science when the study of phenomena in isolation is of primary importance. Work which reveals the connections and linkages between components and phenomena is becoming essential. PRESTO is organized to support this latter type of investigation.

As the development of a Sun to Earth system description proceeds, it is important to keep in mind that much of the solar energy flux is engaged in maintaining the large-scale structures of the atmosphere. Distinguishing the underlying processes which are essential to maintaining these structures from those which manifest short term responses to solar variability is vital to assessing the manners in which our atmosphere will evolve as a result of anthropogenic influences. Establishing these distinctions also presents challenges to the solar-terrestrial community as a whole. For example, the time scales of the forcings and responses in each part of the system vary significantly which can lead to misunderstandings across disciplines. The societal relevance of understanding the role of this coupling is tremendous, not least because of the whole atmosphere impacts of anthropogenic influences and any geo-engineering solutions that might be proposed to counter them.

Abbreviations
CAMS: Climate and Weather of the Sun-Earth System; CCM: Chemistry-Climate model; CE2: Counter electrojet; CFC: Chlorofluorocarbon; CME: Coronal mass ejection; CNIPs: Coupled Model Intercomparison Project Phase 6; EE: Equatorial electrojet; EEP: Energetic electron precipitation; EPP: Energetic particle precipitation; ESI: Elevated stratosphere; EUV: Extreme ultraviolet; FUV: Far ultraviolet; GCM: General Circulation Model; GHG: Greenhouse gas; GOL: Global-scale Observations of the Limb and Disk; GW: Gravity wave; ICON: Ionospheric Observation and Modelling System; IPCC: Intergovernmental Panel for Climate Change; IWC: Ice Water Content; MALTI: Middle Atmosphere/Lower Thermosphere/Ionosphere; MEE: Medium energy electrons; MIL: Mesospheric inversion layer; MLT: Mesosphere lower-thermosphere; MLUV: Middle ultraviolet; NASA: National Aeronautics and Space Administration; NAO: North Atlantic Oscillation; NH: Northern Hemisphere; NIR: Near infrared; NUV: Near ultraviolet; PMC: Polar mesospheric cloud; PRESTO: Predictability of variable Solar-Terrestrial eCoupling; QBO: quasi-biennial oscillation; QDT: Quarterdiurnal tide; SCOSTEP: Scientific Committee on Solar-Terrestrial Physics; SH: Southern Hemisphere; SPE: Solar proton event; SSI: Solar spectral irradiance; SSW: Sudden stratospheric warming; SXR: Soft X-rays; TEC: Total electron content; TSI: Total solar irradiance; UBC: Upper boundary condition; UV: Ultraviolet; VIS: Visible (radiation); VUV: Vacuum ultraviolet

Authors’ contributions
WW and AS coordinated the writing of the manuscript. All authors contributed to the writing or providing comments on the text. All authors read and approved the final manuscript.

Funding
WW and AS gratefully acknowledge funding from the ISEE (Institute for Space-Earth Environmental Research), Nagoya University, for support to attend the VarSITI summarizing workshop where the results discussed in this review paper were presented.

Declarations
Availability of data and materials
Not applicable
Competing interests
The authors declare that they have no competing interests.

Author details
1 Department of Physics, University of New Brunswick, PO Box 4400, Fredericton, Canada. 2 Department of Physics, University of Otago, PO Box 56, Dunedin, New Zealand. 3 George Mason University, 4400 University Drive, Fairfax, USA. 4 National Institute of Polar Research, 10-3, Midori-cho, Tachikawa-shi, Tokyo, Japan. 5 Geoforschungszentrum Potsdam, Telegrafenberg, Potsdam, Germany. 6 Institute of Atmospheric Physics, Czech Academy of Sciences, Bocni II, Prague, Czech Republic. 7 Laboratory for Atmospheric and Space Physics (LASP), University of Colorado (CU), 3665 Discovery Drive, Boulder, USA. 8 Leibniz-Institute of Atmospheric Physics, Schloss-Str. 6, Kühlingborn, Germany. 9 High Altitude Observatory, National Center for Atmospheric Research, PO Box 3000, Boulder, USA. 10 Faculty of Engineering and Physical Sciences, University of Leeds, Leeds, UK. 11 Instituto de Astrofísica de Andalucía, CSIC, Granada, Spain. 12 Space and Atmospheric Sciences Division, Physical Research Laboratory, Ahmedabad, India.

Received: 15 October 2020 Accepted: 8 June 2021
Published online: 18 August 2021

References
Alexander MJ (1998) Interpretations of observed climatological patterns in stratospheric gravity wave variance. J Geophys Res 103:8627–8640
Aiken P, Maute A, Richmond A (2017). The F-region gravity and pressure gradient current systems: a review. Space Sci Rev 206(1):431–450. https://doi.org/10.1007/s11214-016-0266-z
Andersson ME, Verronen PT, Rodger CJ, Clilverd MA, Seppälä A (2014) Missing driver in the Sun-Earth connection from energetic electron precipitation impacts mesospheric ozone. Nat Comm 5(1):401–405
Andersson ME, Verronen PT, Marsh DR, Seppälä A, Clilverd MA, Kalakoski N, van de Kamp M (2018) Polar ozone response to energetic particle precipitation over decadal timescales: the role of medium-energy electrons. J Geophys Res 123(1):607–622
Arsenovic P, Rozanov EV, Stenke A, Funke B, Wissing JM, Mursula K, Tummon F, Peter T (2016) The influence of middle range energy electrons on atmospheric chemistry and regional climate. J Atmos Sol-Terr Phys 149:180–190

https://doi.org/10.1007/s11214-016-0266-z
solar magnetic activity band interaction and instabilities that shape quasi-periodic variability. Nat Commun 6:6491. https://doi.org/10.1038/ncomms7491

McLandress C, Shepherd TG, Polavarapu S, Beagley SR (2012) Is missing orographic gravity wave drag near 60°S the cause of the stratospheric zonal wind biases in chemistry/climate models? J Atmos Sci 69:802–818

Medvedev AS, Klaassen GP (2000) Parameterization of gravity wave-momentum deposition based on nonlinear wave interactions: basic formulation and sensitivity tests. J Atmos Sol-Terr Phys 62:1015–1033

Medvedev AS, Yiğit E, Hartplog P (2017) Ion friction and quantification of the geomagnetic influence on gravity wave propagation and dissipation in the thermosphere-ionosphere. J Geophys Res Space 122(12):12,464–12,475. https://doi.org/10.1002/2017JA024785

Moffat-Griffin T (2019) An introduction to atmospheric gravity wave science in the polar regions and first results from ANGWIN. J Geophys Res Atmos 124(3):1198–1199

Moffat-Griffin T, Taylor M, Nakamura T, Murphy D, Bagesten JV, Lee G (2019) Atmospheric gravity wave science in the polar regions. Eos 100. https://doi.org/10.1029/2019EO100701

Namgaladze A, Korenkov VN, Klimenko V, et al (1988) Global model of the thermosphere-ionosphere-protonosphere system. PACEOPH 127:219–254. https://doi.org/10.1007/BF00679812

Nayak C, Yiğit E (2019) Variation of small-scale gravity wave activity in the ionosphere during the major sudden stratospheric warming event of 2009. J Geophys Res Space 124:470–488. https://doi.org/10.1029/2018JA026048

Nesse Tyssøy H, Sandøen ML, Ødegaard LKG, Stadsnes J, Aasnes A, Zawedde AE (2016) Energetic electron precipitation into the middle atmosphere - constructing the loss cone fluxes from MEPEL POES. J Geophys Res 21(6):5693–5707

Nesse Tyssøy H, Haderlein A, Sandøen ML, Stadsnes J (2019) Intercomparison of the POES/MEPEL loss cone electron fluxes with the CMIP6 parametrization. J Geophys Res Space 124(1):628–642

Neupert WM (1968) Comparison of solar X-ray line emission with microwave emission during flares. Astrophys J Lett 153:159. https://doi.org/10.1086/180220

Newnham DA, Clilverd MA, Rodger CJ, Hendrickx K, Megner L, Kavanagh AJ, Seppälä A, Verronen PT, Andersson ME, Marsh DR, Kovacs T, Feng W, Plane JMC (2018) Observations and modeling of increased nitric oxide in the antarctic polar middle atmosphere associated with geomagnetic storm-driven energetic electron precipitation. J Geophys Res 123(7):6009–6025

 niche AL, Öberheide J, Mlynářík MG, Marsh DR, Gan Q (2019) Solar cycle variability of nonmigrating tides in the 5.3 and 15 μm infrared cooling of the thermosphere (100–150 km) from SABER. J Geophys Res 124(3):2338–2356. https://doi.org/10.1029/2018JA026356

Öberheide J, Shikawa K, Gurubaran S, Ward WE, Fujiwara H, Kosch M, Makela JJ, Takahashi H (2015) The geospace response to variable inputs from the lower atmosphere: a review of the progress made by Task Group 4 of CAWSES-II. Prog Earth Planet Sci 2(1):2

Olsen N, Friis-Christensen E, Floberghagen R, et al. (2013) The swam satellite constellation application and research facility (scarf) and swarm data products. Earth Planets Space 65(1):1189–1200. https://doi.org/10.5047/eps.2013.07.001

Orsolin IJ, Smith-Johnson C, Marsh DR, Stordal F, Rodger CJ, Verronen PT, Clilverd MA (2018) Mesospheric nitric acid enhancements during energetic electron precipitation events simulated by WACCM-X. J Geophys Res 123(13):6984–6998

Orland DA (2017) Daily estimates of the migrating tide and zonal mean temperature in the mesosphere and lower thermosphere derived from SABER data. J Geophys Res Atmos 122(7):3754–3785. https://doi.org/10.1002/2016JD025573

Oyama S, Kero A, Rodger CJ, Clilverd MA, Miyoshi Y, Partamies N, Turunen E, Raita T, Verronen PT, Saito S (2017) Energetic electron precipitation and auroral morphology at the substorm recovery phase. J Geophys Res 122(A4):273

Pallamraju D, Laskar FJ, Singh RP, Baumgardner J, Chakrabarti S (2013) Mise: a multiwavelength imaging spectrograph using echelle grating for daytime optical aeronomy investigations. J Atmos Sol-Terr Phys 103:176–183

Pancheva D, Mukhtarov P, Mitchell NJ, Merzlyakov E, Smith AK, Andonov B, Singer W, Hocking W, Meek C, Manson A, Murayama Y (2008) Planetary waves in coupling the stratosphere and mesosphere during the major stratospheric warming in 2003/2004. J Geophys Res Atmos 113(D12). https://doi.org/10.1029/2007JD009011

Pancheva D, Mukhtarov P, Andonov B, Mitchell NJ, Forbes JM (2009) Planetary waves observed by TIMED/SABER in coupling the stratosphere-mesosphere-lower thermosphere during the winter of 2003/2004 part 1 – comparison with the UKMO temperature results. J Atmos Sol-Terr Phys 71:61–74. https://doi.org/10.1016/j.jastp.2008.09.016

Peck ED, Randall CE, Green JC, Rodriguez JV, Rodger CJ (2015) POES MEPEL differential flux retrievals and electron channel contamination correction. J Geophys Res 120(6):4596–4612

Pedatella NM, Chau JL, Schmidt H, Gershenson LP, Stolle C, Hocker K, Harvey VL, Funke B, Siddiqui TA (2018a) How sudden stratospheric warming affects the whole atmosphere. Eos 99. https://doi.org/10.1029/2018EO092441

Pedatella NM, Liu H-L, Marsh DR, Raeder K, Anderson JL, Chau JL, Gershenson LP, Siddiqui TA (2018b) Analysis and hindcast experiments of the 2009...
Smith KLV, Neely RR, March DR, Polvani LM (2014) The Specified Chemistry Whole Atmosphere Community Climate Model (SC-WACCM). J Adv Mod Earth Sys 6(3):883–901. https://doi.org/10.1002/2014MS000346

Smith-Johnsen C, Nesse Tysysøy H, Hendrickx K, Orsolini YJ, Kumar GK, Ødegaard LKG, Sandanger MI, Stordal F, Megner L (2017) Direct and indirect electron precipitation effect on nitric oxide in the polar middle atmosphere, using a full-range energy spectrum. J Geophys Res Space 122(8):8679–8693

Smith-Johnsen C, Marsh DR, Orsolini YJ, Nesse Tysysøy H, Hendrickx K, Sandanger MI, Ødegaard LKG, Stordal F (2018) Nitric oxide response to the April 2010 electron precipitation event: using WACCM and WACCM-D with and without medium-energy electrons. J Geophys Res Space 123(6):5232–5245

Soares G, Yarnazaki Y, Matzka J, Pinheiro K, Morschhauser A, Stolle C, Allen P (2018) Equatorial counter electrojet longitudinal and seasonal variability in the American sector. J Geophys Res Space 123(11):9906–9920. https://doi.org/10.1002/2018JA025968

Solanki SK, Krivova NA, Haigh JD (2013) Solar irradiance variability and climate. ARA&A 51(1):311–351. https://doi.org/10.1146/annurev-astro-082812-140007

Solomon SC, Woods TN, Didkovsky LV, Emmert JT, Qian L (2010) Anomalously low solar extreme-ultraviolet irradiance and thermospheric density during solar minimum. Geophys Res Lett 37(16):116103. https://doi.org/10.1029/2010GL044468

Solomon SC, Qian L, Didkovsky LV, Viereck RA, Woods TN (2011) Causes of low thermospheric density during the 2007-2009 solar minimum. J Geophys Res Space 116:1A0004. https://doi.org/10.1029/2010JA016508

Solomon SC, Qian L, Burns AG (2013) The anomalous ionosphere between solar cycles 23 and 24. J Geophys Res Space 118(10):6524–6535. https://doi.org/10.1002/2012JA017400

Stober G, Chau JL, Vierinen J, Jacobi C, Wilhelm S (2018) Retrieving horizontally resolved energy spectra of polar mesospheric clouds by the solar backscattered ultraviolet spectral radiometer: evidence of a solar cycle dependence. J Geophys Res 96:927–939

Thomas GE, Oliveira JJ, Deland M, Shettle EPP (2003) Comment on “are noctilucent clouds truly a ‘miner’s canary’ for global change?”. EOS Trans Am Geophys Union 84(36):352–353. https://doi.org/10.1029/2003EO060008

Tomikawa Y, Sato K, Watanabe S, Kawatani Y, Miyazaki T, Takahashi M (2012) Growth of planetary waves and the formation of an elevated stratosphere after a major stratospheric warming in a T21L256 GCM. J Geophys Res 117(D16):017. https://doi.org/10.1029/2011JD017243

Trenberth KE, Fasullo JT, Kiehl J (2009) Earth’s global energy budget. Bull Amer Meteorol Soc 90(3):311–324. https://doi.org/10.1175/2008BAMS2634.1

Tsuchiya S, Shiokawa K, Fujimin H, Otsuba Y, Nakamura T, Yamamoto M (2018) Statistical analysis of the phase velocity distribution of mesospheric and ionospheric waves observed in airmode images over a 16-year period: comparison between Rikubetsu and Shigaraki, Japan. J Geophys Res Space 123(8):6930–6947

Turunen E, Kero A, Verronen, Miyoshi Y, Oyama S, Sato S (2016) Mesospheric ozone destruction by high-energy electron precipitation associated with pulsatling aurora. J Geophys Res 121(19):11,852–11,861

van de Kamp M, Seppälä A, Clilverd MA, Rodger CJ, Verronen PT, Whittaker IC (2016) A model providing long-term datasets of energetic electron precipitation during geomagnetic storms. J. Geophys. Res. Atmos. 121(20):12,520–12,540

van de Kamp M, Rodger CJ, Seppälä A, Clilverd MA, Verronen PT (2018) An updated model providing long-term data sets of energetic electron precipitation including, zonal dependence. J Geophys Res 149(9):180

Verronen PT, Andersson ME, Marsh DR, Kovacs T, Plane JMC (2016) WACCM-D whole atmosphere community climate model with D-region ion chemistry. J Adv Mod Earth Sys 8(2):954–975

Vierinen J, Chau JL, Charuvi H, Urco JM, Clahsen M, Avarksivov V, et al. (2019) Observing mesospheric turbulence with specular meteor radars: a novel method for estimating second-order statistics of wind velocity. Earth Space Sci 1171–1195. https://doi.org/10.1029/2019EA000570

Vincent RA, Hertzog A (2014) The response of superpressure balloons to gravity wave motions. Atmos Meas Tech 7:1043–1055

Vitharana A, Zhu X, Du J, Oberheide J, Ward WE (2019) Statistical modeling of tidal wave motions. Atmos Meas Tech 7:1043–1055

von der Weid LE (2014) A model providing long-term datasetsof energetic electron precipitation during geomagnetic storms. J Geophys Res. Atmos. 121(20):12,520–12,540

von der Weid LE (2014) A model providing long-term datasetsof energetic electron precipitation, including zonal dependence. J Geophys Res 149(9):180

Verronen PT, Andersson ME, Marsh DR, Kovacs T, Plane JMC (2016) WACCM-D whole atmosphere community climate model with D-region ion chemistry. J Adv Mod Earth Sys 8(2):954–975

Vierinen J, Chau JL, Charuvi H, Urco JM, Clahsen M, Avarksivov V, et al. (2019) Observing mesospheric turbulence with specular meteor radars: a novel method for estimating second-order statistics of wind velocity. Earth Space Sci 1171–1195. https://doi.org/10.1029/2019EA000570

Vincent RA, Hertzog A (2014) The response of superpressure balloons to gravity wave motions. Atmos Meas Tech 7:1043–1055

Vitharana A, Zhu X, Du J, Oberheide J, Ward WE (2019) Statistical modeling of tidal wave motions in the mesosphere and lower thermosphere. J Geophys Res Atmos 124(16):9011–9027. https://doi.org/10.1002/2019JD030573

von Schuckmann K, Palmer M, Trenberth K, et al. (2016) An imperative to monitor earth’s energy imbalance. Nature Climate Change 6:138–144. https://doi.org/10.1038/nclimate2876

von Zahn U (2003) Are noctilucent clouds a ‘miner’s canary’ for global change? EOS Trans Am Geophys Union 84(28):261–264. https://doi.org/10.1029/2003EO280001

Watanabe S, Miyahara S (2009) Quantification of the gravity wave forcing of the migrating diurnal tide in a gravity wave–resolving general circulation model. J Geophys Res 114:D01781

Wenzer T, Solanki SK, Krivova NA, Fliri DM (2004) Comparison between KPVT/SPM and SoHO/MDI magnetograms with an application to solar irradiance reconstructions. A&A 427:1031–1043. https://doi.org/10.1051/0004-6361:20041313

Woodman RF (2009) Spread F – an old equatorial aeronomy problem finally resolved? Ann Geophys 27(5):1915–1934. https://doi.org/10.5194/angeo-27-1915-2009

Woods TN, Eparvier FG, Hock R, Jones AR, Woodraska D, Judge D, Didkovsky L, Lean J, Mariska J, Warren H, McMullen D, Chamberlin P, Bertheaume G, Bailey S, Fuller-Rowell T, Syrja J, Tobiska KW, Viereck R (2012) Extreme Ultraviolet Variability Experiment (EVE) on the Solar Dynamics Observatory (SDO): overview of science objectives, instrument design, data products, and model developments. Solar Phys 275(1-2):115–143. https://doi.org/10.1007/s11207-009-9487-6
