Global Sensitivity Analysis and Uncertainty Quantification for Background Solar Wind using the Alfvén Wave Solar Atmosphere Model

Aniket Jivani[1], Nishtha Sachdeva[1], Zhenguang Huang[1], Yang Chen[1], Bart van der Holst[1], Ward Manchester[1], Daniel Iong[1], Hongfan Chen[1], Shasha Zou[1], Xun Huan[1], and Gabor Toth[1]

[1]University of Michigan

November 23, 2022

Abstract

Modeling the impact of space weather events such as coronal mass ejections (CMEs) is crucial to protecting critical infrastructure. The Space Weather Modeling Framework (SWMF) is a state-of-the-art framework that offers full Sun-to-Earth simulations by computing the background solar wind, CME propagation and magnetospheric impact. However, reliable long-term predictions of CME events require uncertainty quantification (UQ) and data assimilation (DA). We take the first steps by performing global sensitivity analysis (GSA) and UQ for background solar wind simulations produced by the Alfvén Wave Solar atmosphere Model (AWSoM) for two Carrington rotations: CR2152 (solar maximum) and CR2208 (solar minimum). We conduct GSA by computing Sobol indices that quantify contributions from model parameter uncertainty to the variance of solar wind speed and density at 1 au, both crucial quantities for CME propagation and strength. Sobol indices also allow us to rank and retain only the most important parameters, which aids in the construction of smaller ensembles for the reduced-dimension parameter space. We present an efficient procedure for computing the Sobol indices using polynomial chaos expansion (PCE) surrogates and space-filling designs. The PCEs further enable inexpensive forward UQ. Overall, we identify three important model parameters: the multiplicative factor applied to the magnetogram, Poynting flux per magnetic field strength constant used at the inner boundary, and the coefficient of the perpendicular correlation length in the turbulent cascade model in AWSoM.

Global Sensitivity Analysis and Uncertainty Quantification for Background Solar Wind using the Alfvén Wave Solar Atmosphere Model

Aniket Jivani[1], Nishtha Sachdeva[2], Zhenguang Huang[2], Yang Chen[3], Bart van der Holst[2], Ward Manchester[2], Daniel Iong[3], Hongfan Chen[1], Shasha Zou[2], Xun Huan[1], Gabor Toth[2]

[1]Dept. of Mechanical Engineering, University of Michigan, Ann Arbor, MI, USA
[2]Dept. of Climate and Space Sciences and Engineering, University of Michigan, Ann Arbor, MI, USA
[3]Dept. of Statistics, University of Michigan, Ann Arbor, MI, USA

Key Points:

- We perform global sensitivity analysis for background solar wind simulations of the Alfvén Wave Solar Atmosphere Model.
- We identify and retain only the most important uncertain parameters from the sensitivity analysis results.
- We carry out the analysis for examples of both solar maximum and solar minimum conditions.

Corresponding author: Aniket Jivani, ajivani@umich.edu

Abstract

Modeling the impact of space weather events such as coronal mass ejections (CMEs) is crucial to protecting critical infrastructure. The Space Weather Modeling Framework (SWMF) is a state-of-the-art framework that offers full Sun-to-Earth simulations by computing the background solar wind, CME propagation and magnetospheric impact. However, reliable long-term predictions of CME events require uncertainty quantification (UQ) and data assimilation (DA). We take the first steps by performing global sensitivity analysis (GSA) and UQ for background solar wind simulations produced by the Alfvén Wave Solar atmosphere Model (AWSoM) for two Carrington rotations: CR2152 (solar maximum) and CR2208 (solar minimum). We conduct GSA by computing Sobol' indices that quantify contributions from model parameter uncertainty to the variance of solar wind speed and density at 1 au, both crucial quantities for CME propagation and strength. Sobol' indices also allow us to rank and retain only the most important parameters, which aids in the construction of smaller ensembles for the reduced-dimension parameter space. We present an efficient procedure for computing the Sobol' indices using polynomial chaos expansion (PCE) surrogates and space-filling designs. The PCEs further enable inexpensive forward UQ. Overall, we identify three important model parameters: the multiplicative factor applied to the magnetogram, Poynting flux per magnetic field strength constant used at the inner boundary, and the coefficient of the perpendicular correlation length in the turbulent cascade model in AWSoM.

Plain Language Summary

Space weather events such as those driven by coronal mass ejections (CMEs) can result in severe geomagnetic storms that impact critical infrastructure. Accurate long-term forecasts are therefore needed together with uncertainty quantification. In this work, we calculate uncertainty and perform sensitivity analysis for the background solar wind that has a major impact on the accuracy of the overall CME simulation. Since these models have many parameters that carry uncertainty, sensitivity analysis allows us to identify the most important ones.

1 Introduction

Coronal mass ejections (CMEs) are large-scale eruptions of the solar coronal plasma and magnetic fields expelled into the solar wind. CMEs can create magnetic storms in the Earth's magnetosphere that are responsible for severe geomagnetic effects ranging from breakdown in radio communications to damage of sensitive electronics on satellites and even disrupting the power grid. Therefore it is imperative to obtain reliable long-term predictions of space weather events driven by CMEs.

Current state-of-the-art modeling capabilities involve numerical simulations using coupled first-principles and/or empirical models. A prominent example is the Space Weather Modeling Framework (SWMF) (Tóth et al. (2005, 2012); Gombosi et al. (2021)) that models domains from the upper solar chromosphere to the Earth's atmosphere and/or the outer heliosphere using efficient coupling between multiple models and is capable of full Sun-to-Earth simulations. Typically, as shown in Figure 1, the model chain consists of obtaining the background solar wind in Stage 1, generating and propagating a CME through the heliosphere to Earth in Stage 2, and finally calculating the magnetospheric impact via geospace models in Stage 3. Along the way, various observation data (in the blue boxes) are also available to calibrate or validate the model. The SWMF offers predictions for several macroscopic plasma quantities, including those that critically impact the magnetosphere and the resulting geomagnetic perturbations, such as the north-south component of the magnetic field, proton density, and solar wind velocity.

These models have seen continued improvements and their predictions have been validated for various phases of the solar cycle against a suite of observations, for instance by Jin et al. (2012); Sachdeva et al. (2019, 2021) and van der Holst et al. (2022). However, reliable long-term predictions of impact as well as the uncertainty surrounding the predictions are crucially needed for informed decision-making in operational settings. Producing a probabilistic forecast in such settings is challenging. The uncertainty space is high-dimensional and the dimensions grow as the simulation is propagated through the model chain (Figure 1). Coupled with the high computational cost of simulations, it becomes costly, even prohibitive, to produce an ensemble of runs that accurately portrays the uncertainty of the overall system. Updating the uncertainty over the course of a simulation with newly acquired remote and in-situ observations of space weather events is also non-trivial but highly important. Consequently, systematic uncertainty quantification (UQ) and data assimilation (DA) are needed to address these challenges.

UQ involves characterizing the uncertainty for a system. Uncertainty may arise due to unknown model parameters (e.g., the Poynting flux emanating from the photosphere and driving and heating the solar wind), incomplete initial and boundary conditions (e.g., the solar magnetograms that greatly impact solar wind solutions and have major uncertainty in estimating the magnetic field near the polar regions), missing or simplified physics (e.g., magnetic reconnection, auroral arcs), etc. We focus on parametric uncertainty in this work. UQ tasks may be broadly divided into two types: forward UQ and inverse UQ (e.g., see (Debusschere et al., 2017)). Forward UQ entails the propagation of uncertainty from inputs to outputs of a model; inverse UQ deals with updating (reducing) the uncertainty of model parameters (and subsequent model predictions and their uncertainty) given new observation data. The key difference is that the former is data-free while the latter incorporates data; the latter is thus also referred to as DA especially in the context of state-space models from geophysical research. Our main goal is to develop the Michigan Sun-to-Earth model with Quantified Uncertainties and Data Assimilation (MSTEM-QUDA) that is capable of forward and inverse UQ (i.e. UQ and DA) for each of the main stages for simulating a CME event from the Sun to Earth. As shown in Figure 1, we will propagate uncertainty from a stage's parameters, update the uncertainty with relevant observation data and generate a more confident ensemble of simulations, before passing them onto the next stage. For this paper, we will focus on the forward UQ part of Stage 1: background solar wind, using simulations produced by the Alfvén Wave Solar atmosphere Model (AWSoM) within the SWMF.

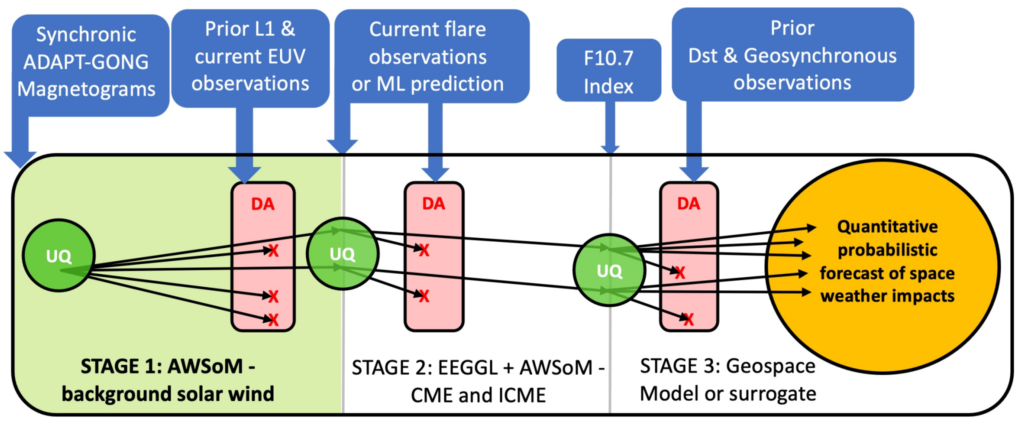

Figure 1: Flow outline of the Michigan Sun-to-Earth model with Quantified Uncertainties and Data Assimilation (MSTEM-QUDA). This paper focuses on forward UQ for the highlighted Stage 1: background solar wind.

Forward UQ is typically carried out using Monte Carlo sampling (i.e. ensemble techniques): first generating samples of input parameters from their uncertainty distribution, then running the model at each sample and lastly analyzing the distribution of the resulting outputs. The number of samples (i.e. simulations) needed to fully explore the parameter space using high-fidelity physical models such as those in the SWMF would be computationally impractical. Strategies for dimension reduction and surrogate modeling are thus highly valuable to mitigate this computational burden. In particular, we will employ techniques of sensitivity analysis to help identify a smaller subset of the most important uncertain parameters, thereby achieving dimension reduction to the parameter space. Since subsequent UQ and DA tasks will be performed jointly on solar wind parameters from Stage 1 together with new parameters associated with the CME and geospace models in Stages 2 and 3, it is crucial to keep the parameter space dimension low.

Sensitivity analysis methods (e.g., (Borgonovo & Plischke, 2016) and various articles under Part IV of (Ghanem et al., 2017)) are concerned with the behavior of a model output quantity of interest (QoI) with respect to changes of model inputs, and can be broadly classified as local sensitivity analysis and global sensitivity analysis (GSA). Local sensitivity analysis studies the impact of output from perturbations of input around a reference point (e.g., local gradient), thus only capture behavior in the neighborhood local to that reference point. In contrast, GSA seeks to quantify the impact on the outputs across the entire domain of all possible values the input parameters can take. Variance-based GSA (Saltelli et al., 2004, 2008) further takes into account the current state of uncertainty of the model input parameters. These effects are formally quantified through the Sobol' sensitivity indices, which decompose the total variance of an output quantity into contributions from the variance of each input parameter. Once the most prominent contributors are identified, the other low-impact parameters may be fixed at nominal values with only small approximation error in representing the overall uncertainty of the system, thereby achieving effective dimension reduction of the parameter space. In addition to dimension reduction, GSA may reveal insight about the physical significance of the parameters, and guide future data acquisition that inform the most important parameters. Being a part of the forward UQ analysis, GSA is performed in an *a priori* fashion—using only model simulations, and not requiring any observation data.

Past efforts related to UQ and sensitivity analysis in solar wind models are summarized here. Poduval et al. (2020) focuses on propagating uncertainties in photospheric flux density synoptic magnetograms to the solar wind speed predictions at 1 au for three different phases of the solar cycle; however uncertainty from other sources (e.g., parametric sources) have yet to be incorporated. Riley et al. (2013) use different combinations of coronal models, the base coronal temperature and the spatial resolution of the numerical grid to generate an ensemble of solar wind speed predictions. In contrast to the data-free nature and uncertainty perspective of GSA, this work focuses on assessing the sensitivity of the model performance (i.e. error measure) when compared to in-situ observations under different input settings. While offering insights on physical significance of the parameters for model performance, only two discrete values for the base coronal temperature are considered in the combinations, and for a single quiescent time period of the solar cycle. Reiss et al. (2020) propose a prediction system that uses an ensemble of solar wind solutions. The ensemble is created by varying the four most important coefficients in the near-sun solar wind speed relation from the Wang-Sheeley-Arge (WSA) model that are identified from sensitivity analysis. Their sensitivities are estimated based on the Elementary Effects Approach (Morris, 1991), which computes a global summary of local estimates extracted at multiple points in the input space. The ensemble, however, is generated using new points specified on a tensor grid of perturbations from the baseline values of the coefficients, which grows exponentially with dimensionality and is not easily scalable.

Our study differs from existing work by employing variance-based GSA for AWSoM that offers sensitivity measure in the context of model parameters' uncertainty contributions. We also assess the sensitivity results for both solar minimum and solar maximum conditions, which correspond respectively to periods of low and high solar magnetic activity. We take an approach to perform GSA by building polynomial chaos expansion (PCE) (Ghanem & Spanos, 1991; Xiu & Karniadakis, 2002; Ernst et al., 2012) surrogate models that are particularly suited for extracting the Sobol' indices. PCE represents a random variable in terms of orthogonal polynomial expansions of other latent variables. This allows us to explicitly associate the randomness in the QoIs to each physical source of uncertainty. In addition to GSA, the PCEs will also allow inexpensive sampling and uncertainty propagation.

The downselect of key parameters from GSA in this work will help mitigate the computational burden of future UQ and DA tasks, where new parameters, features and QoIs will enter in the subsequent stages of the CME model chain. For example, we can vary flux rope parameters while initializing the CME and consider influence of background and flux rope parameters jointly. Inverse UQ on the downselected parameters can help constrain them in order to obtain accurate background conditions of solar wind velocity and density. This is crucial for estimating the propagation speed and strength of the shock wave produced by CMEs launched into the background.

We summarize the key contributions and novelty of our paper as follows.

- We perform GSA for background solar wind simulations of the AWSoM to identify and downselect the most important uncertain parameters.
- We construct PCE surrogate models for time-dependent solar wind QoIs and use them to compute the Sobol' indices and perform uncertainty propagation.
- We assess the uncertainty of sensitivity estimates through a bootstrapping procedure.
- We carry out the analysis for examples of both solar maximum and solar minimum conditions.

The remainder of this paper is organized as follows. Section 2 describes features of AWSoM used for solar wind simulations and discusses the model inputs and outputs as part of the simulation setup. Section 3 provides details on the formulation and computation of Sobol' indices leveraging PCE surrogates and space filling designs. Results and discussions for the overall workflow are presented in Section 4 followed by conclusions and future work in Section 5.

2 The Space Weather Modeling Framework

2.1 SWMF and AWSoM

The Space Weather Modeling Framework (SWMF; Tóth et al. 2012; Gombosi et al. 2021) developed at the University of Michigan couples together different model components that cover various physical domains providing a computational capability of modeling the space-weather environment from the Sun to the Earth and/or outer heliosphere. With over a million lines of code, the SWMF is a fully functional, well documented software for high performance computing. Recently, a major portion of the SWMF source code has been released on Github under a non-commerical open source license (`https://github.com/MSTEM-QUDA`). The full SWMF suite has also been publicly available via registration under a user license (`http://csem.engin.umich.edu/tools/swmf`). The SWMF is also available for runs on request through the Community Coordinated Modeling Center (CCMC) at the NASA Goddard Space Flight Center (GSFC) (`https://ccmc.gsfc.nasa.gov/index.php`).

The Alfvén Wave Solar atmosphere Model (AWSoM; van der Holst et al. 2014; Sokolov et al. 2013; Sokolov et al. 2021; van der Holst et al. 2022) within the SWMF couples the solar corona (SC) and inner heliosphere (IH) components extending from the upper chromosphere, through the transition region into the corona up to 1 au and beyond. AWSoM is a global three-dimensional (3D) extended magnetohydrodynamic (MHD) model based on the Block-Adaptive-Tree Solar wind Roe-type Upwind Scheme (BATSRUS; Powell et al. (1999)). It incorporates coronal heating and solar wind acceleration due to low-frequency Alfvén wave turbulence (see van der Holst et al. (2014) for detailed description of the model equations). The coronal heating is distributed over the isotropic electron temperature and the perpendicular and parallel (with respect to the magnetic field) proton temperatures. AWSoM includes stochastic heating and linear wave damping to heat the electrons and protons (Chandran et al., 2011). The model also incorporates electron heat conduction and radiative losses based on the Chianti model (Dere et al., 1997) for both collisional and collisionless regimes. Recently, the energy partitioning scheme within AWSoM has been improved and been validated with Parker Solar Probe observations (van der Holst et al., 2022).

AWSoM is also a data-driven model that uses the radial component of the observed photospheric magnetic field at the inner boundary. We can use either spherical harmonics or the finite difference iterative potential solver (FDIPS, Tóth et al. (2011)) to extrapolate the observational data to a 3D potential field source surface (PFSS) solution. At the inner boundary, the isotropic electron temperature and anisotropic proton temperature are set to 50,000 K. The density at the inner boundary is set to $2\times10^{17}\,\mathrm{m}^{-3}$. The Poynting flux ($\mathrm{S}_A$) of the outward propagating Alfvén waves at the inner boundary determines the energy flux entering the domain and is proportional to the inner boundary magnetic field strength $B_\odot$ (Fisk, 1996; Fisk & Schwadron, 2001; Sokolov et al., 2013). The coefficient $(S_A/B)_\odot$ is an adjustable parameter with a typical value being $10^6\,\mathrm{Wm^{-2}T^{-1}}$. The Alfvén wave correlation length $\mathrm{L}_\perp$ is another parameter of the equation set solved by AWSoM and is proportional to $\mathrm{B}^{-1/2}$ (Hollweg, 1986). The quantity $\mathrm{L}_\perp\sqrt{B}$ is an adjustable parameter with a typical value of $1.5\times10^5\ \mathrm{m}\sqrt{T}$. The stochastic heating amplitude and exponents (Chandran et al., 2011) that determine the energy partitioning between electrons and protons are typically set to 0.18 and 0.21, respectively.

In this work, we use AWSoM to simulate the solar wind in the SC and IH components of SWMF, which use 3D spherical and Cartesian block-adaptive grids, respectively. The steady-state solution is obtained by solving the MHD equations in co-rotating frames in both SC and IH domains. The spherical buffer grid that couples the SC solution with IH extends from 18 to 20 $\mathrm{R}_\odot$. The SC grid covers 1–24 $\mathrm{R}_\odot$ and the IH component grid covers -250 to 250 $\mathrm{R}_\odot$ with the inner boundary at 20 $\mathrm{R}_\odot$. The grid block size in the SC domain is $6\times8\times8$ grid cells and $8\times8\times8$ grid cells in the IH component. We use adaptive mesh refinement (AMR) to refine the grid where needed, including the heliospheric current sheet and a conical region connecting the Sun and Earth. In this region the angular resolution as low as 0.35° so that the CME propagating towards the Earth is well resolved. The angular resolution is 2.8° everywhere else in the domain. In the IH component, the domain has a smallest cell size of 0.24 $\mathrm{R}_\odot$ in the $-x$ direction and 7.8 $\mathrm{R}_\odot$ at the outer boundary. The simulation uses local time-stepping for 80,000 iterations in SC to relax the solution to a steady state. This is followed by coupling with IH for 1 step. Since the solar wind is super fast magnetosonic in the IH component, it only takes 5,000 iterations to obtain a steady-state solution in IH. Over the years, AWSoM has been extensively validated against remote and in-situ observations during various phases of the solar cycle. AWSoM produces synthetic extreme ultra-violet (EUV) images that have been compared to EUV observations from STEREO/EUVI, SDO/AIA and SOHO/LASCO instruments. (van der Holst et al., 2010; Meng et al., 2015; Jin et al., 2017; Sachdeva et al., 2019, 2021). The AWSoM predicted structure of the solar corona also compares well with the tomographic reconstructions of the density and temperature of electrons near the Sun determined using the Differential Emission Measure Tomography (DEMT) during the quiescent phase (Lloveras et al., 2017, 2020, 2022). In addition, comparisons with Interplanetary Scintillation (IPS) data at various heliospheric distances as well as solar wind plasma observations at 1 au have successfully validated the capability of the AWSoM model to reproduce the solar wind structure near the Sun as well as in the inner heliosphere (Sachdeva et al., 2019).

In this work, we will explore simulations of the background solar wind that are conducted for different values in the parameter space using AWSoM. In particular, we will perform a priori sensitivity analysis. This assessment is a priori in the sense that it is performed without any observation data that would otherwise be needed for DA or model calibration. Hence, the procedure is by design an initial probing on the properties of the model itself. Through this sensitivity analysis, we aim to identify a small subset of only the most impactful uncertain parameters that contribute the most to the overall prediction uncertainty. We can then focus only on these parameters for subsequent compute-intensive tasks, thus achieving a dimension reduction of the uncertainty space.

2.2 Solar Wind Model Input Parameters

We begin by cataloguing the uncertain input parameters (i.e. parametric sources of uncertainty) considered in this study for simulating the background solar wind using AWSoM. We focus on simulating the background solar wind for two Carrington rotation (CR) periods representative of solar maximum (CR2152) and solar minimum (CR2208), using exclusively ADAPT-GONG magnetograms. Shown in Table 1, the parameter list includes variables concerning boundary conditions, sub-model settings, and fitting parameters. Some parameters are categorical, while others are continuous and real-valued. In either case, we specify also the value range each parameter may take in this investigation, which are determined based on assessment from subject matter experts of the study team. In addition to the lower and upper bounds, a constraint is incorporated to restrict the feasible region of `FactorB0` and `PoyntingFluxPerBSi` such that their product is less than $9{\times}10^5$ $\mathrm{Wm^{-2}T^{-1}}$ for solar maximum and less than $1.2{\times}10^6$ $\mathrm{Wm^{-2}T^{-1}}$ for solar minimum (see Figure 2). This constraint is motivated by the underlying physics where the product term is known to be proportional to the total energy injected into the system. Capping the total energy below a reasonable threshold eliminates simulations that are not physically meaningful due to excessive kinetic energy density in the simulated solar wind.

While the parameter list may be expanded more exhaustively, our selection here are based on the prioritization from subject matter experts of the study team. Some choices, such as what type of magnetogram should be used or what version of the model to use, have been decided from prior studies (Sachdeva et al., 2019, 2021). Using ADAPT-GONG maps with the three-temperature AWSoM code provided the best results. The effect of grid resolution was also examined and the choice of grid is based on several exploratory simulations. The grid is fine enough along the Sun-Earth line to capture the essential features impacting Earth, but coarse enough to make hundreds of simulations computationally feasible.

To properly convey the state of uncertainty in these parameters, we endow uniform distributions for all parameters over their feasible region to represent a flat, non-informative state of uncertainty that does not favor any particular area. The choice of uniform distributions appeals to the principle of maximum entropy (Jaynes, 1957), where one can show that given a boundary perimeter, the uniform distribution is formed with the fewest additional assumptions. We will investigate the effects of uncertainty from these input parameters on the model output QoIs.

Parameter	Value Range	Description
	Categorical Parameters	
ADAPT_realization	$\{1,2,\ldots,12\}$	Realization index number from ADAPT
PFSS_method	{HARMONICS, FDIPS}	Method for obtaining the potential field source surface solution
UseSurfaceWaveRefl	{True, False}	Extra reflection for high enough transverse density gradient
	Continuous Parameters	
FactorB0 [–]	$[0.54, 2.7]$	Multiplicative factor for input magnetogram field
PoyntingFluxPerBSI [W m^{-2}T^{-1}]	$[0.3, 1.1] \times 10^6$	Inner boundary Poynting Flux per magnetic field constant of Alfvén waves
LperpTimesSqrtBSI [m T$^{1/2}$]	$[0.3, 3.0] \times 10^5$	Stochastic Heating Profile Perpendicular Correlation Length Coefficient
StochasticExponent [–]	$[0.10, 0.34]$	Ion Stochastic Heating Profile Exponent
nChromoSiAWSoM [m^{-3}]	$[2.0, 50.0] \times 10^{17}$	Inner Boundary Density
rMinWaveReflection [R_s]	$[1.0, 1.2]$	Wave Reflection switched off below this radius

Table 1: Uncertain parameters considered for the AWSoM solar wind model.
An additional constraint is imposed to limit the feasible space of `FactorB0` and `PoyntingFluxPerBSi` such that their product is less than $0.9\,\mathrm{MWm^{-2}T^{-1}}$ for solar maximum and less than $1.2\,\mathrm{MWm^{-2}T^{-1}}$ for solar minimum.

2.3 Solar Wind Model Output Quantities of Interest

The primary prediction output of AWSoM are the macroscopic plasma quantities, such as solar wind velocity, density, ion and electron temperatures, the Alfvén wave turbulence energy densities and the magnetic field vector in the 3D computational domain. These primary output variables can be processed into various QoIs, for example synthetic extreme ultraviolet (EUV) images in the low corona, synthetic Thomson-scattered white light images, or in-situ solar wind and magnetic field values along the Earth orbit. These QoIs can be compared with a comprehensive suite of observations including EUV images from STEREO-A EUVI and the SDO AIA, LASCO observations of electron density, as well as in situ OMNI data obtained at the first Lagrange point (L1) between the Sun and Earth.

Future work on UQ associated with CME events will require accurate predictions of the background solar wind, particularly for the radial velocity U_r and proton number density N_p as these have a major impact on the propagation speed of the CME and the strength of the shock wave produced by fast CMEs. For this reason, we select U_r and N_p as the QoIs. In addition to affecting the CME propagation, U_r and N_p are most important for space weather forecasts while other quantities like plasma temperature or the B_x and B_y components of the magnetic field are less geo-effective. The B_z component is, of course, extremely important, but it typically originates from the flux rope driving the CME. Predicting B_z of the background solar wind is very difficult, as it is dominated by turbulent fluctuations.

To carry out the sensitivity analysis, we will systematically vary the input parameters described in the previous section over their distribution, conduct simulations at the different parameter settings for both CR2152 and CR2208, and extract the QoIs and assess and attribute their variability (detailed in the next section). Representative plots of these QoIs from solar wind simulations can be found in Figure 3.

3 Methodology

3.1 Variance-based Global Sensitivity Analysis

We focus on variance-based GSA (Saltelli et al., 2004, 2008). Variance of a QoI can be decomposed into contributions from the uncertainty of each input parameter. Formally, let $\lambda = [\lambda_1, \lambda_2, \cdots, \lambda_d]$ denote the vector of all input parameters with an associated uncertainty distribution, f_t denote the model, and $f_t(\lambda)$ denote a (time-dependent) model output QoI at time t. The Sobol' indices (Sobol, 2003) (defined below) provide a quantitative measure of all the inputs λ_i in terms of their variance contributions to the total variance of the output QoI $f_t(\lambda)$. The key task in GSA is therefore to compute these Sobol' indices. Once computed, these indices can be used for dimension reduction, where low-sensitivity parameters may be fixed at their nominal values without significantly underrepresenting the QoI's variance. The reduced dimension can bring computational savings for downstream tasks such as UQ and DA for subsequent CME and geospace simulations.

The *main effect (first-order)* Sobol' index measures variance contribution solely due to the ith parameter:

$$S_i^t = \frac{\mathrm{Var}_{\lambda_i}\left(\mathbb{E}_{\lambda_{\sim i}}\left[f_t(\lambda)|\lambda_i\right]\right)}{\mathrm{Var}\left(f_t(\lambda)\right)} \tag{1}$$

where $\lambda_{\sim i}$ refers to all components of λ *except* the ith component, $\mathbb{E}_{\lambda_{\sim i}}$ then denotes the expectation with respect to all λ compnents except for the ith, and Var_{λ_i} denotes the variance with respect to only the λ_i component; $\mathbb{E}$ and Var without any subscript indicates expectation and variance involving all components. The main effect index is always between 0 and 1, and a high value indicates that the ith parameter is an important variance (uncertainty) contributor to the QoI. However, a small main effect index by itself does not automatically imply low importance for λ_i, since additional variability may be induced from the interaction of λ_i with other parameters.

The *joint effect (second-order)* Sobol' index measures variance contribution due to the interaction of ith and jth parameters:

$$S_{ij}^t = \frac{\mathrm{Var}_{\lambda_i,\lambda_j}(\mathbb{E}_{\lambda_{\sim ij}}[f_t(\lambda)|\lambda_i, \lambda_j])}{\mathrm{Var}(f_t(\lambda))} - S_i^t - S_j^t. \tag{2}$$

In a similar manner, sensitivity indices for even higher order interactions (e.g., from simultaneous interactions among multiple parameters) can be defined, and the total variance of a QoI can be decomposed into fractional contributions through the relation:

$$1 = \sum_i S_i^t + \sum_i \sum_{j>i} S_{ij}^t + \sum_i \sum_{j>i} \sum_{k>i} S_{ijk}^t + \ldots + S_{123\ldots d}^t. \tag{3}$$

Furthermore, the effect hierarchy principle (Sec. 4.6 of (Wu & Hamada, 2009)) states that only the lower order effects are the most significant. If the main effect and joint effect sensitivity indices sum close to 1, then we can conclude that the higher order interactions among parameters are negligible.

A key assumption behind the above definitions of Sobol' indices is that the input parameters are mutually independent, i.e. their joint distribution can be factored into the products of individual marginal distributions $p(\lambda_i, \lambda_j) = p(\lambda_i)p(\lambda_j)$. While this is satisfied for a uniform distribution over a rectangular domain formed from the various parameter ranges described in Table 1, it is violated when imposing the constraint on the product of `FactorB0` and `PoyntingFluxPerBSi`: e.g., knowing the value of one parameter provides information about what the other parameter could be owing to the constraint, hence they are not independent. There are efforts to formulate a generalized GSA for dependent inputs (Da Veiga et al., 2009; Chastaing et al., 2012), but they are generally difficult to exercise or requires parameter transformations that are not interpretable compared to their original forms. Therefore, we retain the definition derived for the independent setting, but acknowledging that some approximation errors are incurred.

The Sobol' indices cannot be computed in closed-form except for very simple models, and generally they need to be approximated numerically. While different flavors of efficient Monte Carlo (MC) methods have been developed to estimate these indices (Sobol, 1990; Jansen, 1999; Saltelli et al., 1999; Sobol, 2001; Saltelli, 2002; Saltelli et al., 2010), the MC nature means they still require a large number of model evaluations and can become impractical when each model simulation is already expensive: a single AWSoM simulation takes about 7,000 CPU core hours. An alternative strategy is then to trade off model fidelity and accuracy for speed, by first building a surrogate model and then using this approximate but fast surrogate model to estimate the sensitivity indices. We introduce next a surrogate model form that is particularly well suited for estimating the Sobol' indices.

3.2 Polynomial Chaos Expansions

A common surrogate model used for UQ is the PCE. A PCE is a spectral expansion of a random variable, and is particularly attractive for GSA as it has a form that allows convenient estimates of the Sobol' sensitivity indices. We provide a brief introduction of PCE below, and refer readers to several books and review papers for detailed discussions (Ghanem & Spanos, 1991; Najm, 2009; Xiu, 2009; Le Maître & Knio, 2010).

A real-valued random variable u with finite variance (such as an input parameter or an output QoI) can be represented by the following expansion (Ernst et al., 2012):

$$u = \sum_{\|\beta\|_1=0}^{\infty} b_\beta \Psi_\beta(\xi_1, \ldots, \xi_d), \tag{4}$$

where ξ_j are independent and identically distributed (i.i.d.) reference (latent) variables; d is the number of stochastic degrees of freedom in the system (typically the number of uncertain input parameters); b_β are the expansion coefficients; $\beta = (\beta_1, \ldots, \beta_d)$, $\forall \beta_j \in \mathbb{N}_0$, is a multi-index; and Ψ_β are (normalized) multivariate orthogonal polynomials (basis functions) that are products of univariate orthonormal polynomials:

$$\Psi_\beta(\xi_1, \ldots, \xi_d) = \prod_{j=1}^{d} \psi_{\beta_j}(\xi_j). \tag{5}$$

The univariate functions ψ_{β_j} are polynomials of degree β_j in ξ_j, and orthonormal with respect to the probability density of ξ (i.e., $p(\xi)$):

$$\mathbb{E}[\psi_k(\xi)\psi_n(\xi)] = \int \psi_k(\xi)\psi_n(\xi)p(\xi)\, d\xi = \delta_{k,n}, \tag{6}$$

where $\delta_{k,n}$ is the Kronecker delta. While different choices of ξ and ψ_β are available under the generalized Askey family (Xiu & Karniadakis, 2002), we employ uniformly distributed ξ and Legendre polynomials in this study to conveniently mirror the uniform distributions of the input parameters from Table 1. Lastly, the infinite sum in is truncated in practice:

$$u \approx \sum_{\beta \in \mathcal{J}} b_\beta \Psi_\beta(\xi_1, \ldots, \xi_{n_t}), \tag{7}$$

where $\mathcal{J}$ is some finite index set. For example, one popular choice for $\mathcal{J}$ is the "total-order" expansion of degree p, where $\mathcal{J} = \{\beta : \|\beta\|_1 \leq p\}$.

Under this formulation, we can write the PCE for input parameter and output QoI at a time t as

$$\lambda_i \approx \sum_{\beta \in \mathcal{J}} c_\beta \Psi_\beta(\xi_1, \ldots, \xi_d), \quad f_t \approx \sum_{\beta \in \mathcal{J}} b_{t,\beta} \Psi_\beta(\xi_1, \ldots, \xi_d). \tag{8}$$

Since the distribution of ξ is strategically chosen to match the type as our input parameters (i.e. uniform distributions in our case), the PCE for λ_i can be determined easily as a linear expansion (i.e. c_β are simply the scale and shift terms acting on ξ_i). The main task is then to compute the PCE coefficients $c_{t,\beta}$ for the output QoI. We take a regression approach to estimate these coefficients, by solving the following linear system:

$$\begin{bmatrix} \Psi_{\beta^1}(\xi^{(1)}) & \cdots & \Psi_{\beta^{(n_t)}}(\xi^{(1)}) \\ \vdots & & \vdots \\ \Psi_{\beta^1}(\xi^{(N)}) & \cdots & \Psi_{\beta^N}(\xi^{(n_t)}) \end{bmatrix} \begin{bmatrix} b_{t,\beta^1} \\ \vdots \\ b_{t,\beta^{n_t}} \end{bmatrix} = \begin{bmatrix} f(t, \lambda(\xi^{(1)})) \\ \vdots \\ f(t, \lambda(\xi^{(n_t)})) \end{bmatrix}, \tag{9}$$

where Ψ_{β^n} refers to the nth polynomial basis function, b_{t,β^n} is the coefficient corresponding to that term, and $\xi^{(m)}$ is the mth regression (training) point. Ψ is thus the regression matrix where each column corresponds to a basis function and each row corresponds to a regression point. To prevent overfitting, we can include an ℓ_2 (ridge regression) or an ℓ_1 (LASSO) regularization.

Once the PCE for the QoIs is constructed, we can extract the Sobol indices analytically from their expansion coefficients via the formulae:

$$S_i^t = \frac{1}{\mathrm{Var}(f_t(\lambda))} \sum_{\beta \in \mathcal{J}_i} b_{t,\beta}^2, \text{ where } \mathcal{J}_i = \{\beta \in \mathcal{J} : \beta_i > 0, \beta_k = 0, k \neq i\} \tag{10}$$

$$S_{ij}^t = \frac{1}{\mathrm{Var}(f_t(\lambda))} \sum_{\beta \in \mathcal{J}_{ij}} b_{t,\beta}^2 \text{ where } \mathcal{J}_{ij} = \{\beta \in \mathcal{J} : \beta_i > 0, \beta_j > 0, \beta_k = 0, k \neq i, k \neq j\}$$

The QoI total variance can be calculated as

$$\mathrm{Var}(f(\lambda)) = \sum_{0 \neq \beta \in \mathcal{J}} b_{t,\beta}^2. \tag{11}$$

3.3 Design of Computer Experiments

We briefly describe how to select the training points $\xi^{(m)}$ to form the regression system for constructing the PCEs in the previous section. Since each AWSoM simulation is computationally expensive (about 7,000 total CPU hours per run taking 4 hours of wall time on 32 nodes with 56 cores per node on the Frontera computing system (Stanzione et al., 2020)) a judicious selection of the simulation input values can be quite beneficial. While one may approach this task by defining and optimizing some criteria that reflects the quality of estimated Sobol' sensitivity indices, such a goal-oriented approach is non-trivial to formulate. Instead, we take an explorative strategy and seek space-filling designs (Joseph, 2016) that can "cover" the parameter space well.

One popular space-filling approach is the Latin Hypercube design (LHD) (McKay et al., 1979), which can be constructed using a maximin design criterion that maximizes the minimum distance between all pairs of points (Morris & Mitchell, 1995). The maximin LHD for a multi-dimensional space can retain good space-filling properties when projected onto any single dimension, but not when projecting onto multi-dimensional subspaces (i.e. when focusing on a subset of multiple parameters) (Joseph, 2016). We thus adopt an improved Maximum Projection (MaxPro) design (Joseph et al., 2015, 2020) that uses a weighted distance measure to account for projections to all possible subspaces. Another notable advantage of using MaxPro designs is that new samples can be added in a sequential manner where the importance for different factor levels based on sensitivity results can be incorporated into the objective function (Wang et al., 2018).

MaxPro design is typically defined for a box domain. With the only non-rectangular domain in our study being the constraint on the product of `FactorB0` and `PoyntingFluxPerBSi` (see Figure 2), we simply generate MaxPro sample in the bounding rectangle and then reject the points that lay outside the constraint.

4 Results and Discussion

4.1 AWSoM Solar Wind Simulations

We perform solar wind simulations using the AWSoM model for CR2152 (solar maximum) and CR2208 (solar minimum). The model input parameter values are generated from their feasible ranges summarized in Table 1 using the MaxPro design described in Section 3.3. Scatter plots of these samples for select pairs of input parameters are shown in Figure 2, with the left-most panel showing the constraint on the product of `FactorB0` and `PoyntingFluxPerBSi`. Given our computational budget, 200 runs are conducted for each of the two CR periods.

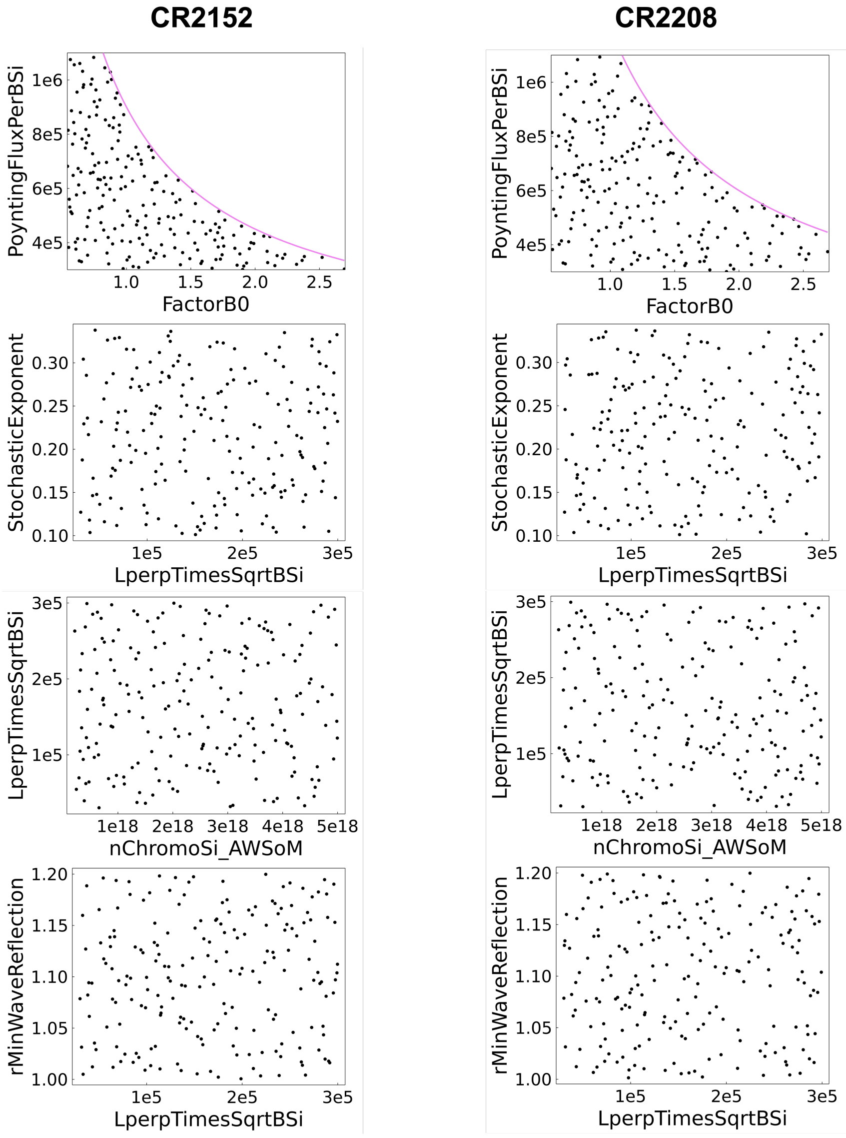

Figure 2: Scatter plots of MaxPro design samples to perform AWSoM simulations for select pairs of input parameters for CR2152 (solar maximum, left column) and CR2208 (solar minimum, right column). Each AWSoM run is initiated at each point for a total of 200 runs.

From the 200 simulations for each CR, 5 of CR2152 and 1 of CR2208 did not converge while all others succeeded. The results of all successful runs are analyzed to filter out those that are not physically meaningful. We extract the plasma state along Earth's orbit and a simulation is discarded if *both* of the following exclusion criteria are triggered:

- the radial velocity exceeds 900 km/s *or* falls below 200 km/s, *and*
- the number density exceeds 100 cm^{-3}.

In the end, 174 runs are retained for CR2152 and 199 for CR2208. The final ensemble of select predicted QoIs at 1 au are shown in Figure 3.

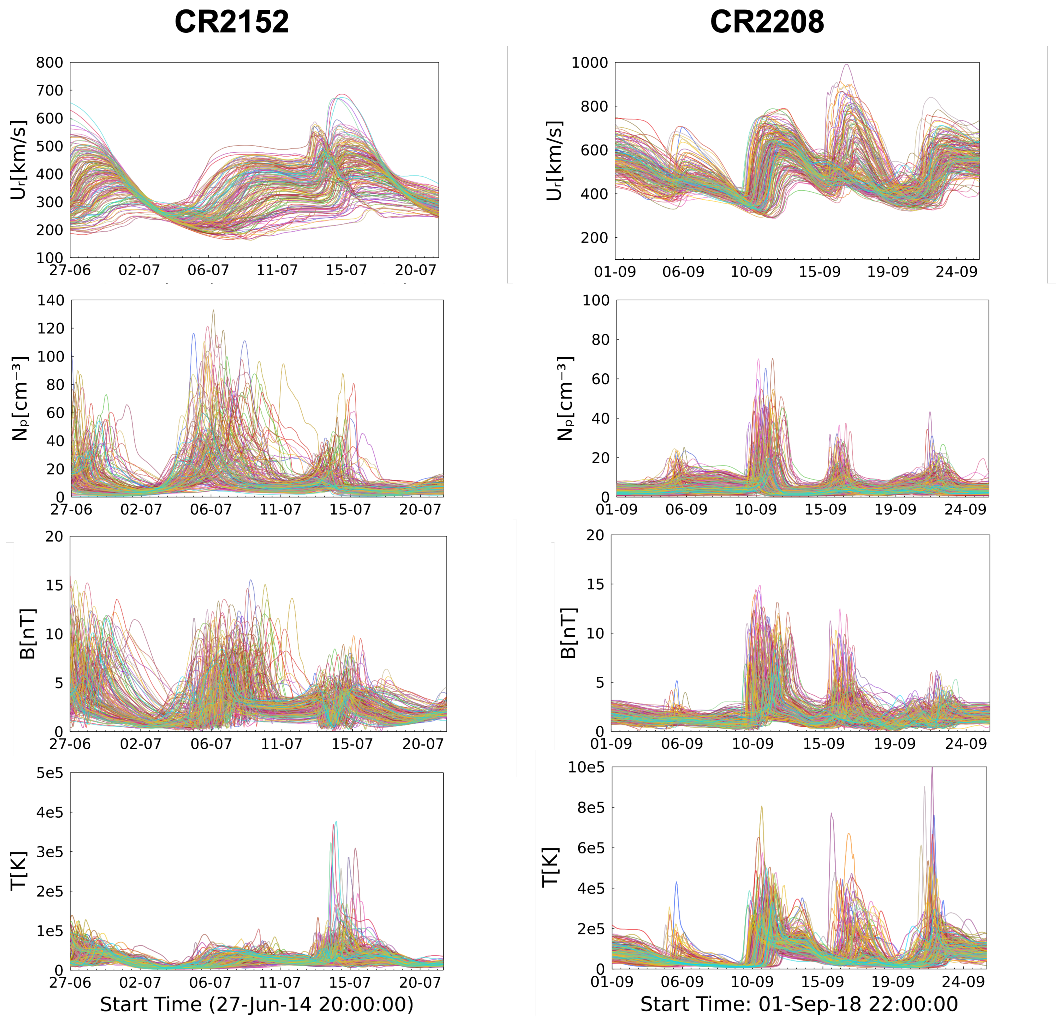

Figure 3: Ensemble of AWSoM simulation results for CR2152 (solar maximum, left column) and CR2208 (solar minimum, right column). Each line is from a different simulation.

4.2 UQ and GSA using PCE Surrogate

We use the set of AWSoM simulations to construct PCE surrogates following Section 3.2. In particular, we construct a separate PCE at 577 time points of each QoIs: radial velocity (U_r) and number density (N_p), for both CR2152 and CR2208. The input space of each PCE is 6 dimensional, encompassing all the continuous input parameters from the second half of Table 1.

The parameters in the first half of Table 1 are categorical (i.e. not ordinal), and they do not have any intrinsic ordering of their values or a notion of distance. They are not true random variables and quantities such as mean and variance are undefined. Therefore, the concept of sensitivity for categorical variables is ill-posed altogether. As a result, we consider sensitivity only for the six continuous parameters. Note that uncertainty contributions from the three categorical parameters are still captured since they are varied in generating the AWSoM simulation set. Our PCEs are built by marginalizing out (averaged over) the three categical parameters and trained to predict the QoI values based on the six continuous input parameters. Ridge regression is adopted for computing the

PCE coefficients in the regression system Equation (9), with regularization parameter selected through cross-validation. We employ PCEs with total order expansions of degree 2. While higher degree polynomials may be attempted, the increased number of unknown coefficients is more prone to overfitting given our small sample size (around 200). We also verified that increasing to degree 3 does not lead to substantial differences of the surrogate predictions, supporting that degree 2 is sufficient. All PCE constructions are carried out using `PolyChaos.jl` (Mühlpfordt et al., 2020), an open source package available in the Julia programming language (Bezanson et al., 2017). Once the PCE surrogates are available, we can use them to inexpensively perform MC-based uncertainty propagation by first drawing samples from the uncertainty distribution of the input and then using the PCEs to evaluate the output QOIs. Figure 4 presents the predictive uncertainty on the QoIs highlighting their mean (solid red line) $\pm$ 2 standard deviations (red shaded area), and overlaid with boxplots to illustrate more details of the distribution at different time-slices. The sample mean (blue dashed line) is identical to the surrogate mean (red solid line).

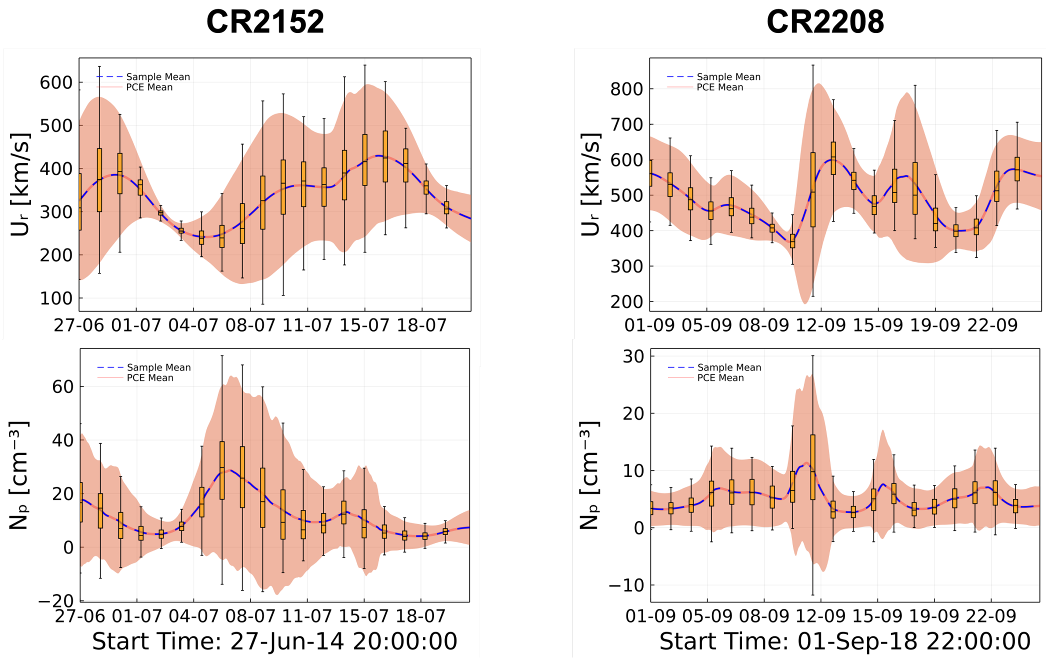

Figure 4: Predictive mean (red line) ± 2 standard deviation (shaded area) for QoIs U_r and N_p using PCE surrogates for CR2152 (solar maximum, left column) and CR2208 (solar minimum, right column). The boxplots at selected locations give additional information about the distributions, showing the median and interquartile range (IQR); whiskers extend to 1.5 IQR on either side. The sample mean (dashed blue line) is essentially equal to the predictive mean.

Using Equations 10–11, we can calculate the Sobol sensitivity indices directly from the PCE coefficients. In particular, we focus on the sensitivity for radial velocity U_r and number density N_p with respect to all the continuous input parameters from Table 1. The main effect indices S_i^t for U_r and N_p are plotted over time during CR2152 and CR2208 in Figure 5. At any particular time instant, S_i^t represents the relative variance contribution from the ith parameter. For CR2152 (solar maximum), overall `FactorB0` and `LperpTimesSqrtBSI` appear to be most dominating followed by `PoyntingFluxPerBSi`, while `rMinWaveReflection`, `StochasticExponent`, and `nChromoSi_AWSoM` have much smaller contributions. For CR2208 (solar minimum), `LperpTimesSqrtBSi` has a much smaller contribution than it is in CR2152 (solar maximum). This agrees with our expectations: the `LperpTimesSqrtBSi` parameter has the most impact along open magnetic field lines coming from coronal holes, which are more likely to be at low latitude during solar maximum and therefore have an impact at Earth orbit. `FactorB0` and `PoyntingFluxPerBSi` appear to be the most influential, especially for the number density N_p. For U_r, `StochasticExponent` also has significant contributions particularly for solar minimum. The sum of main effect indices from all parameters at a time instant can also provide an indication regarding the interaction effects among parameters. If the sum is much less than 1, the interactions between parameters is non-negligible. For example the sum of N_p's sensitivity indices for CR2152 is close to 0.5 around July 3, 2014, suggesting there is significant parameter interactions at that time.

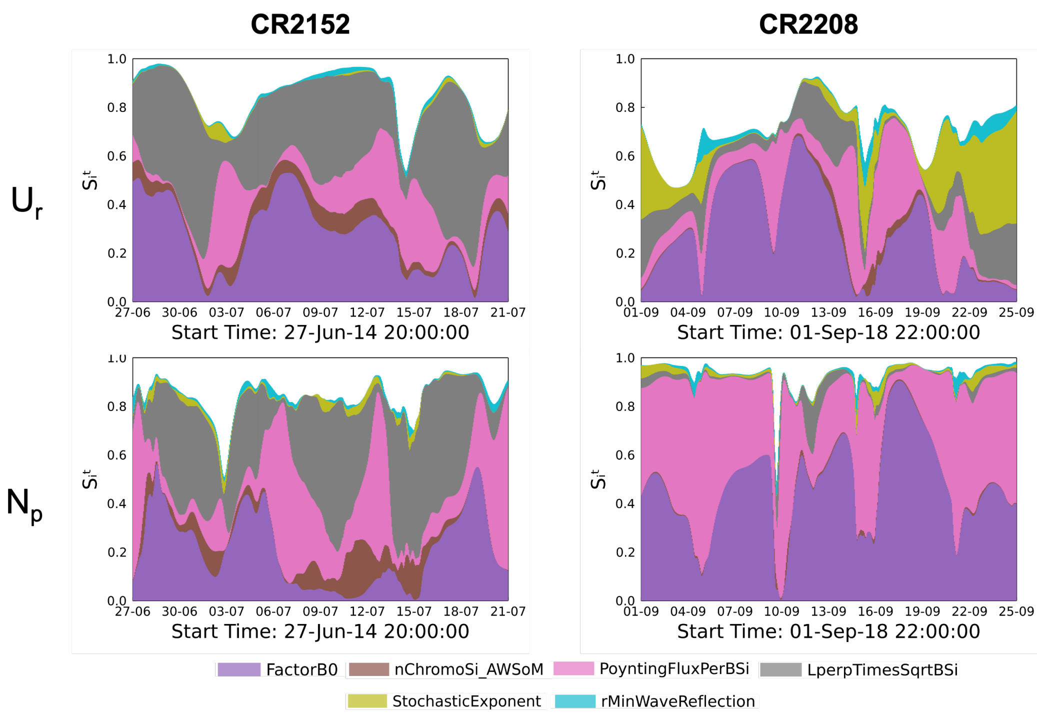

Figure 5: Time-varying main effect S_i^t for CR2152 (solar maximum, left column) and CR2208 (solar minimum, right column), for QoIs U_r and N_p.

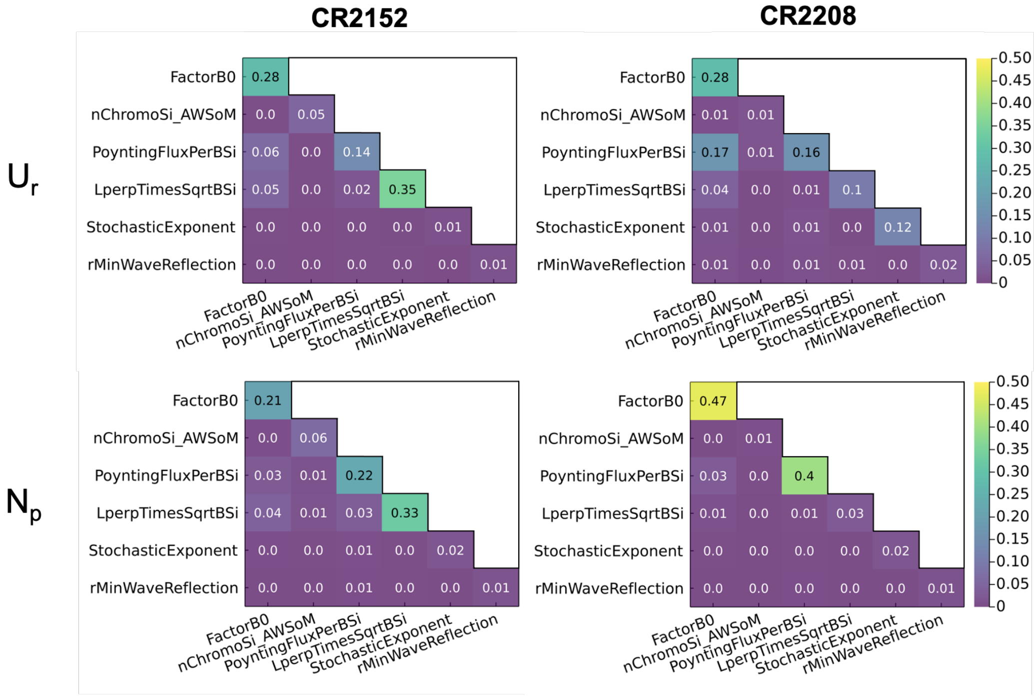

Figure 6: Time-averaged main effect S_i and joint effect S_{ij} for CR2208 (solar maximum, left column) and CR2208 (solar minimum, right column) for QoIs U_r and N_p.

Lastly, we summarize the time-dependent Sobol' sensitivity indices by computing the time-averaged main effect and joint effect indices in Figure 6, where the (i,j)th element indicates the time-averaged value of $S_{i,j}$ and the diagonal elements represent S_i. The time-averaged sensitivity indices confirm the observations from the time-dependent results that the most important variance contributors for QoIs U_r and N_p in CR2152 (with a threshold chosen as $S_i > 0.2$ for either QoI) are `FactorB0`, `PoyntingFluxPerBSi` and `LperpTimesSqrtBSi`, and for CR2208 are `FactorB0` and `PoyntingFluxPerBSi`. The remaining parameters' contributions, when time-averaged, are very small. As a sanity check, we can also see that the averaged main and joint sensitivities approximately sum to 1, as suggested by Equation (3).

We note that PCE in general cannot constrain its output value to be positive only, whereas the number density N_p can only be positive. As a result, we have occasionally encountered negative N_p predictions from the the PCE surrogates. One possible technique to guarantee positivity is to build PCEs for predicting logarithm of the QoIs (i.e. $\log N_p$), and then extract the non-logarithm values by taking the exponent. However, subsequently computed Sobol' sensitivity indices then indicate the parameter contributions on the variance of $\log N_p$ and not of N_p, which may alter the ranking of parameters (Borgonovo et al., 2014). In our testing with the log-QoIs setup, we see example from Figure 7 that indeed N_p (CR2152) is now guaranteed to be always positive and its corresponding Sobol' indices support the same conclusion of the most sensitive parameters, but with different rankings (similar results for CR2208 are omitted for brevity).

4.3 Uncertainty of the Sobol' Index Estimates

Given that our Sobol' indices are estimated using small sample size (around 200), it is important to assess the uncertainty of these estimates. Ideally, one can repeat the

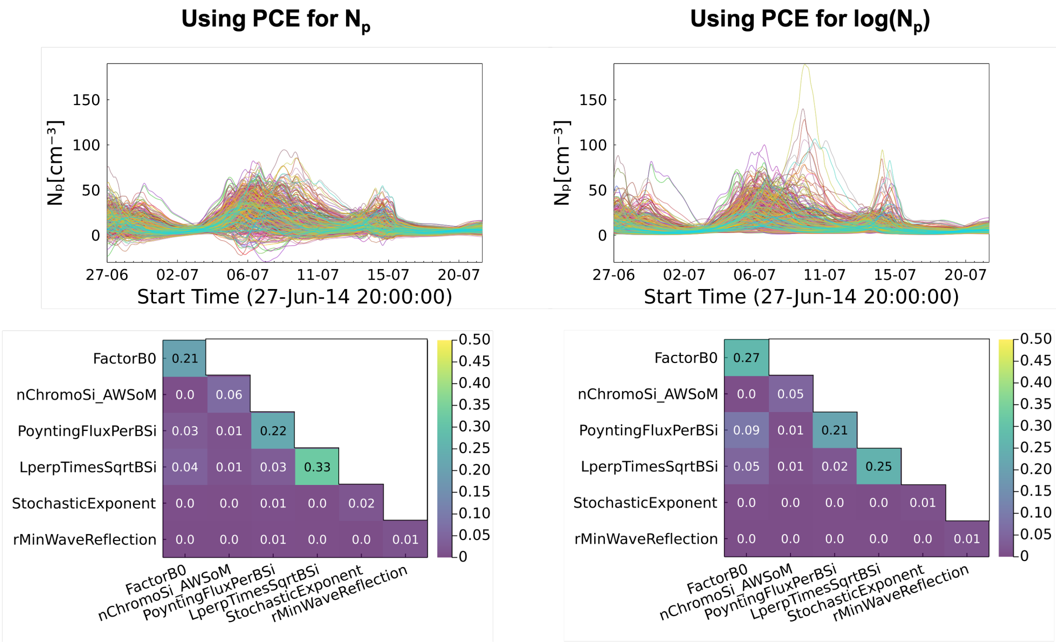

Figure 7: Comparison of results as an illustration when building surrogate on the original N_p (left column) and $\log(N_p)$ (right column). Predictions of N_p at 400 test points from trained PCE surrogates are shown on the top row, while the bottom row shows the time-averaged sensitivity heatmaps.

GSA procedure with new batches of samples and compute the variance of the repeated trials, but such a process would be prohibitively expensive. Therefore, we use a bootstrapping technique that only uses existing and available samples, summarized in Algorithm 1.

Algorithm 1: Procedure for Bootstrapped GSA

Input: Input parameters λ at N design points, N QoI simulations $f(\lambda)$, bootstrap sample sizes n_{start}, n_{end}, step size Δ, number of replications K for each sample size

$n = [n_{\text{start}}, n_{\text{start}} + \Delta, \cdots, n_{\text{end}}]$;
`/* The outer loop runs through different sample sizes */`
for $i = 1 : length(n)$ **do**
  $n_{\text{Samples}} = n[i]$;
  `/* The inner loop runs` K `replications per sample size */`
  for $k = 1, 2, \cdots K$ **do**
    Sample indices $i_k \in \{1, \cdots, N\}$ with replacement (n_{Samples} in all);
    Build PCEs with input parameters λ and outputs f indexed by i_k;
    Calculate and store time-averaged main effects $S_i[:, k, i]$;
  end
end
return S_i ;
Calculate mean and standard deviations over K replications

We carry out the bootstrapping analysis for $n = \{20, 40, \ldots, 140\}$. For each n, $K = 1000$ trials of estimating the time-averaged S_i are repeated, and their mean and standard deviation are computed. To save time, we did not re-optimize the regularization parameters and opted for ordinary least squares for Equation (9) in building the new PCEs. The results are plotted in Figure 8 for U_r and N_p in both CR2152 and CR2208. The figures suggest that $n = 20$ or 40 would carry significant errors and present challenges for distinguishing the influential parameters, and estimates start to stabilize after $n = 60$. The mean S_t values at $n = 120$ and $n = 140$ are quite close to the values we obtained with the full sample set, and the rankings of the most influential parameters are reasonably robust over multiple replications.

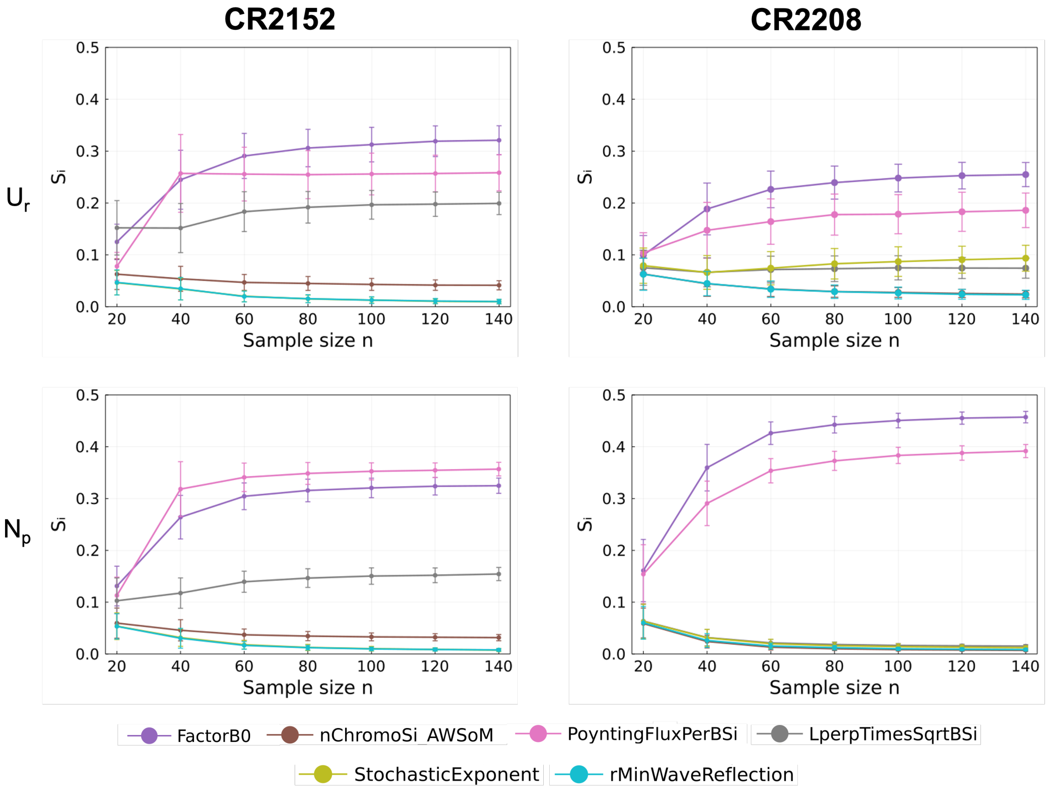

Figure 8: Mean ± standard deviation (over $K = 1000$ replications) time-averaged S_i for U_r and N_p under different bootstrap set size n in CR2152 (solar maximum, left column) and CR2208 (solar minimum, right column)

.

Our bootstrapping analysis carries several limitations. First, the uncertainty on Sobol' indices obtained from bootstrapping is an underestimate especially as n becomes closer to the size of the full dataset. This is because the repeated subset trials would have larger overlaps among each other and therefore lowered variability, since each subset would necessarily contain a large portion of the full dataset (e.g., if subsampling $n = 190$ from a full dataset of 200 runs, then the repeated subsets must have many overlap samples). Second, in order to reduce the substantial computational burden from the many repeated PCE and GSA during bootstrapping, ordinary least squares is used without regularization for the new PCE regression problems. Lastly, we only perform the bootstrapping analysis on the time-averaged Sobol' indices for brevity, and do not consider their uncertainty at each time point; although this may be done easily.

5 Conclusions and Future Work

We conducted variance-based GSA for background solar wind during CR2152 (a solar maximum period) and CR2208 (a solar minimum period) simulated using the AWSoM from the SWMF. We computed the main and joint effect Sobol' sensitivity indices for output QoIs of radial velocity and proton number density at 1 au, with respect to the uncertainty of a number of input parameters including `FactorB0`, `nChromoSi_AWSoM`, `PoyntingFluxPerBSi`, `LperpTimesSqrtBSI`, `StochasticExponent`, and `rMinWaveReflection`. The Sobol' indices quantify the fractional contribution of and individual input parameter's uncertainty towards the total variance of the QoIs, and therefore provide sensitivity information that reflects the current state of parameter uncertainty. Furthermore, this GSA can be performed in a data-free manner, without needing any observation data at 1 au.

We presented an efficient computational procedure for estimating the Sobol' indices by creating PCE surrogate models from a dataset of AWSoM simulations selected through space-filling designs of the model parameters. Once these PCEs became available, the Sobol' indices were calculated analytically from the expansion coefficients. At the same time, forward UQ was also achieved by sampling the PCEs to obtain predictive uncertainty for the QoIs. The uncertainty of the estimated Sobol' indices were also estimated through a bootstrapping procedure. Overall, we found the most impactful parameters to be `FactorB0`, `PoyntingFluxPerBSi`, and `LperpTimesSqrtBSI` for CR2152 (solar maximum); and `FactorB0` and `PoyntingFluxPerBSi` for CR2208 (solar minimum). For future tasks, only these parameters need to be kept as uncertain while the other low-impact parameters may be fixed at nominal values, thereby achieving dimension reduction of the parameter space.

There are several limitations of our current work that warrant interesting future studies. Our results are obtained from two specific CR periods, and the generalizability of the high-sensitive parameters to other solar maximum and solar minimum periods needs to be tested. On a more technical side, the Sobol' indices definitions employed are for input parameters with independent uncertainty distributions. However our constraint between `FactorB0` and `PoyntingFluxPerBSi`, while justified from a physical understanding of the system, violates the independent assumption. As a result, our computed Sobol' indices incur additional error due to this effect, and generalized GSA techniques that may accommodate dependent parameter distributions may be explored (Chastaing et al., 2012).

Lastly, while we have taken the first step towards an overall probabilistic forecast framework of space weather events by focusing on UQ of the background solar wind, the next parts of our work will involve DA and the CME and geospace stages in completing the Sun-to-Earth model. Computations for these future tasks will benefit from the reduced dimension of the solar wind parameter space from this paper.

Acronyms

QoI Quantity of Interest
UQ Uncertainty Quantification
DA Data Assimilation
PCE Polynomial Chaos Expansion
GSA Global Sensitivity Analysis
CME Coronal Mass Ejection
SWMF Space Weather Modeling Framework
AWSoM Alfvén Wave Solar atmosphere Model
SC Solar Corona
IH Inner Heliosphere

6 Open Research

The scripts and routines used to produce the results in this manuscript are available at the University of Michigan (UM) Library Deep Blue Data Repository here: Results for "Global Sensitivity Analysis and Uncertainty Quantification for Background Solar Wind in the Alfvén Wave Solar Atmosphere Model":

https://deepblue.lib.umich.edu/data/concern/file_sets/41687h82q/anonymous_link/b5aefa23760a609fd9eb5b53fc6cb91f0fca3dd0f7a378503fea7cdf84b4e622

A major portion of the SWMF source code has been released on Github under a non-commerical open source license (https://github.com/MSTEM-QUDA). The full SWMF suite is publicly available via registration under a user license (http://csem.engin.umich.edu/tools/swmf).

Acknowledgments
This work is supported by the National Science Foundation (NSF) under grant number PHY-2027555: "SWQU: NextGen Space Weather Modeling Framework Using Data, Physics and Uncertainty Quantification". The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC resources under the "LRAC: NextGen Space Weather Modeling Framework Using Data, Physics and Uncertainty Quantification" allocation on the Frontera supercomputer.

References

Bezanson, J., Edelman, A., Karpinski, S., & Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM Review*, *59*(1), 65–98. doi: 10.1137/141000671

Borgonovo, E., & Plischke, E. (2016). Sensitivity analysis: A review of recent advances. *European Journal of Operational Research*, *248*(3), 869–887. doi: 10.1016/j.ejor.2015.06.032

Borgonovo, E., Tarantola, S., Plischke, E., & Morris, M. D. (2014). Transformations and invariance in the sensitivity analysis of computer experiments. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(5), 925–947. doi: 10.1111/rssb.12052

Chandran, B. D. G., Dennis, T. J., Quataert, E., & Bale, S. D. (2011). Incorporating kinetic physics into a two-fluid solar-wind model with temperature anisotropy and low-frequency Alfvén-wave turbulence. *Astrophys. J.*, *743*. doi: 10.1088/0004-637X/743/2/197

Chastaing, G., Gamboa, F., & Prieur, C. (2012). Generalized Hoeffding-Sobol decomposition for dependent variables - application to sensitivity analysis. *Electronic Journal of Statistics*, *6*(none). doi: 10.1214/12-EJS749

Da Veiga, S., Wahl, F., & Gamboa, F. (2009). Local Polynomial Estimation for Sensitivity Analysis on Models With Correlated Inputs. *Technometrics*, *51*(4), 452–463. doi: 10.1198/TECH.2009.08124

Debusschere, B., Sargsyan, K., Safta, C., & Chowdhary, K. (2017). Uncertainty Quantification Toolkit (UQTk). In *Handbook of Uncertainty Quantification* (pp. 1807–1827). Cham: Springer International Publishing.

Dere, K. P., Landi, E., Mason, H. E., Fossi, B. C. M., & Young, P. R. (1997). CHIANTI - an atomic database for emission lines. *Astron. Astrophys. Suppl. Ser.*, *125*, 149–173.

Ernst, O. G., Mugler, A., Starkloff, H.-J., & Ullmann, E. (2012). On the convergence of generalized polynomial chaos expansions. *ESAIM: Mathematical Modelling and Numerical Analysis*, *46*(2), 317–339. doi: 10.1051/m2an/2011045

Fisk, L. A. (1996). Motion of the footpoints of heliospheric magnetic field lines at the Sun: Implications for recurrent energetic particle events at high heliographic lattitudes. *J. Geophys. Res.*, *101*(A7), 15,547–15,553.

Fisk, L. A., & Schwadron, N. A. (2001). The behavior of the open magnetic field of the Sun. *Astrophys. J.*, *560*, 425–438.

Ghanem, R., Higdon, D., & Owhadi, H. (Eds.). (2017). *Handbook of uncertainty quantification.* Cham: Springer International Publishing. doi: 10.1007/978-3-319-12385-1

Ghanem, R., & Spanos, P. D. (1991). *Stochastic Finite Elements: A Spectral Approach* (1st ed.). New York, NY: Springer New York. doi: 10.1007/978-1-4612-3094-6

Gombosi, T. I., Chen, Y., Glocer, A., Huang, Z., Jia, X., Liemohn, M. W., … Gombosi, T. I. (2021). What sustained multi-disciplinary research can achieve: The space weather modeling framework. *Journal of Space Weather and Space Climate*, *11*. doi: 10.1051/swsc/2021020

Hollweg, J. V. (1986). Transition region, corona, and solar wind in coronal holes. *J. Geophys. Res.*, *91*(A4), 4111-4125. doi: 10.1029/JA091iA04p04111

Jansen, M. J. W. (1999). Analysis of variance designs for model output. *Computer Physics Communications*, *117*(1), 35–43. doi: 10.1016/S0010-4655(98)00154-4

Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical Review*, *106*(4), 620–630. doi: 10.1103/PhysRev.106.620

Jin, M., Manchester, W. B., van der Holst, B., Gruesbeck, J. R., Frazin, R. A., Landi, E., … Gombosi, T. (2012). A global two-temperature corona and inner heliosphere model: A comprehensive validation study. *Astrophys. J.*, *7745*, 6. doi: 10.1088/0004-637X/745/1/6

Jin, M., Manchester, W. B., van der Holst, B., Sokolov, I., Tóth, G., Mullinix, R. E., … Gombosi, T. I. (2017). Data-Constrained Coronal Mass Ejections in a Global Magnetohydrodynamics Model. *The Astrophysical Journal*, *834*(2), 173. doi: 10.3847/1538-4357/834/2/173

Joseph, V. R. (2016). Space-filling designs for computer experiments: A review. *Quality Engineering*, *28*(1), 28–35. doi: 10.1080/08982112.2015.1100447

Joseph, V. R., Gul, E., & Ba, S. (2015). Maximum projection designs for computer experiments. *Biometrika*, *102*(2), 371–380. doi: 10.1093/biomet/asv002

Joseph, V. R., Gul, E., & Ba, S. (2020). Designing computer experiments with multiple types of factors: The MaxPro approach. *Journal of Quality Technology*, *52*(4), 343–354. doi: 10.1080/00224065.2019.1611351

Le Maître, O. P., & Knio, O. M. (2010). *Spectral Methods for Uncertainty Quantification: with Applications to Computational Fluid Dynamics.* Houten, Netherlands: Springer Netherlands. doi: 10.1007/978-90-481-3520-2

Lloveras, D. G., Vásquez, A. M., Nuevo, F. A., & Frazin, R. A. (2017). Comparative study of the three-dimensional thermodynamical structure of the inner corona of solar minimum carrington rotations 1915 and 2081. *Solar Physics*, *292*(10), 153. doi: 10.1007/s11207-017-1179-z

Lloveras, D. G., Vásquez, A. M., Nuevo, F. A., Frazin, R. A., Manchester, W., Sachdeva, N., … Gilardy, H. (2022). Three-Dimensional Structure of the Corona During WHPI Campaign Rotations CR-2219 and CR-2223. *Journal of Geophysical Research (Space Physics)*, *127*(6), e30406. doi: 10.1029/2022JA030406

Lloveras, D. G., Vásquez, A. M., Nuevo, F. A., Mac Cormack, C., Sachdeva, N., Manchester, W., … Frazin, R. A. (2020). Thermodynamic Structure of the Solar Corona: Tomographic Reconstructions and MHD Modeling. *Solar Physics*, *295*(6), 76. doi: 10.1007/s11207-020-01641-z

McKay, M. D., Beckman, R. J., & Conover, W. J. (1979). A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics*, *21*(2), 239–245. doi: 10.2307/1268522

Meng, X., van der Holst, B., Tóth, G., & Gombosi, T. I. (2015). Alfven wave solar model (AWSoM): Proton temperature anisotropy and solar wind acceleration. *Mon. Not. Roy. Ast. Soc.*, *454*, 3697. doi: 10.1093/mnras/stv2249

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, *33*(2), 161-174. doi: 10.1080/00401706.1991.10484804

Morris, M. D., & Mitchell, T. J. (1995). Exploratory designs for computational experiments. *Journal of Statistical Planning and Inference*, *43*(3), 381–402. doi: 10.1016/0378-3758(94)00035-T

Mühlpfordt, T., Zahn, F., Hagenmeyer, V., & Faulwasser, T. (2020). PolyChaos.jl – A Julia Package for Polynomial Chaos in Systems and Control. *arXiv e-prints*. https://arxiv.org/abs/2004.03970

Najm, H. N. (2009). Uncertainty Quantification and Polynomial Chaos Techniques in Computational Fluid Dynamics. *Annual Review of Fluid Mechanics*, *41*(1), 35–52. doi: 10.1146/annurev.fluid.010908.165248

Poduval, B., Petrie, G., & Bertello, L. (2020). Uncertainty Estimates of Solar Wind Prediction Using HMI Photospheric Vector and Spatial Standard Deviation Synoptic Maps. *Solar Physics*, *295*(10), 138. doi: 10.1007/s11207-020-01704-1

Powell, K., Roe, P., Linde, T., Gombosi, T., & De Zeeuw, D. L. (1999). A solution-adaptive upwind scheme for ideal magnetohydrodynamics. *J. Comput. Phys.*, *154*, 284-309. doi: 10.1006/jcph.1999.6299

Reiss, M. A., MacNeice, P. J., Muglach, K., Arge, C. N., Möstl, C., Riley, P., ... Amerstorfer, U. (2020). Forecasting the Ambient Solar Wind with Numerical Models. II. An Adaptive Prediction System for Specifying Solar Wind Speed near the Sun. *The Astrophysical Journal*, *891*(2), 165. doi: 10.3847/1538-4357/ab78a0

Riley, P., Linker, J. A., & Mikić, Z. (2013). On the application of ensemble modeling techniques to improve ambient solar wind models: Ensemble modeling of the solar wind. *Journal of Geophysical Research: Space Physics*, *118*(2), 600–607. doi: 10.1002/jgra.50156

Sachdeva, N., Tóth, G., Manchester, W. B., van der Holst, B., Huang, Z., Sokolov, I. V., ... Vásquez, A. M. (2021). Simulating Solar Maximum Conditions Using the Alfvén Wave Solar Atmosphere Model (AWSoM). *The Astrophysical Journal*, *923*(2), 176. doi: 10.3847/1538-4357/ac307c

Sachdeva, N., van der Holst, B., Manchester, W. B., Tóth, G., Chen, Y., Lloveras, D. G., ... Henney, C. J. (2019). Validation of the Alfvén Wave Solar Atmosphere Model (AWSoM) with Observations from the Low Corona to 1 au. *The Astrophysical Journal*, *887*(1), 83. doi: 10.3847/1538-4357/ab4f5e

Saltelli, A. (2002). Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, *145*(2), 280–297. doi: 10.1016/S0010-4655(02)00280-1

Saltelli, A., Annoni, P., Azzini, I., Campolongo, F., Ratto, M., & Tarantola, S. (2010). Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer Physics Communications*, *181*(2), 259–270. doi: 10.1016/j.cpc.2009.09.018

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., ... Tarantola, S. (2008). *Global Sensitivity Analysis: The Primer*. Chichester, United Kingdom: John Wiley & Sons, Ltd.

Saltelli, A., Tarantola, S., Campolongo, F., & Ratto, M. (2004). *Sensitivity Analysis in Practice: A Guide to Assessing Scientific Models*. Chichester, United Kingdom: John Wiley & Sons.

Saltelli, A., Tarantola, S., & Chan, K. P.-S. (1999). A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output. *Technometrics*, *41*(1), 39–56. doi: 10.1080/00401706.1999.10485594

Sobol, I. M. (1990). On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe Modelirovanie*, *2*(1), 112–118.

Sobol, I. M. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, *55*(1-3), 271–280. doi: 10.1016/S0378-4754(00)00270-6

Sobol, I. M. (2003). Theorems and examples on high dimensional model representation. *Reliability Engineering & System Safety*, *79*(2), 187–193. doi: 10.1016/S0951-8320(02)00229-6

Sokolov, I. V., Holst, B. v. d., Manchester, W. B., Su Ozturk, D. C., Szente, J., Taktakishvili, A., ... Gombosi, T. I. (2021, February). Threaded-field-line Model for the Low Solar Corona Powered by the Alfvén Wave Turbulence. *Astrophys. J.*, *908*(2), 172. doi: 10.3847/1538-4357/abc000

Sokolov, I. V., van der Holst, B., Oran, R., Downs, C., Roussev, I. I., Jin, M., ... Gombosi, T. I. (2013). Magnetohydrodynamic waves and coronal heating: Unifying empirical and MHD turbulence models. *Astrophys. J.*, *764*, 23. doi: 10.1088/0004-637X/764/1/23

Stanzione, D., West, J., Evans, R. T., Minyard, T., Ghattas, O., & Panda, D. K. (2020). Frontera: The Evolution of Leadership Computing at the National Science Foundation. In *Practice and Experience in Advanced Research Computing* (pp. 106–111). New York, NY, USA: Association for Computing Machinery. doi: 10.1145/3311790.3396656

Tóth, G., Meng, X., Gombosi, T. I., & Ridley, A. (2011). Reducing numerical diffusion in magnetospheric simulations. *J. Geophys. Res.*, *116*, A07211. doi: 10.1029/2010JA016370

Tóth, G., Sokolov, I. V., Gombosi, T. I., Chesney, D. R., Clauer, C., Zeeuw, D. L. D., ... Kóta, J. (2005). Space Weather Modeling Framework: A new tool for the space science community. *J. Geophys. Res.*, *110*, A12226. doi: 10.1029/2005JA011126

Tóth, G., van der Holst, B., Sokolov, I. V., Zeeuw, D. L. D., Gombosi, T. I., Fang, F., ... Opher, M. (2012). Adaptive numerical algorithms in space weather modeling. *J. Comput. Phys.*, *231*, 870–903. doi: 10.1016/j.jcp.2011.02.006

van der Holst, B., Huang, J., Sachdeva, N., Kasper, J. C., Manchester, I., W. B., Borovikov, D., ... MacDowall, R. J. (2022). Improving the Alfvén Wave Solar Atmosphere Model Based on Parker Solar Probe Data. *Astrophys. J.*, *925*(2), 146. doi: 10.3847/1538-4357/ac3d34

van der Holst, B., Manchester, W., Frazin, R., Vásquez, A., Tóth, G., & Gombosi, T. (2010). A data-driven, two-temperature solar wind model with alfvén waves. *Astrophys. J.*, *725*, 1373-1383. doi: 10.1088/0004-637X/725/1/1373

van der Holst, B., Sokolov, I., Meng, X., Jin, M., Manchester, W. B., Tóth, G., & Gombosi, T. I. (2014). Alfvén wave solar model (AWSOM): Coronal heating. *Astrophys. J.*, *782*, 81. doi: 10.1088/0004-637X/782/2/81

Wang, D., Ba, S., & Myers, W. (2018). A Sequential Maximum Projection Design Framework for Computer Experiments with Inert Factors. *Statistica Sinica*. doi: 10.5705/ss.202016.0165

Wu, C.-F., & Hamada, M. (2009). *Experiments: planning, analysis, and optimization* (2nd ed ed.). Hoboken, N.J: Wiley. (OCLC: ocn276140904)

Xiu, D. (2009). Fast Numerical Methods for Stochastic Computations: A Review. *Communications in Computational Physics*, *5*(2-4), 242–272. http://www.global-sci.com/intro/article_detail/cicp/7732.html

Xiu, D., & Karniadakis, G. E. (2002). The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations. *SIAM Journal on Scientific Computing*, *24*(2), 619–644. doi: 10.1137/S1064827501387826