Posted on 30 Nov 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.1002/essoar.10508267/v2 — This a preprint and has not been peer reviewed. Data may be preliminary.

Global sensitivity analysis using the ultra-low 1 resolution Energy Exascale Earth System Model 2

Irina Kalashnikova Tezaur[1,1], Kara Peterson[1,1], Amy Powell[1,1], John Jakeman[1,1], and Erika Roesler[1,1]

[1]Sandia National Laboratories

November 30, 2022

Abstract

For decades, the Arctic has been warming at least twice as fast as the rest of the globe. As a first step towards quantifying parametric uncertainty in Arctic feedbacks, we perform a variance-based global sensitivity analysis (GSA) using a fully-coupled, ultra-low resolution (ULR) configuration of version 1 of the Department of Energy's Energy Exascale Earth System Model (E3SMv1). The study randomly draws 139 realizations of ten model parameters spanning three E3SMv1 components (sea ice, atmosphere and ocean), which are used to generate 75 year long projections of future climate using a fixed pre-industrial forcing. We quantify the sensitivity of six Arctic-focused quantities of interest (QOIs) to these parameters using main effect, total effect and Sobol sensitivity indices computed with a Gaussian process emulator. A sensitivity index-based ranking of model parameters shows that the atmospheric parameters in the CLUBB (Cloud Layers Unified by Binormals) scheme have significant impact on sea ice status and the larger Arctic climate. We also use the Gaussian process emulator to predict the response of varying each variable when the impact of other parameters are averaged out. These results allow one to assess the non-linearity of a parameter's impact on a QOI and investigate the presence of local minima encountered during the spin-up tuning process. Our study confirms the necessity of performing global analyses involving fully-coupled climate models, and motivates follow-on investigations in which the ULR model is compared rigorously to higher resolution configurations to confirm its viability as a lower-cost surrogate in fully-coupled climate uncertainty analyses.

Global sensitivity analysis using the ultra-low resolution Energy Exascale Earth System Model

Irina Tezaur[1], Kara Peterson[2], Amy Powell[3], John Jakeman[4], Erika Roesler[5]

[1]Quantitative Modeling & Analysis Dept, Sandia National Laboratories, Livermore, CA 94551 U.S.A.
[2]Computational Mathematics Dept, Sandia National Laboratories, Albuquerque, NM 87123 U.S.A.
[3]Systems Design & Architecture Dept, Sandia National Laboratories, Albuquerque, NM 87123 U.S.A.
[4]Optimization & UQ Dept, Sandia National Laboratories, Albuquerque, NM 87123 U.S.A.
[5]Geosciences Engineering Dept, Sandia National Laboratories, Albuquerque, NM 87123 U.S.A.

Key Points:

- We perform the first global sensitivity analysis using the fully-coupled ultra-low resolution Energy Exascale Earth System Model (E3SM).
- Uncertainty in cloud physics parameters is found to most greatly impact Arctic climate predictions.
- Our inferred quantity of interest-parameter correlations uncover key physical feedbacks and can guide model tuning.

Corresponding author: Irina Tezaur, `ikalash@sandia.gov`

Abstract

For decades, the Arctic has been warming at least twice as fast as the rest of the globe. As a first step towards quantifying parametric uncertainty in Arctic feedbacks, we perform a variance-based global sensitivity analysis (GSA) using a fully-coupled, ultra-low resolution (ULR) configuration of version 1 of the Department of Energy's Energy Exascale Earth System Model (E3SMv1). The study randomly draws 139 realizations of ten model parameters spanning three E3SMv1 components (sea ice, atmosphere and ocean), which are used to generate 75 year long projections of future climate using a fixed pre-industrial forcing. We quantify the sensitivity of six Arctic-focused quantities of interest (QOIs) to these parameters using main effect, total effect and Sobol sensitivity indices computed with a Gaussian process emulator. A sensitivity index-based ranking of model parameters shows that the atmospheric parameters in the CLUBB (Cloud Layers Unified by Binormals) scheme have significant impact on sea ice status and the larger Arctic climate. We also use the Gaussian process emulator to predict the response of varying each variable when the impact of other parameters are averaged out. These results allow one to assess the non-linearity of a parameter's impact on a QOI and investigate the presence of local minima encountered during the spin-up tuning process. Our study confirms the necessity of performing global analyses involving fully-coupled climate models, and motivates follow-on investigations in which the ULR model is compared rigorously to higher resolution configurations to confirm its viability as a lower-cost surrogate in fully-coupled climate uncertainty analyses.

Plain Language Summary

Feedbacks associated with Arctic warming are consequential for both the region and the strongly coupled global climate system. With the goal of assessing the variability of the impacts of global warming and associated feedbacks in model-based predictions, we study the sensitivity of the Arctic climate state to ten uncertain model parameters using a computationally inexpensive ultra-low resolution (ULR) configuration of the Department of Energy's global climate model, the Energy Exascale Earth System Model (E3SM). We can confidently conclude that, of the ten parameters considered, the atmospheric parameters in E3SM's cloud physics model are the most influential, having a strong influence on atmosphere, sea ice and ocean quantities of interest. Since identifying such cross-component influences is impossible without running a fully-coupled climate model, our study demonstrates the importance of fully-coupled climate analyses. To the best of our knowledge, this is the first global sensitivity study using E3SM's current release, and the first known scientific study involving the ULR configuration. Our study suggests that the ULR E3SM shows some promise in being used as a relatively inexpensive surrogate for higher resolution climate models, and demonstrates a need for rigorous quantitative follow-on studies involving this model configuration.

1 Introduction

Understanding the impact of warming on the Arctic is important because regional events can lead to high-consequence global changes (Lenton, 2008, 2012; Bathiany et al., 2016) including tipping points (irreversible changes in the global climate system (Lenton, 2008; Peterson et al., 2020)). Melting of the Greenland ice sheet will result in global sea level rise with risks to coastal infrastructure (Graeter et al., 2018). Sea ice loss will lead to increased maritime activity and possibly geopolitical conflict as more nations vie for access to the region (L. C. Smith & Stephenson, 2013). In addition, there is evidence that loss of sea ice and Arctic warming can induce changes in mid-latitude weather and precipitation (Cohen, Zhang, et al., 2018; Cohen, Pfeiffer, & Francis, 2018; Cvijanovic et al., 2017) potentially leading to food and water shortages (Parry et al., 2001).

According to recent data (*Snow, Water, Ice, and Permafrost in the Arctic (SWIPA)*, 2017; Richter-Menge et al., 2019; IPCC, 2021), the Arctic is warming at more than twice the rate of the rest of the globe. This accelerated Arctic warming leads to changes in a variety of physical systems influencing Arctic climate. For instance, the well-known ice-albedo feedback effect has been shown to contribute to sea ice loss. As highly reflective sea ice is lost, the surface albedo is reduced and solar radiation absorption in the darker ocean water is increased (Goosse et al., 2018). This positive feedback is counteracted by a negative feedback mechanism whereby thinner sea ice grows more quickly in response to thermodynamic forcing from the ocean and atmosphere. Permafrost thaw is increasing greenhouse gas release, thereby increasing warming (Parazoo et al., 2018; Schuur et al., 2015). Both sea ice and land ice melt are increasing freshwater flux into the North Atlantic, which can lead to ocean current disruptions and further changes to climate (Sevellec et al., 2017).

As a first step towards identifying possible tipping events stemming from climate change-driven processes in the Arctic with quantified uncertainty, we present a global sensitivity analysis of climate projections of version 1 of the U.S. Department of Energy's (DOE's) fully-coupled Energy Exascale Earth System Model (E3SMv1). To motivate the main contributions of this paper, we first provide a brief overview of related past work, focusing on studies aimed at addressing the sensitivity of Earth System Model (ESM) components and coupled models to various model parameters.

1.1 Overview of related work

Recent years have seen a number of studies aimed at understanding the sensitivity of various climate models to relevant parameters. The vast majority of this work has focused on individual components of a global ESM, e.g., the ocean, sea ice and atmosphere components. Several authors have investigated the sensitivity of ocean models to parameters, most of them examining subgrid mixing parameterizations, wind drag, model domain and grid resolution, numerical formulations and topography (Alexanderian et al., 2012; Bernard et al., 2006; M. Hecht & Smith, 2008; M. W. Hecht et al., 2008; Hurlburt & Hogan, 2000; Maltrud & McClean, 2005; Asay-Davis et al., 2018; Reckinger et al., 2015). A handful of studies have examined the sensitivity of model predictions to model parameters in stand-alone configurations of sea ice models, including (Kim et al., 2006; Peterson et al., 2010; Uotila et al., 2012; Urrego-Blanco et al., 2016). In the most recent of these works (Urrego-Blanco et al., 2016), Urrego-Blanco *et al.* conducted a comprehensive sensitivity analysis of sea ice thickness and area to 39 sea ice model parameters using Sobol sequences together with a fast emulator for the Los Alamos sea ice model, CICE (Community Ice CodE) (Hunke et al., 2015). Similar sensitivity studies have been done for stand-alone atmosphere models, e.g. (Zhao et al., 2013; Covey et al., 2013; Qian et al., 2018; Rasch et al., 2019; Guo et al., 2014). Zhao *et al.* (Zhao et al., 2013) evaluated the sensitivity of radiative fluxes at the top of the atmosphere to various cloud microphysics and aerosol parameters. Covey *et al.* (Covey et al., 2013) used Morris one-at-a-time (MOAT) screening to estimate sensitivity with respect to 27 atmospheric parameters. Qian *et al.* (Qian et al., 2018), estimated the sensitivity of the model fitness of generalized linear model (GLMs) of response variables obtained from short (three day) simulations of a 1° resolution E3SM atmosphere model (EAM) with respect to 18 parameters from various parts of the atmospheric dycore, including parameterizations of deep convection, shallow convection and cloud macro/microphysics. Guo *et al.* (Guo et al., 2014), used GLMs to determine the most influential parameters of the Cloud Layers Unified by Binormals (CLUBB) physics parameterization within the single-column version of the Community Atmosphere model version 5 (SCAM5). In related recent work focused on the EAM, Rasch *et al.* (Rasch et al., 2019) demonstrated the utility of using lower-resolution versions of the EAM atmospheric component and short-term hindcasts to guide tuning and sensitivity analysis of higher-resolution models.

While the aforementioned studies provide much insight into individual ESM components, without considering a fully-coupled ESM, it is impossible to identify the interaction among various climate components. Hence, studies focusing on a single climate component have the danger of significantly overlooking relevant climate feedbacks. Performing sensitivity studies on fully-coupled climate models is far more challenging than considering an individual climate component. The main hurdle is the fact that running a fully-coupled ESM is far more computationally expensive than running a single climate component. Since sensitivity studies typically require many simulation ensembles, sensitivity analyses using fully-coupled models are typically intractable without the use of efficient surrogates, especially at "production" grid resolutions. The authors are aware of only one reference focusing on a sensitivity study involving several climate components using a fully-coupled ESM, namely (Urrego-Blanco et al., 2019). In (Urrego-Blanco et al., 2019), Urrego-Blanco *et al.* use the 1° resolution of the E3SM v0-HiLAT (EHV0) fully coupled climate system (developed for the simulation of high-latitude processes) to identify emerging relationships between sea ice area, net surface longwave radiation and atmospheric circulation over the Beaufort gyre. The authors consider five model parameters, two from the atmosphere model (version 5 of the Community Atmosphere Model, or CAM5 (Dennis et al., 2012)), two from the sea ice model (version 5 of the Los Alamos Sea Ice Model, or CICE5 (Hunke et al., 2015)) and one from the ocean model (version 2 of the Parallel Ocean Program, or POP2 (R. Smith et al., 2010)), and initialize their model using pre-industrial forcing. By employing an elementary effects or MOAT method (Morris, 1991) for their sensitivity analysis (an approach that perturbs one input parameter at a time, rather than all parameters together), the authors are able to keep the number of ensemble members (or E3SM simulations) required down to just 24.

It is worthwhile to note that there are other works utilizing global climate models for sensitivity analyses targeting a single climate component. For instance, the authors of (Rae et al., 2014) perform a sensitivity study of the sea ice simulation within the global coupled climate model HadGEM3. Here, both the Arctic and Antarctic are considered. In a similar vein, Uotila *et al.* (Uotila et al., 2012) explore the sensitivity of the global sea ice distribution of the Australian Climate Ocean Model (AusCOM) to a range of sea ice physics-related parameters within a global ocean-ice model comprised of AusCOM coupled with the Los Alamos CICE model. While studies such as these have the advantage of incorporating feedbacks from the global climate system, they have a similar limitation of single-component sensitivity studies in that they preclude the identification of cross-component parameter interactions.

1.2 Contributions and organization

Our present work is primarily motivated by the recent study in (Urrego-Blanco et al., 2019), but differs in several important ways. First, we consider version 1 of the E3SM (referred to herein as E3SMv1), the newest available release of the E3SM to date. Second, we employ a much lower spatial resolution grid than those considered in (Urrego-Blanco et al., 2019). We will refer to our resolution model as the "Ultra-Low Resolution" (ULR) model, which corresponds to a 7.5° grid resolution in the atmosphere and 240 kilometer grid resolution for the ocean and sea ice. By using a lower resolution grid, we are able to afford far more simulations, allowing us to consider more parameters and employ more sophisticated sensitivity analysis approaches than the MOAT method used in (Urrego-Blanco et al., 2019). Specifically, we perform a variance-based global sensitivity analysis which uses Gaussian process emulators constructed using the `PyApprox` library (J. D. Jakeman, 2021). To the best of our knowledge, this is the first global sensitivity analysis involving the fully-coupled E3SMv1, and the first scientific study involving the ULR configuration of this model.

In our variance-based GSA, we study the effect of ten parameters, spanning three E3SM components, the sea ice model (MPAS-SeaIce (M. R. Petersen et al., 2019)), the E3SM atmosphere model (EAM (Rasch et al., 2019)) and the ocean model (MPAS-Ocean (M. Petersen et al., 2018)), on six Arctic-focused quantities of interest (QOIs). We construct fast Gaussian process emulators for these QOIs using 139 75-year ensemble runs of the fully-coupled ULR E3SMv1. Each simulation is initialized from a spun-up initial condition generated specifically for this study (a spun-up initial condition was not readily available at the considered resolution) and forced with pre-industrial control conditions. Using each emulator we calculate Sobol sensitivity, main effect and total effect indices of our ten parameters. Main effect indices quantify the effects of single parameters acting in isolation, and Sobol and total effect indices are useful for identifying strong parameter interactions.

The 139 ensemble runs comprising this study exhibited significant variability, with several runs resulting in complete loss of Arctic sea ice and several other runs exhibiting an apparent exponential growth in the amount of Arctic sea ice. The main takeaway from our study is that the parameters in the cloud physics parameterization within the atmosphere component of the E3SMv1 have the most impact on the Arctic climate state. Our study identified several relationships between QOI, which match physics-based intuition (e.g., ensemble members with low sea ice extent had high surface air temperature), and led to plausible conclusions regarding feedback processes important to the Arctic climate state (e.g., seasonal cloud convective regimes can create a feedback that affects Arctic sea ice extent). These results suggest that the ULR model may serve as a viable lower-cost surrogate for sensitivity analysis and uncertainty quantification workflows, and motivate a follow-on evaluation/validation study in which the ULR model is compared rigorously to higher resolution configurations of the E3SM for the purpose of identifying its current limitations. By constructing univariate functions by marginalizing all but a single parameter, we are additionally able to determine whether increasing/decreasing a given parameter will increase or decrease a given QOI. These results are useful in guiding model spin-ups, and are consistent with the parameter-QOI correlations uncovered by our manual spin-up of the ULR E3SMv1.

The remainder of this paper is organized as follows. We detail the methods employed in this study in Section 2. This includes a description of our coupled model, E3SMv1, and our tuning of this model, together with a discussion of the design and implementation of our global sensitivity study using this coupled model. In Section 3, we present the main results of our global sensitivity study applied to the ULR E3SMv1, and provide a discussion of their significance. We end with a concluding summary (Section 4).

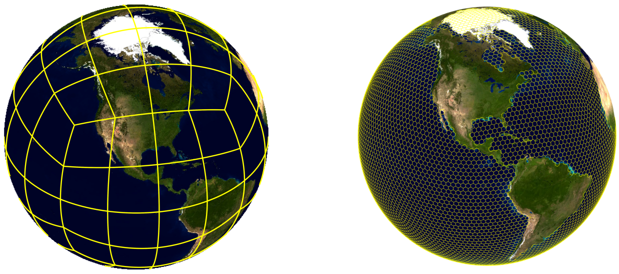
2 Methods

2.1 E3SMv1 coupled climate model

In the present study, E3SMv1 is used to investigate changes in Arctic sea ice in response to internal variability related to ocean and atmosphere modes as well as in response to perturbations in the model parameters. E3SM consists of component models for atmosphere, ocean, ice, land, and river transport. The E3SM Atmosphere Model (EAM) (Rasch et al., 2019) has a spectral element dynamical core discretized on a cubed sphere grid using 72 vertical levels. The standard resolution E3SM configuration uses a 1° grid for both EAM and the E3SM Land Model (ELM) (Bisht et al., 2018), which corresponds to approximately 110 km at the equator. The ocean and sea ice models are based on the Model for Prediction Across Scales (MPAS) framework (Heinzeller et al., 2016). MPAS-Ocean (M. Petersen et al., 2018) uses a finite volume discretization on an unstructured Voronoi grid, which is shared with MPAS-SeaIce (M. R. Petersen et al., 2019). At the standard resolution, the ocean and sea ice grid has a resolution varying between 60 km at midlatitudes and 30 km at the poles. The Model for Scale Adaptive River Transport (MOSART) (Cornette, 2012) is also employed, and has a resolution of 50 km.

The present study is based on an ULR configuration of E3SMv1. We chose an ULR configuration that would provide a computationally tractable way to generate larger numbers of ensemble runs to explore the parameter space in the coupled model. This ULR configuration has a grid resolution of approximately 7.5 degrees for EAM and ELM and 240 km or approximately 2.2 degrees for MPAS-Ocean and MPAS-SeaIce. A plot of the ULR grid employed in this study is provided in Figure 1. It is noted that the ULR is not an officially supported E3SM resolution. While this resolution is commonly used for testing purposes by E3SM developers, this paper is the first work (to the authors' knowledge) that investigates the use of the ULR configuration for scientific studies. To quantify the computational advantages of the ULR configuration, we note that it achieves approximately 4 simulated years per day per node on the Skybridge cluster (described in Section 2.5) in comparison to 0.035 simulated years per day per node for the 1° standard resolution configuration of E3SM. This results in an estimate that the ULR configuration is more than 100 times faster than the standard resolution configuration.

In the following section, we assess the predictive performance of the ULR E3SM. We find that ULR predictions capture the large scale features of the 1° model, which suggests that the ULR model can help inform sensitivity analysis and uncertainty quantification of higher resolution models.



(a) (b)

Figure 1. Ultra-low resolution grid for atmosphere (a) and ocean (b) used in our E3SMv1 study.

.

2.2 E3SMv1 ultra-low configuration tuning

For our ULR simulations, we first performed a spin-up (i.e., running the model until equilibrium state is achieved) using pre-industrial control (piControl) forcing for 500 simulated years with default parameter values. It is desirable at the end of the spin-up to have a near-zero long-term average net top-of-atmosphere (TOA) energy flux, constant global average mean surface air temperature and stable yearly sea ice coverage in order to initialize the perturbed runs with a stable state. Our original 500-year spin-up simulation exhibited a warm bias, with surface temperature elevated, compared with observations and declining sea ice over the 500-year period. (See Figure 2). To improve the model tuning, we ran an additional 180 years starting from year 500 of the spin-up simulation using atmospheric parameter values modified to match the final tuning from the Golaz *et al.* paper (Golaz et al., 2019). Parameter values are given in Table 1.

The branch run with the Golaz *et al.* values did result in a more realistic climate with stable surface temperature, net TOA flux and sea ice extent. In Figure 2, time series plots of these quantities for the 500-year spin-up using default parameter values are shown in blue with the final 180-years from the simulation with modified parameter values shown in red. Note that slopes of the quantities are near zero for the branched run indicating that the simulations have reached an equilibrium. The final year of this branch run was used as the initial condition for all perturbed sensitivity analysis simulations as well as for a baseline simulation that continued with the same parameter values for an additional 75 years. Investigations of the impact of the equilibrium values of the initial state on sensitivity

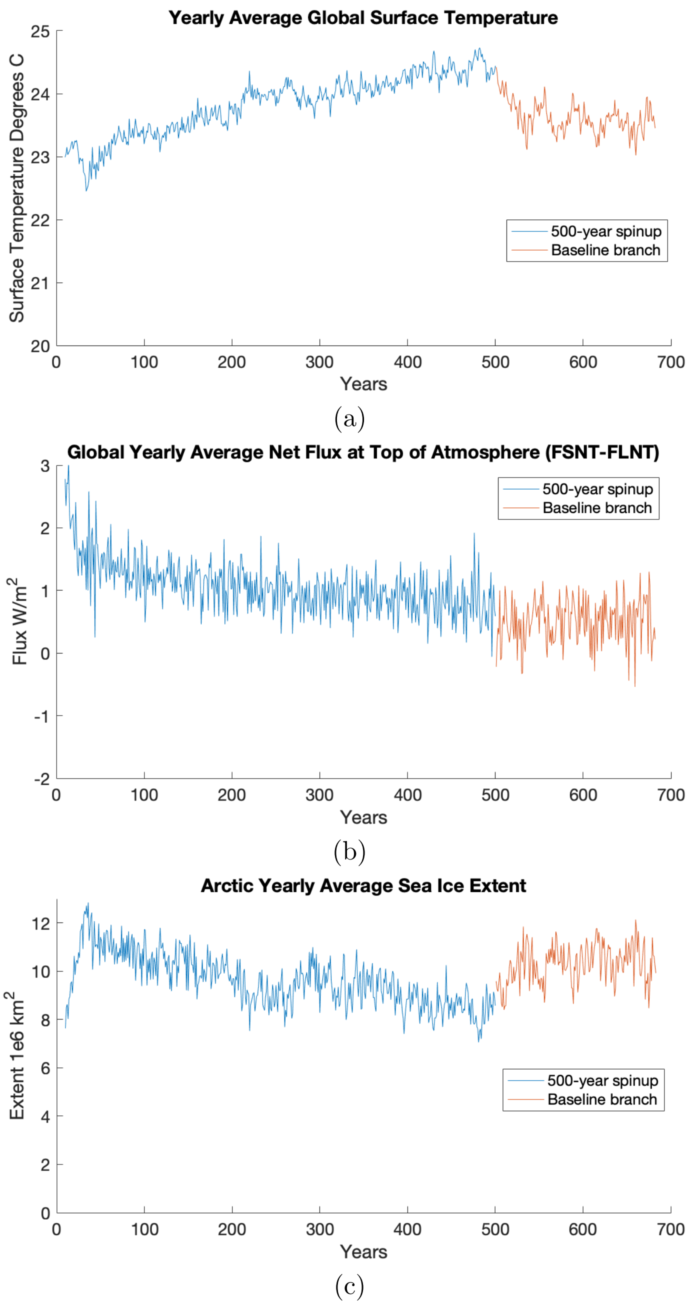


Figure 2. Yearly averaged global surface air temperature (°C) (a), yearly averaged net flux at TOA (W/m^2) (b), and yearly averaged sea ice extent (10^6 km^2) (c). The blue line is from the 500-year ULR model spin-up with default parameter values and the red line is from the 180-year branch run with modified parameters values as shown in Table 1.

.

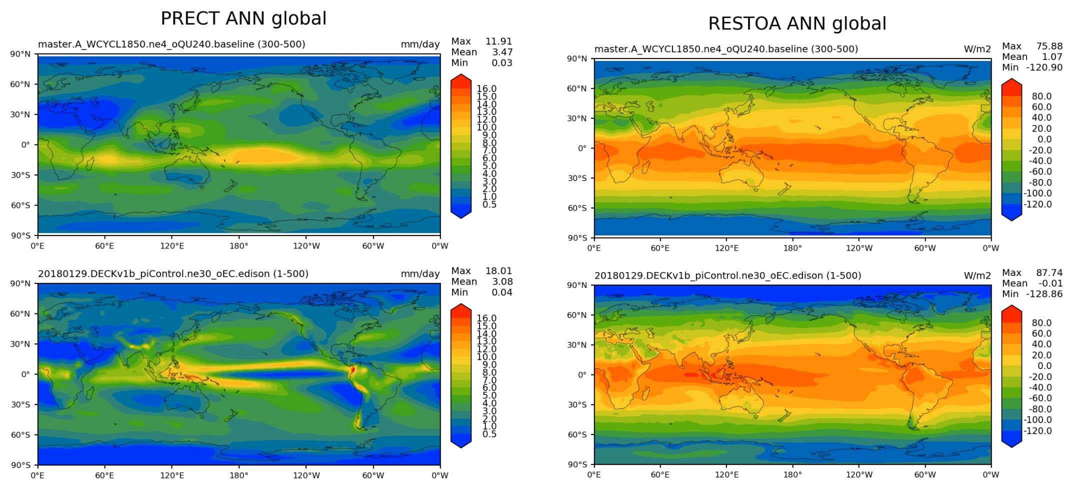
analysis results are beyond the scope of this study, but this could be addressed in future work using additional tunings of the ULR model informed by our results involving the marginalized main effect indices (Section 3.5).

To confirm that the ULR simulation is able to capture large-scale spatial-variations, we computed annual climatologies from the final 200 years in the initial spin-up. Plots of

Table 1. Default atmospheric parameter values for ULR configuration and corresponding values from Golaz *et al.* (Golaz et al., 2019). In this table, `zmconv_ke` is the coefficient for evaporation of convective precipitation, `so4_sz_thresh_icenuc` is the Aitken mode SO_4 size threshold used for homogeneous ice nucleation, and `clubb_c14` is the damping coefficient for u'^2 and v'^2 in the CLUBB (Larson, 2020) aerosol physics parameterization.

Parameter	Default value	Golaz *et al.* value
`zmconv_ke`	1.5×10^{-6}	5.0×10^{-6}
`so4_sz_thresh_icenuc`	7.53×10^{-8}	5.0×10^{-8}
`clubb_c14`	1.3	1.06

the annual average global precipitation, and TOA flux are shown in Figure 3, where the final 200 years of the ULR baseline spin-up is compared to the average of the 500 year pre-industrial control simulation from the 1° E3SMv1 CMIP6 simulations. The 1° resolution E3SMv1 simulations have been scientifically validated and provide a reference for these quantities in the ULR simulation (Golaz et al., 2019). In Figure 4, zonal mean values of surface temperature and zonal wind comparisons between the ULR and 1° show vertical variation in the atmosphere. Figures 3 and 4 demonstrate that the ULR simulation does capture the large-scale features of the flow providing support that the ULR configuration can be an effective surrogate for the standard resolution and provide useful information to guide targeted higher-resolution modeling.

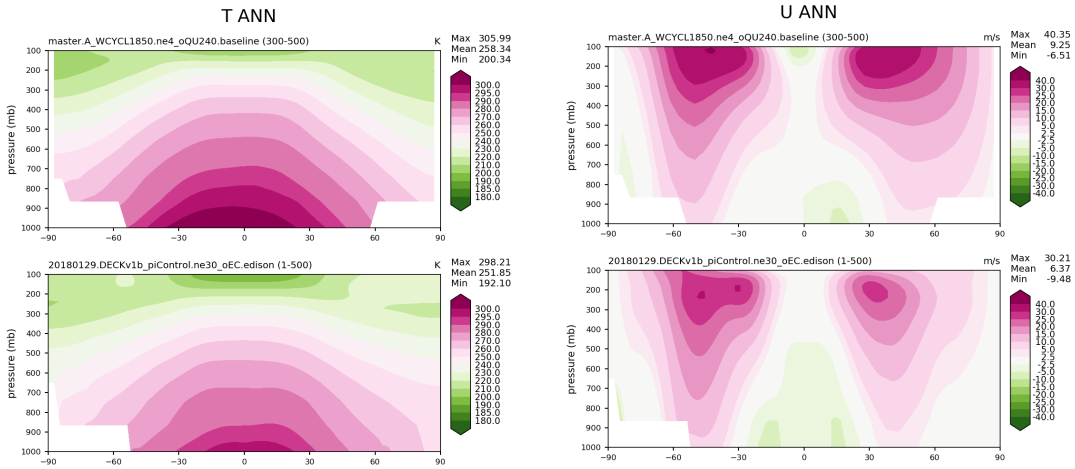


(a) (b)

Figure 3. Precipitation (mm/day) (a) and TOA (W/m^2) (b) for years 300-500 of the ULR pre-industrial control spin-up (top) and for the 1° standard resolution pre-industrial control run (bottom).

2.3 Design of global sensitivity study (GSA)

The first step in designing a sensitivity study, given a spun-up initial condition, is selecting the set of parameters (which will be denoted by $\{z_i\}$) to be perturbed, together with the set of relevant QOIs on which the parameters are expected to have an effect. A description of the parameters, their baseline values, and the range of their perturbed values is given in Table 2. The parameters were chosen based on their significance in previous sensitivity studies involving both individual component as well as fully-coupled climate



(a) (b)

Figure 4. Zonal temperature (°C) (a), and zonal winds (m/s) (b) for years 300-500 of the ultra-low resolution pre-industrial control spin-up (top) and for the 1° standard resolution pre-industrial control run (bottom).

simulations, most notably (Urrego-Blanco et al., 2016, 2019; Reckinger et al., 2015; Asay-Davis et al., 2018; Qian et al., 2018; Rasch et al., 2019). Of the ten parameters, three are from the sea ice component (MPAS-SeaIce), three are from the ocean component (MPAS-Ocean) and four are from the atmosphere model (EAM) – more specifically, the CLUBB (Larson, 2020) aerosol physics parameterization within EAM.

Our global sensitivity analysis (GSA) is based upon random realizations of the ten parameters randomly selected from a uniform distribution over the ranges defined by the "Min" and "Max" values given in Table 2. The sampling and associated model evaluations were managed using the `DAKOTA` library (Adams et al., 2013), an open-source software package for optimization, uncertainty quantification and advanced parametric analysis. Much like the parameters themselves, the selection of the parameter ranges was guided by past analyses (Urrego-Blanco et al., 2016, 2019; Reckinger et al., 2015; Asay-Davis et al., 2018; Qian et al., 2018; Rasch et al., 2019). It is worthwhile to note that the three MPAS-SeaIce parameters selected in our GSA were hard-coded to their default values in the `master` branch of the E3SM[1]. In order to enable the straightforward specification of these parameters in the relevant input file, a fork of the E3SM was created[2] and used in the present study. Instructions for cloning this fork as well as building the code and submitting a perturbed run are provided in Appendix B of (Peterson et al., 2020).

In the present study, we report sensitivity metrics for a set of six QOIs, summarized in Table 3. This set of QOIs is selected for several reasons, including: (1) their overlap with QOIs considered in similar past works (Rasch et al., 2019; Urrego-Blanco et al., 2019) (to enable comparisons), (2) their importance and relevance to studying the Arctic climate state (e.g., the CLDLOW QOI, which represents low cloud coverage, is selected because low clouds are particularly important in the Arctic and may impact sea ice coverage), and (3) the fact that they span the three climate components targeted by this study (sea ice, ocean, atmosphere). Following the approach in (Urrego-Blanco et al., 2016, 2019), we look at the QOIs in Table 3 annually as well as seasonally.

[1] Available at: `https://github.com/E3SM-Project/E3SM`.
[2] Available at: `https://github.com/karapeterson/E3SM` (`add_namelist_params` branch).

Table 2. Global sensitivity analysis parameters.

Component	Variable	Parameter	Baseline	Min	Max	Description [Units]
MPAS-SeaIce	z_1	`ksno`	3.0×10^{-1}	2.0×10^{-1}	6.0×10^{-1}	Snow conductivity [$\mathrm{Wm^{-1}K^{-1}}$]
	z_2	`lambda_pond`	1.1574×10^{-6}	1.15×10^{-8}	1.15×10^{-4}	Drainage timescale of ponds [$\mathrm{s^{-1}}$]
	z_3	`dragio`	5.36×10^{-3}	2.0×10^{-4}	1.6×10^{-1}	Ocean-ice drag [−]
EAM	z_4	`cldfrc_dp1`	4.5×10^{-2}	2.0×10^{-2}	1.0×10^{-1}	Deep convection cloud fraction parameter in CLUBB [−]
	z_5	`clubb_c1`	1.335	1.0	5.0	Constant associated with dissipation of variance w^2 in CLUBB [−]
	z_6	`clubb_c8`	4.3	2.0	8.0	Constant associated with Newtonian damping of w^3 in CLUBB [−]
	z_7	`gamma_coeff`	3.2×10^{-1}	1.0×10^{-1}	5.0×10^{-1}	Constant width of PDF in w coord in CLUBB [−]
MPAS-Ocean	z_8	`standardgm_tracer_kappa`	1.8×10^{3}	6.0×10^{2}	1.8×10^{3}	Bolus coefficient of GM parameterization of eddy transport [$\mathrm{m^2/s}$]
	z_9	`cvmix_kpp_critical _bulkrichardsonnumber`	2.5×10^{-1}	2.0×10^{-1}	1.0	Bulk Richardson number used in KPP vertical mixing scheme [−]
	z_{10}	`salinity_restoring constant_piston_velocity`	1.585×10^{-6}	7×10^{-7}	3×10^{-6}	Rate at which salinity is restored to a monthly climatology [m/s]

Note that we originally obtained results for a larger set of QOIs than those summarized in Table 3, as discussed in (Peterson et al., 2020). Specifically, we considered five additional QOIs: the surface air specific humidity averaged over 60-90° (QS), the large-scale snow precipitation averages over 60-90° (PRECSL), and the mean sea level pressure over the Beaufort Sea, the Aleutian Low and the Siberian High (BH, AL and SH, respectively). We omit these results here largely for the sake of brevity. The former two QOIs (QS and PRECSL) were highly correlated with other QOIs, so including those results would not add much to the discussion. Additionally, our sensitivity analysis results for the latter three QOIs (BH, AL and SH) precluded us from making strong conclusions about the impact of parameter variations on these QOIs, as the relevant ensemble trajectories resembled white noise (indicating there was no clear signal) and high errors in the sensitivity indices were observed.

Each perturbed simulation in our study was run up to time T_{final}, and was given a spin-up period of $T_{\text{spin-up}} < T_{\text{final}}$ to equilibrate the simulation (that is, to get past the inevitable transient period that occurs when the run commences). Here, we prescribed a spin-up period of 50 years ($T_{\text{spin-up}} = 50$ years), and each perturbed model configuration was run until time $T_{\text{final}} = 75$ years. In general, it is not expected for all the perturbed simulations to run to completion, and indeed crashes (discussed in more detail in Section 3) occurred for a handful of our runs. For the successful runs (runs that made it to year 75), our six QOIs were calculated by averaging annually and seasonally over the last 25 years of the simulations (i.e., between times $t = T_{\text{spin-up}} + 1$ and T_{final}).

Table 3. Global sensitivity analysis quantities of interest (QOIs).

QOI	Units	Description	Component
SIE	km^2	Total Arctic sea ice extent	sea ice
SIV	km^3	Total Arctic sea ice volume	sea ice
SST	°C	Sea surface temperature averaged over 60-90° N	ocean
TS	°C	Surface air temperature averaged over 60-90° N	atmosphere
FLNS	W/m^2	Net longwave flux at surface over 60-90° N	atmosphere
CLDLOW	–	Low cloud coverage below 700 hPa averaged over 60-90° N	atmosphere

As discussed earlier in Sections 2.1 and 2.2, the GSA study performed herein used the ULR configuration of the E3SMv1 and pre-industrial (piControl) forcing. Repeating the study with a different forcing, such as one of the forcings in (Golaz et al., 2019), would be an interesting and useful follow-on exercise to the present study.

2.4 Variance-based global sensitivity analysis

In this section, we describe the variance-based GSA used to determine the relative sensitivity of model predictions to uncertain model parameters.

2.4.1 Sobol indices

In this paper, Sobol sensitivity indices (Sobol, 2001) are used to quantify the relative importance of parameter combinations on a given QOI. With this goal, let f denote a model output QOI that depends on some model parameters $z = [z_1, ..., z_d]^T$. Any function f with finite variance parameterized by a set of independent variables z with probability distribution $\rho(z) = \prod_{j-1}^{d} \rho(z_j)$ and support $\Gamma = \bigotimes_{j=1}^{d} \Gamma_j$ can be decomposed into the following finite sum, referred to as the Analysis of Variance (ANOVA) decomposition,

$$f(z) = \hat{f}_0 + \sum_{i=1}^{d} \hat{f}_i(z_i) + \sum_{i,j=1}^{d} \hat{f}_{i,j}(z_i, z_j) + \cdots + \hat{f}_{1,\ldots,d}(z_1, \ldots, z_d), \tag{1}$$

or more compactly

$$f(z) = \sum_{\mathbf{u}\subseteq\mathcal{D}} \hat{f}_{\mathbf{u}}(z_{\mathbf{u}}), \tag{2}$$

where $\hat{f}_{\mathbf{u}}$ quantifies the dependence of the function f on the variable dimensions $i \in \mathbf{u}$ and $\mathbf{u} = (u_1, \ldots, u_s) \subseteq \mathcal{D} = \{1, \ldots, d\}$.

The functions $\hat{f}_{\mathbf{u}}$ can be obtained by integration, specifically

$$\hat{f}_{\mathbf{u}}(z_{\mathbf{u}}) = \int_{\Gamma_{\mathcal{D}\setminus\mathbf{u}}} f(z) d\rho_{\mathcal{D}\setminus\mathbf{u}}(z) - \sum_{\mathbf{v}\subset\mathbf{u}} \hat{f}_{\mathbf{v}}(z_{\mathbf{v}}), \tag{3}$$

where $d\rho_{\mathcal{D}\setminus\mathbf{u}}(z) = \prod_{j\notin\mathbf{u}} d\rho_j(z)$ and $\Gamma_{\mathcal{D}\setminus\mathbf{u}} = \bigotimes_{j\notin\mathbf{u}} \Gamma_j$. The first-order terms $\hat{f}_{\mathbf{u}}(z_i)$, $||\mathbf{u}||_0 = 1$ represent the effect of a single variable acting independently of all others. Similarly, the second-order terms $||\mathbf{u}||_0 = 2$ represent the contributions of two variables acting together, and so on.

The terms of the ANOVA expansion are orthogonal, i.e. the weighted L^2 inner product $(\hat{f}_{\mathbf{u}}, \hat{f}_{\mathbf{v}})_{L^2_\rho} = 0$, for $\mathbf{u} \neq \mathbf{v}$. This orthogonality facilitates the following decomposition of the variance of the function f

$$\mathbb{V}[f] = \sum_{\mathbf{u}\subseteq\mathcal{D}} \mathbb{V}\left[\hat{f}_{\mathbf{u}}\right], \qquad \mathbb{V}\left[\hat{f}_{\mathbf{u}}\right] = \int_{\Gamma_{\mathbf{u}}} \hat{f}_{\mathbf{u}}^2 d\rho_{\mathbf{u}}, \tag{4}$$

where $d\rho_{\mathbf{u}}(z) = \prod_{j\in\mathbf{u}} d\rho_j(z)$.

Two popular measures of sensitivity are the main effect and total effect indices given respectively by

$$S_i^M = \frac{\mathbb{V}\left[\hat{f}_{\mathbf{e}_i}\right]}{\mathbb{V}[f]}, \qquad S_i^T = \frac{\sum_{\mathbf{u}\in\mathcal{J}} \mathbb{V}\left[\hat{f}_{\mathbf{u}}\right]}{\mathbb{V}[f]}, \tag{5}$$

where $\mathbf{e}_i$ is the unit vector, with only one non-zero entry located at the i^{th} element, and $\mathcal{J} = \{\mathbf{u} : i \in \mathbf{u}\}$. Main effect values represent the expected decrease in variance obtained from observing z_i. The total effects measure the variance that remains after learning the values of every variable except z_i. In the following, we also report Sobol indices (Sobol, 2001)

$$S_{\mathbf{u}} = \frac{\mathbb{V}\left[\hat{f}_{\mathbf{u}}\right]}{\mathbb{V}[f]},$$

which measure the contribution of the interaction between the parameter subset $\mathbf{u}$ on the variance of the function f.

Note that three aforementioned quantities (Sobol indices, main effect indices and total effect indices), measure some aspect of *global* sensitivity. In particular, they reflect a variance attribution over the range of the input parameters, as opposed to the local sensitivity reflected by a derivative.

2.4.2 Gaussian process

The Sobol indices (4) can be computed using a number of different methods, for example via (Quasi) Monte Carlo sampling (Saltelli et al., 2010), using surrogates (such as polynomial chaos expansions (Sudret, 2008)), or with sparse grids (J. Jakeman et al., 2019). Herein, we employ the software library `PyApprox` (J. D. Jakeman, 2021), a flexible and efficient open-source[3] tool for high-dimensional approximation and uncertainty quantification, which utilizes Gaussian processes (Rasmussen & Williams, 2006; Harbrecht et al., 2020).

Gaussian processes are well-suited to computing approximations of high-dimensional computationally-expensive models, such as the one we consider in this paper. They have a number of desirable properties. First, Gaussian processes can accurately approximate the output of a complex model with limited training data. Second, sensitivity indices can be computed easily from the Gaussian process. Finally, the surrogate and the Sobol indices are endowed with probabilistic error estimates which measure the influence of using a finite set of training data. These error estimates can be used to weight the confidence placed in decisions made from the output of the Gaussian process.

Building a Gaussian process requires specifying a correlation function, $C(z, z')$ and a trend function. The Gaussian process leverages the correlation between training samples to approximate the residuals between the training data and the trend function. In this work we set the trend function to zero and consider the squared exponential kernel

$$C(z, z') = \exp\left(-\sum_{i=1}^{d} \frac{1}{2l_i^2}\left(z_i - z_i'\right)^2\right),$$

where $l = [l_1, \ldots, l_d]^T$ is a vector hyper-parameters that determine the exact nature of the correlation function.

A Gaussian process is a distribution over a set of possible functions. Given a set of training samples $\mathcal{Z} = \{z^{(i)}\}_{m=1}^{M}$, and associated function values $y = [f(z^{(1)}), \ldots, z^{(M)}]^\top$ (realizations of the random output Y) the posterior mean and variance of the Gaussian process are

$$m^\star(z) = t(z)^T A^{-1} y, \qquad C^\star(z, z') = C(z, z') - t(z)^T A^{-1} t(z'),$$

respectively, where

$$t(z) = [C(z, z^{(1)}), \ldots, C(z, z^{(N)})]^T,$$

and A is a matrix with elements $A_{ij} = C(z^{(i)}, z^{(j)})$ for $i, j = 1, \ldots, M$. In this work, we use `Scikit-learn` (Pedregosa et al., 2011) to construct the Gaussian process and estimate the hyper-parameters. Because of the differing magnitudes of the ranges of the training samples and values, we found it essential to normalize the training data. Specifically, we transformed the training samples to $[-1, 1]^d$ and normalized the training values to have mean zero and unit variance. Once the Gaussian process is constructed, we post-process the approximation using `PyApprox` to obtain main effect functions and sensitivity indices. Because the Gaussian process is itself random, the aforementioned quantities are also random.

2.4.2.1 Marginalized main effect functions. The main effect functions $\hat{f}_i(z_i) = \mathbb{E}\left[Y \mid z_i\right] - \mathbb{E}\left[Y\right]$ are linear functionals of the Gaussian process and thus the posterior distributions of $\hat{f}_i(z_i)$ are also Gaussian. For tensor-product densities ρ and separable kernels of the form $C(z, z') = \prod_{i=1}^{d} C_i(z_i, z_i')$, such as the squared-exponential used here, we can compute the posterior mean and variance of the main effect functions using one-dimensional (1D) quadrature rules (Oakley & O'Hagan, 2004). Specifically, the posterior mean of $\hat{f}_i(z_i)$ is $\mathbb{E}^\star\left[\mathbb{E}\left[Y \mid z_i\right]\right] - \mathbb{E}^\star\left[\mathbb{E}\left[Y\right]\right]$ where

$$\mathbb{E}^\star\left[\mathbb{E}\left[Y \mid z_i\right]\right] = t_i(z_i) \prod_{\substack{j=1 \\ j \neq i}}^{d} \int_\Gamma t_j(z_j) \rho_j(z)\, dz \tag{6}$$

[3] Available at: `https://github.com/sandialabs/pyapprox`.

Here the superscript $\star$ indicates we are taking the expectation over the posterior distribution of the Gaussian process and we have used the separability of the kernel to write $t(z) = \prod_{i=1}^{d} t_i(z_i)$. We use 100 point Gaussian quadrature rules to compute the 1D integrals in (6). We use a similar technique to compute the posterior variance of the main effect functions.

$$\mathbb{V}^\star \left[\mathbb{E}\left[Y \mid z_i\right]\right] = C_i(z_i, z_i')u - \left(t(z_i) \circ \tau\right)^\top A^{-1} \left(t(z_i) \circ \tau\right),$$

where $\circ$ is the Hadamard (element-wise) product and

$$u = \prod_{\substack{j=1\\j\neq i}}^{d} \int_\Gamma \int_\Gamma C_j(z_j, z_j')\rho_j(z)\rho_j(z')\, dzdz', \qquad \tau = \prod_{\substack{j=1\\j\neq i}}^{d} \int_\Gamma t_j(z_j)\rho_j(z)\, dz.$$

The left expression above requires a two-dimensional (2D) tensor-product quadrature, but since we are not evaluating the simulation model, this is inexpensive to apply. In Figures 15 and 16, we plot the normalized posterior mean of the main effect functions marginalized over one parameter at a time, plus or minus two standard deviations, that is

$$\mathbb{V}^\star \left[Y\right]^{-\frac{1}{2}} \left(\mathbb{E}^\star \left[\mathbb{E}\left[Y \mid z_i\right]\right] - \mathbb{E}^\star \left[\mathbb{E}\left[Y\right]\right]\right) \pm 2\mathbb{V}^\star \left[Y\right]^{-\frac{1}{2}} \mathbb{V}^\star \left[\mathbb{E}\left[Y \mid z_i\right]\right]^{\frac{1}{2}}.$$

2.4.2.2 Sensitivity indices. Given the presentation above, the posterior distribution of Sobol, main effect and total effect indices can not be obtained analytically. Following (Oakley & O'Hagan, 2004), we compute the posterior mean and variance as the sample average of the estimates of the indices obtained using 1000 different random realizations of the Gaussian process. For each realization we compute the sensitivity indices accurately (close to machine precision) using a procedure similar to that used for constructing the main effect functions. We omit the exact expressions used because they are overly complex. In Figures 8-13 we plot the median sensitivity indices (red line), the interquartile range (box) and the minimum and maximum values (whiskers).

2.5 Global sensitivity analysis workflow

Figure 5 summarizes our GSA workflow. First, an appropriate initial condition is obtained by spinning up the E3SM to equilibrium, as discussed in Section 2.2. Next, after selecting $T_{\text{spin-up}}$ and T_{final} (ensuring that these values are large enough to avoid initial transients in the ensemble runs), we employ the `DAKOTA` library (Adams et al., 2013) to generate N random samples of the parameters $\{z_i\}$ from the selected parameter ranges or probability distributions (Table 2). We then create namelist files for each of our E3SM runs, corresponding to each of the N randomly selected parameter sets (for our study, the relevant namelist files are `user_nl_cam`, `user_nl_mpaso`, `user_nl_mpascice`), and set off N runs of the E3SM, branching off the spun-up initial condition. Finally, we post-process the perturbed runs to extract from them the relevant QOIs (see Table 3), and perform the GSA by providing M QOI-parameter pairs to `PyApprox`, where $M \leq N$ is the number of runs that completed successfully (simulated the global climate state to time T_{final}). The workflow depicted in Figure 5 was largely automated through the creation of shell scripts that execute the relevant commands comprising these steps. These scripts are stored in a repository containing the E3SM fork used for this study[4]. All of our runs were performed on the Skybridge high-capacity cluster located at Sandia National Laboratories, which contains 1848 nodes, each having 16 2.6 GHz Intel Sandy Bridge processors.

3 E3SM simulation results

In the present study, a total of $N = 212$ sets of parameter combinations were generated, assuming uniform probability distributions given by the "Min" and "Max" values found in

[4] Available at: `https://github.com/karapeterson/E3SM`.

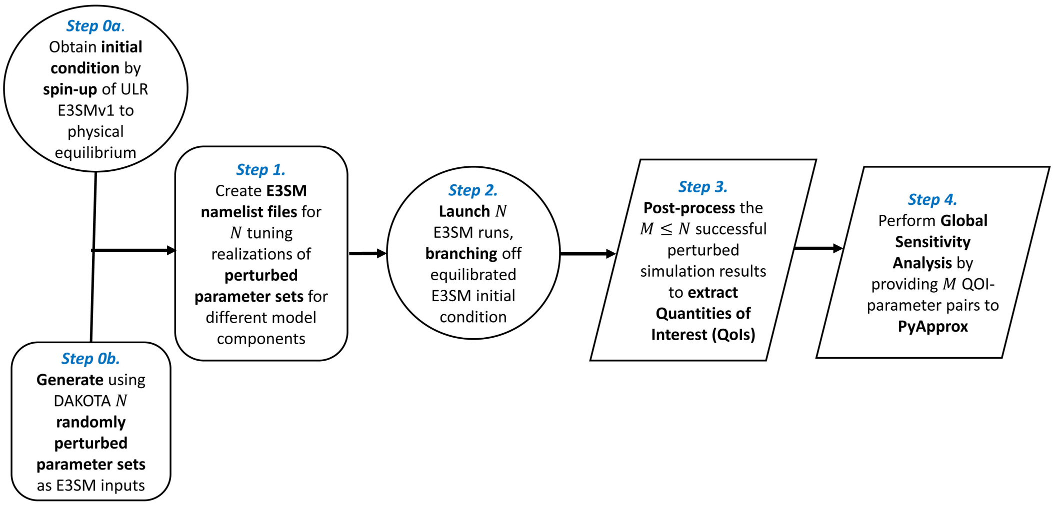


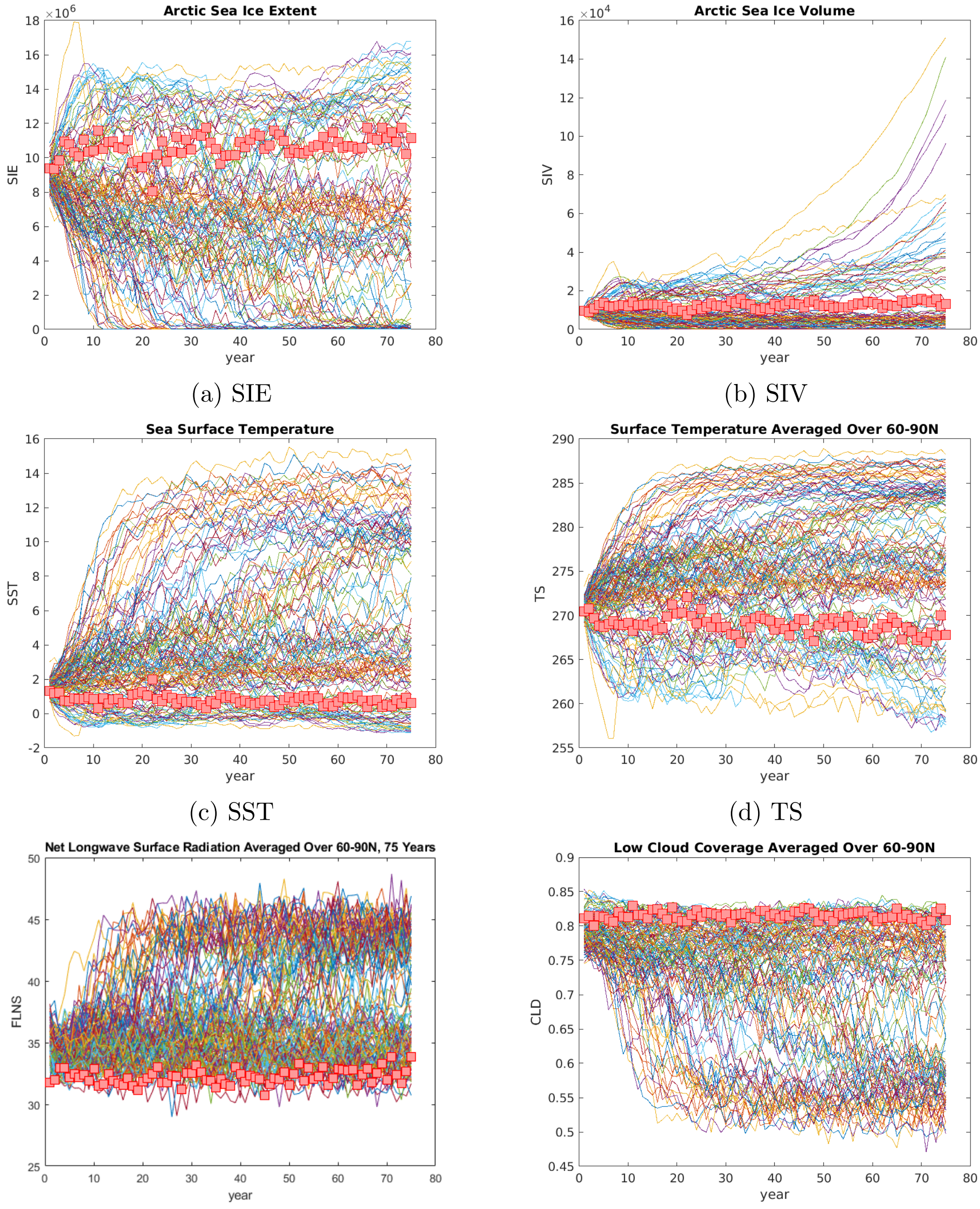
Figure 5. GSA workflow. Here N denotes the total number of perturbed E3SM simulations launched, and $M \leq N$ is the number of runs that completed successfully (simulated the global climate state to time T_{final}).

Table 2 for each parameter. We then set off 212 75-year perturbed runs of E3SMv1, one for each set of parameter values using pre-industrial control forcing. In addition to perturbing the values of the parameters in Table 2, modified parameter values from (Golaz et al., 2019), which are given in Table 1, were used for all of the perturbed runs for consistency with the final model spin-up, discussed in Section 2.2. The values of all 212 perturbed sets of parameters are given in Appendix C of (Peterson et al., 2020). Parameter values for the so-called "baseline" run, which was a continuation of the final spin-up run and included in our ensemble set, are given in Table 2. All of our simulations were run on 96 processors (6 nodes) of Sandia's Skybridge high-capacity cluster described earlier in Section 2.5.

Of the $N = 212$ perturbed runs, a total of 138 runs made it to $T_{\text{final}} = 75$ years. The baseline run also made it to $T_{\text{final}} = 75$ years, totaling $M = 139$ successful runs. As described earlier in Section 2.3, in calculating the QOIs in Table 3, we performed averaging both annually and seasonally over years 51-75, so as to allow each perturbed run a spin-up/equilibration period of 50 years.

3.1 Ensemble trajectories

Figure 6 shows the trajectories of all six QOIs considered (Table 3) for each of the 139 successful ensemble runs (runs that made it to year $T_{\text{final}} = 75$). The QOIs are averaged over each year and plotted as a function of the year since the start of each perturbed run. The baseline run is distinguished from the others by the red markers. All six QOIs are effectively in equilibrium at all times for the baseline run, as expected. A careful inspection of the trajectories in Figure 6 reveals that the relationships between the QOIs are also as expected, i.e., runs giving rise to a large sea ice area also give rise to a smaller surface air temperature. Additionally, one can see from Figure 6 that most of the perturbed runs appear to have reached equilibrium by year 40. This justifies the selection of $T_{\text{spin-up}} = 50$ years. It is interesting to remark that significant changes to the QOIs are seen in the perturbed runs, with several runs resulting in a complete loss of Arctic sea ice and several runs exhibiting an apparent exponential growth in Arctic sea ice. This suggests that our parameter choices and ranges produced a sufficiently wide range of possible climate outcomes, as intended.



(a) SIE (b) SIV

(c) SST (d) TS

(e) FLNS (f) CLDLOW

Figure 6. Ensemble trajectories of the QOIs in Table 3 for the ensemble members that made it to year 75. The baseline run is distinguished from the others by the red markers.

3.2 Ensemble statistics

We now look at some statistics for the perturbed runs that made it to year 75. Figure 7 shows box-and-whiskers plots for each of the six QOIs considered, calculated by season. Here, the seasons are defined as follows: "Winter" is comprised of the months of January to March, "Spring" is comprised of the months April to June, "Summer" is comprised of the months July to September, and "Autumn" is comprised of the months October to December. The red central mark indicates the median of the data, whereas the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The whiskers extend to the most extreme data points not considered outliers, and the outliers[5] are plotted using the '+' symbol.

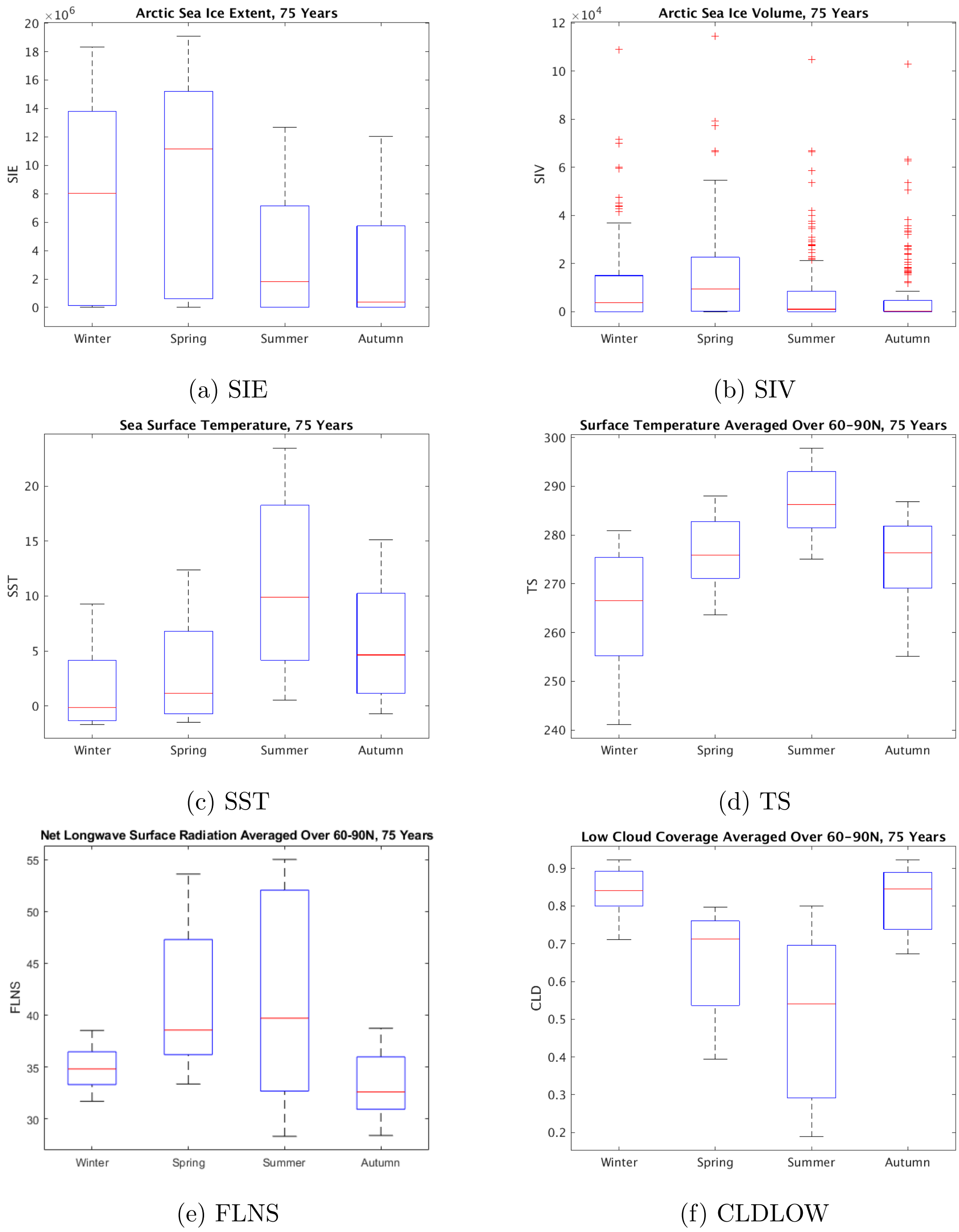
Figure 7 shows that the maximum and minimum sea ice extent is observed in the "Spring" and "Autumn" seasons, respectively. This result may seem surprising, as observational data and standard 1° resolution E3SM simulations (see (Peterson et al., 2020)) have shown that the maximum and minimum sea ice extent in general occur in March and September, respectively, which would correspond to the "Winter" and "Autumn" seasons based on our definition. A closer inspection reveals that, for the majority of our ULR runs, including the baseline run, the maximum and minimum sea ice extent occurs in April and October (for a plot showing this, the reader is referred to (Peterson et al., 2020)). Similarly, the maximum and minimum sea ice volume occurs in May and October, respectively. The cause of this shift in the month of maximum and minimum sea ice extent and volume in the ULR configuration is uncertain at this time, but these results motivate follow-on work to understand the behavior in more detail.

It is interesting to look at the relative spreads of the box-and-whiskers plots in Figure 7. This spread can be viewed as a measure of uncertainty. One can see from Figure 7(a) that the SIE QOI has the smallest uncertainty in the melting seasons (during which it is particularly relevant for trans-Arctic shipping routes), summer and autumn. The only QOI with significant outliers is the SIV. Referring to the ensemble trajectory plots, namely Figure 6(b), the reader can observe that the SIV QOI (an estimator of older, multi-year ice) is the only QOI with a significant number of trajectories so anomalous that they predict an apparent exponential growth in Arctic sea ice volume. It is likely that these trajectories translate to the outliers in the box-and-whiskers plot for SIV (Figure 7(b)); however, it is unclear what mechanism within the ULR E3SM is causing a feedback of this type. The SIV QOI has the same uncertainty trends as the SIE QOI if outliers are excluded; however, if outliers are included, the uncertainty in SIV is comparable across all four seasons, a result similar to the one obtained in (Urrego-Blanco et al., 2019). The remaining four QOIs have the largest uncertainty during the seasons in which they are either minimal (for TS and CLDLOW) or maximal (for SST and FLNS), on average. Certain expected correlations in uncertainties between the QOIs are observed. For example, the box-and-whisker plot spreads for the FLNS and CLDLOW mimic each other across all four seasons, which can be explained by the fact that FLNS is in general strongly determined by cloud variations and cloud cover (Schweiger et al., 2008).

3.3 Correlations in QOIs

Tables 4 – 7 give the correlation coefficients between our six QOIs, averaged seasonally. In general, the relationships between the QOIs are consistent with expectations. SIE and SIV, as well as SST and TS, have a strong positive correlation across all four seasons. SIE/SIV are negatively correlated with SST/TS, again as expected: larger sea ice volumes occur under lower air and sea surface temperatures. One can additionally observe a general negative correlation between CLDLOW and FLNS, especially during the warmer spring, summer and autumn seasons. This relationship can be explained by the fact that clouds absorb and re-emit the longwave radiation emitted by the surface. Since FLNS is defined as upward positive, one expects to see an increase in longwave flux down at the surface in the presence of clouds, which has the effect of decreasing net longwave flux. There is virtually no correlation between the following pairs of QOIs in the winter season: (SIE, FLNS), (SIV, CLDLOW) and (TS, FLNS). While the lack of correlation between (SIE, FLNS) and (SIV, CLDLOW) in winter may be surprising, this result is consistent with recent studies (Kay &

[5] Outliers are defined as values that are more than 1.5 times the interquartile range away from the top or bottom of the box in a given box plot.



(a) SIE (b) SIV

(c) SST (d) TS

(e) FLNS (f) CLDLOW

Figure 7. Box-and-whiskers plots showing ensemble statistics for the first six QOIs from Table 3. The red central mark indicates the median of the data, whereas the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. Outliers are plotted using the '+' symbol.

Gettelman, 2009) that have suggested that there is no clear relationship between cloud cover and sea ice extent/area/volume during the freezing season. The reader can observe negative relationships between CLDLOW and the surface temperature QOIs (SST and TS) across all four seasons. In the spring and summer seasons, when the sun is above the horizon, clouds will generally reflect solar (shortwave) radiation, which would potentially decrease surface temperature. This interpretation is consistent with our results in all seasons but winter. In the winter season, the general expectation is that cloud coverage would increase surface temperature. This trend is not observed in our data. It is possible that the fact that our data set contains a number of runs without any sea ice coverage is biasing the results. Since, at the present time, there do not exist observational data for the case of no sea ice (especially in winter), it may not be possible to interpret the CLDLOW correlations with the surface temperatures.

Table 4. Table of correlation coefficients between the six QOIs considered (Table 3), averaged during the winter season (January–March) over the last 25 years, for the successful ensemble runs. Large positive correlation coefficients (≥ 0.75) are colored blue. Large negative correlation coefficients (≤ -0.75) are colored yellow.

	SIE	SIV	SST	TS	CLDLOW	FLNS
SIE	1.0	0.77	−0.90	−0.98	0.44	−0.039
SIV		1.0	−0.57	−0.86	−0.0545	0.38
SST			1.0	0.87	−0.67	0.28
TS				1.0	−0.30	−0.096
CLDLOW					1.0	−0.77
FLNS						1.0

Table 5. Table of correlation coefficients between the six QOIs considered (Table 3), averaged during the spring season (April–June) over the last 25 years, for the successful ensemble runs. Large positive correlation coefficients (≥ 0.75) are colored blue. Large negative correlation coefficients (≤ -0.75) are colored yellow.

	SIE	SIV	SST	TS	CLDLOW	FLNS
SIE	1.0	0.79	−0.97	−0.98	0.97	−0.89
SIV		1.0	−0.69	−0.86	0.70	−0.50
SST			1.0	0.95	−0.99	0.94
TS				1.0	−0.95	0.83
CLDLOW					1.0	−0.95
FLNS						1.0

Table 6. Table of correlation coefficients between the six QOIs considered (Table 3), averaged during the summer season (July–September) over the last 25 years, for the successful ensemble runs. Large positive correlation coefficients (≥ 0.75) are colored blue. Large negative correlation coefficients (≤ -0.75) are colored yellow.

	SIE	SIV	SST	TS	CLDLOW	FLNS
SIE	1.0	0.85	−0.90	−0.92	0.89	−0.87
SIV		1.0	−0.66	−0.73	0.66	−0.59
SST			1.0	0.99	−1.0	0.97
TS				1.0	−0.99	0.95
CLDLOW					1.0	−0.98
FLNS						1.0

Table 7. Table of correlation coefficients between the six QOIs considered (Table 3), averaged during the autumn season (October–December) over the last 25 years, for the successful ensemble runs. Large positive correlation coefficients (≥ 0.75) are colored blue. Large negative correlation coefficients (≤ -0.75) are colored yellow.

	SIE	SIV	SST	TS	CLDLOW	FLNS
SIE	1.0	0.84	−0.78	−0.95	0.68	−0.51
SIV		1.0	−0.57	−0.81	0.43	−0.20
SST			1.0	0.93	−0.95	0.83
TS				1.0	−0.83	0.65
CLDLOW					1.0	−0.94
FLNS						1.0

3.4 Main effects, total effects and Sobol indices

Finally, we present and discuss the results of the GSA study using the methodology and workflow described in Sections 2.4 and 2.5, respectively. Our main results are summarized in Figures 8–13 below. For each row of each figure, three plots are reported, which show the main effect, Sobol and total effect indices (from left to right, respectively) corresponding to each of the ten parameters considered (Table 2). As discussed in more detail in Section 2.4.2, the main effect indices measure the effect of individual parameters acting alone and can sum to at most 1. As the sum approaches 1, the contribution of all parameter combinations involving two or more variables decreases. A value of 1 indicates that the function is purely additive and there is no interaction between any parameters. Total effect indices measure the total contribution of each parameter to the variance of a given QOI; specifically, they measure the contributions of all interactions involving a specific parameter. Consequently the total effect index of a single variable will always be at least as large as the main effect index of that variable. Furthermore, the sum of all total effect indices can be greater than 1, because Sobol indices for parameter interactions involving at least two variables can be used to compute the total effects of multiple variables, i.e., the Sobol index of $S_{ij} = \mathbb{V}[\hat{f}_{ij}]$ will contribute to the total effect indices of both the ith and jth variables. Comparing main effect and total effect indices can be used to determine the strength of high-order (involving more than two parameter) interactions. For example, in Figure 10(b), the main effect of `clubb_c1` (z_5) is less than 3% of the total variance, yet the total effect of this variable is over 20% of the total variance. While main and total effect indices summarize the contributions of a single parameter to the variance of a QOI, Sobol indices can be used to identify the contribution of specific parameter interactions to the total variance. Sobol indices involving just one parameter are labeled "(z_k)" and indices involving two parameters are labeled "(z_i, z_j)" with $i \neq j$. Contributions by miscellaneous pairs of parameters in which the percent contribution was $< 1\%$ were omitted from the plots. We found that there were no strong interactions involving three or more variables. The confidence intervals provided in the plots provide a more goal oriented means to determining the confidence in parameter rankings. Overlapping intervals of sensitivity indices suggest that we cannot rank parameters confidently.

Figures 8–13 also report the predictivity coefficient Q^2, which is a measure of the mean square error (MSE) of the Gaussian process model using cross-validation (Marrel et al., 2008). A value of $Q^2 = 1$ is indicative of a perfect cross-validation fit for the given data. Larger values of Q^2 imply greater confidence can be placed in the sensitivity results; however, the value of Q^2 that engenders sufficient confidence is subjective.

3.4.1 Atmospheric parameters

From Figures 8–13, we can credibly conclude that the atmospheric parameters `cldfrc_dp1` (z_4), `clubb_c1` (z_5), `clubb_c8` (z_6), and `gamma_coeff` (z_7) are the most sensitive for all seasons and QOI. The minimum values (bottom whisker) of the total effects of these parameters are all larger than the maximum values (top whisker) of the other parameters. This result is consistent with results obtained in earlier sensitivity studies, namely the fully-coupled study of (Urrego-Blanco et al., 2019). Although there are uncertainty bounds that make it difficult to rigorously pick the most important parameter, based on the median values of the main and total effect indices obtained from Gaussian process emulator approximations, the parameter z_6 (`clubb_c8`) is consistently the most important parameter for all six QOIs and across all four seasons, followed by z_7 (`gamma_coeff`). In fact, for most seasons and QOIs, the minimum total effect values of these two parameters are greater than the maximum values for all other parameters. The main effects trends for parameters `clubb_c1` (z_4) and `clubb_c8` (z_5) are not as clear cut, but seem to follow similar correlation patterns for the QOI as `clubb_c8` (z_6) and `gamma_coeff` (z_7) respectively (i.e., `clubb_c1` has similar trends to `clubb_c8`, and `clubb_c1` has similar trends to `gamma_coeff`).

To streamline and consolidate some of the presentation, we introduce and analyze Figure 14, which plots the seasonal variation of the median total sensitivity (total effects) indices of the four most influential (atmospheric) parameters. In this plot, the box represents 25-75% confidence intervals, the red line denotes the median of the data and the blue dot denotes the mean of the data. Whiskers designate the minimal and maximum values of the total effects indices.

The `cldfrc_dp1` (z_4) parameter. The `cldfrc_dp1` (z_4) CLUBB parameter, which controls cumulus cloud-formation convective regimes in the E3SM (Larson, 2020; Qian et al., 2018), has a significant impact on four of the six QOIs considered here, namely SIE, SST, CLDLOW and FLNS. Figure 14 shows that CLDLOW is most sensitive to this parameter in winter. In contrast SIE and FLNS are most sensitive to `cldfrc_dp1` in spring (Figures 8, 13 and 14). The sensitivities of SIE and SIV have strong cyclic seasonal trends. In addition, non-cyclical seasonal variation is present in SIV and CLDLOW. Seasonal variation in the median values of the sensitivity indices of some other QOI are also present; due to large confidence intervals that overlap, these trends may be considered plausible, but, without higher accuracy, not credible. With this being said, it is interesting to note that the seasonal trend in the median total effect indices of SIV and SIE differ significantly. These differences could reflect the difference between relatively stable multi-year ice (measured by SIV) and young, seasonal ice (measured by SIE).

The `clubb_c1` (z_5) parameter. The `clubb_c1` (z_5) parameter controls the balance of cumulus versus stratocumulus clouds, as discussed in (Larson, 2020). Large positive values of this parameter favor cumulus clouds, while small or negative values are associated with stratocumulus clouds. Stratocumulus clouds are hybrids of the layered stratus and cellular cumuli clouds, and are believed to have a planet-wide surface cooling effect, but earlier investigations have hypothesized that this cloud type in the Arctic has surface warming effects over most of the year (Eastman & Warren, 2010). Figure 14 shows that the SIE, TS and FLNS QOIs exhibit a strong sensitivity to `clubb_c1` (z_5) during the autumn season. These results are consistent with previous observational and modeling studies (Huang et al., 2019; Philipp et al., 2020; Kay & Gettelman, 2009; Eastman & Warren, 2010; Taylor et al., 2015), which have reported a correlation between cloud type, Arctic surface temperature and Arctic sea ice extent during the October–November months. Interestingly, our CLDLOW QOI does not show as strong a sensitivity to `clubb_c1` (z_5) in the autumn as seen for the FLNS QOI. This indicates that while `clubb_c1` (z_5) influences cloud type (cumulus or stratocumulus (Larson, 2020)), it may not strongly influence the fraction of general low cloud cover. That FLNS is responsive to `clubb_c1` (z_5) in autumn is not surprising, given that this season represents the period of maximum interannual variation in SIE, which both reflects and influences the atmosphere/cloud-ocean-sea ice feedback.

The clubb_c8 (z_6) parameter. The clubb_c8 (z_6) parameter was developed to achieve radiative balance in atmospheric models (Larson, 2020; Qian et al., 2018). Specifically, increasing clubb_c8 (z_6) brightens clouds, resulting in Earth surface cooling, as brighter clouds reflect more incoming solar radiation. Figure 14 reports that the clubb_c8 (z_6) has significant influence over all six QOIs considered across all four seasons, with a median main effect of at least 0.4. It is interesting to observe that the CLDLOW and FLNS responses to clubb_c8 (z_6) trend similarly across the four seasons. Even accounting for errors in sensitivity indices, Figure 14 suggests that FLNS has the strongest seasonal response to perturbation of clubb_c8 (z_6) in winter. The SIE QOI shows a strong response to clubb_c8 (z_6) in autumn, with a median total effect of approximately 0.6 and a lower bound of the confidence interval above 0.5. This seems to suggest that cloud brightening has the potential to control the degree to which sea ice is lost towards the end of the melting season (autumn). The impact of clubb_c8 (z_6) perturbation relative to the other atmospheric parameters with the exception of the significantly less influential clubb_c1 (z_5) parameter on the SST QOI is difficult to separate due to overlapping uncertainty bounds for these QOIs (Figure 10). In contrast, clubb_c8 (z_6) is very clearly the most dominant parameter when it comes to its influence over the TS QOI for all seasons (Figure 11).

The gamma_coeff (c_7) parameter. Like clubb_c8 (z_6), gamma_coeff (z_7) parameter is a tunable parameter in the CLUBB shallow convection parameterization scheme that can brighten or dim low clouds, developed to achieve global radiative balance in E3SM (Larson, 2020). Our results show both relatively strong (SIE, SIV, CLDLOW, FLNS) and moderate (TS, SST) seasonal responsiveness to gamma_coeff (z_7) (Figure 14). SIE shows greatest response to gamma_coeff (z_7) in spring, the period of both the onset of melt season and the annual maximum, with mean total effects of 0.50, and minimum/maximum total effects of approximately 0.40/0.60, respectively. In spring, the season during which SIE is most responsive to gamma_coeff (z_7), the Arctic is moving into longer days, as both the annual SIE maximum is reached, and the melt season is beginning. In this context, cloud brightening potentially influences surface energy balance, because brighter clouds reflect more incoming solar radiation. Interestingly, SIV, an estimator of multi-year ice, shows a markedly different response to perturbation of this parameter than SIE, a proxy for seasonal and marginal ice; however, these results should be interpreted with some caution due to the large confidence intervals. While the gamma_coeff (z_7) and clubb_c8 (z_6) parameters both have ostensible control on cloud brightness, their impacts upon SIE differ markedly: the greatest mean total effects for the clubb_c8 (z_6) parameter were observed in autumn (≈ 0.60), compared to spring for the gamma_coeff (z_7) (≈ 0.40). The different responses are explained by the fact that the parameters represent distinct terms in CLUBB (Larson, 2020).

Interactions between atmospheric parameters. It is important to note that while the present study reveals that significant parameter interactions generally involve the four atmospheric parameters, our study demonstrates the effect of these parameters on QOIs from E3SM components other than the atmosphere model. These results would be impossible to obtain without a global fully-coupled ESM. Despite non-trivial errors in the sensitivity indices, we can also conclude that certain parameter interactions involving the four most sensitive parameters contribute more to the variability of all QOI than any of the six insensitive parameters. For example, the Sobol index labeled (z_5, z_6) in Figure 9, which quantifies the strength of the interactions between clubb_c1 and clubb_c8 for the QOI SIV in spring, is much stronger than the total effects of the six insensitive parameters. Indeed in this case the interaction contributes more than cldfrc_dp1 (z_4) acting alone. Additionally, for the CLDLOW and FLNS QOIs (Figures 12 and 13, respectively), a number of parameter interactions involving the various atmospheric parameters are at least comparable to the effect of clubb_c1 (z_5) acting alone.

3.4.2 Sea ice and ocean parameters

While we see little impact from the sea ice and ocean parameters relative to the atmospheric parameters, there are a few cases for which the total effects for these parameters are non-zero. Of the sea ice parameters, `ksno` (z_1) had the largest total effect for several QOIs in several seasons. Non-zero total effect indices associated with `ksno` (z_1) for the SST and FLNS QOIs are shown in Figures 10 and 13, respectively. This result is consistent with the observation that the snow conductivity can affect ocean temperature since it impacts the amount of heat flux (solar radiation) that reaches the ocean in ice-covered waters. During the late spring, summer and early autumn seasons, this solar radiation input would primarily come from short-wave solar radiation. Of the sea ice parameters, the reader can observe that the `salinity_restoring_constant_piston_velocity` (z_{10}) parameter had a total effect of approximately 0.05-0.13 in summer and 0.03-0.08 in winter for the SIV QOI. It is well-known that salinity of the upper ocean has an impact on the column thermodynamics of the sea ice, so it is not surprising that the salinity would influence the SIV. In general, in the melting season (summer), the upper ocean is less saline, as the older sea ice releases freshwater during melt; conversely, in the freezing season (winter), the upper ocean is more saline due to brine rejection during freezing. Our results suggest that salinity changes during these two changes have some influence on the sea ice volume due to ocean-sea ice feedbacks such as these. The sea ice and ocean parameters do not show up in the parameter pairs appearing in our Sobol indices results.

3.5 Marginalized main effect indices

In this section we present the univariate marginalized main effect functions (equation (6)) described in Section 2.4.2. These main effect functions enable us to determine *a priori* whether increasing/decreasing a given parameter will increase or decrease a given QOI. These results are particularly useful for model spin-up/tuning, which can be an *ad hoc* trial-and-error process. For the sake of brevity, we provide the marginalized main effects results for only two of our QOIs averaged annually[6], SIE and TS (Figures 15 and 16, respectively), as these are the QOIs most relevant for model spin-ups.

The results presented below demonstrate that, as expected, the four atmospheric parameters considered herein have the greatest influence when it comes to model spin-up/tuning. The reader can observe by examining Figures 15 and 16 that there are clear-cut parameter-QOI correlations for the `clubb_c8` (z_6) and `gamma_coeff` (z_7) parameters. The parameter `clubb_c8` (z_6) has a strong positive correlation with SIE and a strong negative correlation with TS, whereas the parameter `gamma_coeff` (z_7) has a strong negative correlation with SIE and a strong positive correlation with TS. The fact that SIE and TS have opposite trends is consistent with the QOI correlations uncovered earlier (Section 3.3). It is interesting that the marginalized main effects plots for the remaining two atmospheric parameters, `cldfrc_dp` (z_4) and `clubb_c1` (z_5), have inflection points and some level of convexity/concavity, meaning that determining whether increasing/decreasing these parameters will increase/decrease a QOI depends on the parameter value. In our manual spin-up of the ULR E3SMv1, we found by trial-and-error that `cldfrc_dp1` (z_4) had a significant effect on tuning the model, in particular, increasing `cldfrc_dp1` within the range $[0.075, 0.5]$ decreased TS and increased SIE (Peterson et al., 2020). This provides some corroboration of the results in Figures 15 and 16.

Reconciling the results discussed above with the relevant physical processes requires discussion of the physical effects our four atmospheric parameters. Without loss of generality, we will focus on the surface air temperature, or TS, QOI. From Table 2, `clubb_c1` (z_5) and `clubb_c8` (z_6) have an effect on the skewness of the Probability Density Function (PDF) of the vertical velocity w' within the CLUBB parameterization (Qian et al., 2018;

[6] Identical conclusions were obtained from the analogous seasonal plots.

Larson, 2020; Guo et al., 2014). High skewness in the vertical velocity causes the production of cumulus-like layers of clouds with a low cloud fraction, whereas low skewness results in stratocumulus clouds having a high cloud fraction. Increasing `clubb_c8` (z_6) and `clubb_c1` (z_5) is known to lead to cloud brightening and cooling at the Earth surface (Larson, 2020). This result is consistent with our analysis. Additionally, with low values of `clubb_c1` (z_5), which favor insolation-reducing stratiform clouds, SIE is relatively high and TS is low, a result consistent with observational studies on the general surface-cooling effects of this cloud type. Like stratocumulus clouds, cumuli can reflect incident solar radiation, or trap heat, depending on the cloud height and optical density. Since SIE is relatively low and TS is relatively high for larger values of `clubb_c1` (z_5), our results point to the heat-trapping effects of the cumulus species. The parameter `gamma_coeff` (z_7), which controls the width of the individual Gaussians within the CLUBB parameterization (Larson, 2020), has broad effects within CLUBB, influencing not only shallow convection but also stratocumulus cloud formulation. As discussed in (Qian et al., 2018), increasing `gamma_coeff` (z_7) has a similar effect to increasing skewness, which leads to a smaller cloud fraction. Thus, the parameter `gamma_coeff` (z_7) is expected to have a similar effect on the surface air temperature as `clubb_c1` (z_5), which is in general consistent with our results. Finally, we turn our attention to the last atmospheric parameter, `cldfrc_dp1` (z_4), CLUBB's deep convection cloud parameter. Increasing this parameter results in the movement (convection) of hotter and therefore less dense material upward, causing colder and denser material to sink under the gravity, cooling the Earth's surface. Yet again, the negative `cldfrc_dp1` (z_4)-TS correlation uncovered by our results is consistent with this physical mechanism.

While the subplots in Figures 15 and 16 corresponding to the ocean and sea ice parameters are flat compared to the subplots corresponding to the atmospheric parameters, the reader can observe a slight curvature in the plots for sea ice parameters `ksno` (z_1) and `dragio` (z_3). It is interesting to remark that the trends present in these parameter-QOI correlations are similar to the trends uncovered using an alternate marginalization technique for the stand-alone sea ice model GSA of (Urrego-Blanco et al., 2016) (see Figure 11 in this reference).

4 Summary

We have performed a GSA involving ten parameters and six QOIs spanning three climate components (atmosphere, ocean, sea ice) using a fully-coupled ULR configuration of E3SMv1. To the best of our knowledge, this is the first GSA involving the fully-coupled E3SMv1, and the first scientific study involving the ULR configuration. In order to perform the sensitivity analysis, we created a fast Gaussian process emulator from 139 75-year runs of the ULR E3SMv1, which included pre-industrial control forcing and were initialized from a spun-up initial condition developed for the purpose of this study. The runs exhibited a great deal of variability, spanning the gamut from complete loss of Arctic sea ice to apparent exponential growth in Arctic sea ice. Our Gaussian process emulator was used to determine Sobol indices, main effect indices and total effect indices for each QOI-parameter combination, and provided uncertainty bounds for each set of indices. While the sometimes large uncertainty bounds made it difficult to rigorously pick out the most influential parameter for each QOI, the study enabled a definitive ranking of the most dominant parameters affecting each QOI annually and seasonally. We found the atmospheric parameters related to cloud physics within the CLUBB model in EAM (and their interactions) had by far the greatest impact on the Arctic climate state. While our study demonstrated that the most significant parameter-parameter interactions involved the atmospheric parameters with each other, it enabled us to investigate the effect of these parameters on QOIs from E3SM components different than the atmosphere model. The fact that this investigation would not be possible with a stand-alone atmosphere model reinforces the need for coupled analyses when studying climate model uncertainties/sensitivities. We performed a careful study of QOI-QOI correlations and parameter-parameter interactions using our sensitivity indices, and were able to reconcile these relationships with several well-known Arctic feedback processes. By approximating univariate main effect functions (Oakley & O'Hagan, 2004), we were able to determine the sensitivity of individual QOIs on individual parameters, thereby inferring QOI-parameter correlations, useful for model spin-up/tuning. We performed a careful study of the marginalized main effect functions for the most influential (atmospheric) parameters, and demonstrated that the trends uncovered by the study are consistent with both our manual spin-up of the ULR E3SMv1 as well as the physical processes underlying the CLUBB parameterization (e.g., the formation of cumulus vs. stratocumulus clouds, the relative amount of shortwave cloud forcing/cloud brightening).

The GSA described herein motivates several future research endeavors. While the ULR model's ability to identify and explain a variety of physical processes and feedbacks present within the global climate system suggests that the model has the potential for serving as a low-cost surrogate for higher resolution configurations of the E3SM, a rigorous quantitative study in which the ULR model is evaluated and compared to more standard (higher resolution) configurations is needed to confirm the model's viability in scientific studies, and is a logical next step motivated by the present work. For completeness, this evaluation/validation endeavor could include not only the ULR model but also data-driven or machine-learned surrogates trained using high-resolution E3SM data, towards determining which class of surrogates provides the best trade-off when it comes to computational cost, accuracy and robustness. As discussed above, the present study used a simple pre-industrial control forcing; a worthwhile follow-on study is one in which the analysis described in this work would be repeated, but under alternate (more realistic, e.g. those with a prescribed CO_2 increase) forcings such as those in (Golaz et al., 2019). The marginalized main effect functions we produced could be used to generate improved initial conditions (recall that the initial condition used in this study had a warm bias) for such follow-on targeted studies using the ULR E3SMv1. We additionally envision augmenting the present study with higher-fidelity ensemble data (e.g., using a medium-low resolution, or MLR, of the E3SMv1 having a resolution of approximately 2.7° for the atmosphere component (Peterson et al., 2020)), towards a multi-fidelity global sensitivity analysis study. The GSA results described herein have the potential of informing targeted studies and spin-ups at higher resolutions, such as the MLR E3SM. Determining to what extent the marginalized main effects results presented herein can be used to tune higher-resolution models would be a valuable future exercise.

Acknowledgments

This work was funded by the Laboratory Directed Research & Development (LDRD) program at Sandia National Laboratories. The writing of this manuscript was funded by the first author's Presidential Early Career Award for Scientists and Engineers (PECASE). Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the U.S. Government.

Per the Enabling FAIR Data Project guidelines, we have made the data used in the GSA performed herein publicly available. These data can be downloaded from the following `github` repository: `https://github.com/karapeterson/E3SM\_ULR\_GSA\_Data`.

References

Adams, B., Bauman, L., Bohnhoff, W., Dalbey, K., Ebeida, M., Eddy, J., ... Vigil, D. (2013). *DAKOTA, A Multilevel Parallel Object-Oriented Framework for Design Opti-*

mization, Parameter Estimation, Uncertainty Quantification, and Sensitivity Analysis: Version 5.4 User's Manual (Tech. Rep.). Livermore, CA and Albuquerque, NM.

Alexanderian, A., Winokur, J., Sraj, I., Srinivasan, A., Iskandarani, M., Thacker, W., & Knio, O. (2012). Global sensitivity analysis in an ocean general circulation model: a sparse spectral projection approach. *Computational Geosciences*, *16*, 757–778.

Asay-Davis, X., Comeau, D., Fyke, J., Hoffman, M., Petersen, M., Pricec, S., … Jr., P. W. (2018). *Antarctic ice shelf-ocean interactions in high-resolution, global simulations using the Energy Exascale Earth System Model (E3SM) Part 2: Sensitivity studies and model tuning.* 2018 Ocean Sciences Meeting, American Geophysical Union, Portland, Oregon.

Bathiany, S., Dijkstra, H., Crucifix, M., Dakos, V., Brovkin, V., Williamson, M. S., … Scheffer, M. (2016, 11). Beyond bifurcation: using complex models to understand and predict abrupt climate change. *Dynamics and Statistics of the Climate System*, *1*(1). Retrieved from `https://doi.org/10.1093/climsys/dzw004` doi: 10.1093/climsys/dzw004

Bernard, B., Madec, G., Penduff, T., Molines, J.-M., Treguier, A.-M., Sommer, J. L., … Cuevas, B. D. (2006). Impact of partial steps and momentum advection schemes in a global ocean circulation model at eddy-permitting resolution. *Ocean Dynamics*, *56*, 543–567.

Bisht, G., Riley, W., Hammond, G., & Lorenzetti, D. (2018). Development and evaluation of a variably saturated flow model in the Global E3SM Land Model (ELM) version 1.0. *Geoscientific Model Development Discussions*, *11*(10), 1–40. doi: 10.5194/gmd-2018-44

Cohen, J., Pfeiffer, K., & Francis, J. A. (2018). Warm Arctic episodes linked with increased frequency of extreme winter weather in the United States. *Nature Communications*, *9*(869).

Cohen, J., Zhang, X., Francis, J., Jung, T., Kwok, R., Overland, J., & Smith, D. (2018). *Arctic change and possible influence on mid-latitude climate and weather: A US CLIVAR white paper* (Tech. Rep. No. 2018-1).

Cornette, W. M. (2012). Mosart: Modeling the radiative environment of earth's atmosphere, terrain, oceans, and space. *Journal of the Washington Academy of Sciences*, *98*(4), 27–46. Retrieved from `http://www.jstor.org/stable/24536506`

Covey, C., Lucas, D. D., Tannahill, J., Garaizar, X., & Klein, R. (2013). Efficient screening of climate model sensitivity to a large number of perturbed input parameters. *Journal of Advances in Modeling Earth Systems*, *5*(3), 598-610. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/jame.20040` doi: 10.1002/jame.20040

Cvijanovic, I., Santer, B. D., Bonfils, C., Lucas, D. D., Chiang, J. C. H., & Zimmerman, S. (2017). Future loss of Arctic sea-ice cover could drive a substantial decrease in california's rainfall. *Nature Communications*, *8*(1947).

Dennis, J. M., Edwards, J., Evans, K. J., Guba, O., Lauritzen, P. H., Mirin, A. A., … Worley, P. H. (2012). CAM-SE: A scalable spectral element dynamical core for the Community Atmosphere Model. *The International Journal of High Performance Computing Applications*, *26*(1), 74-89. Retrieved from `https://doi.org/10.1177/1094342011428142` doi: 10.1177/1094342011428142

Eastman, R., & Warren, S. G. (2010). Interannual variations of arctic cloud types in relation to sea ice. *Journal of Climate*, *23*(15), 4216–4232. Retrieved from `https://journals.ametsoc.org/view/journals/clim/23/15/2010jcli3492.1.xml` doi: https://doi.org/10.1175/2010JCLI3492.1

Golaz, J.-C., Caldwell, P. M., Van Roekel, L. P., Petersen, M. R., Tang, Q., Wolfe, J. D., … Zhu, Q. (2019). The DOE E3SM coupled model version 1: Overview and evaluation at standard resolution. *Journal of Advances in Modeling Earth Systems*, *0*(ja). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001603` doi: 10.1029/2018MS001603

Goosse, H., Kay, J. E., Armour, K. C., Bodas-Salcedo, A., Chepfer, H., Docquier, D., …

Vancoppenolle, M. (2018). Quantifying climate feedbacks in polar regions. *Nature Communications*, *9*(1919).

Graeter, K. A., Osterberg, E. C., Ferris, D. G., Hawley, R. L., Marhsall, H. P., Lewis, G., ... Birkel, S. D. (2018). Ice core records of West Greenland melt and climate forcing. *GRL*, *45*, 3164–3172.

Guo, Z., Wang, M., Qian, Y., Larson, V. E., Ghan, S., Ovchinnikov, M., ... Zhou, T. (2014). A sensitivity analysis of cloud properties to CLUBB parameters in the single-column Community Atmosphere Model (SCAM5). *Journal of Advances in Modeling Earth Systems*, *6*(3), 829-858. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2014MS000315` doi: https://doi.org/10.1002/2014MS000315

Harbrecht, H., Jakeman, J., & Zaspel, P. (2020). Weighted greedy-optimal design of computer experiments for kernel-based and gaussian process model emulation and calibration. *Communications in Computational Physics*. (In press)

Hecht, M., & Smith, R. (2008). *Towards a physical understanding of the North Atlantic: A review of mode studies.* In M. Hecht and H. Hasumi (Eds.), Ocean Modeling in an Eddying Regime Geophysical Monograph Series (Section 3, pp. 339–352). Washington, DC: American Geophysical Union.

Hecht, M. W., Hunke, E., Maltrud, M. E., Petersen, M. R., & Wingate, B. A. (2008). *Lateral mixing in the eddying regime and a new broad-ranging formulation.* In M. Hecht and H. Hasumi (Eds.), Ocean Modeling in an Eddying Regime Geophysical Monograph Series, (Section 2, pp. 213–240). Washington, DC: American Geophysical Union.

Heinzeller, D., Duda, M. G., & Kunstmann, H. (2016). Towards convection-resolving, global atmospheric simulations with the model for prediction across scales (mpas) v3.1: an extreme scaling experiment. *Geoscientific Model Development*, *9*(1), 77–110. Retrieved from `https://gmd.copernicus.org/articles/9/77/2016/` doi: 10.5194/gmd-9-77-2016

Huang, Y., Dong, X., Bailey, D. A., Holland, M. M., Xi, B., DuVivier, A. K., ... Deng, Y. (2019). Thicker clouds and accelerated arctic sea ice decline: The atmosphere-sea ice interactions in spring. *Geophysical Research Letters*, *46*(12), 6980-6989. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019GL082791` doi: https://doi.org/10.1029/2019GL082791

Hunke, E., Lipscomb, W., Turner, A., Jeffery, N., & Elliot, S. (2015). *CICE: The Los Alamos sea ice model. Documentation and software user's manual version 5.1* (Tech. Rep.). Los Alamos, New Mexico.

Hurlburt, H. E., & Hogan, P. J. (2000). Impact of 1/8 degree to 1/64 degree resolution on gulf stream model-data comparisons in basin-scale subtropical atlantic ocean models. *Dynamics of Atmospheres and Oceans*, *32*(3), 283 - 329. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0377026500000506` doi: https://doi.org/10.1016/S0377-0265(00)00050-6

IPCC. (2021). *Climate change 2021: the physical science basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (Tech. Rep.).

Jakeman, J., Eldred, M., Geraci, G., & Gorodetsky, A. (2019). Adaptive multi-index collocation for uncertainty quantification and sensitivity analysis. *International Journal for Numerical Methods in Engineering*. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/nme.6268` doi: 10.1002/nme.6268

Jakeman, J. D. (2021). *Pyapprox: Approximation and probabilistic analysis of data.* https://sandialabs.github.io/pyapprox/index.html.

Kay, J. E., & Gettelman, A. (2009). Cloud influence on and response to seasonal arctic sea ice loss. *Journal of Geophysical Research: Atmospheres*, *114*(D18). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2009JD011773` doi: https://doi.org/10.1029/2009JD011773

Kim, J. G., Hunke, E. C., & Lipscomb, W. H. (2006). Sensitivity analysis and parameter tuning scheme for global sea-ice modeling. *Ocean Modelling*, *14*(1), 61 - 80. Retrieved from http://www.sciencedirect.com/science/article/pii/S146350030600028X doi: https://doi.org/10.1016/j.ocemod.2006.03.003

Larson, V. (2020). *CLUBB-SILHS: A parameterization of subgrid variability in the atmosphere* (Tech. Rep.). Retrieved from https://arxiv.org/pdf/1711.03675.pdf

Lenton, T. M. (2008). Tipping elements in the Earth climate system. *PNAS*, *105*(6), 1786–1793.

Lenton, T. M. (2012). Arctic climate tipping points. *Ambio*, *41*(1), 10–22.

Maltrud, M. E., & McClean, J. L. (2005). An eddy resolving global 1/10 degree ocean simulation. *Ocean Modelling*, *8*(1), 31 - 54. Retrieved from http://www.sciencedirect.com/science/article/pii/S1463500303000684 doi: https://doi.org/10.1016/j.ocemod.2003.12.001

Marrel, A., Iooss, B., Van Dorpe, F., & Volkova, E. (2008). An efficient methodology for modeling complex computer codes with Gaussian processes. *Computational Statistics and Data Analysis*, *52*(10), 4731-4744. Retrieved from https://www.sciencedirect.com/science/article/pii/S0167947308001758 doi: https://doi.org/10.1016/j.csda.2008.03.026

Morris, M. D. (1991). Factorial sampling plans for preliminary computational experiments. *Technometrics*, *33*(2), 161-174. Retrieved from https://www.tandfonline.com/doi/abs/10.1080/00401706.1991.10484804 doi: 10.1080/00401706.1991.10484804

Oakley, J. E., & O'Hagan, A. (2004). Probabilistic sensitivity analysis of complex models: a bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *66*(3), 751-769. Retrieved from https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2004.05304.x doi: https://doi.org/10.1111/j.1467-9868.2004.05304.x

Parazoo, N. C., Koven, C. D., Lawrence, D. M., Romanovsky, V., & Miller, C. E. (2018). Detecting the permafrost carbon feedback: talik formation and increased cold-season respiration as precursors to sink-to-source transitions. *The Cryosphere*, *12*(1), 123–144. doi: 10.5194/tc-12-123-2018

Parry, M., Arnell, N., McMichael, T., Nicholls, R., Martens, P., Kovats, S., … Fischer, G. (2001). Millions at risk: defining critical climate change threats and targets. *Global Environmental Change*, *11*(3), 181-183. Retrieved from https://www.sciencedirect.com/science/article/pii/S0959378001000115 doi: https://doi.org/10.1016/S0959-3780(01)00011-5

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Petersen, M., Asay-Davis, X., Jacobsen, D., Maltrud, M., Ringler, T., Roekel, L. V., … Wolfram, P. (2018). *MPAS-Ocean Model User's Guide Version 6.0* (Tech. Rep.). Los Alamos, New Mexico.

Petersen, M. R., Asay-Davis, X. S., Berres, A. S., Chen, Q., Feige, N., Hoffman, M. J., … Woodring, J. L. (2019). An Evaluation of the Ocean and Sea Ice Climate of E3SM Using MPAS and Interannual CORE-II Forcing. *Journal of Advances in Modeling Earth Systems*, *11*(5), 1438-1458. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018MS001373 doi: https://doi.org/10.1029/2018MS001373

Peterson, K., A, P., Tezaur, I., Roesler, E., Nichol, J., Peterson, M., … Bull, D. (2020). *Arctic tipping points triggering global change LDRD final report* (Tech. Rep.). Sandia National Laboratories Report, SAND2020–9932, Livermore, CA and Albuquerque, NM.

Peterson, K., Bochev, P., & Paskaleva, B. (2010). *Development, sensitivity analysis, and uncertainty quantification of high-fidelity Arctic sea ice models.* Sandia National Laboratories Report, SAND2010–6218, Albuquerque, NM and Livermore, CA.

Philipp, D., Stengel, M., & Ahrens, B. (2020). Analyzing the arctic feedback mechanism between sea ice and low-level clouds using 34 years of satellite observations. *Journal of Climate*, *33*(17), 7479-7501. Retrieved from https://journals.ametsoc.org/view/

journals/clim/33/17/jcliD190895.xml doi: https://doi.org/10.1175/JCLI-D-19-0895.1

Qian, Y., Wan, H., Yang, B., Golaz, J.-C., Harrop, B., Hou, Z., ... Zhang, K. (2018). Parametric sensitivity and uncertainty quantification in the version 1 of E3SM atmosphere model based on short perturbed parameter ensemble simulations. *Journal of Geophysical Research Atmospheres*, *123*, 13046–13073.

Rae, J. G. I., Hewitt, H. T., Keen, A. B., Ridley, J. K., Edwards, J. M., & Harris, C. (2014). A sensitivity study of the sea ice simulation in the global coupled climate model, HadGEM3. *Ocean Modeling*, *74*, 60-76.

Rasch, P. J., Xie, S., Ma, P. L., Lin, W., Wang, H., Tang, Q., ... Yang, Y. (2019). An overview of the atmospheric component of the Energy Exascale Earth System Model. *Journal of Advances in Modeling Earth Systems*, *11*, 2377-2411.

Rasmussen, C. E., & Williams, C. (2006). *Gaussian processes for machine learning.* MIT Press.

Reckinger, S. M., Petersen, M. R., & Reckinger, S. J. (2015). A study of overflow simulations using MPAS-Ocean: vertical grids, resolution, and viscosity. *Ocean Modeling*, *96*, 291-313.

Richter-Menge, J., Druckenmiller, M. L., & Jeffries, M. (Eds.). (2019). *Arctic report card 2019* (Tech. Rep.). Retrieved from https://www.arctic.noaa.gov/Report-Card

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. Retrieved from https://www.sciencedirect.com/science/article/pii/S0010465509003087 doi: https://doi.org/10.1016/j.cpc.2009.09.018

Schuur, E., McGuire, A., Schadel, C., & et al. (2015). Climate change and the permafrost carbon feedback. *Nature*, *520*, 171–179. doi: 10.1038/nature14338

Schweiger, A. J., Lindsay, R. W., Vavrus, S., & Francis, J. A. (2008). Relationships between arctic sea ice and clouds during autumn. *Journal of Climate*, *21*(18), 4799 - 4810. Retrieved from https://journals.ametsoc.org/view/journals/clim/21/18/2008jcli2156.1.xml doi: 10.1175/2008JCLI2156.1

Sevellec, F., Fedorov, A. V., & Liu, W. (2017). Arctic sea-ice decline weakens the atlantic meridional overturning circulation. *Nature Climate Change*, *7*, 604–610.

Smith, L. C., & Stephenson, S. R. (2013). New trans-Arctic shipping routes navigable by midcentury. *PNAS*, *110*(13), 4871–4872.

Smith, R., Jones, P., Briegleb, B., Bryan, F., Danabasoglu, G., Dennis, J., & Others. (2010). *, The Parallel Ocean Program (POP) Reference Manual. Ocean component of the Community Climate System Model (CCSM) and Community Earth System Model (CESM)* (Tech. Rep.). Los Alamos, New Mexico.

Snow, water, ice, and permafrost in the Arctic (SWIPA) (Tech. Rep.). (2017). Oslo.

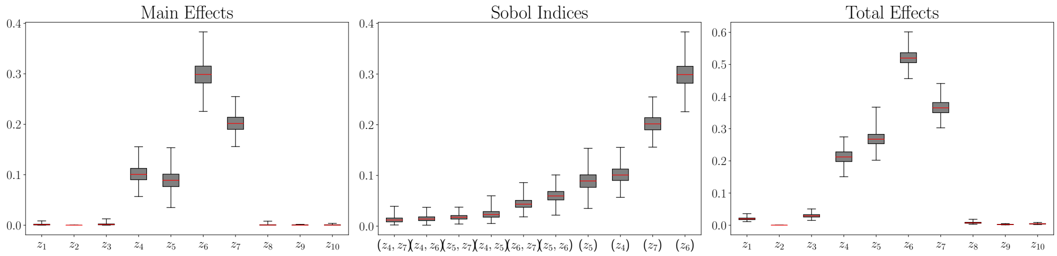
Sobol, I. (2001). Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, *55*(1), 271 - 280. Retrieved from http://www.sciencedirect.com/science/article/pii/S0378475400002706 (The Second IMACS Seminar on Monte Carlo Methods) doi: https://doi.org/10.1016/S0378-4754(00)00270-6

Sudret, B. (2008, JUL). Global sensitivity analysis using polynomial chaos expansions. *Reliability Engineering & System Safety*, *93*(7), 964–979. doi: {10.1016/i.ress.2007.04.002}

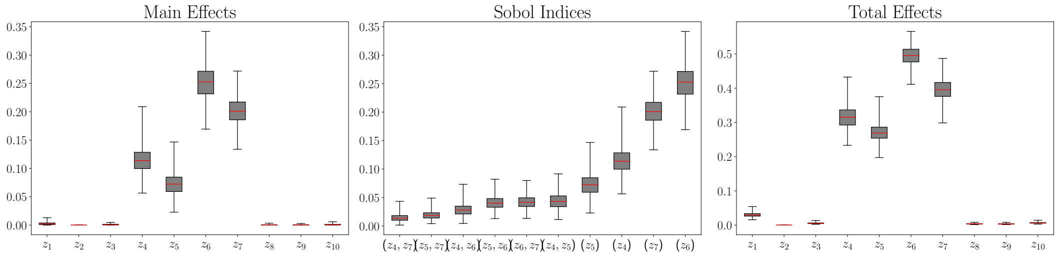
Taylor, P. C., Kato, S., Xu, K.-M., & Cai, M. (2015). Covariance between arctic sea ice and clouds within atmospheric state regimes at the satellite footprint level. *Journal of Geophysical Research: Atmospheres*, *120*(24), 12656-12678. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2015JD023520 doi: https://doi.org/10.1002/2015JD023520

Uotila, P., O'Farrell, S., Marsland, S., & Bi, D. (2012). A sea-ice sensitivity study with a global ocean-ice model. *Ocean Modelling*, *51*, 1 - 18. Retrieved from http://www.sciencedirect.com/science/article/pii/S1463500312000625 doi: https://

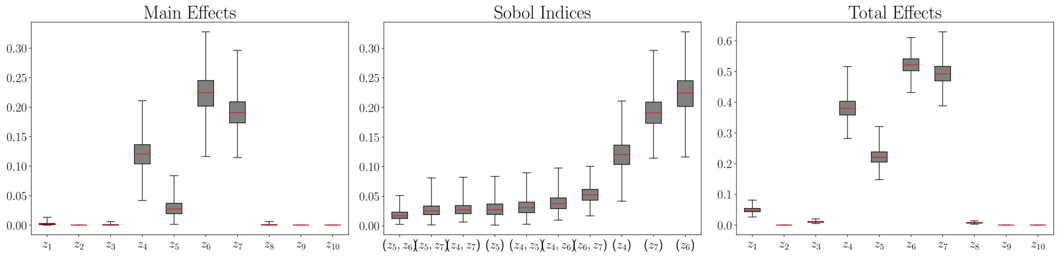
doi.org/10.1016/j.ocemod.2012.04.002
Urrego-Blanco, J. R., Hunke, E. C., & Urban, N. (2019). Emergent relationships among sea ice, longwave radiation, and the Beaufort high circulation exposed through parameter uncertainty analysis. *JGR Oceans*, *124*, 9572-9589. doi: 10.1029/2019JCO14979
Urrego-Blanco, J. R., Urban, N. M., Hunke, E. C., Turner, A. K., & Jeffery, N. (2016). Uncertainty quantification and global sensitivity analysis of the Los Alamos sea ice model. *JGR Oceans*, *121*, 2709-2732. doi: 10.1029/2015JCO11558
Zhao, C., Liu, X., Qian, Y., Yoon, J., Hou, Z., Lin, G., ... Bao, J. (2013). A sensitivity study of radiative fluxes at the top of atmosphere to cloud-microphysics and aerosol parameters in the community atmosphere model cam5. *Atmospheric Chemistry and Physics*, *13*(21), 10969–10987. Retrieved from `https://acp.copernicus.org/articles/13/10969/2013/` doi: 10.5194/acp-13-10969-2013



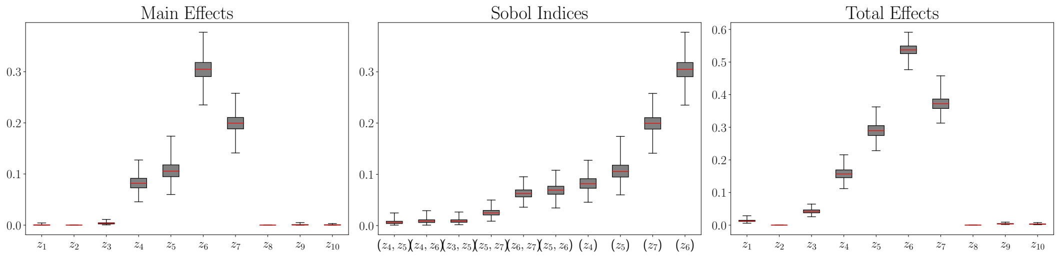
(a) Annual ($Q^2 = 0.866$)



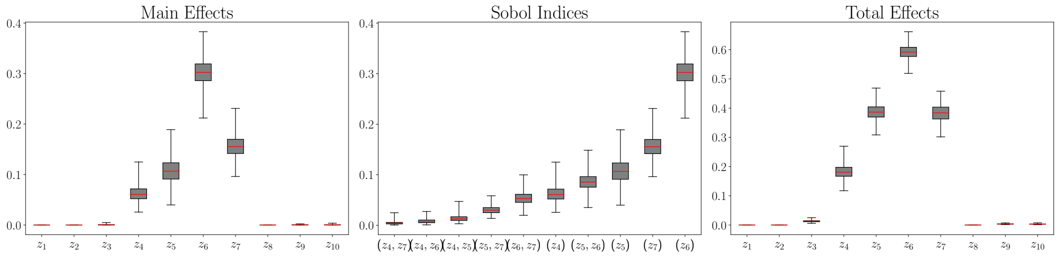
(b) Winter ($Q^2 = 0.819$)



(c) Spring ($Q^2 = 0.737$)

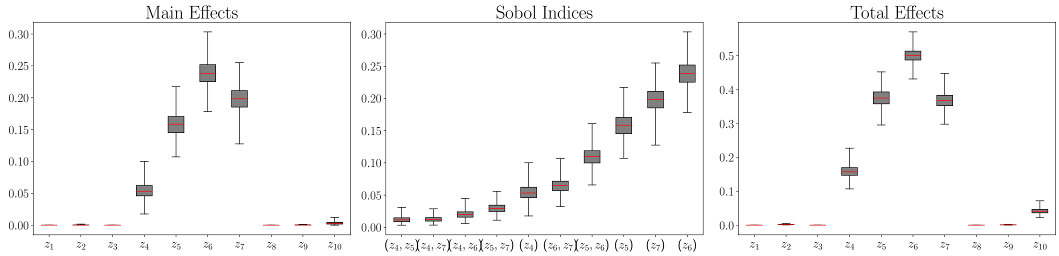


(d) Summer ($Q^2 = 0.769$)

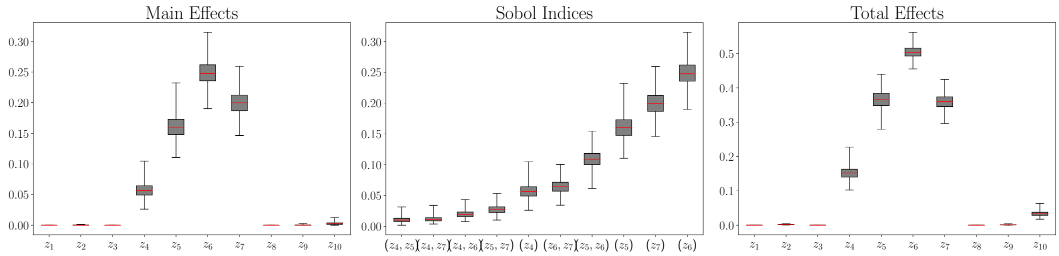


(c) Autumn ($Q^2 = 0.688$)

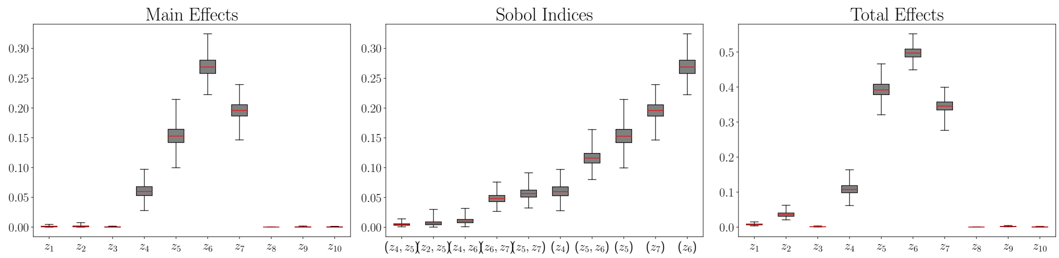
Figure 8. GSA results: main effects, Sobol and total effects indices (from left to right) for the Sea Ice Extent (SIE) QOI calculated annually and by season. The box-and-whiskers plots depict GSA results obtained using a Gaussian process emulator, which provides uncertainty bounds: the red central mark indicates the median of the data, the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The parameters $\{z_i\}$ are described in Table 2.



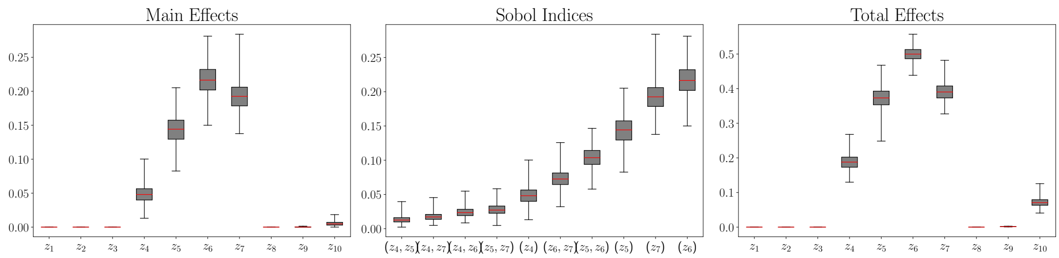
(a) Annual ($Q^2 = 0.518$)



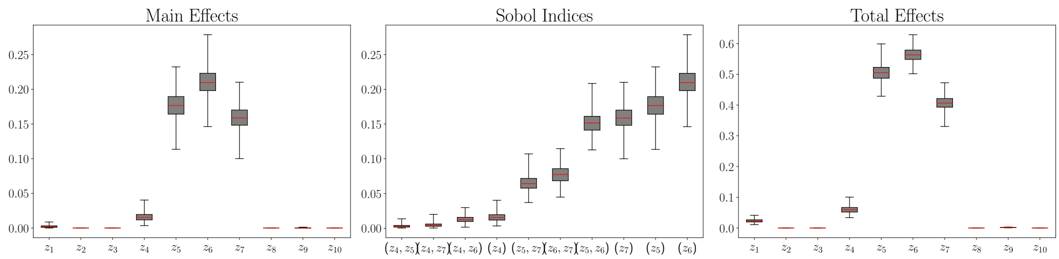
(b) Winter ($Q^2 = 0.562$)



(c) Spring ($Q^2 = 0.713$)

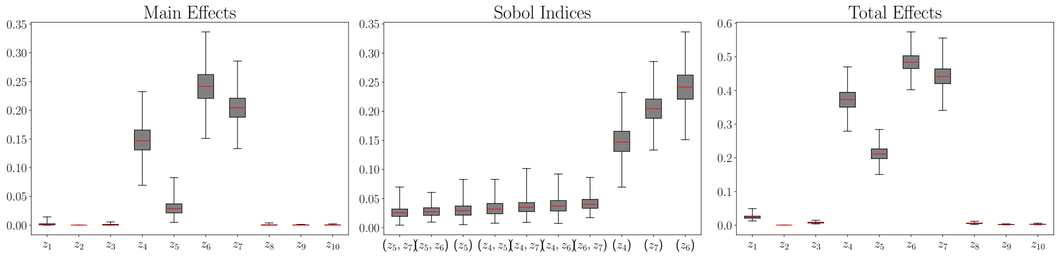


(d) Summer ($Q^2 = 0.425$)

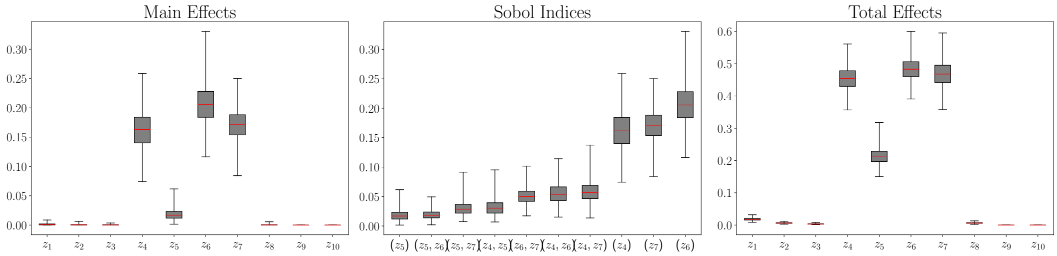


(c) Autumn ($Q^2 = 0.481$)

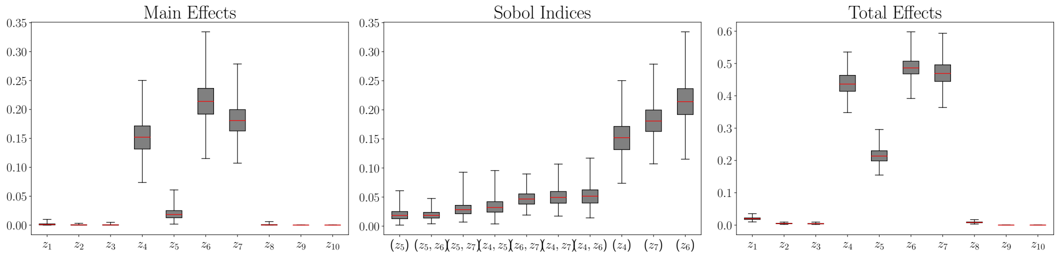
Figure 9. GSA results: main effects, Sobol and total effects indices (from left to right) for the Sea Ice Volume (SIV) QOI calculated annually and by season. The box-and-whiskers plots depict GSA results obtained using a Gaussian process emulator, which provides uncertainty bounds: the red central mark indicates the median of the data, the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The parameters $\{z_i\}$ are described in Table 2.



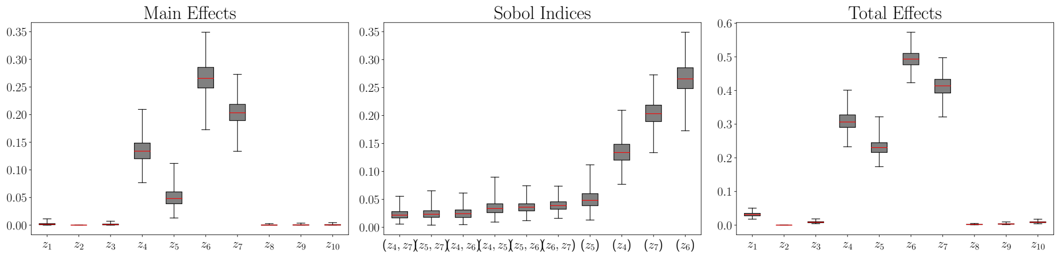
(a) Annual ($Q^2 = 0.822$)



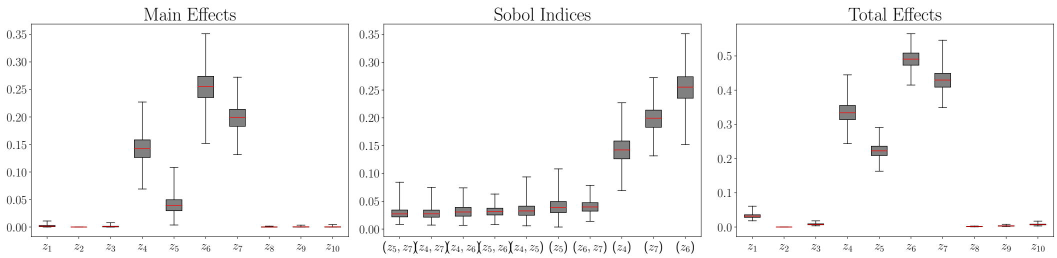
(b) Winter ($Q^2 = 0.778$)



(c) Spring ($Q^2 = 0.773$)

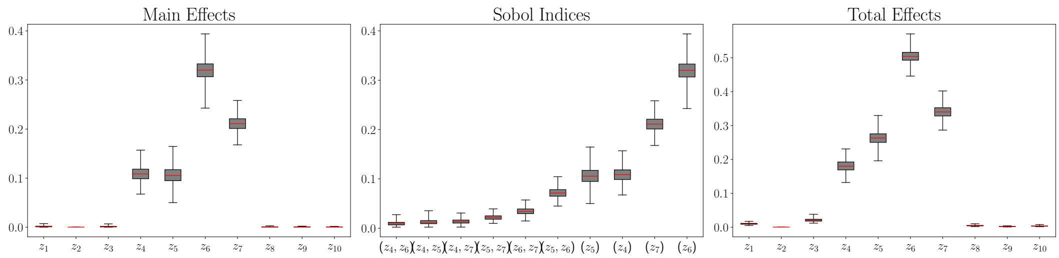


(d) Summer ($Q^2 = 0.848$)

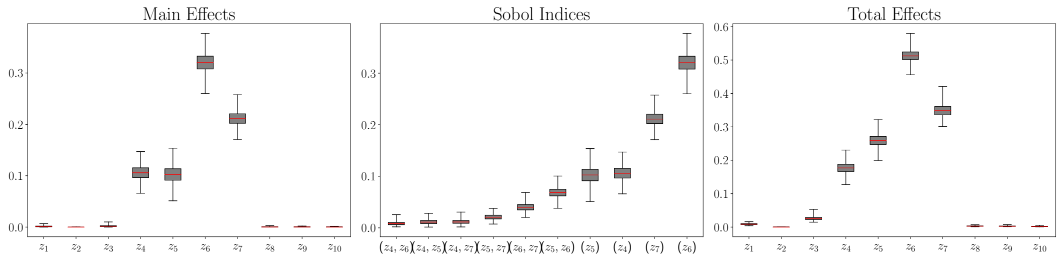


(c) Autumn ($Q^2 = 0.837$)

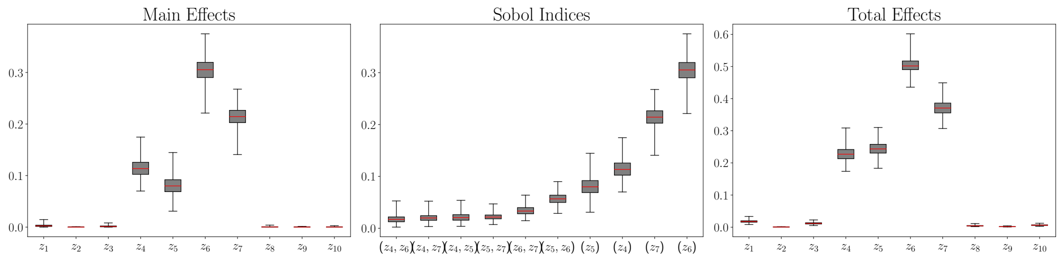
Figure 10. GSA results: main effects, Sobol and total effects indices (from left to right) for the Sea Surface Temperature Averaged Over 60-90° (SST) QOI calculated annually and by season. The box-and-whiskers plots depict GSA results obtained using a Gaussian process emulator, which provides uncertainty bounds: the red central mark indicates the median of the data, the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The parameters $\{z_i\}$ are described in Table 2.



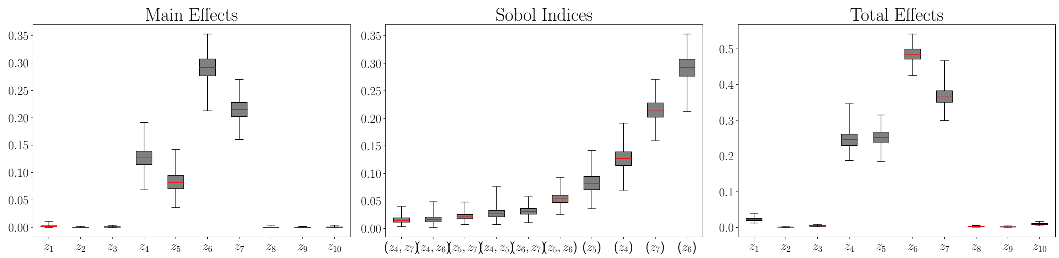
(a) Annual ($Q^2 = 0.909$)



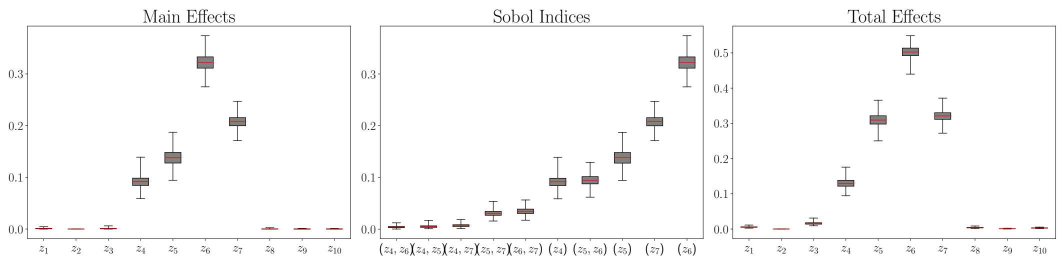
(b) Winter ($Q^2 = 0.895$)



(c) Spring ($Q^2 = 0.884$)

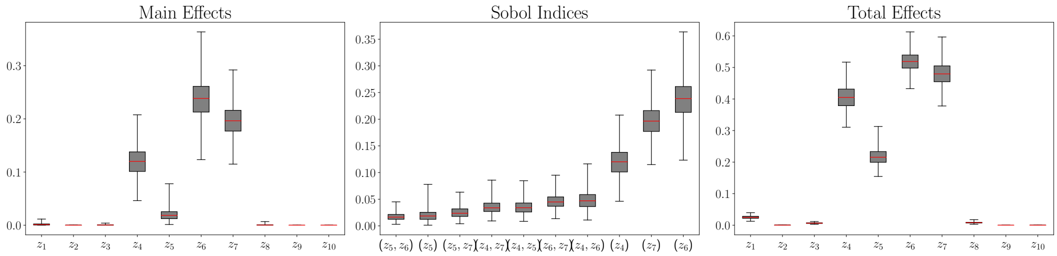


(d) Summer ($Q^2 = 0.877$)

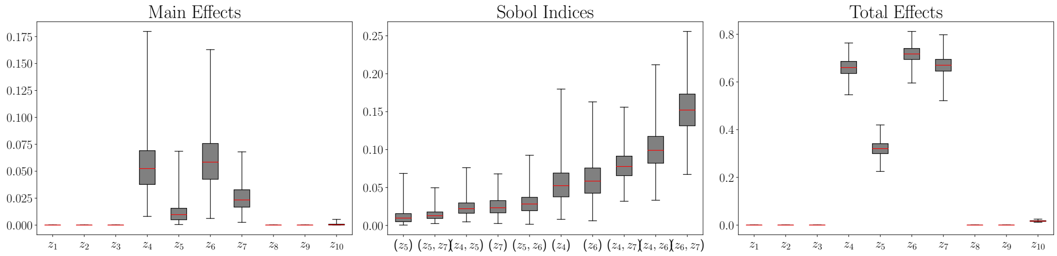


(c) Autumn ($Q^2 = 0.912$)

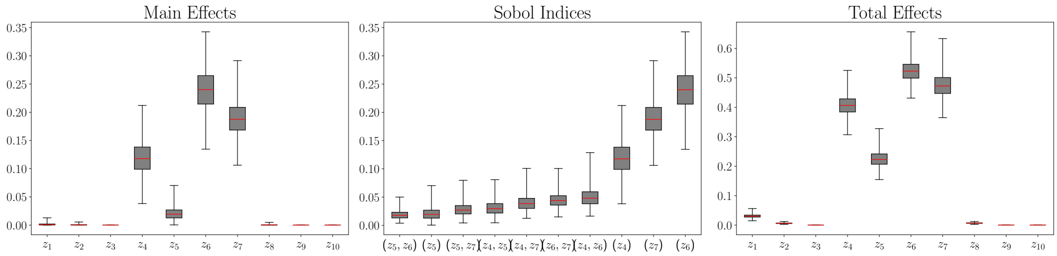
Figure 11. GSA results: main effects, Sobol and total effects indices (from left to right) for the Surface Temperature Averaged Over 60-90° (TS) QOI calculated annually and by season. The box-and-whiskers plots depict GSA results obtained using a Gaussian process emulator, which provides uncertainty bounds: the red central mark indicates the median of the data, the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The parameters $\{z_i\}$ are described in Table 2.



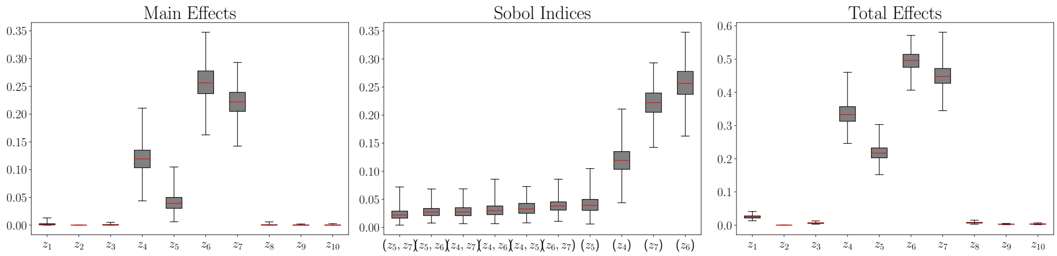
(a) Annual ($Q^2 = 0.782$)



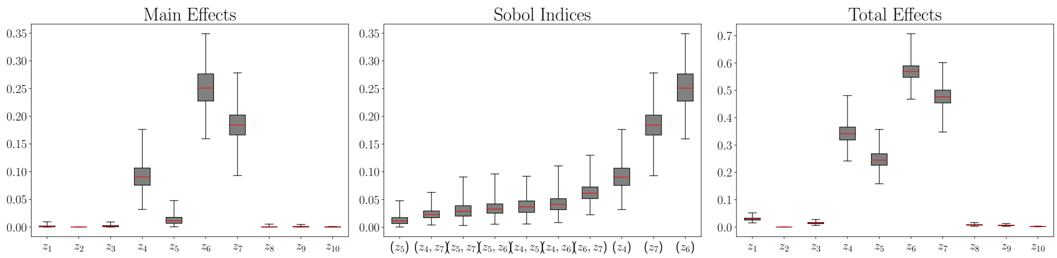
(b) Winter ($Q^2 = 0.295$)



(c) Spring ($Q^2 = 0.759$)

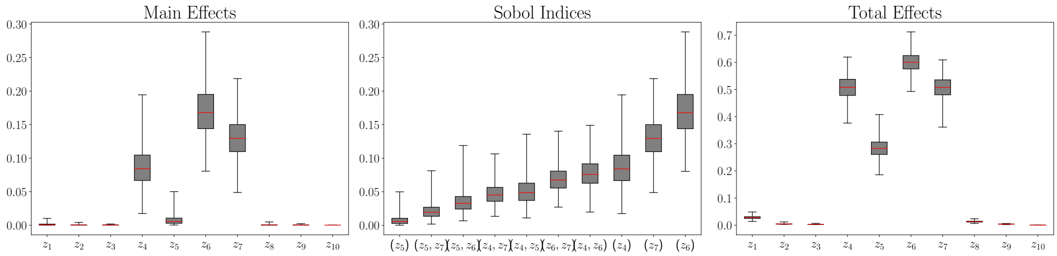


(d) Summer ($Q^2 = 0.792$)

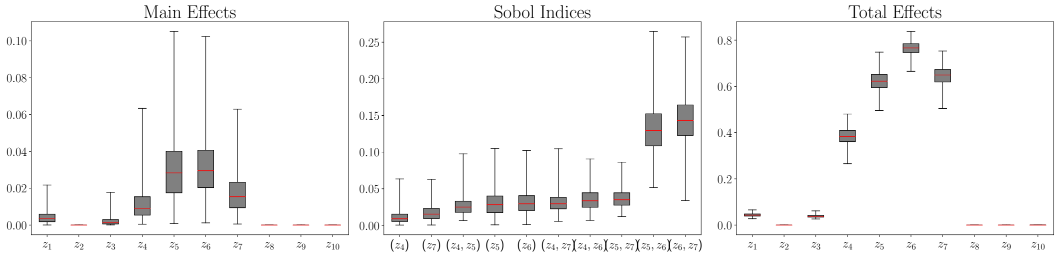


(c) Autumn ($Q^2 = 0.747$)

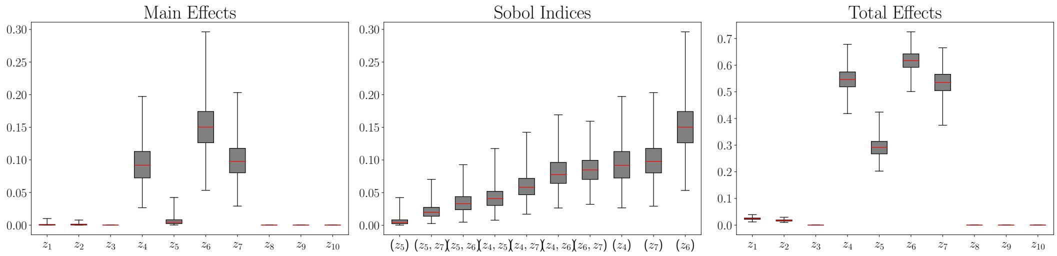
Figure 12. GSA results: main effects, Sobol and total effects indices (from left to right) for the Low Cloud Coverage Averaged Over 60-90° (CLDLOW) QOI calculated annually and by season. The box-and-whiskers plots depict GSA results obtained using a Gaussian process emulator, which provides uncertainty bounds: the red central mark indicates the median of the data, the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The parameters $\{z_i\}$ are described in Table 2.



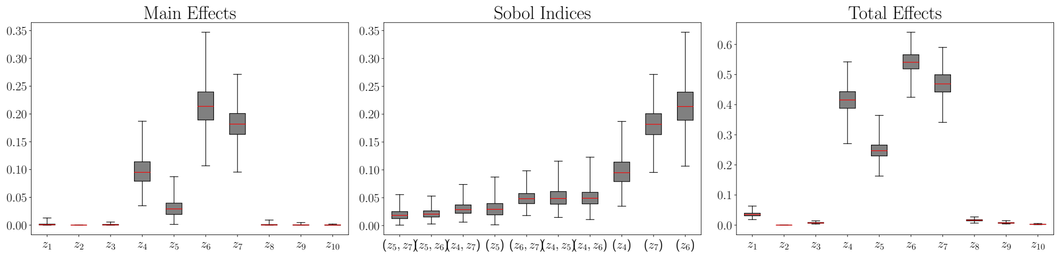
(a) Annual ($Q^2 = 0.676$)



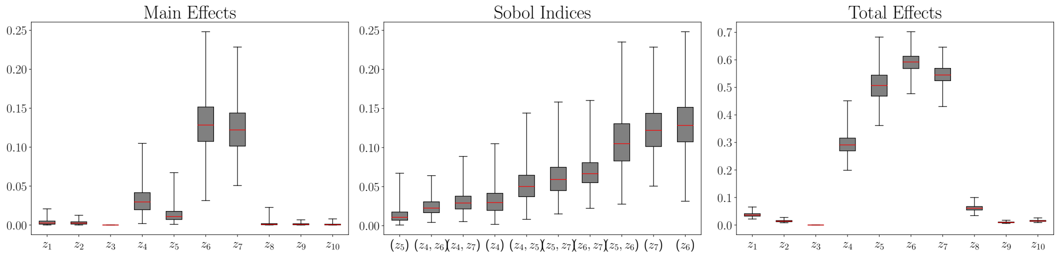
(b) Winter ($Q^2 = 0.153$)



(c) Spring ($Q^2 = 0.584$)



(d) Summer ($Q^2 = 0.734$)



(c) Autumn ($Q^2 = 0.576$)

Figure 13. GSA results: main effects, Sobol and total effects indices (from left to right) for the Net Longwave Surface Radiation Averaged Over 60-90° (FLNS) QOI calculated annually and by season. The box-and-whiskers plots depict GSA results obtained using a Gaussian process emulator, which provides uncertainty bounds: the red central mark indicates the median of the data, the bottom and top edges of the box indicate the 25th and 75th percentiles, respectively. The parameters $\{z_i\}$ are described in Table 2.

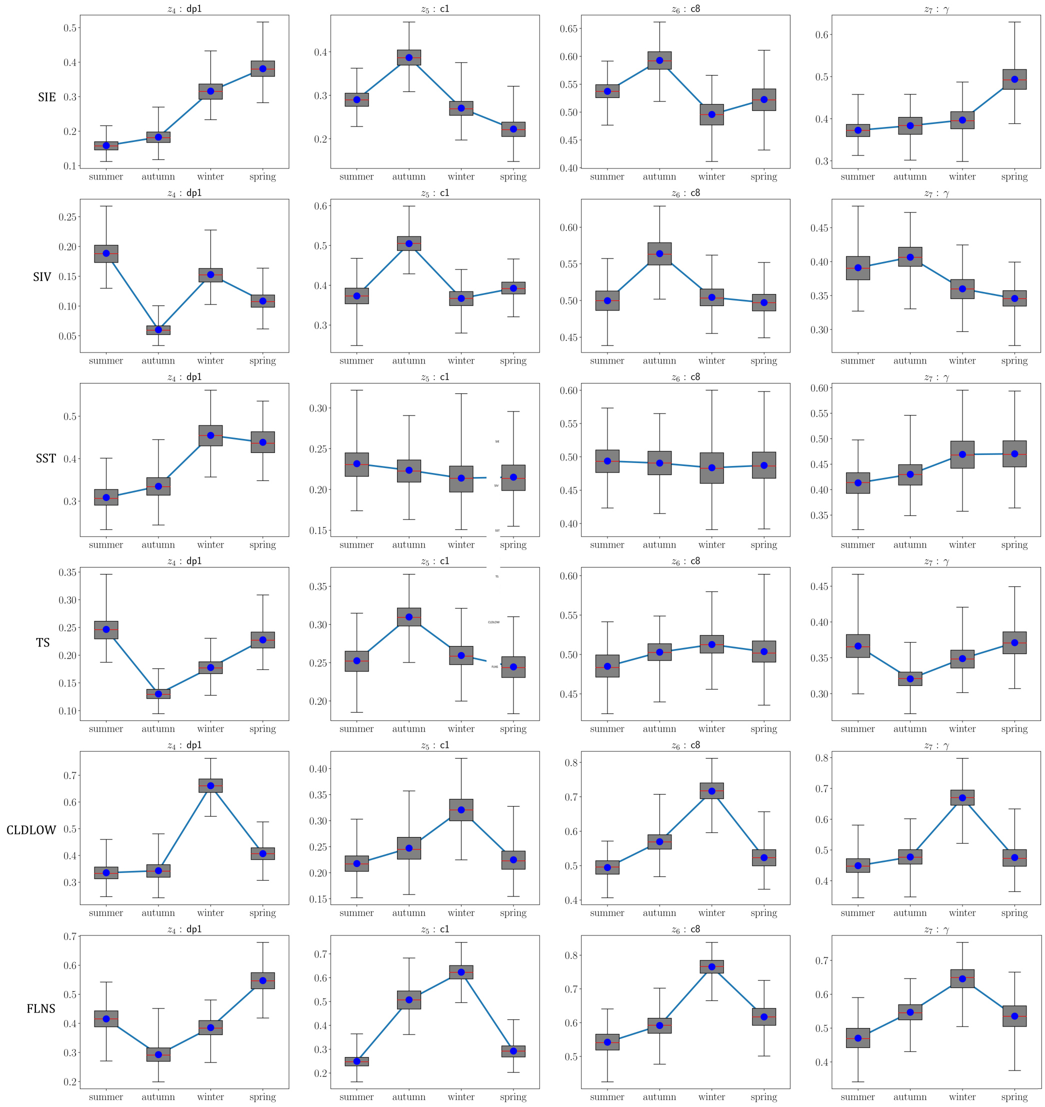


Figure 14. GSA results: seasonal variation of the mean total sensitivity (total effects) indices of the four most influential parameters. The box represents 25-75% confidence intervals. The median of the data is denoted by the red line. The mean of the data is denoted by the blue dot. Whiskers designate the minimal and maximal values of the total effects indices.

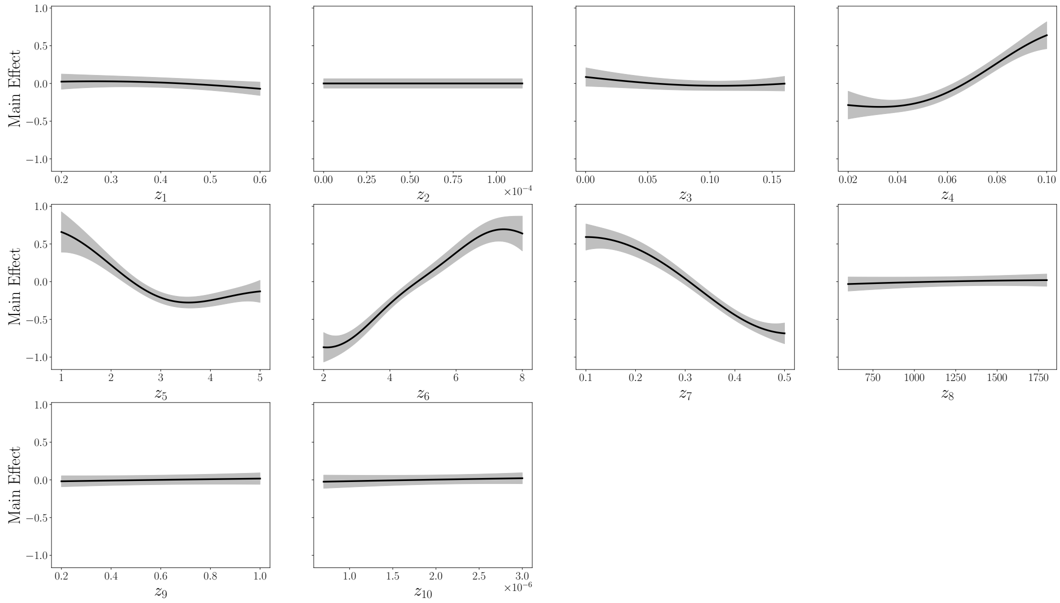


Figure 15. Marginalized main effects of the most important parameters affecting annual sea ice extent (SIE). The black solid line represents the median of the main effects calculated using a Gaussian process and the gray shading represents the 95% confidence intervals of the main effects calculated using the Gaussian process emulator.

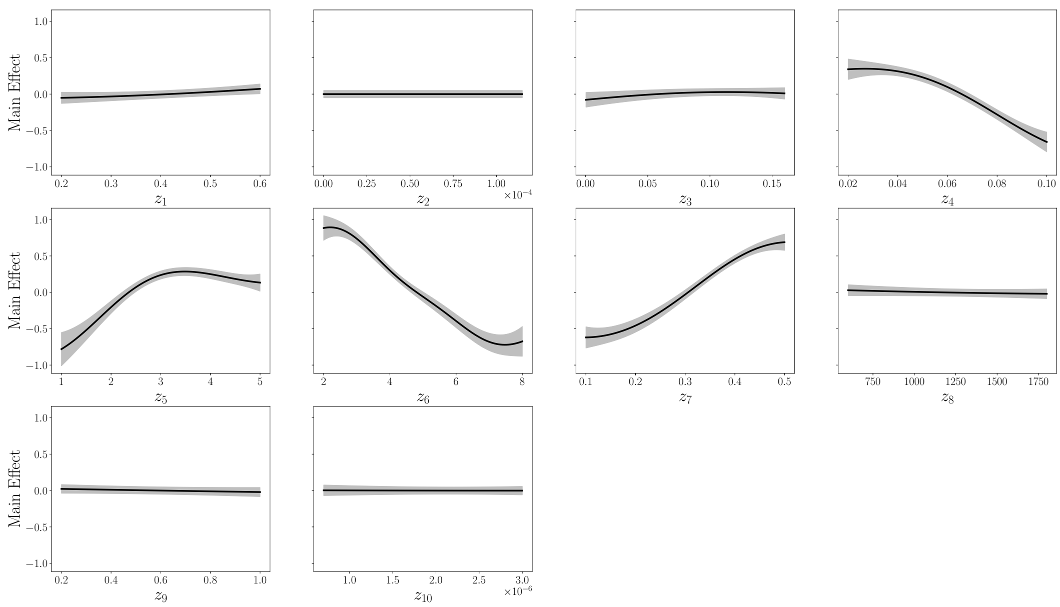


Figure 16. Marginalized main effects of the most important parameters affecting annual surface temperature (TS). The black solid line represents the median of the main effects calculated using a Gaussian process. The gray shading represents the 95% confidence intervals of the main effects calculated using the Gaussian process emulator.