5-2-2024

Second-Order Permutation Variable Importance: Quantifying and Attributing Uncertainty in Future Coastal Risk

Dwight Dinkins
dmd4883@rit.edu

Second-Order Permutation Variable Importance: Quantifying and Attributing Uncertainty in Future Coastal Risk

by

Dwight Dinkins

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of Master of Science in Applied and Computational Mathematics

School of Mathematics and Statistics

College of Science

Rochester Institute of Technology

Rochester, NY

May 2, 2024

Abstract

Sensitivity analysis (SA) systematically assesses and quantifies a model's uncertainty by examining the impact of parameter variations on the model's output. An objective of sensitivity analysis is to ascertain model-input parameters that contribute the most to the uncertainty of a model's output and could serve as a gauge to determine whether a parameter's variation is suitable, expected, or a realistic representation of the phenomenon being modeled. Incorporating sensitivity analysis for assessing the importance of a model feature is imperative for decision-making. Therefore, as a statistical methodology, SA's application in investigating coastal regions that are vulnerable to sea-level rise and extreme weather events—exacerbated by climate change—is crucial to effectively plan for and mitigate the consequences that follow. Quantifying the uncertainty of future forcings of coastal hazards and their impacts on risk estimates would better inform decision and policy making. Using random forest (RF) regression, this research provides a framework for assessing uncertainty in coastal risk predictions by proposing an extension of permutation variable importance for second-order interactions of model parameters. The second-order permutation variable importances assesses the importance of a combination of features by computing the mean-squared-error between the model's true value and its prediction computed over randomly permuted spaces that include the combination of features itself. The second-order permutation variable importances are tested on a simple mechanistically-motivated emulator, using a semi-empirical model and dataset for global mean sea-level (GMSL) change and a coupled sea level-coastal impacts model for the U.S. Gulf Coast. Our results demonstrate and

provide another relationship between permutation variable importances and Sobol sensitivity indices.

1 Introduction

The capacity of humans and the rate at which they can alter and affect the climate grows as technology grows. Investigating how they're contributing to climate change (i.e., how much?) and its resulting impacts on sea-level rise has been of interest since the 1960s.[29] Like any phenomenon, the climate accompanies considerable uncertainty as a result of the number of phenomenal interactions there are, such as the increase in greenhouse gasses and, consequently, global temperatures (e.g., global mean surface temperature), ice sheet disintegration, ozone depletion and a variety of other climate forces and drivers. In order to understand these phenomenal interactions and human contribution thereof, many climate scientists and modelers have considered special scenarios—specifically climate trajectories—to gain insight into potential futures and their accompanying model drivers.[18][19] [1] These model-based climate trajectories capture potential climate conditions, permit the quantification of uncertainty of human contribution to climate change, and illuminate the practicality of proposed mitigation and adaptation strategies.[26]

Model-based climate trajectories (i.e.,scenarios) can be categorized as being either emission, environmental, vulnerability, or earlier scenario work, each of which consists of analyzing a multitude of climate forcings in isolation or in conjunction with economic, technological, political, etc. trends.[27] As a result of the variety of climate forcings, their uncertainties, and the growing complexity of climate models, the additional uncertainty

[1] To be clear here, I define a "model driver" to be a variable within a mathematical model that makes large or minuscule contributions to the uncertainty of the model output.

accompanying climate models, resulting from assumptions reflected within representations of climate processes and, generally, differences in the models developed, pose the issue of discrepancy among climate risk estimations and projections. Moreover, the time expense required for developing model-based climate scenarios have caused delays in availability of these climate scenarios and employment within climate modeling, resulting in further inconsistencies among risk assessments. [28]

The process by which climate scenarios were developed for understanding uncertainties associated with potential futures followed assumptions or relied upon a posteriori knowledge germane to socioeconomic influences of greenhouse gas emissions. [26] proposed a *parallel* approach in which climate scenarios were developed nonsequentially by identifying characteristics of radiative forcings that permit broad climate modelling, and captures various combinations of climate forcings that effectuated that level of radiative forcing. In contrast to the *sequential* approach (i.e. the original climate scenarios), which began with socioeconomic influences and, then, progressively linking climate forcings in a linear fashion and using the resultant climate projections for risk assessment, the parallel approach starts with a desired radiative forcing characteristic(s) that is representative of a potential climate scenario. These climate scenarios were given the name: representative concentration pathways (RCP), since each pathway (i.e., trajectory) corresponds to a single radiative forcing characteristic, whose concentration levels overtime are, again, a result of a plethora of combinations of climate forcings.

Representative Concentration Pathways (RCP) are a new collection of model-based climate scenarios that substitute climate scenarios originally proposed by the Intergovernmental Panel of Climate Change (IPCC).[2] This approach allows for climate modellers to produce new climate models, experiments, and climate scenarios using the emissions generated by RCPs, effectively reducing the time between the availability of the climate scenario and employment in climate modelling.[28] What's unique about RCPs are that there are only four to select from: RCP2.6, RCP4.5, RCP6.0, and RCP8.5, each of which being distinct and offer a broad range of potential climate scenarios for exploration. Two such radiative forcing trajectories, RCP2.6 and RCP8.5, were explored by [18] to ascertain how climate drivers affect global mean sea-level rise overtime, using the Building blocks for Relevant Ice and Climate Knowledge (BRICK) model from [45] in conjunction with Random Forest—in the context of quantifying uncertainty in future coastal hazards. Radiative forcing trajectories, RCP2.6 and RCP8.5, correspond to climate scenarios in which the difference between incoming solar radiation absorbed by the Earth and energy radiated back to space achieves a maximum value of $\sim 3\mathrm{Wm}^{-2}$ before the year 2100[3] and any radiative forcing retaining values $> 8.5\mathrm{Wm}^{-2}$ in the year 2100 for RCP2.6 and RCP8.5, respectively. In the year 2020, [18] initial findings reveal that thermal expansion is predominately responsible for the increase in global mean sea-level rise, contributing 13.5 % in importance but saw a gradual reduction to 4.9% by the year 2055 under the RCP2.6 climate scenario. In contrast, [18] findings of RCP8.5 climate scenario revealed that the importance of thermal expansion was 12.3% for the year 2020 and, then,

[2] RCPs are *potential* futures and are not true or actual climate projections and/or estimates

[3] RCP2.6, as explained by [28], declines after the year 2100

decreases to 7.2% by the year 2040. Between the years 2020 and 2040, the relative importance of glacier ice-sheet disintegration is 6.2% to 4.2% and 6.8% to an apparent 0% for RCP2.6 and RCP8.5, respectively.[18] The former findings suggest that while thermal expansion remained a contributing factor of climate change (specifically, the rise in global mean sealevel), the decline of importance of thermal expansion is due to other contributions of global mean sealevel: ice sheet disintegration, despite its relative importance diminishing, in the case of RCP8.5.

To expound, as greenhouse gases accumulate, heat becomes trapped within Earth's atmosphere, causing a differential between the Earth's absorption and conversion of solar radiation to thermal radiation and its subsequent emission out into space. This heat not only melts glaciers and ice-caps and, in the long-term, Antarctic ice-sheet but also is absorbed by the Earth's ocean, causing it to expand and increase the global mean sealevel (i.e., thermal expansion). The uncertainty of thermal expansion equipped with the uncertainty in global mean temperature rise aggrandizes the overall uncertainty accompanying projections of global mean sea-level rise—twofold.

Furthermore, in both climate scenarios, there's substantial agreement: thermal expansion, as a consequence of low-high greenhouse gas emissions, and ice-sheet disintegration are climate, model drivers of global mean sea-level rise. This is consistent with IPCC's Sixth Assessment Report (AR6), "Impacts, Adaptation, and Vulnerability", in which they present an assessment of climate risks within the near-term, analyzing

various climate trajectories and emphasizing that the increase in global temperature and, consequently, sealevel will result in the inundation of major coastal cities and degradation of ecosystems.[19] The latter and [18] research and findings, by themselves, stresses the significance of UQ and SA as indispensable tools for informed decision-making, and the potential impact it may have if absent for assessing climate risk projections—for modeling in general. This emphasis is prevalent among the climate science community as newer model configurations are publicized, consisting of a combination of representative concentration pathways, shared-socioeconomic pathways (SSP), and regional sea-level rise (RSLR) scenarios for robust and accurate climate risk projections. One such publication by [35], who assessed the uncertainties resulting from combinations of RCP and SSP scenarios and their influence on risk estimates of the expected annual damage and adaptation cost using the Dynamic Interactive Vulnerability and Assessment (DIVA) modelling framework in conjunction with Random Forest regression as its substitution. Briefly, like RCPs, SSPs are postulated alternative futures of global mean sealevel, in terms of socio-economic challenges, intended for assessing the efficacy of proposed mitigation and adaptation strategies. To quantify the uncertainty of expected annual damage and adaptation cost, [35] used the predictions of RF model to obtain sensitivity indices for the output variables of interest. They found that the time evolution of the uncertainties associated with the expected annual damage and adaptation costs, as a result of the rise of global mean sealevel, increased during the 21^{st}-century. Specifically, the relative contribution to the uncertainty of the expected annual damage was largely due to SSP by the year of 2030 with a sensitivity exceeding 50%, whereas RCP

largely contributed to the uncertainty of the adaptation cost with a sensitivty of 80%. By the year 2050-2060, SSP scenarios' influence on the uncertainty of the expected annual damage declined in relative importance as a result of the emergence of RCP scenarios' influence on the global mean sea-level rise (i.e., rise in coastal flood damages). Initially, the relative contribution to the uncertainty of the adaption cost were due to RSLR with a sensitivity index exceeding 40%. While this pertains to local or *regional*, coastal sea-level contributions, the uncertainty accompanying the rise in greenhouse gas emissions globally dominated, resulting in a rise in sealevel and accordingly the rise in adaptation cost. By the year 2050, RCP exceeded RSLR in relative contribution to the uncertainty of adaptation cost, attaining a sensitivity index $>60\%$.[35] To this end, [35] noted that improved knowledge acquisition regarding RSLR prior to the year 2050 would aid in climate mitigation strategies by reducing uncertainty accompanying adaptation costs, as well as reduce the uncertainty associated with the expected annual damage provided the capability to unambiguously identify the specific SSP scenario "...when accounting for the occurrence likelihood of SSP-RCP combinations."[36]

These research findings demonstrate the power and utility of incorporating SA and UQ into models whose intended use is for decision and policy making.

Uncertainty is a characteristic inherent within all phenomenon. It refers to the lack of assurance or the degree of uncertainty germane to accuracy and reliability of results. Uncertainty comprises of two subcategories of uncertainty: *aleatoric* and *epistemic*

uncertainty. Epistemic uncertainty arises when a modeller lacks a comprehensive understanding of the phenomenon being modelled. This could come in the form of having an incomplete understanding of the underlying physics, insufficient or imprecise data, or simplifications and assumptions made during the model development process. Aleatoric uncertainty refers to the inherent stochasticity or randomness within the system being modelled, for example, climate processes.

In this context, uncertainty quantification (UQ) is a process that involves assessing the range and likelihood of potential futures in climate projections. This approach systematically addresses many of the inherent uncertainties accompanying climate modeling, namely variability in climate systems, model inadequacy, residual variability etc.[22] Sensitivity analysis (SA) complements uncertainty quantification by examining how variations in model inputs affect the model output. The latter statistical methodology (i.e., SA) is crucial for identifying key drivers of climate change and determining the robustness of model projections. Exploring how different assumptions and parameters influence model results, SA provides insight into the relative importance of various climate processes, forcings, and climate model drivers. It is essential for prioritized research and policy efforts, and for assessing the confidence in model projections and their implications for adaptation and mitigation strategies.

In the realm of climate modeling and managing climate risks, the pervasive nature of uncertainty demands meticulous attention and thoughtful analysis. This is crucial not

just for understanding current climate dynamics but also for predicting future climate conditions and informing policy and decision-making processes aimed at mitigating climate change impacts. The identification of sources of uncertainty within climate-risk projections and understanding the roles these uncertainties play—both individually and in combination with other variables—is a fundamental aspect of climate science. Climate models are simplifications of the Earth's climate system, which are either developed using mathematical representations of that physical process or are data driven—or both. They, at their incipience, introduce uncertainty within climate projections and is exacerbated by assumptions made during the model development stage. Model calibration is a quintessence of how climate models "imbibe" uncertainty, by estimating model parameters using empirical data that may possess measurement errors due to human and technological limitations; this is an example of parameter uncertainty and usually arises within the initial phases of the model development stage, in the case of a prognostic model (*see* **§2.1**). The stochastic nature of the climate system means that small differences in initial conditions could result in divergent outcomes which, as aforementioned, is an example of aleatoric uncertainty and *residual* uncertainty.[23] The presence of these uncertainties complicates the assessment and management of climate risks. It affects the ability to accurately predict the frequency and intensity of extreme weather events, the rate of sea-level rise, the impacts on biodiversity and ecosystems, and socioeconomic impacts, among others. This, in turn, poses challenges for developing effective adaptation and mitigation strategies.

Understanding the roles of different processes and their associated uncertainties in climate modeling is crucial for several reasons: improving model fidelity, risk assessment, and planning. To manage and understand uncertainty in climate modeling and risk management, the employment of statistical methodologies for uncertainty quantification and sensitivity analysis are paramount, and research and development of novel methods for attributing and characterizing that uncertainty to model inputs are necessary. By placing uncertainty at the forefront of climate science, researchers and policymakers can better navigate the complexities of climate modeling and risk management. This approach facilitates a more nuanced understanding of potential future conditions and supports the development of flexible, adaptive strategies for mitigating and adapting to climate change.

The objective of SA is to quantify and attribute model-output uncertainties to its input parameters. There are various SA methodologies that can be used for this task and any one of them fall in one of the two categories: local or global methods.

Local methods of SA examine the effects of small perturbations within a region of space on the output of a model. Specifically, it explores the sensitivity of the model near a specific point within the parameter space by differentiating the model or performing a "one-at-a-time" (OAT) method, which entails changing the value of a parameter OAT while keeping all other model parameters fixed.[32] On the other hand, global methods of SA examines the influence of input parameters over their entire range of possible values,

considering the entire input space rather than just a local region. These methods include variance-based sensitivity analysis (i.e., Sobol indices, more on these later),Fourier Amplitude Sensitivity Testing (FAST), Pearson, Spearman, and Kendall correlation coefficients, which can quantify the contribution of each input parameter to the output variance or (in the case of the latter three) the linear dependence between features.[24] A major distinction between local and global methods of SA are the assumptions made about the model—about the phenomenon being studied. Local methods assume that the modelled phenomenon (or phenomena) behaves linearly within the proximity of a specific point being analyzed and, thus, unable to detect nonlinearity if present within a model, whereas global methods can. Another difference is the computational complexity between the two categories, as local methods happen to be less computationally expensive than global methods by virtue of the fact that it does *not* examine the entire parameter space. The choice between local and global sensitivity analysis depends on the specific goals of the analysis, the nature of the model, and available computational resources. Local sensitivity analysis is useful for a quick, initial assessment of sensitivity around a specified point, which can be particularly useful in the early stages of model development or when computational resources are limited. Global sensitivity analysis, while more computationally intensive, provides a comprehensive view of the importance and influence of input parameters across their entire range, offering deeper insights into the model's behavior under uncertainty.

In light of these differences, global methods are more appropriate and suitable for

assessing climate models since climate forcings are not isolated phenomena but rather affect and interact with each other—a kind of compound uncertainty.[37] Therefore, global methods, specifically variance-based SA, will be the primary statistical methodology of focus.

Variance–based sensitivity analysis is a global SA method that's interested in ascertaining how perturbations of a model input parameter affect a model's output of interest or metric, which uses variance as the primary metric of uncertainty for this method. The variance of the model output can be decomposed as the sum of individual and combinations of model-input partial variances. Sobol sensitivity analysis, a variance-based SA method, operationalizes this concept through the application of the Hoeffding Decomposition, also known as high-dimensional model representation (HDMR), facilitating the isolation and aggregation of model-input uncertainties. It provides a robust quantitative measure of a model parameter's relative importance and can handle nonlinearity among inputs. Subsequently, Sobol SA provides what are called first, second, third etc., and total-order indices (or measures) which are the quantifiers of uncertainty within a model. These measures are computed under the assumption that each model input is independent and identically distributed, using a predefined *design of experiments.*[33] Indeed, the model undergoes a simulation in which it is computed on pseudo-random or stratified-random low-discrepancy samples, to explore parameter relationships and complex behaviors. The choice of sampling strategy is contingent on the model design and the modeller's choice of method for quantifying and attributing uncertainty

among input—which is largely dependent upon the phenomenon being studied. Sobol first-order measures quantifies individual, model-input contributions to our model's output variance, which is represented as a percentage of the total variance of the model. Similarly, the second and total-order Sobol measures quantifies variable combinations' and total effect (i.e. the sum of first, second, etc. order measures) contributions to the model output variance. It quantifies the relative influence that each model parameter has on the prediction of the model output (i.e. factor fixing and mapping), etc and enables the modeller to pay selective attention on more influential model parameters, as they are contributing the most to the uncertainty of the model output.[40] This aspect may incentivize the modeller to either simplify the proposed model by the removal of a non-influencial model parameter or the inclusion of a influential model parameter for analysis.[1]

Many complex, modeled phenomena are expressed either vectorially or as a system of equations. In this context, multivariate Sobol SA is a generalization of Sobol SA for models of the latter type, in which each component or equation is a function that may or may not be expressed as a combination of other functional outputs present within a system or vectorial equation. It relies on a multivariate Hoeffding Decomposition, like traditional Sobol SA, and provides insight into the relative influence of model-inputs on a model output of interest.[17]

Despite the advantages of this approach,*viz.*, can handle nonlinearity and estimate a

model parameter's global influence on the output of the model, a drawback of Sobol SA is the computational complexity of computing the Sobol indices, since one would have to compute it for each model feature and combinations of features, which can quickly become computationally expensive if the proposed model is highly parameterized. For instance, in the case of Sobol first and total-order indices, a model with d-variables and N model runs, would require $N(d+2)$ model evaluations.

Researchers have used a variety of approaches, in order to overcome this challenge. For example, the aforementioned [18] used Random Forest classifiers for quantifying the uncertainty associated with model features overtime, which uses an impurity-based approach (specifically, gini impurity for classification) for assessing model-input uncertainties and, more importantly, its impacts on global, mean sea-level rise.

Random Forest is a concoction of predictive models in the form of finitely many decision trees to make predictions for classification and regression models, ensuring[4] robustness and accuracy. A decision tree is a hierarchical model in which a randomly chosen feature is initialized and successively split based upon a threshold that offers the greatest reduction of impurity (*see* **Figure** 1). [5][8] Random Forest provides, as a tool, a 'generalized' model for the purpose of classification and regression problems but, also, offers an intuitive approach for quantifying model-input uncertainty. It is nearly an 'all-in-one' model in the sense that it provides a model with hyperparameters for calibration and allows for the modeling of complex interactions between features and observational

[4]This assumes that sufficient hyperparameter tuning was performed for improved accuracy

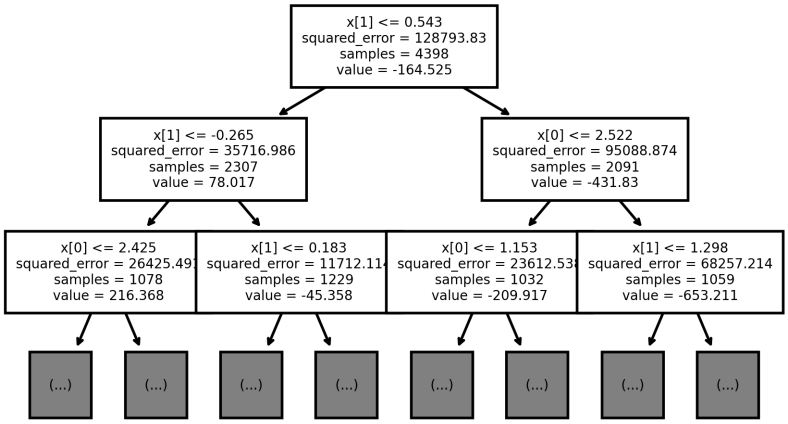

Figure 1: A Random Forest decision tree

data, making it applicable to a wide range of real-world phenomena, provided that relevant data is available. To expound, during machine learning model development, the modeller partitions the available dataset into three separate datasets: training, validation, and testing datasets using a desired split. This procedure is done to iteratively assess the model's performances on the validation set, after repeatedly learning and calibrating the model on training data, in order to determine the model's generalizability (and/or accuracy) on unseen data (i.e., test dataset). The validation set allows for the tuning of model parameters without using the test set. This way, the test set remains an unbiased, independent measure of model performance—preventing data leakage and over-fitting.

An important aspect of random forests is its ability to assess the importance of different input variables when making predictions, which is achieved through measures such as mean impurity decrease or gini importance that aid in determining the im-

portance of a feature by examining its contribution to the predictability of the model's output. By analyzing variable importances, modellers can gain valuable insights into the underlying relationships within their data and prioritize features accordingly for further analysis or feature selection. This enables a better understanding of the driving factors behind the model's decisions and enhances interpretability, making it a popular choice for both predictive modeling and exploratory data analysis. In contrast to conventional modeling wherein the modeller develops a mathematical representation of a real-world phenomenon then quantifies the relative importance of model features as an independent procedure, RF uses the calibrated model and adjoins SA by traversing the decision trees, accumulating node purities corresponding to features that offer the greatest reduction of impurity. In other words, it uses the resultant impurity-decrease as well as the frequency of a feature's split, to quantify how important such a feature is regarding the predictability of the model; RF utilizes an impurity-based methodology for assessing model-input uncertainty. This procedure alleviates the necessity of computing Sobol indices for each model feature, as a result of its computational efficiency and presents an alternative, complementary methodology to traditional SA—with a focus on understanding model uncertainty through the lens of predictability.

Although random forests mitigate the computational expense accompanying Sobol Indices, RF's variance[5] reduction for regression isn't effective at capturing model-input interactions (that is, its algorithm see **§2.2**), particularly the attribution of uncertainties

[5]the modeller has the option to choose different criterions.

for nonlinear models.[41] This underscores the need for refinement in methodologies to balance computational efficiency with analytical precision.

Developing a procedure that retains the predictive power of Random Forests and incorporates global sensitivity analysis is the motivation of this thesis.

To rectify RF's inability to detect and quantify combined features' contributions to a model's output variance, I propose a second-order permutation variable importance as an appendage to ascertain the presence of nonlinear features within a model and can be employed for both data-driven (e.g., Random Forests) and deterministic models. Second-order permutation variable importances uses a model's prediction of perturbed features and estimates its mean-squared deviation from the model's true values, yielding a variance decomposition congruous with the Hoeffding-Variance Decomposition—which is fundamental for estimating Sobol variance-based indices. These measures, therefore, bare a relationship with Sobol SA (specifically, its indices), provide an alternative formulation for variance-based SA, and are generalizable in application–despite its intended remedial use for RF.

To this end, the paper is outlined in the following fashion: in Section 1, I provide a brief, yet detailed, introduction of sensitivity analysis, specifically Sobol sensitivity analysis and RandomForest's variance reduction and their motivation; in Section 2, I introduce the concept behind Random Forest, its "under-the-hood" mechanics in preventing overfitting, predictive power, but with a focus on its ability to select features of importance

in predicting the models output. Moreover, in Section 2, I provide a formulation of Sobol Sensitivity Analysis, recent development of total and first-order permutation variable importances and their relationship with Sobol Indices, as well as introduce a new proposition of second-order permutation variable importances. In Section 2, I introduce test functions: Ishigami Function and Rahmstorf (2007) Global Mean Sea-level (GMSL) model along with a real-world application using the BRICK-CIAM dataset; in Section 3.1,3.2,3.3, I compare the results of the second order measures of permutation variable importances to corresponding Sobol second-order measures for the Ishigami function and Rahmstorf model and present their respective confidence intervals;Conclusion, Acknowledgements, Appendix, Bibliography.

2 Methods

2.1 Sobol

Prognostic models are developed using well-founded mathematical concepts for gaining insight into an observed phenomenon. But, occasionally, a modeller may overlook underlying phenomenon or be inclined to consider a plethora of potential model variables that are believed to play a role in the observed phenomenon, resulting in either under or over-parameterized models. Even if the developed model aligns with observational data, the potential that a perturbation of a model parameter results in a considerable or negli-

gible deviation from the predicted model output could be a consequence of overlooking underlying phenomenon or the inclusion of frivolous parameters (or variables), respectively, necessitating further investigation prior to its use for decision-making. The ability to select model features and parameters that are pertinent to the observed phenomenon at-hand would aid in curtailing over-parameterized models. When modelling real-word phenomena, the consideration of the unforeseen—especially in climate modeling—is worthy of investigation. Investigating relevant (including unforeseen) features and their roles' played, ascertaining known 'model-drivers' are components of the "checks and balances" process and is essential for model development. In other words, this is the modeler's chance to validate or confirm that their prescriptions and assumptions were correctly encoded, relevant, and appropriate (i.e. realistic) for the observed phenomenon. Therefore, developing a quantitative correspondence between model-input and output uncertainties is of paramount importance for validating the suitability of a model's design.

Sensitivity analysis is the study of the variation in a model's output, Y, as a function of uncertainties in its input, X.[38] Suppose we have a generalized model $f(X) = Y$, where $X = [X_{0k}, X_{1k}, \ldots, X_{nm}]$ (n and m denote the n^{th}-column and m^{th}-row) are assumed to be independent, identically distributed random variables (i.e. each feature, X_i, is similar ($\sim$ in probability distribution). One can decompose the model output, Y,

as the sum of partial functions of X by

$$Y = f(X) = f_\circ + \sum_{i=0}^{n} f_i(X_i) + \sum_{i<j}^{n} f_{ij}(X_{ij}) + \cdots + f_{0,1,\ldots,n}(X_{0,1,\ldots,n}), \tag{1}$$

where

$$f_\circ = \mathbb{E}[Y] \tag{2}$$

$$f_i(X_i) = \mathbb{E}[Y \mid X_i] - f_\circ \tag{3}$$

$$f_{ij}(X_{ij}) = \mathbb{E}[Y \mid X_i,\, X_j] - f_i - f_j - f_\circ. \tag{4}$$

This functional decomposition is known as high dimensional model representation (HDMR) or ANOVA (i.e., Analysis of Variance) decomposition, which is a statistical method used in SA for isolating individual and combinations of model-input contributions to the model output.[39][30] The model, f, must be a quadratically integrable function (i.e., L^2-function), and the integral of each term (namely f_i, f_{ij}, etc.) equals zero; that is,

$$\int_\Omega f_{i_0\ldots,i_r}(X_{i_0\ldots,i_r}) dX_l = 0, \qquad \text{for } l = i_0 \ldots, i_r \tag{5}$$

where Ω is the input space for which $X_i \in \Omega$. Using equation **(1)**, one can decompose the model output variance as the sum of individual and combinations of model-input partial

variances of the former quantities (*see* Equations (2)-(4)), yielding

$$\mathbb{V}(Y) = \sum_{i=0}^{n} V_i + \sum_{i<j}^{n} V_{ij} + \cdots + V_{0,\ldots,n}, \tag{6}$$

where the notation, $\mathbb{V}$, denotes the variance operator and indexed V_i and V_{ij} are the partial variances of individual and combinations of model inputs. These partial variances are, mathematically,

$$\begin{aligned} V_i &= \mathbb{V}_{X_i}(\mathbb{E}_{X_{\sim i}}[Y \mid X_i]) \\ V_{ij} &= \mathbb{V}_{X_{ij}}(\mathbb{E}_{X_{\sim ij}}[Y \mid X_i,\, X_j]) \\ &\vdots \\ V_{0,\ldots,n} &= \mathbb{V}_{X_{0,\ldots,n}}(\mathbb{E}_{X_{\sim 0,\ldots,n}}[Y \mid X_0, \ldots, X_n]), \end{aligned} \tag{7}$$

in which the notation $X_{\sim i}$ is shorthand for the set of all model inputs, X_N, $N = \{0, \ldots, n\}$, except input X_i (i.e., $X_{\sim i} \equiv X_{N \setminus \{i\}}$), and the sub-scripted operators denotes the space over which the operators are computed. To determine the relative influence that model-inputs have on the variance of the model out, the ratio of the partial variances with respect to the total model variance (i.e., $\mathbb{V}(Y)$ yields the quantities: Sobol Indices (or measures) of the first, second, and total order measures. These indices are represented, respectively, as

$$S_i = \frac{V_i}{\mathbb{V}(Y)} \qquad S_{ij} = \frac{V_{ij}}{\mathbb{V}(Y)} \qquad S_{\tau_i} = \frac{\mathbb{E}_{X_{\sim i}}[\mathbb{V}_{X_i}(Y \mid X_i)]}{\mathbb{V}(Y)}, \tag{8}$$

requiring that the summation

$$1 = \sum_{i=0}^{n} S_i + \sum_{i<j}^{n} S_{ij} + \cdots + S_{0,\ldots,n} \tag{9}$$

The first-order Sobol index could be used for factor prioritization wherein a model-input is regarded as influential provided that it has the largest percentage-variance. Likewise, the total-order Sobol index, which quantifies the total uncertainty attributed to an individual model input, may be discarded from analysis provided that its contribution is minuscule (i.e., ≈ 0) and possesses no interactions.[1] The latter is an example of factor fixing, in which variations of a model parameter has no significant contribution to the variance of the model-output.[40] Second-order Sobol indices captures the contribution of model-input combinations (that is, their interaction[6]) to the variance of the model output.

In summary, Sobol sensitivity indices quantify the relative influence of each model parameter in the variance of the model output (i.e. factor fixing and mapping, factor prioritization)[40]. Sobol SA isn't limited in application. However, the computational expense becomes encumbering and slows the modeling and analytical progress, especially since law-driven models are often highly parameterized. In the case of high-dimensional or highly parameterized models, applying Sobol SA could further confound the interpretation of model results. This computational challenge has been overcome by employing

[6] A combination of features are said to interact, if the sum of their individual effects on the model output cannot be expressed as a sum [39]

alternative global or local methodologies: Shapley Effects for machine learning and the Method of Morris, to name a few. The former method is nearly as computationally demanding as Sobol SA, whereas the latter (i.e., Method of Morris) is more computationally efficient, especially for models with a large number of model-inputs since it requires fewer model evaluations than relatively comprehensible methods like Sobol SA. The Method of Morris (MoM) works by systematically and sequentially varying one input at a time (i.e., OAT sampling) across a predefined grid, quantifying the change in a model's output. In contrast to Sobol SA, MoM estimates what are called "elementary effects," which are global sensitivity measures.[34]

2.2 Sobol Indices: Implementation

The computation of Sobol first and total-order measures entails several model evaluations, using a Monte-Carlo based numerical procedure. First, the modeller must generate low-discrepancy, stratified, pseudo or quasi-random samples following an a priori assumption germane to the distribution of each feature (i.e., ***design of experiment***). For simplicity, assume that the model $f(X) = Y$, where each $X_i \in X$ are independent, uniformly distributed random variables within the unit interval (i.e., $X_i \in \mathbb{U}[0, 1]$) and Y is the output of interest; we'd like to know how variation of our model features affect the variance of our model output. The procedure experiment by [39] goes as follows:

i. Generate a matrix with dimension $(2k, N)$ with the appropriate feature transformation, where k denotes the number of random samples and N the total number of features, and divide the matrix into two, distinct sub-samples: $\mathbf{A}$ and $\mathbf{B}$.

ii. Create matrices, $\mathbf{C}_i$, by interchanging a column (i.e., feature) of $\mathbf{B}$ with that of a column of $\mathbf{A}$

iii. Evaluate the model over the samples spaces, yielding

$$Y_{\mathbf{A}} = f(\mathbf{A}) \qquad Y_{\mathbf{B}} = f(\mathbf{B}) \qquad Y_{\mathbf{C}_i} = f(\mathbf{C}_i) \tag{10}$$

These measures suffice in computing the Sobol first and total-order measures whose formulas are given by [39]

$$S_i = \frac{V_i}{\mathbb{V}(Y)} = \frac{\frac{1}{N}\sum Y_{\mathbf{A}}^j Y_{\mathbf{C}_i}^j - \frac{1}{N^2}\sum Y_{\mathbf{A}}^j \sum Y_{\mathbf{B}}^j}{\frac{1}{N}(Y_{\mathbf{A}}^j)^2 - f_\circ^2} \tag{11}$$

$$S_{\tau_i} = \frac{\mathbb{E}_{X_{\sim i}}[\mathbb{V}_{X_i}(Y \mid X_i)]}{\mathbb{V}(Y)} = 1 - \frac{\frac{1}{N}\sum Y_{\mathbf{B}}^j Y_{\mathbf{C}}^j - f_\circ^2}{\frac{1}{N}(Y_{\mathbf{A}}^j)^2 - f_\circ^2} \tag{12}$$

where the mean

$$f_\circ = \left(\frac{1}{N}\sum_{j=1}^{N} Y_{\mathbf{A}}^j\right)^2.$$

2.3 Random Forest Feature Importance

Random Forest is an ensemble machine learning model that [5][8] proposed as a solution for curtailing over-fitting by introducing randomized node optimization (RNO), out-of-bag (OOB) bootstrapping and permutation variable importances (PVi) due to decision tree algorithms' (i.e., Classification and Regression Trees (CART)) proclivity to over-fit training data during the model development stage. Again, a decision tree is a model with a hierarchical structure in which features are randomly chosen and then successively split (producing child nodes) based on a threshold that offers the greatest reduction of a criterion (*see* **Figure** 1). Throughout this process, RNO introduces a degree of randomness in the selection of split points and features at each node of a decision tree.[9] Instead of exhaustively searching for the most optimal split, RNO randomly selects a subset of features and split points, thereby reducing computational complexity. This approach not only accelerates the training process but also contributes in diversifying the decision trees within the forest, potentially leading to improved model robustness and generalization capabilities. The randomization aspect aids in mitigating over-fitting, as it ensures that the trees in the ensemble are not overly sensitive to specific features in the training data. In the background, OOB samples data points with replacement to create multiple subsets, each of which is used to train a separate model in the ensemble.[10] The key feature of OOB bootstrapping is that some data points are left out (i.e., not sampled) for each subset. These 'out-of-bag' samples are then used to evaluate the model's performance, providing an internal mechanism for cross-validation and aiding with estimating the

generalization error of the ensemble model without the need for a separate test set. OOB bootstrapping is particularly advantageous in scenarios with limited data, as it maximizes the use of available data for both training and validation. Therefore, given its ensemble nature wherein multiple decision trees are consolidated to make a prediction, RF's structure can be exploited to gauge the importance of each feature by observing how frequently a feature is used to split data across the numerous trees. Consequently, this frequency provides, as proxy, a measure for a variable's importance, permitting us to ascertain model-input contribution to the prediction of our model by assigning more weight to a feature that reduces the variance (or more generally, a criterion) at each split.

Mathematically, assume we have a regression model $f(\mathbf{X}) = Y$, where each $\mathbf{X}_k \in \mathbf{X}$, $\mathbf{X}_k = (X_{0,k}, \ldots, X_{l,k})$, denotes a model feature of the dataset $\mathbf{X}$ and Y is the model's output of interest. Let $\mathcal{O}_n = \{(\mathbf{X}_k, Y_k) \mid 0 \leqslant k \leqslant m\}$ be n-bootstraps of the training data. For each tree, Random Forests selects a bootstrap $O_i \in \mathcal{O}_n$ to train.[2] That is, trees are grown by first generating bootstrap samples of the training samples $\mathcal{O}_n$, performs a split on the k-features, and chooses the feature among them that reduces the variance between the parent and child nodes.[3] To this end, random forest makes a prediction by averaging over the T-decision trees. The prediction of a point $\mathbf{x}_l^{\mathsf{c}}$ from the test dataset is, therefore

$$\hat{f}_T(\mathbf{x}_l^{\mathsf{c}}; R_T, \mathcal{O}_n) = \frac{1}{T} \sum_{t=1}^{T} \hat{f}(\mathbf{x}_l^{\mathsf{c}}; R_t, \mathcal{O}_n), \tag{13}$$

in which each $R_t \in R_T$ is independent, independent of $\mathcal{O}_n$ and randomize the T-decision trees by resampling the bootstrap training data $\mathcal{O}_n$, denoted as $\mathcal{O}_n(R_t) \subset \mathcal{O}_n$; heretofore, I will refer to decision trees as estimators.[3]

After all T-estimators are fully grown, the impurity of a node is simply a measure that maximizes how well or certain a random forest model prediction is at that node. Let S be a set of splitting nodes $t \in T$ and define a split function $\phi(s,t)$ for every $s \in S$ and $t \in T$. Then for every parent node t, there is a split s^* that maximizes the goodness of the split function $\phi(s,t)$. Suppose we have a learning sample $\mathcal{O}_i \in \mathcal{O}_n$ for an RF-estimator to train and that contains a node corresponding to a feature $\mathbf{X}_k \in t$ and satisfying a condition c (i.e., $\mathbf{X}_k \leqslant c$) and $c \in C$. Let N be the total number of elements in $\mathcal{O}_i$ (i.e., $\left\|\mathcal{O}_i\right\|$), N_c be the total number of cases satisfying the condition $\leqslant c$, and $N_c(t)$ be the total number of elements satisfying the condition $\mathbf{X}_k \leqslant c$ at the node t.[7] Breiman (i.e., [5]) defines the goodness of a split as one that reduces the overall impurity of a node t. In order to determine the best split s^*, [6] represents $\phi(s,t)$ in terms of the impurity function for any node t, namely

$$\phi(p_{\mathbf{c_t}}) = \phi(p(c_1 \mid t), p(c_2 \mid t), \ldots p(c_n \mid t)) \tag{14}$$

where $p_{\mathbf{c_t}} = p(c \mid t) = \frac{p(c,t)}{p(t)}$ are estimated, frequency-probabilities and satisfy

$$\sum_{c \in C} p(c \mid t) = 1 \tag{15}$$

$$p(c,t) = \frac{\pi(c) N_c(t)}{N_c} \tag{16}$$

$$p(t) = \sum_{c \in C} p(c,t) \tag{17}$$

Thus, $p(c \mid t)$ are the proportions of cases for which $\mathbf{X}_k \leqslant c$ in $\mathcal{O}_i$ for node t, and $\pi(c)$ are prior probabilities, which are either supplied by the analyst or are $\pi(c) = \frac{N_c}{N}$; in the latter case

$$p(c \mid t) = \frac{N_c(t)}{N}.$$

For each parent node t, a split is performed, allotting a proportion $p_{R,\mathbf{c_t}}$ to a child-right node and a proportion $p_{L,\mathbf{c_t}}$ to a child-left node. Then, the best split s^* is determined by the minimum difference of impurity function at node t, $\Delta\phi(s,t) = \Delta\phi(p_{\mathbf{c_t}})$, which is given by

$$\Delta\phi(s,t) = \phi(p_{\mathbf{c_t}}) - p_{R,\mathbf{c_t}}\phi(p_{\mathbf{c_{t_R}}}) - p_{L,\mathbf{c_t}}\phi(p_{\mathbf{c_{t_L}}}), \tag{18}$$

where, for clarity, $\phi(p_{\mathbf{c_{t_R}}})$ and $\phi(p_{\mathbf{c_{t_L}}})$ are the impurities at the child-right and child-left nodes $t = t_R$ and $t = t_L$, satisfying the condition $\mathbf{X}_k \leqslant c$ and $\mathbf{X}_k > c$, respectively. Let $\tilde{T}$ denote set of *ordered* splits and terminal nodes (i.e., leaf nodes), and let $\Phi(t) = \phi(s,t)p(t)$.

Then, the total impurity of an estimator

$$\Phi(T) = \sum_{t \in \tilde{T}} \Phi(t) \tag{19}$$

$$= \sum_{t \in T} \phi(s,t) p(t) \tag{20}$$

$$= \sum_{t \in \tilde{T}} \phi(p_{c_t}) p(t) \tag{21}$$

.[7] The choice of impurity function (i.e., gini criterion, mean-squared-error, etc.) depends on the choice of random forest model: classification or regression. The above impurity function, taking as argument probabilities p_{c_t}, is a generalization for computing the impurity for regression problems; nevertheless, the formulation is generalizable for random forest classification problems as well.

To summarize, this procedure entails traversing a collection of estimators and computing the reduction of impurity at each decision node. Since each decision node corresponds to a feature, the reduction of impurity corresponds to the amount of uncertainty associated with that feature; the accumulated impurities for the features are the total uncertainties apportioned among the features—for that estimator. For a collection of estimators, a feature's importance is the average total reduction of impurity over all estimators—for all nodes.

Despite the apparent attractiveness of this approach at quantifying uncertainty, impurity-based sensitivity measures for variable importances suffer from the bias in-

duced by the proportion of samples allotted for each child node. For example, suppose you have a classification model wherein a feature(s) have a sizable amount of unique values (e.g.[11]). That feature will have a greater chance to split in a way that reduces the impurity at its decision node. In other words, that feature's impurity will be allotted a higher proportion of samples (i.e. a higher weight), thereby biasing the true importance of other features and the feature itself. In the case of a regression model, if the model possesses interaction between features, the impurity-based procedure will yield misleading results. The latter was proven by Scornet (2010) and Baptiste (2017), demonstrating that if each model-input feature is independent and interactions are absent, then the impurity-based procedure yields a variance decomposition similar to that of Sobol.[41][16] Most real-world phenomena often necessitate complex models that have interactions (*see* §.

At present, there isn't a consensus as to the interpretability of the impurity-based measures—nor a theoretical substantiation for the justification of impurity-based measures as quantities of uncertainty for variable importance. Although the aforementioned methodology fails for models wherein interactions are present, their absence yields a variance decomposition that reduces the accompanying computational expense, overcoming a disadvantage of Sobol SA.

Nevertheless, Breiman(2001) suggested the adoption of permutation importance in which a feature is permuted so as to break the relationship between it and the model's

output, then evaluated on the random forest out-of-bag samples for which the prediction error is calculated before and after the permutation—averaged over all estimators. This, too, suffers as a variable importance measure, since it overestimates the importance of features that possess interactions. Strobl (2008), using the present framework of permutation importances, thought to define a conditional permutation grid for conditioning on groups of model features.[42] The values of a model's features are randomly permuted for each estimator, using the samples that were generated for the tree-growing process and the binary splitting. Using these binary splits, one can create a grid by bisecting the sample space within which feature values are permuted; the prediction error is similarly computed before and after the permutation in order to assess individual variable contribution to the prediction of the model's output. The conditional permutation importances are computed individually for each estimator and then averaged.[43] Strobl's procedure for estimating variable importances also failed in quantifying variable importances, as it had overestimated the importance of variables that possessed interactions. The former and latter are merely examples in attempting to overcome the artifact of Random Forest feature importances and aforementioned permutation importances proposed by Breiman (2001).The reader is directed to other considerations proposed by authors to mitigate this downfall. [47][14][48][25][13]

2.4 Permutation Variable Importances

In this section, we will first start off discussing recent developments of permutation variable importances and their relationship to Sobol (variance-based) sensitivity indices. We shall, again, consider a model with random variables Y, X, defined as $f(X) = Y$, where for each $i \in \mathbb{N}$ $X_i \in X = [X_{0k}, X_{1k}, \ldots, X_{nm}]$ are independent and identically distributed random variables. As aforementioned, random forest makes a prediction by computing the expectation of the predictions made by a pre-initialized collection of estimators (*see* **§0.2.2**).

Let E be such a collection, whereby $E = \{(X_k,\, Y_k) \mid 0 \leq k \leq m\}$ be a collection of bootstrap training data of the data X, and let $\overline{E}_k := E \setminus E_k$, $E_k \subset E$, represent the ***out-of-bag*** bootstrap samples. After training the estimator with random sample E_m, the out-of-bag prediction error, B_k, of a Random Forest predictive model $\hat{f}(X_k) = Y_k$ is defined as

$$\mathrm{B}_k = \frac{1}{\left|\left|\overline{E}_k\right|\right|} \sum_{(X_j,\, Y_j) \in \overline{E}_k} \left(Y_j - \hat{f}(X_j)\right)^2 \tag{22}$$

where the quantity on the right is simply the mean-squared-error (MSE) and $\left|\left|\cdot\right|\right|$ is the magnitude of the sample. Using MSE as the measure of model uncertainty, it is reasonable to expect that a variable's importance is contingent on the error that amounts after perturbing model features; in other words, higher error implies more importances.

Let B_k^i be the resultant out-of-bag error after evaluating the model, $\hat{f}(X_{(i)})$, where $X_{(i)}$ is an independent, identically distributed random copy of $X_j \in \overline{E}_k$ such that the i^{th}-feature of the parameter data is randomly permuted; this is done so as to break the relationship between it and the model's output, similar to that proposed by Breiman (2001).[2] The expected out-of-bag error for each estimator is defined as

$$\hat{\mathcal{I}}_{\hat{f}}(X_{(i)}) = \frac{1}{K}\sum_{k=1}^{K}\left[\mathrm{B}_k^i - \mathrm{B}_k\right] \tag{23}$$

Zhu et al., (2015) demonstrated that the latter measure converged to the difference between the expected squared error over $X_{\sim i}$ and the squared-error of the model noise; that is,

$$\hat{\mathcal{I}}_{\hat{f}}(X_{(i)}) = \mathbb{E}\big[(Y - f(X_{(i)}))^2\big] - \mathbb{E}\big[(Y - f(X))^2\big]. \tag{24}$$

[2][49] As we are considering a deterministic model, equation **(18)** can be simplified by substituting the equation $f(X) = Y$, yielding

$$\hat{\mathcal{I}}_{\hat{f}}(X_{(i)}) = \mathbb{E}[(f(X) - f(X_{(i)}))^2]. \tag{25}$$

2.5 Sobol & PVi Relationships

We are now in a position to introduce the work of [15], who established the first relationship between Sobol sensitivity indices and permutation variable importance, deserving of the title

Lemma 2.1. *Let $X_i \in X = [X_{0k}, X_{1k}, \ldots, X_{nm}]$ be independent and identically distributed subgroup of random variables, and let $X_{\bar{i}}$ denote the group of variables that does not appear in X_i . Assume that we observe Y and X in the following additive regression model:*

$$\begin{aligned} Y &= f(X) + \varepsilon \\ &= f_i(X_i) + f_{\bar{i}}(X_{\bar{i}}) + \varepsilon, \end{aligned}$$

in which $\mathbb{E}[\varepsilon] = 0$. Then the total-order Sobol-PVi relationship is

$$S_{\tau_i} = \frac{\hat{\mathcal{I}}_{\hat{f}}(X_{(i)})}{2\mathbb{V}(Y)}, \tag{26}$$

*where $\hat{\mathcal{I}}_{\hat{f}}(X_{(i)})$ is from equation **(19)***

Since the latter quantity returns the total importance of feature X_i–over the space $X_{\bar{i}}$–, a natural extension is the importance of a feature $X_{\bar{i}}$ (i.e., $X_{\sim i}$) over the space X_i. With this in mind, [4] established the first-order Sobol-PVi relationship, which we introduce as

Lemma 2.2. *Let $X = [X_0, \ldots, X_i, \ldots, X_n]$ be a random-vector of independent components, let $X_{\sim i}$ be a independent copy of X such that X is randomly permuted except feature X_i, and assume $Y = f(X)$. Then, for each component X_i*

$$S_i = 1 - \frac{\mathcal{I}_{\hat{f}}(X_i)}{2\mathbb{V}(Y)}, \tag{27}$$

where the new quantity

$$\mathcal{I}_{\hat{f}}(X_i) = \mathbb{E}\big[(f(X) - \hat{f}(X_{\sim i}))^2\big]. \tag{28}$$

In a similar fashion, it is sensible to consider grouped variable importances, which I proffer in the form of a proposition

Proposition 2.3. *Suppose $Y = f(X)$ in which $X_{\sim ik} \in X$ is defined as in **Lemma 2.2**. Then the second-order Sobol-PVi relationship, for each combination of components X_{ik},*

$$S_{ik} = 1 - \frac{\mathcal{I}_{\hat{f}}(X_{ik})}{2\mathbb{V}(Y)} - S_i - S_k, \tag{29}$$

*where the quantities S_i and S_k are defined as in equation **(21)**.*

The proof for the latter proposition is in Appendix **C**. It should be clear that the relationships between Sobol and PVi *are not unique*, as PVi employs the same technique as Sobol SA and utilizes an alternate variance measure: mean-squared-error. The mean-

squared-error of a models prediction from its true value can be expressed as the sum of the model's variance, squared bias, and irreducible error. Since **Lemma 2.2** and **Proposition 2.3** considers an unbiased model (i.e., the bias term is zero) with a sufficiently large number of samples, the estimators $\mathcal{I}_{\hat{f}}(X_i)$ and $\mathcal{I}_{\hat{f}}(X_{ik})$ quantify how much the predictions vary between different spaces (e.g., $X_{\sim i}$, etc.) and can decrease or increase but will not converge to the model's variance (i.e., $\mathbb{V}(Y)$). Instead, the estimators yield individual and combinations of model-input variances plus some irreducible error, which is the noise inherent in the data that no perfect model can eliminate.

Proposition 2.3 is a mere extension of [4]'s first-order permutation variable importances, which was developed by considering the "first-order" PVi of a combination of variables (i.e., originally $\mathcal{I}_{\hat{f}}(X_{(ik)})$).

2.6 Test Functions

I tested the latter findings using RandomForestRegressor as an emulator for both the Ishigami function and Rahmstorf model [31], using sea-level data provided by [12] and the National Oceanic and Atmospheric Administration (NOAA) historical temperature data set. The Rahmstorf model is a semi-empirical model, consisting of three parameters: α ($\frac{\text{mm}}{^\circ C}$), T_{eq} ($^\circ C$), and $S_\circ$. The α-parameter functions as the rate at which sea-level changes with respect to changes in global temperature, T_{eq}-parameter is the temperature

at equilibrium (i.e., no change in observed temperature), and $S_\circ$-parameter is simply the initial condition of global mean sea-level, relative to the years 1951-1980. The Rahmstorf model and Ishigami function is given as follows, respectively:

$$\frac{dS}{dt} = \alpha(T(t) - T_{eq}) \tag{30}$$

$$f(\mathbf{x}) = \sin(x_1) + a\sin^2(x_2) + bx_3^4\sin(x_1), \tag{31}$$

where $a = 7$ and $b = 0.1$. Moreover, for the Rahmstorf model, I centered the temperature and sea-level data relative to 1951-1980 by computing the mean corresponding to those years and subtracting it from the former and latter datasets. I used first-order Euler's Method for solving equation **(30)**, using $\alpha = 3.6$, $T_{eq} = -0.5$, and $S_\circ = -118.15\overline{6}$ on the centered dataset (*see* Figure 2).

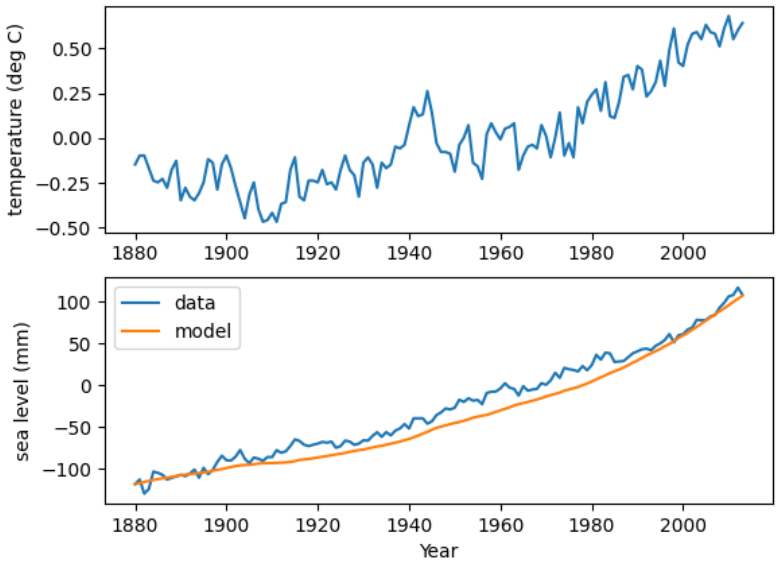

Figure 2: Fitted Rahmstorf model on using NOAA temperature and Church and White (2011) sea-level datasets.

Furthermore, I used random forest regression to model the BRICK-CIAM dataset as a real-world application to demonstrate the generality of PVi for sensitivity analysis. The BRICK-CIAM dataset has 57 features (tables of features are provided in Appendix §**A.3** Tables **13** & **14**), each of which contain sampled statistics of ensemble members corresponding to different U.S. Gulf coast segments; the model's output of interest is the aggregated cost of *protection*, *retreat*, *inundation* (loss of unprotected land), *wetlands*, and *flooding* (i.e., ensemble) for each member associated with distinct U.S. Gulf coast segments.[46]

2.7 Test Function: Rahmstorf Model Experimental Design

In order to test our findings, I assumed that each model-input were independently, uniformly distributed random variables; thus, each parameter of Rahmstorf model had uniform prior distributions. The model-inputs of the Rahmstorf model were scaled, having distributions $\alpha \sim \mathbb{U}(0, 5)$, $T_{eq} \sim \mathbb{U}(-1, 2)$, and $S_\circ \sim \mathbb{U}(-125, 100)$ whose ranges were determined based on empirical observations—a posteriori; I generated the uniform distributions using 60,000 Latin-hypercube samples (i.e., stratified, low-discrepancy, random sampling). For each sample, the Rahmstorf model-output returned the last element of its prediction, using the available data then loaded into a separate dataset of the same length. I partitioned the dataset into training and test datasets and performed bayes-search 4-fold cross-validation using the training data (*see* Appendix **A.1** Table **11**).

2.8 Test Function: Ishigami Function Experimental Design

In the case of the Ishigami function, I adopted the assumptions in [20], in which the prior distributions of the model variables were $x_i \sim \mathbb{U}(-\pi, \pi)$ (i.e., i.i.d uniform priors) and were generated using a $60{,}000$ Latin-Hypercube samples. I partitioned the dataset into training and testing and performed bayes-search 4-fold cross-validation using the training dataset.

2.9 BRICK-CIAM: Random Forest Regression Model

No assumptions were made germane to the prior distribution of each feature, as BRICK-CIAM is of the form of a dataset whose model is random forest regression itself; therefore, hyperparameter tuning was the only performed task. Therefrom, the BRICK-CIAM dataset was partitioned into two subsets: 80% for training the model and 20% for testing the model's performance. A bayes-search, 4-fold cross-validation was performed, and the best hyperparameters are reported in Appendix §**A.3** Table **15**.

Both grid-search and bayes-search, 4-fold cross-validation used a negative mean-squared-error (MSE) scoring method—for both test functions and the BRICK-CIAM RF model (i.e., the highest MSE corresponds to the best hyperparameters). All code, pertaining PVi estimation, were implemented in Python programming language and can

be found in Appendix **D**.

3 Results: Sobol Indices & PVi Estimation and Comparison

3.1 Rahmstorf Model Test Function

Using the best hyperparameters in Appendix **A.1**, Tables **1** & **2** and Tables **3** & **4** represent the first and second-order PVi and Sobol measures, respectively; like combinations denote first-order importances, otherwise they represent second-order importances. There's a strong similarity between the importances reported in Tables **1** & **2** and Tables **3** & **4**. For example, the second-order Sobol indices and PVi estimates for $(\alpha, S_\circ)$ and $(T_{eq}, S_\circ)$ are negative. In principle, partial variance measures shouldn't be negative, since variance is the squared-deviation from the mean. Recall from **§2.1** that the sum of the first, second, etc. order importances for variance-based measures must be 1. The sum of the first and second-order PVi is 0.9951, whereas the first and second-order Sobol indices sum to 0.993. This may be a consequence of rounding to four significant figures or (in the case of PVi) the squared bias that accompanies computing the mean-squared-error. Notwithstanding their minor differences, there is clear agreement between the two importance measures: $S_\circ$ contribute little to no variance of the model's output.

	P_i	5th	95th	Quantile Difference
α	0.0844	0.0714	0.0966	0.0252
T_{eq}	0.6701	0.6650	0.6751	0.0101
$S_\circ$	0.0381	0.0285	0.0481	0.0196

Table 1: First-Order Permutation Variable Importances and corresponding quantile difference. 95% Confidence Interval using 1,000 bootstraps.

		P_{ik}	5th	95th	Quantile Difference
α	T_{eq}	0.2124	0.2123	0.2123	0.0000
	$S_\circ$	-0.0050	-0.0013	-0.0013	0.0000
T_{eq}	$S_\circ$	-0.0049	-0.0032	-0.0032	0.0000

Table 2: Second-Order Permutation Variable Importances and corresponding quantile difference. 60,000 Latin-Hypercube Samples were used and 1,000 bootstraps for 95% confidence interval.

	S_i	5th	95th	Quantile Difference
α	0.0849	0.0743	0.0947	0.0204
T_{eq}	0.6738	0.6666	0.6804	0.0138
$S_\circ$	0.0399	0.0318	0.0479	0.0162

Table 3: First-Order Sobol sensitivity measures for the Rahmstorf model with corresponding 1,000 bootstrap 95% Confidence Intervals.

		S_{ik}	5th	95th	Quantile Difference
α	T_{eq}	0.2082	0.9650	0.9690	0.0039
	$S_\circ$	-0.0069	0.1073	0.1285	0.0212
T_{eq}	$S_\circ$	-0.0069	0.7002	0.7138	0.0136

Table 4: Second-Order Sobol Indices for the Rahmstorf Model with corresponding 95% Confidence Intervals computed using 1,000 bootstrap samples.

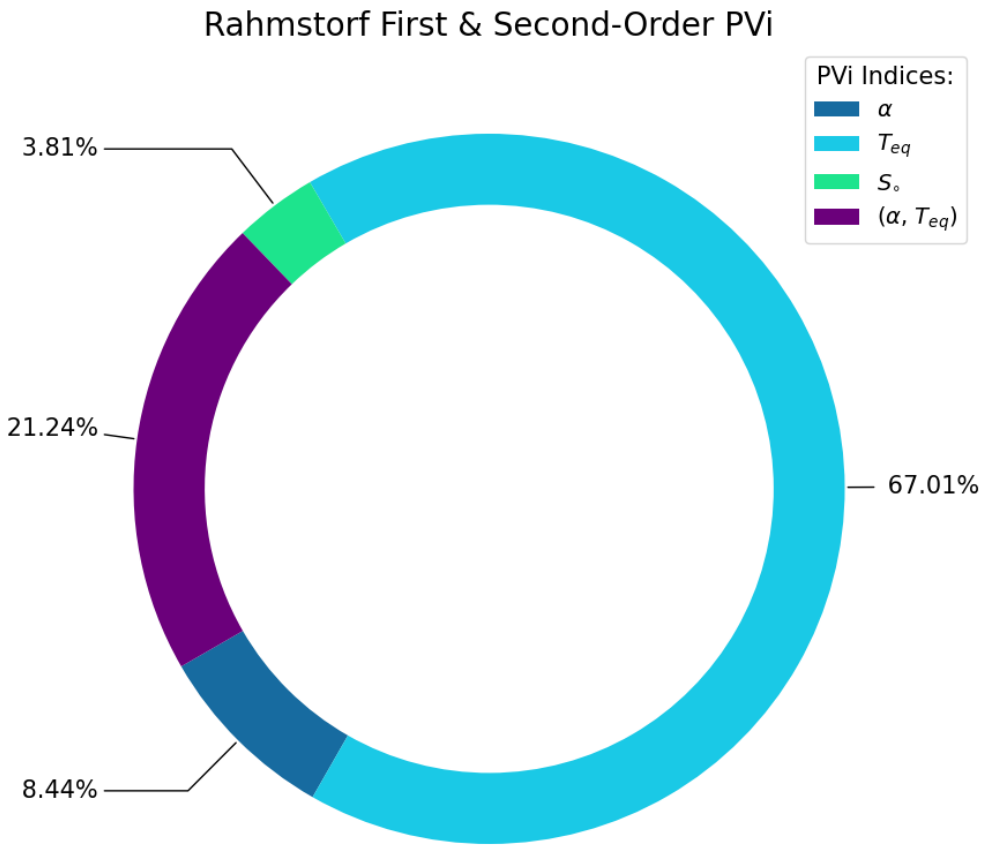

Figure 3: Rahmstorf Model First & Second-Order PVi Pie chart

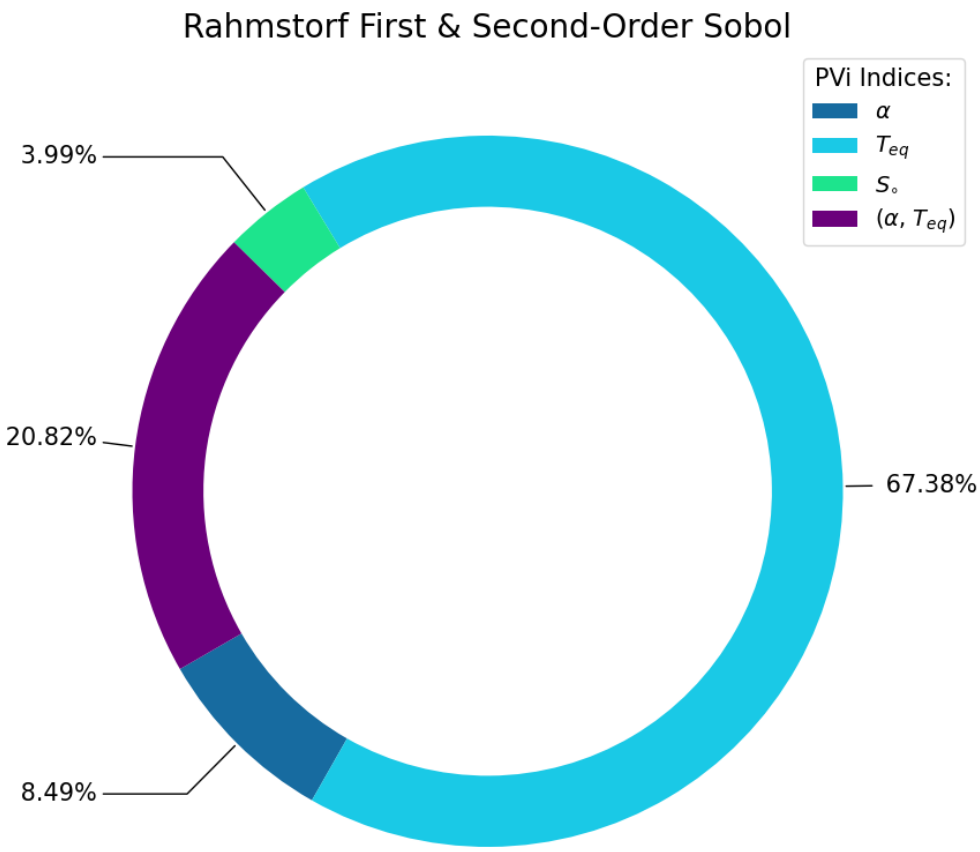

Figure 4: Sobol First & Second-Order Indices

3.2 Ishigami Test Function

I used the best hyperparameters reported in Appendix §**A.2** Table **12** for first and second-order PVi and Sobol indices. For the Ishigami function, the first and second-order PVi were underestimated (or over, in the case of the first-order importance of x_1 and x_3) potentially due to the nonlinearity (or, simply, the nature) of the function, despite the apparent convergence in Table **6**. The sum of the permutation variable importances is exactly 0.9878, whereas the sum of the Sobol indices is 1.0014; so, Sobol over-approximated the contribution of features (x_1, x_2), (x_2, x_3), and (x_1, x_3). The over or under-estimation of parameter importances could be a consequence of an insufficient amount of samples supplied to random forest and Sobol SA or the accumulated squared bias accompanying the mean-squared-error.[7]

	P_i	5th	95th	Quantile Difference
x_1	0.3228	0.3153	0.3298	0.0145
x_2	0.4500	0.4432	0.4577	0.0145
x_3	0.0123	-0.0015	0.0264	0.0279

Table 5: Ishigami First-Order PVi and corresponding quantile difference. 95% Confidence Interval were computed using 1,000 bootstrap samples.

		P_{ik}	5th	95th	Quantile Difference
x_1	x_2	-0.0123	-0.0126	-0.0126	0.0000
	x_3	0.2316	0.2320	0.2320	0.0000
x_2	x_3	-0.0166	-0.0142	-0.0142	0.0000

Table 6: Second-Order PVi and corresponding quantile difference. 95% Confidence Intervals were computed using 1,000 bootstrap samples.

[7]Note that the PVi were computed using the mean-squared-error between the model's true value and random forest's prediction.

	S_i	5th	95th	Quantile Difference
x_1	0.3096	0.3004	0.3189	0.0186
x_2	0.4402	0.4310	0.4493	0.0182
x_3	-0.0037	-0.0159	0.0093	0.0252

Table 7: Ishigami First-Order Sobol sensitivity measures with corresponding 1,000 bootstrap sampled 95% Confidence Interval Widths.

		S_{ik}	5th	95th	Quantile Difference
x_1	x_2	0.0013	0.7439	0.7582	0.0144
	x_3	0.2527	0.5503	0.5660	0.0157
x_2	x_3	0.0013	0.4258	0.4478	0.0221

Table 8: Second-Order Sobol Indices with corresponding 95% Confidence Intervals computed using 1,000 bootstrap samples

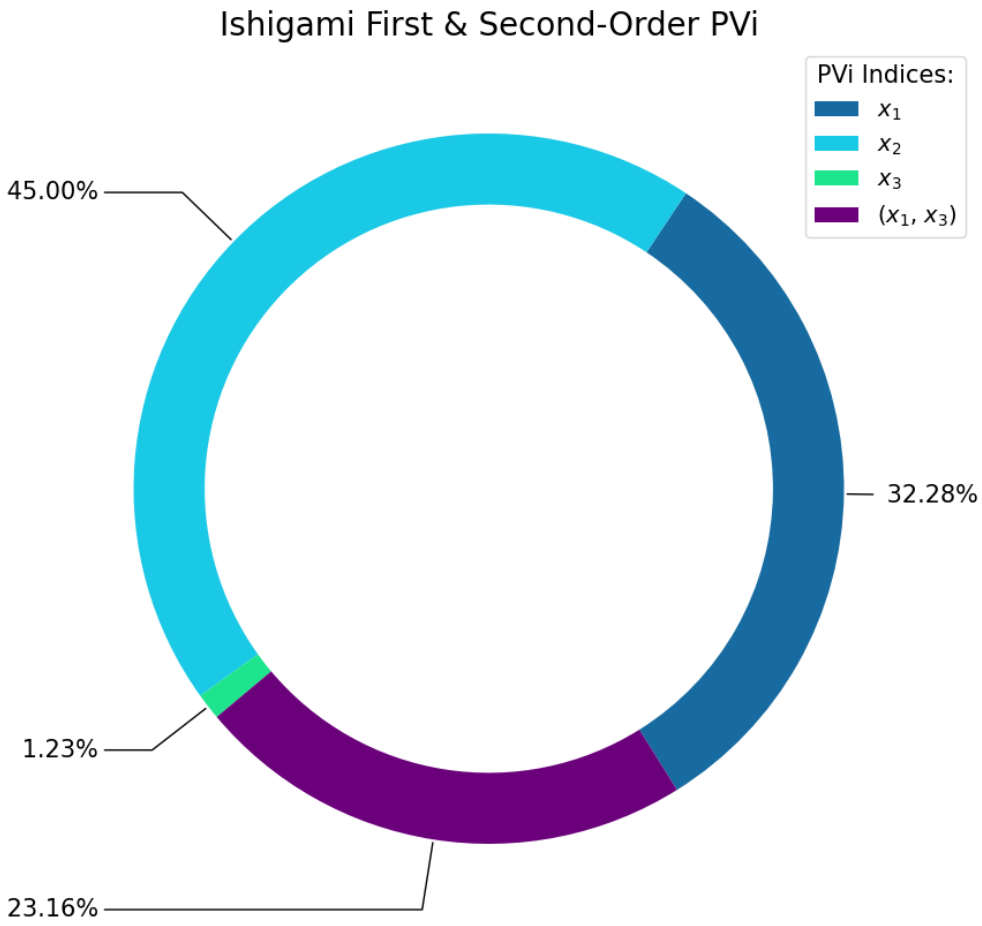

Figure 5: Ishigami First & Second-Order PVi Pie chart

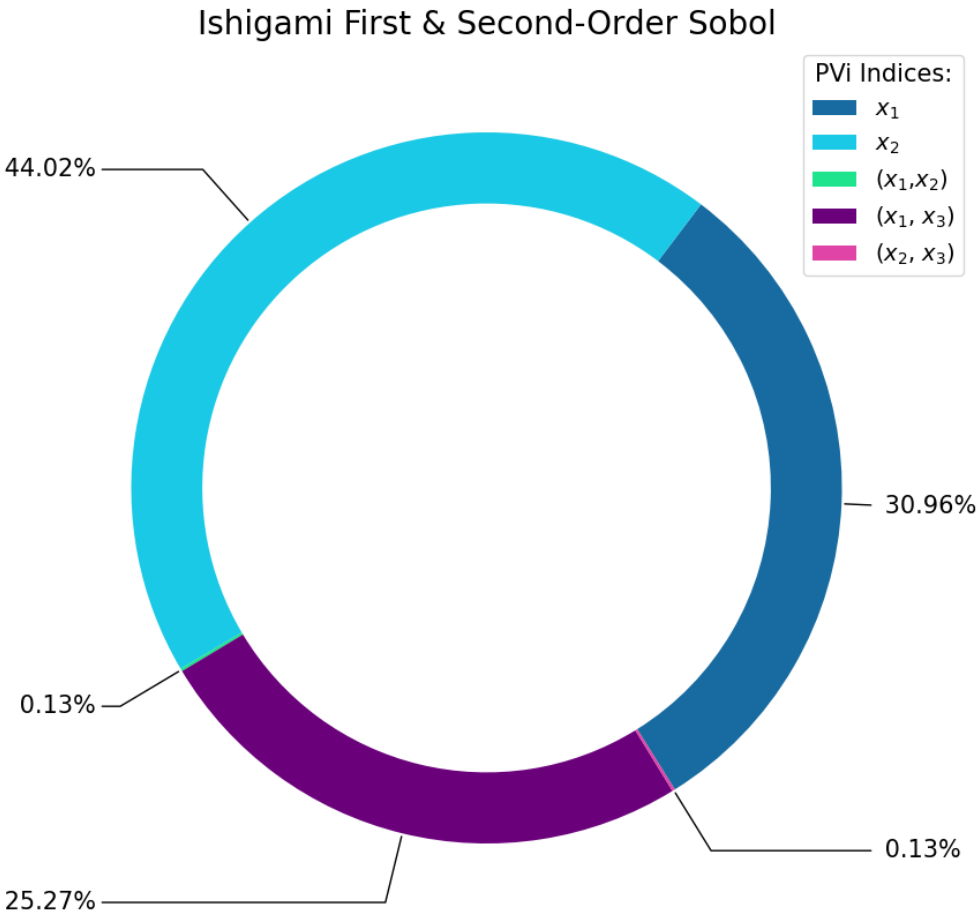

Figure 6: Sobol First & Second-Order Indices

3.3 BRICK-CIAM Model

I estimated the first and second-order PVi using the best hyperparameters in Appendix §**A.3** Table **15**. The BRICK-CIAM dataset is parameter-less and lacks a *design of experiment*, so the Sobol first and second-order indices couldn't be computed. Since the BRICK-CIAM dataset has 57 features, the first and second-order PVi estimates are reported in the Appendix §**B**. The second-order PVi estimates were all negative with low confidence intervals widths. As a result, this I've determined that there were an insufficient number of samples to attain the desired level of accuracy. Figure **8** illustration the total-order importances of categorized features.

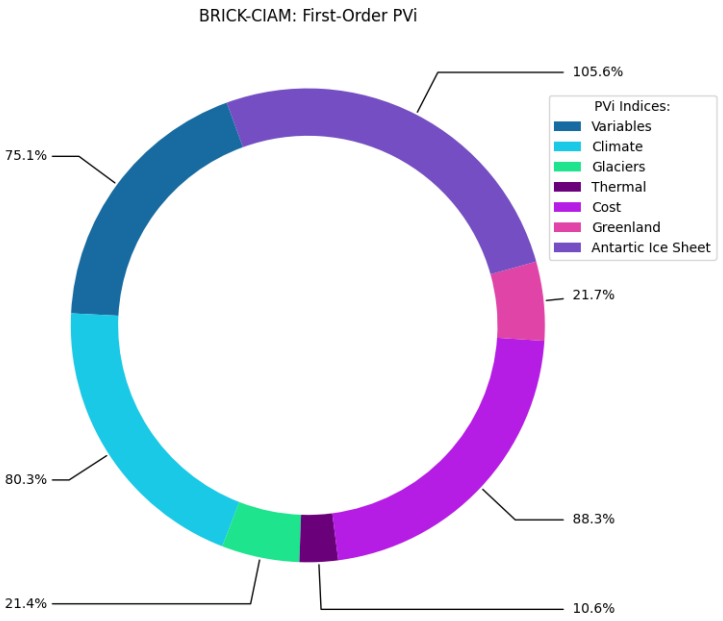

Figure 7: BRICK-CIAM Categorized First-Order PVi

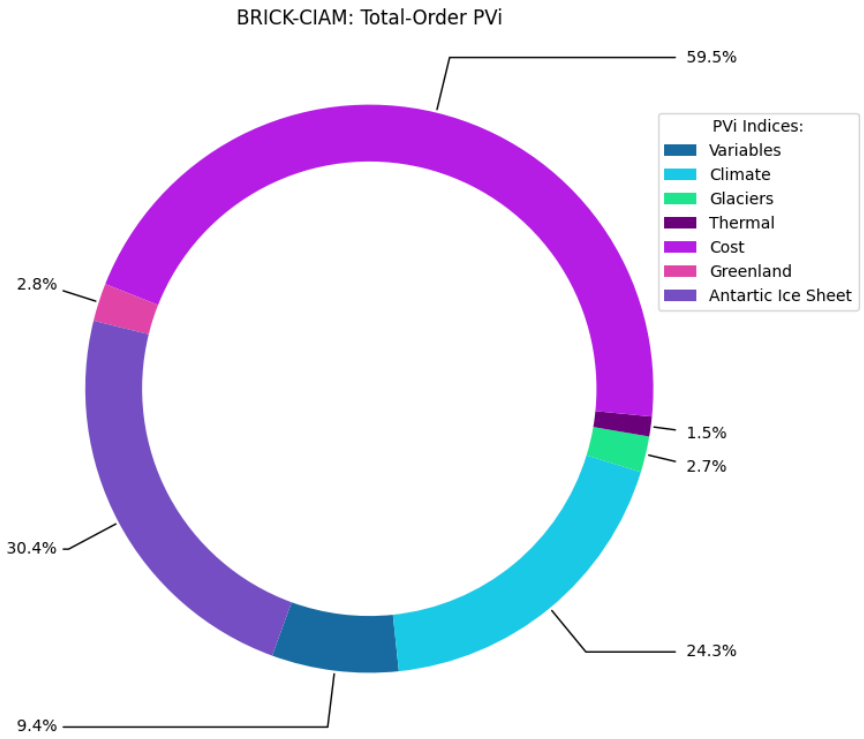

Figure 8: BRICK-CIAM (Summed) Total-Order Importances of Categorized Features

3.4 BRICK-CIAM: Identifying Important Features

As a result of the absence of second-order PVi, I sought to improve the predictive performance of the model by removing features that contributed little to the variance of the model's output. Using the BRICK-CIAM dataset, I appended a feature, introducing "white-noise", to the dataset to accentuate the importances of other features present within the dataset. This feature (with moniker "SignalVar") was Gaussian with a mean of zero and variance of one and was trained (in conjunction with the other features of the dataset) using the best hyperparameters reported in Appendix **A**.**3** Table **15**. A feature was deemed important if the total-order PVi contribution to the model's output variance were $\geqslant 0.0068$ (*see* Table **9**). I reduced the number of features from 57 to 22 total features. Thereafter, I used these model features as the new dataset and re-performed

Parameters	P_τ
dvbm s	0.5397
antarctic temp threshold	0.1946
climate sensitivity	0.1644
rf scale aerosol	0.0276
antarctic lambda	0.0246
movefactor s	0.0195
thermal alpha	0.0085
greenland beta	0.0077
antarctic precip0	0.0075
wvpdl s	0.0074
greenland b	0.0072
Q10	0.0071
temperature 0	0.0070
antarctic mu	0.0070
heat diffusivity	0.0069
rho gmsl	0.0068
greenland alpha	0.0068
ocean heat 0	0.0068
rho greenland	0.0068
antarctic runoff height0	0.0067
std. antarctic	0.0067
anto alpha	0.0067

Table 9: BRICK-CIAM: Feature Selection

bayes-search, 4-fold cross-validation with the same specifications as the previous partitioning (i.e., 80/20 split for training and testing, respectively); the best hyperparameters are reported in Appendix §**A.3** Table **16**.

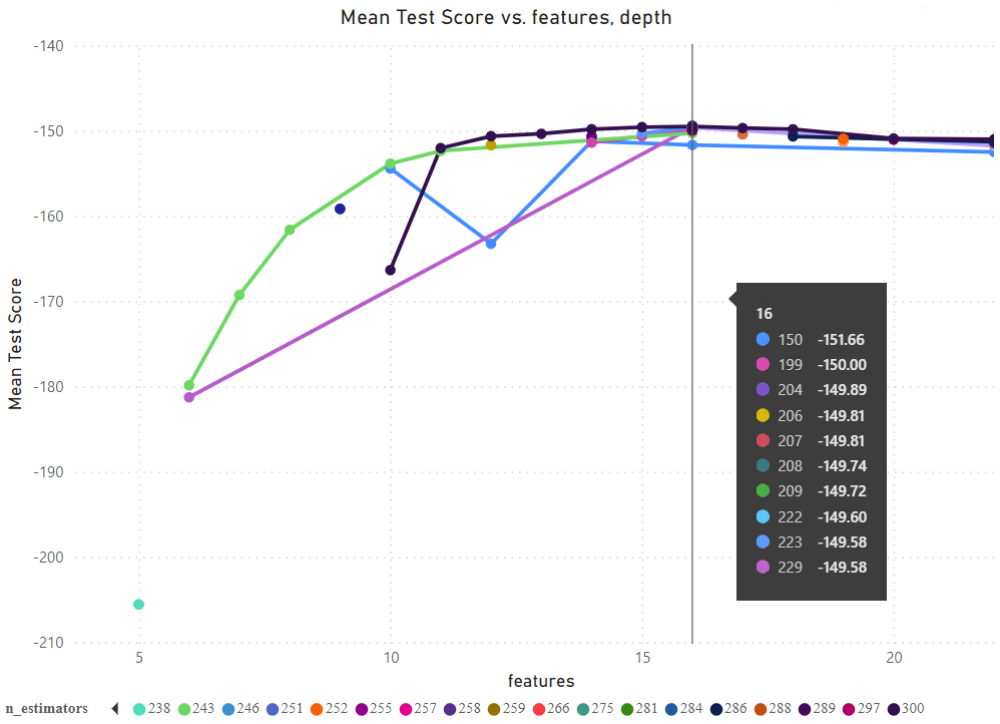

Figure 9: BRICK-CIAM Bayes-Search Cross-Validation using the newly selected features.

After feature selection, the new first and total-order PVi measures are[8]

[8]Second-Order PVi are reported in Appendix §**B** Table **20**.

	P_i	5th	95th	Quantile Difference
dvbm s	0.5809	0.5697	0.5927	0.0230
climate sensitivity	0.2377	0.2173	0.2583	0.0410
antarctic temp threshold	0.2084	0.1872	0.2276	0.0403
rf scale aerosol	0.1300	0.1067	0.1531	0.0464
movefactor s	0.0712	0.0481	0.0963	0.0482
antarctic lambda	0.0689	0.0442	0.0942	0.0500
temperature 0	0.0530	0.0275	0.0787	0.0512
greenland beta	0.0520	0.0260	0.0767	0.0507
antarctic precip0	0.0517	0.0254	0.0770	0.0516
ocean heat 0	0.0515	0.0263	0.0758	0.0495
rho greenland	0.0504	0.0273	0.0757	0.0485
wvpdl s	0.0504	0.0253	0.0755	0.0502
rho gmsl	0.0501	0.0260	0.0747	0.0487
antarctic mu	0.0496	0.0262	0.0744	0.0482
std. antarctic	0.0495	0.0230	0.0739	0.0509
anto alpha	0.0494	0.0258	0.0746	0.0488
antarctic runoff height0	0.0493	0.0241	0.0731	0.0490
greenland b	0.0492	0.0229	0.0746	0.0517
greenland alpha	0.0489	0.0248	0.0751	0.0503
heat diffusivity	0.0488	0.0237	0.0739	0.0502
Q10	0.0482	0.0243	0.0736	0.0493
thermal alpha	0.0463	0.0208	0.0729	0.0521

	P_τ	5th	95th	Quantile Difference
dvbm s	0.5397	0.5244	0.5543	0.0300
antarctic temp threshold	0.1946	0.1878	0.2014	0.0136
climate sensitivity	0.1644	0.1588	0.1702	0.0113
rf scale aerosol	0.0276	0.0266	0.0285	0.0019
antarctic lambda	0.0246	0.0231	0.0262	0.0031
movefactor s	0.0195	0.0188	0.0202	0.0013
thermal alpha	0.0085	0.0082	0.0088	0.0006
greenland beta	0.0077	0.0074	0.0080	0.0006
antarctic precip0	0.0075	0.0073	0.0078	0.0005
wvpdl s	0.0074	0.0071	0.0077	0.0006
greenland b	0.0072	0.0069	0.0074	0.0005
Q10	0.0071	0.0068	0.0074	0.0005
temperature 0	0.0070	0.0067	0.0073	0.0005
antarctic mu	0.0070	0.0067	0.0073	0.0006
heat diffusivity	0.0069	0.0066	0.0072	0.0006
rho gmsl	0.0068	0.0065	0.0070	0.0005
greenland alpha	0.0068	0.0065	0.0070	0.0005
ocean heat 0	0.0068	0.0065	0.0070	0.0005
rho greenland	0.0068	0.0065	0.0070	0.0005
antarctic runoff height0	0.0067	0.0065	0.0070	0.0005
std. antarctic	0.0067	0.0064	0.0069	0.0005
anto alpha	0.0067	0.0064	0.0069	0.0005

Table 10: BRICK-CIAM First & Total-Order PVi with 95% Confidence Intervals computed using 1,000 bootstrap samples.

4 Conclusion

I have demonstrated the relationship between second-order permutation variable importances and Sobol second-order indices, as well as their generality and adaptability—for both deterministic and data-driven models—for estimating the importance of a individual and combinations of features. The BRICK-CIAM random forest regression model revealed an apparent shortcoming of Sobol sensitivity analysis: when the proposed model is data-driven (e.g., a machine learning models), performing Sobol SA would be a meticulous task, as it would require generating prior, distributional data—most of which likely absent within the original dataset—alongside a prior distribution for the model's output which may or may not align with the generated, *observational* statistics of our model's input; second-order permutation variable importances effectively remedied that. By merely breaking the relationship between individual and combinations of model inputs, one can determine how variations of model inputs—and resulting prediction—affect the expected-squared-deviation from the model's true value.

The second-order permutation variable importances is an extension of [4]'s first-order PVi for estimating second-order PVi. It was designed as an auxiliary, computationally efficient methodology for ascertaining major contributors of the model input to the variance of the model's output.

5 Acknowledgement

First, I thank Dr. Tony Wong for his guidance during this thesis process and equipping me with the skills and techniques of statistical inference.

I thank Kelly Feke and Prasanna Ponfilio Rodrigues for running simulations to generate the Gulf Coast coastal impacts (i.e., BRICK-CIAM) dataset that I analyzed in §3.3.

And, finally, I thank my parents, Kevin & Daina Burke, and former professors, Dr. Christopher Leary & Dr. Olympia Towsley/Nicodemi for always believing in my ability, for their guidance, and emphasis on adhering to mathematical rigor.

A Appendix: Hyperparameter Tuning

A.1 Rahmstorf Model

To verify that RF Rahmstorf model had an adequate number of samples, Figure **10** illustrates gradual increments in sample size with respect to the mean-squared-error. Furthermore, I partitioned the dataset into two subsets in preparation of random forest hyper-parameter tuning. Specifically, the data was split into 80% for training and 20% for testing the model. I performed a bayes-search, 4-fold cross validation using the training data so that the model was trained on 60%, validated on 20%, and left the remaining 20%—untouched—for testing.

Table 11: Rahmstorf Hyperparameter Tuning 1: Top 10 hyperparameters for RandomForestRegressor

rank test score	max. depth	max. features	num. estimators	mean test score	std. test score
1	18	2	500	-29.5664	0.6932
2	30	2	500	-29.6813	0.5597
2	28	2	500	-29.6813	0.5597
2	29	2	500	-29.6813	0.5597
5	27	2	500	-29.6838	0.5605
6	25	2	500	-29.6858	0.5852
7	26	2	500	-29.6867	0.5571
8	23	2	500	-29.7030	0.5989
9	24	2	500	-29.7248	0.5317
10	27	2	479	-29.7364	0.5469

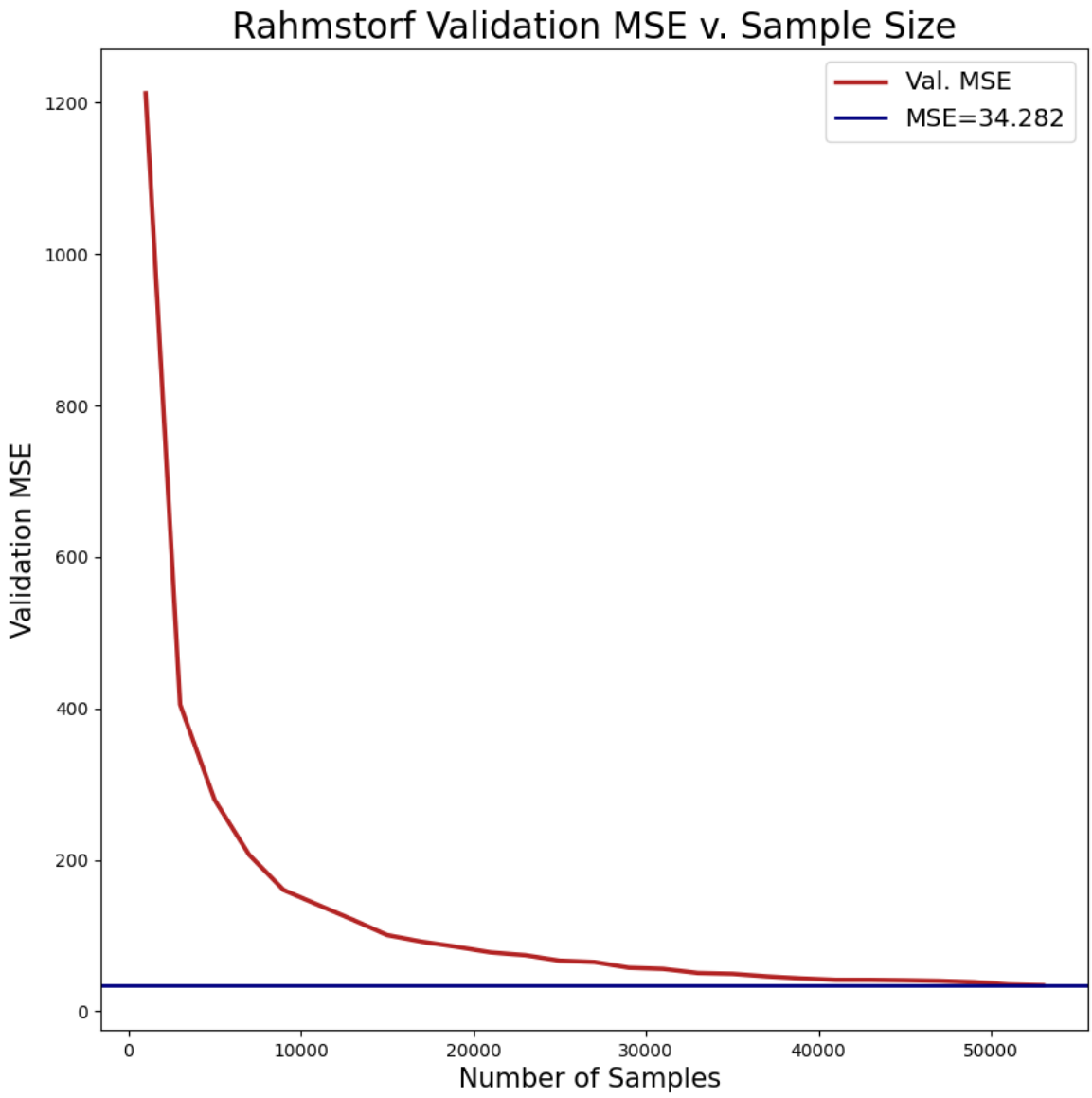

Figure 10: 4-Fold Cross-Validation

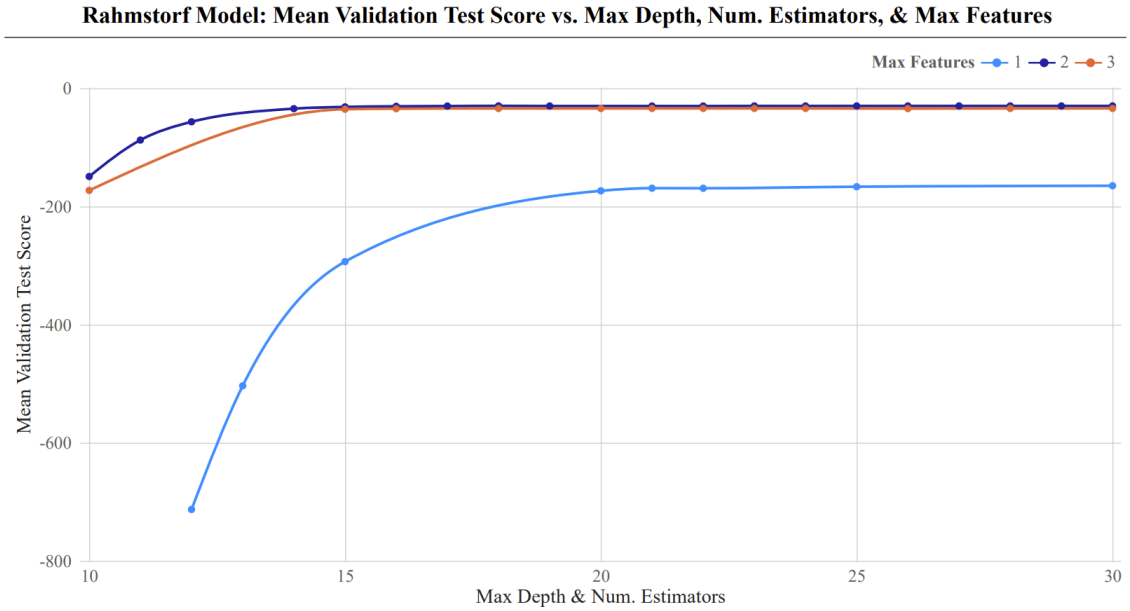

Figure 11: Bayes-Search, 4-Fold Cross-Validation Hyperparameter Tuning Paths

A.2 Ishigami Function

I verified that the RF Ishigami model had a sufficient sample size by gradually incrementing the sample size and reporting the model's performance using the test data; Figure **12** illustrates the decrease in the model's mean-squared-error. Furthermore, I partitioned the dataset into two subsets for a bayes-search, 4-fold cross-validation whereby 80% of the training data was supplied and further partitioned into 60% for training the RF model and 20% for validation; the remaining 20% was left untouched for testing the model's predictive performance (*see* Table **12**).

Table 12: Ishigami Hyperparameter Tuning: Top 10 hyperparameters for RandomForestRegressor

rank test score	max. depth	max. features	num. estimators	mean test score	std. test score
1	33	2	500	-0.0657	0.0021
2	29	2	500	-0.0658	0.0022
3	39	2	500	-0.0658	0.0024
4	27	2	500	-0.0658	0.0022
5	38	2	500	-0.0658	0.0021
6	37	2	500	-0.0658	0.0023
7	40	2	500	-0.0659	0.0022
8	29	2	398	-0.0659	0.0023
9	32	2	324	-0.0660	0.0019
10	36	2	500	-0.0660	0.0021

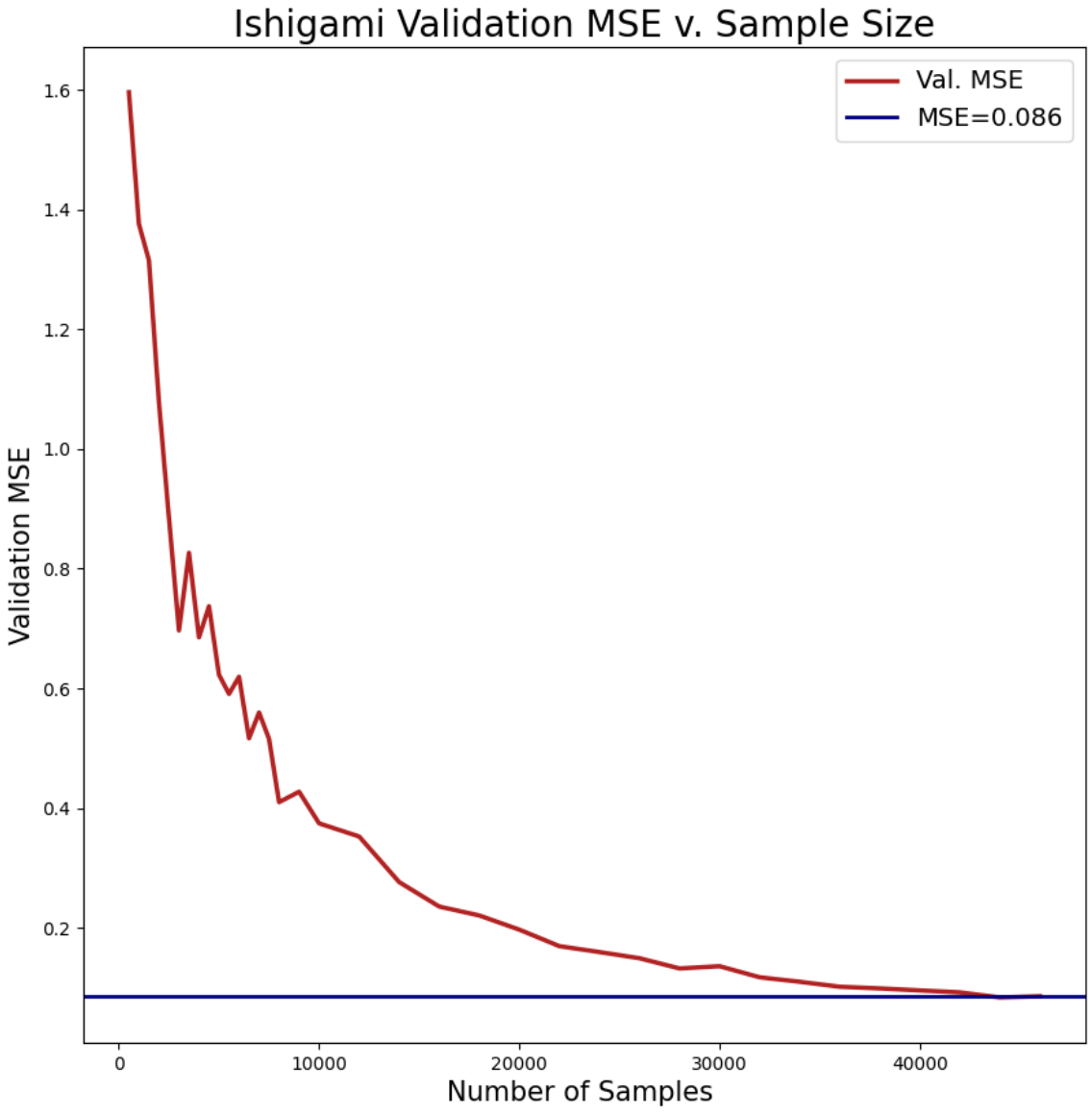

Figure 12: 4-Fold Cross-Validation

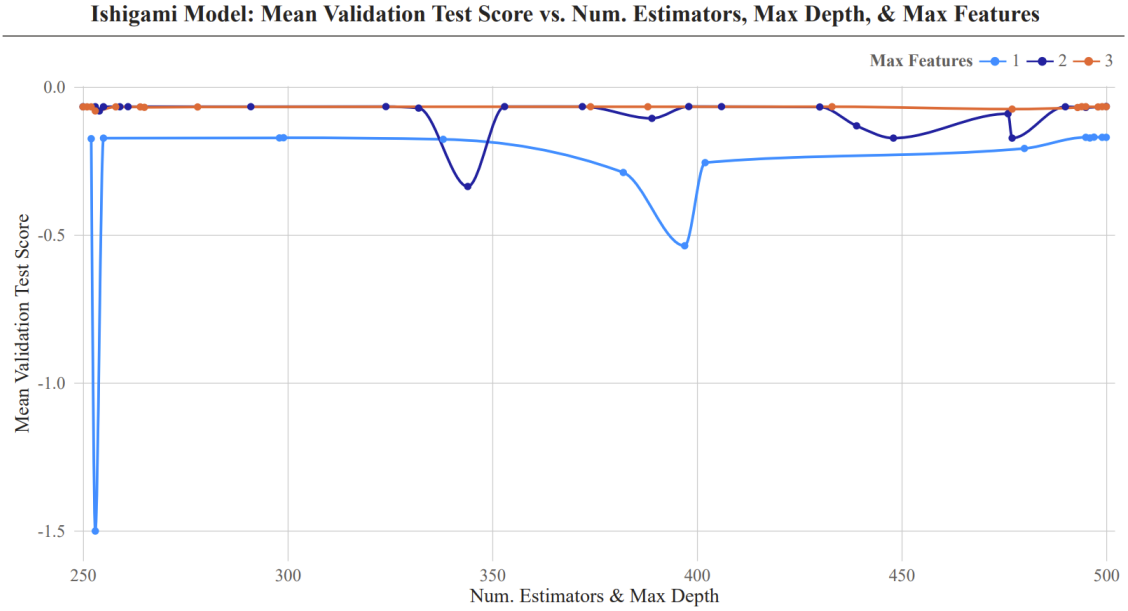

Figure 13: Bayes-Search, 4-Fold Cross-Validation Hyperparameter Tuning Paths

A.3 BRICK-CIAM Random Forest Regression Model

Model Parameters	Parameter Description
std. temp	standard deviation of IPCC AR1 residuals for temperature (°C)
std. ocean heat	—-"—-ocean heat uptake (10^{22}J)
std. glaciers	—-"—-glacial sea-level contribution (m)
std greenland	—-"—-Greenland sea-level contribution (m)
std antarctic	—-"—-Antarctic sea-level contribution (m)
std gmsl	—-"—-global mean sea-level rise (m)
sigma whitenoise co2	White noise standard deviation for CO_2
rho temperature	Lag-1 (i.e., single time step) autocorrelation coefficient for temperature
rho ocean heat	—-"—-ocean heat uptake
rho glaciers	—-"—-glacial sea-level contribution
rho greenland	—-"—-Greenland sea-level contribution
rho antarctic	—-"—-Antarctic sea-level contribution
rho gmsl	—-"—-global mean sea-level rise
alpha0 CO2	Measure of autocorrelation memory for CO_2
CO2 0	Initial CO_2 concentration (ppm)
N2O 0	—-"—-N_2O concentration (ppb)
temperature 0	—-"—-temperature anomaly (°C
ocean heat 0	—-"—-ocean heat uptake (10^{22}J)
thermal s0	Initial condition for thermal expansion sea-level contribution (m SLE)
greenland v0	—-"—-Greenland ice sheet sea-level contribution (m SLE)
glaciers v0	—-"—-glacier/ice cap volume (m sea level equivalent (SLE))
glaciers s0	—-"—-glacier/ice cap sea-level contribution (m SLE)
antarctic s0	—-"—-Antarctic sea-level contribution (m SLE)
Q10	Respiration sensitivity
CO2 fertilization	Carbon fertilization factor
CO2 diffusivity	Ocean carbon diffusivity ($\frac{m}{y}$)
heat diffusivity	Ocean vertical diffusivity ($\frac{cm^2}{s}$)
rf scale aerosol	Aerosol radiative forcing scaling factor
climate sensitivity	Equilibrium climate sensitivity to doubling CO_2 (°C)
thermal alpha	Global ocean-averaged thermal expansion coefficient ($\frac{kg}{m^3 °C}$)
greenland a	Equilibrium Greenland ice sheet volume temperature sensitivity ($\frac{m}{°C}$)
greenland b	Equilibrium Greenland ice sheet volume (m SLE)
greenland alpha	Greenland ice sheet response timescale temperature sensitivity ($\frac{1}{°C y}$)
greenland beta	—-"—-($\frac{1}{y}$)
glaciers beta0	Initial glacial/ice cap mass balance temperature sensitivity ($\frac{m}{y °C}$)
glaciers n	Exponent for glacier/ice cap area-volume scaling
anto alpha	Antarctic ocean temperature sensitivity to global temperature ($\frac{°C}{°C}$)
anto beta	Equilibrium Antarctic Ocean Temperature (°C)
antarctic gamma	Power for relation of Antarctic ice flow speed to water depth
antarctic alpha	Effect of ocean subsurface temperature on ice flux partition parameter
antarctic mu	Parabolic ice surface parameter ($m^{0.5}$)
antarctic nu	Antarctic runoff and precipitation proportionality constant ($\frac{1}{(my)^{0.5}}$)
antarctic precip0	Annual Antarctic precipitation for surface temperature 0°C (m)
antarctic kappa	Coefficient, dependency of Antarctic precipitation on temperature ($\frac{1}{°C}$)
antarctic flow0	Antarctic ice flow at grounding line proportionality constant ($\frac{m}{y}$)
antarctic runoff height0	Antarctic runoff line height at 0°C surface temperature (m)
antarctic c	Antarctic runoff line height temperature sensitivity ($\frac{m}{°C}$)
antarctic bed height0	Undisturbed bed height at Antarctic continent center (m)
antarctic slope	Slope of Antarctic ice sheet bed before ice loading
antarctic lambda	Fast Antarctic dynamic disintegration rate (m)
antarctic temp threshold	Fast Antarctic dynamic disintegration trigger temperature (°C)

Table 13: BRICK-CIAM Feature Description (No.1)

Model Parameter	Parameter Description	Units	Central Value	Distribution	Source
dvbm s	Benchmark land Value	Million 2010 ($\frac{\$}{km^2}$)	5.376	$\mathcal{N}[\mu = 5.376, \sigma = 2.688][0, \infty)$	FUND; Darwin et al. (1995)
movefactor s	Relocation cost as fraction of income	$(0, 1)$	1	$\mathcal{N}[\mu = 1, \sigma = 1][0.5, 3]$	Anthoff & Tol (2014) and Dias (2016)
vslel s	Elasticity of value of statistical life (VSL)	Unitless	0.47	$\mathcal{N}[\mu = 0.47, \sigma = 0.15][0, \infty)$	FUND; Viscussi & Aldy (2003)
vslmult s	VSL multiplier on US GDP	Unitless	200	$\mathcal{N}[\mu = 200, \sigma = 100][0, \infty)$	FUND; Cline (1992)
wvel s	Income elasticity of wetland value	Unitless	1.16	$\mathcal{N}[\mu = 1.16, \sigma = 0.46][0, \infty)$	Brander et al. (2006)
wvpdl s	Population density elasticity of wetland value	Unitless	0.47	$\mathcal{N}[\mu = 0.47, \sigma = 0.12][0, 1]$	Brander et al. (2006)

Table 14: BRICK-CIAM Feature Description (No.2)

I partitioned the dataset into two subset: 80% for training the model and 20% for

testing the predictive performance of the model. I performed bayes-search, 4-fold cross-validation using the training dataset, with hypereparameters `n_estimators: [390,425]` and `max_features: [20,57]` (*see* Table **15**) The scoring method used on the validation set was the negative mean-squared-error (i.e., the highest MSE is the best).

rank test score	max. features	num. estimators	mean test score	std test score
1	39	425	-169.173677	5.294332
2	39	406	-169.204068	5.223874
3	39	405	-169.210433	5.221445
4	39	407	-169.212157	5.214966
5	45	407	-169.212598	5.380295
6	45	425	-169.235793	5.379624
7	45	422	-169.237222	5.404831
8	45	390	-169.247635	5.448536
9	39	398	-169.268758	5.234401
10	39	390	-169.283428	5.281460

Table 15: BRICK-CIAM Random Forest Regression Bayes-Search Hyperparameter Tuning

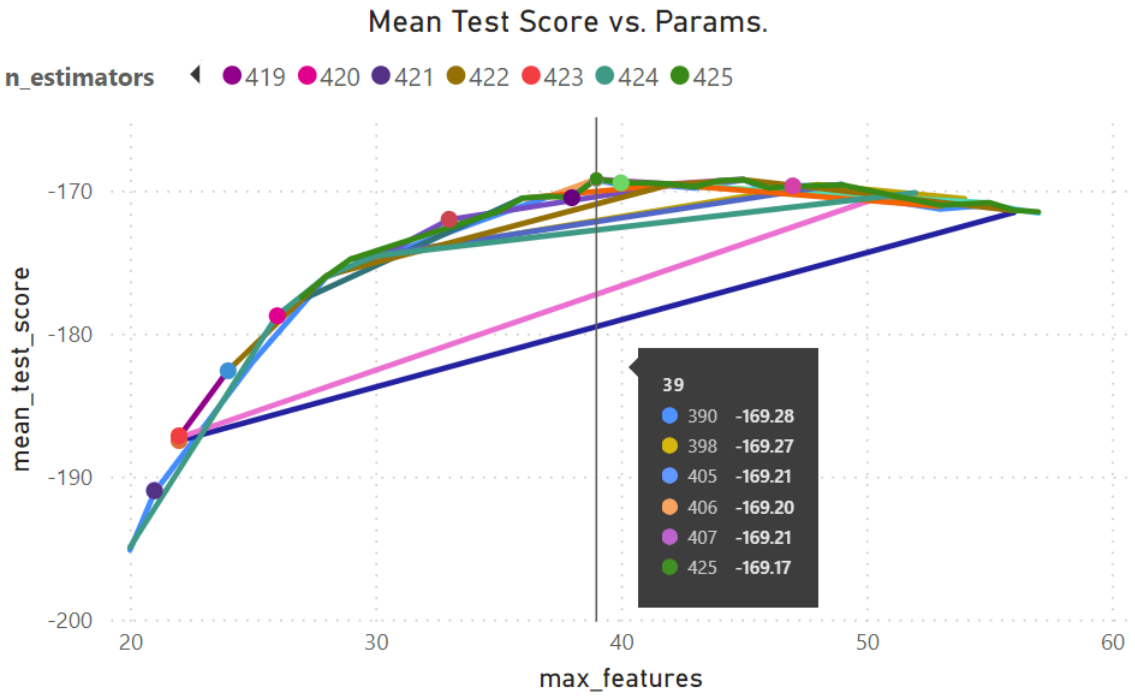

Figure 14: BRICK-CIAM Bayes-Search Hyperparameter Tuning. The y-axis is the cross-validation mean-squared-error and the x-axis is the number of features.

rank test score	max. depth	max. features	num. estimators	mean test score	std test score
1	20	16	284	-149.423725	4.296720
2	20	16	281	-149.429844	4.354963
3	20	16	300	-149.474921	4.311994
4	21	16	300	-149.529267	4.084101
5	18	16	275	-149.532301	4.214400
6	25	15	300	-149.559345	3.460700
7	20	16	251	-149.570159	4.402607
8	20	16	235	-149.573129	4.529260
9	20	16	223	-149.581624	4.459024
10	20	16	229	-149.583658	4.449387

Table 16: BRICK-CIAM Bayes-Search CV Best Hyperparameters for newly selected features

B BRICK-CIAM: First & Second-Order PVi estimates

Parameters	P_i	5th	95th	Quantile Difference
std. temp	0.052	0.026	0.077	0.051
std. ocean heat	0.052	0.029	0.078	0.048
std. glaciers	0.052	0.029	0.077	0.048
std. greenland	0.052	0.027	0.078	0.051
std. antarctic	0.053	0.028	0.077	0.050
std. gmsl	0.053	0.029	0.078	0.049
sigma whitenoise co2	0.052	0.028	0.077	0.050
rho temperature	0.052	0.026	0.078	0.051
rho ocean heat	0.052	0.028	0.076	0.048
rho glaciers	0.052	0.026	0.078	0.052
rho greenland	0.054	0.028	0.079	0.051
rho antarctic	0.053	0.029	0.079	0.050
rho gmsl	0.053	0.027	0.078	0.051
alpha0 CO2	0.052	0.028	0.077	0.049
CO2 0	0.052	0.029	0.077	0.048
N2O 0	0.052	0.026	0.077	0.052
temperature 0	0.056	0.031	0.079	0.049
ocean heat 0	0.055	0.029	0.082	0.053
thermal s0	0.052	0.028	0.077	0.049
greenland v0	0.052	0.027	0.078	0.050
glaciers v0	0.052	0.028	0.078	0.050
glaciers s0	0.052	0.027	0.077	0.050
antarctic s0	0.052	0.028	0.077	0.049
Q10	0.052	0.027	0.077	0.049
CO2 fertilization	0.052	0.027	0.077	0.049
CO2 diffusivity	0.052	0.026	0.078	0.052
heat diffusivity	0.053	0.028	0.078	0.050
rf scale aerosol	0.134	0.111	0.158	0.047
climate sensitivity	0.240	0.220	0.261	0.041
thermal alpha	0.051	0.027	0.077	0.050
greenland a	0.053	0.026	0.077	0.050
greenland b	0.053	0.027	0.079	0.051
greenland alpha	0.052	0.026	0.078	0.052
greenland beta	0.055	0.027	0.079	0.052
glaciers beta0	0.053	0.028	0.079	0.051
glaciers n	0.052	0.028	0.076	0.048
anto alpha	0.053	0.028	0.077	0.049
anto beta	0.052	0.027	0.077	0.050
antarctic gamma	0.052	0.028	0.077	0.049
antarctic alpha	0.052	0.028	0.077	0.049
antarctic mu	0.053	0.027	0.078	0.051
antarctic nu	0.052	0.028	0.077	0.049
antarctic precip0	0.055	0.028	0.079	0.051
antarctic kappa	0.052	0.027	0.078	0.052
antarctic flow0	0.052	0.028	0.079	0.051
antarctic runoff height0	0.053	0.027	0.078	0.051
antarctic c	0.053	0.028	0.079	0.051
antarctic bed height0	0.052	0.028	0.078	0.051
antarctic slope	0.053	0.028	0.078	0.049
antarctic lambda	0.072	0.048	0.095	0.047
antarctic temp threshold	0.213	0.191	0.234	0.043
dvbm s	0.587	0.576	0.598	0.023
movefactor s	0.074	0.050	0.099	0.049
vslel s	0.052	0.027	0.076	0.049
vslmult s	0.052	0.025	0.076	0.051
wvel s	0.052	0.027	0.077	0.050
wvpdl s	0.054	0.027	0.077	0.050

Table 17: BRICK-CIAM First-Order PVi and 95% Confidence Intervals using 1,000 bootstrap samples.

Parameters	P_τ	5th	95th	Quantile Difference
std. temp	0.007	0.006	0.007	0.001
std. ocean heat	0.007	0.006	0.007	0.001
std. glaciers	0.007	0.006	0.007	0.001
std. greenland	0.007	0.006	0.007	0.000
std. antarctic	0.007	0.007	0.007	0.001
std. gmsl	0.007	0.006	0.007	0.001
sigma whitenoise co2	0.007	0.006	0.007	0.000
rho temperature	0.007	0.006	0.007	0.001
rho ocean heat	0.007	0.006	0.007	0.000
rho glaciers	0.007	0.006	0.007	0.000
rho greenland	0.007	0.007	0.007	0.001
rho antarctic	0.007	0.006	0.007	0.001
rho gmsl	0.007	0.007	0.007	0.001
alpha0 CO2	0.007	0.006	0.007	0.001
CO2 0	0.007	0.006	0.007	0.001
N2O 0	0.007	0.006	0.007	0.001
temperature 0	0.007	0.007	0.008	0.001
ocean heat 0	0.007	0.007	0.007	0.001
thermal s0	0.007	0.006	0.007	0.001
greenland v0	0.007	0.006	0.007	0.001
glaciers v0	0.007	0.006	0.007	0.001
glaciers s0	0.007	0.006	0.007	0.001
antarctic s0	0.007	0.006	0.007	0.001
Q10	0.007	0.007	0.007	0.001
CO2 fertilization	0.007	0.006	0.007	0.001
CO2 diffusivity	0.007	0.007	0.007	0.001
heat diffusivity	0.007	0.007	0.007	0.001
rf scale aerosol	0.029	0.028	0.030	0.002
climate sensitivity	0.158	0.152	0.162	0.010
thermal alpha	0.008	0.008	0.008	0.001
greenland a	0.007	0.007	0.007	0.001
greenland b	0.007	0.007	0.007	0.001
greenland alpha	0.007	0.007	0.007	0.001
greenland beta	0.008	0.007	0.008	0.001
glaciers beta0	0.007	0.006	0.007	0.001
glaciers n	0.007	0.006	0.007	0.001
anto alpha	0.007	0.007	0.007	0.001
anto beta	0.007	0.006	0.007	0.001
antarctic gamma	0.007	0.006	0.007	0.001
antarctic alpha	0.007	0.006	0.007	0.000
antarctic mu	0.007	0.007	0.007	0.001
antarctic nu	0.007	0.006	0.007	0.001
antarctic precip0	0.007	0.007	0.008	0.001
antarctic kappa	0.007	0.006	0.007	0.001
antarctic flow0	0.007	0.006	0.007	0.001
antarctic runoff height0	0.007	0.007	0.007	0.001
antarctic c	0.007	0.006	0.007	0.001
antarctic bed height0	0.007	0.006	0.007	0.001
antarctic slope	0.007	0.006	0.007	0.000
antarctic lambda	0.023	0.022	0.025	0.003
antarctic temp threshold	0.193	0.186	0.199	0.013
dvbm s	0.540	0.525	0.554	0.029
movefactor s	0.018	0.017	0.019	0.001
vslel s	0.007	0.006	0.007	0.001
vslmult s	0.007	0.006	0.007	0.001
wvel s	0.007	0.006	0.007	0.001
wvpdl s	0.007	0.007	0.008	0.001

Table 18: BRICK-CIAM Mode Total-Order PVi with 5^{th} and 95^{th}-percentile Confidence Intervals using 1,000 bootstrap samples

		Categorized Second-Order PVi
Variables	Climate	-0.0512
	Glaciers	-0.0512
	Thermals	-0.0512
	Cost	-0.0512
	Greenland	-0.0512
	Antarctic	-0.0512
Climate	Glaciers	-0.0512
	Thermals	-0.0512
	Cost	-0.0512
	Greenland	-0.0512
	Antarctic	-0.0512
Glaciers	Thermals	-0.0512
	Cost	-0.0512
	Greenland	-0.0512
	Antarctic	-0.0512
Thermals	Cost	-0.0513
	Greenland	-0.0512
	Antarctic	-0.0512
Cost	Greenland	-0.0512
	Antarctic	-0.0512
Greenland	Antarctic	-0.0512

Category	(Averaged) Categorized P_i
Variables	0.0525
Climate	0.0799
Glaciers	0.0523
Thermals	0.0517
Cost	0.1453
Greenland	0.0530
Antarctic	0.0638

		5th	95th	Quantile Difference
Variables	Climate	-0.0531	-0.0495	0.0036
	Glaciers	-0.0536	-0.0487	0.0049
	Thermals	-0.0538	-0.0484	0.0054
	Cost	-0.0534	-0.0495	0.0040
	Greenland	-0.0536	-0.0487	0.0050
	Antarctic	-0.0526	-0.0498	0.0028
Climate	Glaciers	-0.0546	-0.0483	0.0062
	Thermals	-0.0553	-0.0482	0.0071
	Cost	-0.0546	-0.0497	0.0049
	Greenland	-0.0547	-0.0483	0.0064
	Antarctic	-0.0531	-0.0498	0.0033
Glaciers	Thermals	-0.0590	-0.0433	0.0157
	Cost	-0.0556	-0.0481	0.0075
	Greenland	-0.0569	-0.0452	0.0116
	Antarctic	-0.0533	-0.0489	0.0044
Thermals	Cost	-0.0569	-0.0475	0.0095
	Greenland	-0.0587	-0.0431	0.0155
	Antarctic	-0.0537	-0.0486	0.0052
Cost	Greenland	-0.0557	-0.0481	0.0076
	Antarctic	-0.0536	-0.0499	0.0038
Greenland	Antarctic	-0.0534	-0.0489	0.0045

Category	(Averaged) Categorized P_τ
Variables	0.0067
Climate	0.0241
Glaciers	0.0067
Thermals	0.0074
Cost	0.0975
Greenland	0.0071
Antarctic	0.0195

Table 19: BRICK-CIAM: Categorized Permutation Variable Importance. 95% Confidence Intervals were computed using 1,000 bootstrap samples.

	std. antarctic	rho greenland	rho gmsl	temperature 0	ocean heat 0	Q10	heat diffusivity	rf scale aerosol	climate sensitivity	thermal alpha	greenland b	greenland beta	anto alpha	antarctic mu	antarctic precipl0	antarctic lambda	antarctic temp threshold	dvbm s	movefactor s	wvpdl s
std. antarctic	-0.0198	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0466	-0.0478	-0.0466	-0.0456	-0.0456	-0.0456	-0.0456	-0.0466	-0.0466	-0.0478	-0.0198	-0.0466	-0.0456
rho greenland	-0.0456	-0.0419	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0466	-0.0498	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0466	-0.0478	-0.0198	-0.0466	-0.0456
rho gmsl	-0.0456	-0.0456	-0.0499	-0.0456	-0.0456	-0.0456	-0.0456	-0.0466	-0.0478	-0.0466	-0.0456	-0.0456	-0.0456	-0.0456	-0.0466	-0.0466	-0.0478	-0.0198	-0.0466	-0.0456
temperature 0	-0.0456	-0.0456	-0.0456	-0.0500	-0.0456	-0.0456	-0.0456	-0.0466	-0.0469	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0466	-0.0478	-0.0198	-0.0466	-0.0456
ocean heat 0	-0.0456	-0.0456	-0.0456	-0.0456	-0.0501	-0.0456	-0.0456	-0.0466	-0.0469	-0.0466	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0466	-0.0478	-0.0198	-0.0466	-0.0456
Q10	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0478	-0.0456	-0.0466	-0.0478	-0.0466	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0466	-0.0478	-0.0198	-0.0466	-0.0456
heat diffusivity	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0498	-0.0466	-0.0466	-0.0466	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0466	-0.0478	-0.0198	-0.0466	-0.0456
rf scale aerosol	-0.0466	-0.0466	-0.0466	-0.0478	-0.0478	-0.0466	-0.0466	-0.1222	-0.1078	-0.0478	-0.0466	-0.0478	-0.0466	-0.0466	-0.0478	-0.0466	-0.0478	-0.0198	-0.0466	-0.0478
climate sensitivity	-0.0478	-0.0478	-0.0478	-0.0469	-0.0469	-0.0478	-0.0466	-0.1078	-0.2306	-0.0478	-0.0466	-0.0498	-0.0466	-0.0466	-0.0478	-0.0466	-0.0478	-0.0198	-0.0466	-0.0456
thermal alpha	-0.0466	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0478	-0.0478	-0.0478	-0.0466	-0.0466	-0.0466	-0.0466	-0.0466	-0.0466	-0.0478	-0.0198	-0.0466	-0.0456
greenland b	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0478	-0.0478	-0.0466	-0.0478	-0.0466	-0.0466	-0.0466	-0.0466	-0.0466	-0.0478	-0.0198	-0.0466	-0.0456
greenland beta	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0478	-0.0478	-0.0466	-0.0466	-0.0478	-0.0466	-0.0466	-0.0466	-0.0466	-0.0478	-0.0198	-0.0466	-0.0456
anto alpha	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0478	-0.0478	-0.0466	-0.0466	-0.0466	-0.0478	-0.0466	-0.0466	-0.0466	-0.0478	-0.0198	-0.0466	-0.0456
antarctic mu	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0478	-0.0478	-0.0466	-0.0466	-0.0466	-0.0466	-0.0478	-0.0466	-0.0466	-0.0478	-0.0198	-0.0466	-0.0456
antarctic precipl0	-0.0466	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0478	-0.0478	-0.0466	-0.0466	-0.0466	-0.0466	-0.0466	-0.0478	-0.0466	-0.0478	-0.0198	-0.0466	-0.0456
antarctic lambda	-0.0466	-0.0466	-0.0466	-0.0466	-0.0466	-0.0466	-0.0466	-0.0478	-0.0622	-0.0466	-0.0466	-0.0466	-0.0466	-0.0466	-0.0466	-0.0478	-0.0478	-0.0198	-0.0466	-0.0466
antarctic temp threshold	-0.0478	-0.0478	-0.0478	-0.0478	-0.0478	-0.0478	-0.0498	-0.0655	-0.0490	-0.0490	-0.0478	-0.0490	-0.0478	-0.0478	-0.0478	-0.0478	-0.0478	-0.0198	-0.0506	-0.0498
dvbm s	-0.0198	-0.0498	-0.0498	-0.0478	-0.0478	-0.0498	-0.0478	-0.0658	-0.0969	-0.0490	-0.0498	-0.0490	-0.0498	-0.0498	-0.0498	-0.0498	-0.0498	-0.0198	-0.0577	-0.0478
movefactor s	-0.0466	-0.0466	-0.0466	-0.0466	-0.0466	-0.0466	-0.0466	-0.1051	-0.1054	-0.1047	-0.0490	-0.0490	-0.0466	-0.0466	-0.0466	-0.0466	-0.0506	-0.0577	-0.0710	-0.0490
wvpdl s	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0478	-0.0498	-0.0490	-0.0456	-0.0456	-0.0456	-0.0456	-0.0456	-0.0478	-0.0498	-0.0478	-0.0490	-0.0490

Table 20: BRICK-CIAM Second-Order PVi (Full Matrix).

B.1 BRICK-CIAM: Categories of Features

Variables
std. temp
std. ocean heat
std. glaciers
std. greenland
std. antarctic
std. gmsl
sigma whitenoise co2
rho temperature
rho ocean heat
rho glaciers
rho greenland
rho antarctic
rho gmsl
alpha0 CO2

Antarctic
antarctic gamma
antarctic alpha
antarctic mu
antarctic nu
antarctic precip0
antarctic kappa
antarctic flow0
antarctic runoff height0
antarctic c
antarctic bed height0
antarctic slope
antarctic lambda
antarctic temp threshold
antarctic s0
anto alpha
anto beta

Climate
CO2 0
N2O 0
temperature 0
ocean heat 0
Q10
CO2 fertilization
CO2 diffusivity
heat diffusivity
rf scale aerosol
climate sensitivity

Glaciers
glaciers v0
glaciers s0
glaciers beta0
glaciers n

Thermals
thermal s0
thermal alpha

Cost
dvbm s
movefactor s
vslel s
vslmult s
wvel s
wvpdl s

Greenland
greenland a
greenland b
greenland alpha
greenland beta

C Appendix: Formal Proof of Proposition 2.3

Proof: Assume $f(X) = Y$ in which $X = [X_0, \ldots, X_i, \ldots, X_n]$ is a random vector of independent components and $X_{\sim ik}$ is an independent copy of X such that X is randomly permuted except for variables X_i and X_k. Then

$$\begin{aligned}\mathcal{I}_{\hat{f}}(X_{ik}) &= \mathbb{E}\big[\big(f(X) - \hat{f}(X_{\sim ik})\big)^2\big] \\ &= \mathbb{E}\big[\big(f(X) + \mathbb{E}[Y] - \mathbb{E}[Y] - \hat{f}(X_{\sim ik})\big)^2\big] \\ &= \mathbb{E}\big[\big(f(X) - \mathbb{E}[Y]\big)^2\big] + \mathbb{E}\big[\big(\hat{f}(X_{\sim ik}) - \mathbb{E}[Y]\big)^2\big] - 2\mathbb{E}\big[(f(X) - \mathbb{E}[Y])(\hat{f}(X_{\sim ik}) - \mathbb{E}[Y])\big]\end{aligned}$$

By independence, $\mathbb{E}\big[\big(\hat{f}(X_{\sim ik}) - \mathbb{E}[Y]\big)^2\big] = \mathbb{V}[Y]$ (i.e., the variance of Y), therefrom,

$$\mathcal{I}_{\hat{f}}(X_{ik}) = 2\mathbb{V}[Y] - 2\mathbb{E}\big[(f(X) - \mathbb{E}[Y])(\hat{f}(X_{\sim ik}) - \mathbb{E}[Y])\big]. \tag{32}$$

The second term on the right-hand side of equation **(32)** can be rewritten as an integral, with probability density function p, such that

$$\begin{aligned}= \int_{\mathbb{R}}\int_{\mathbb{R}}\Bigg[\int_{\mathbb{R}^{n-2}}(f(\mathbf{x}) - \mathbb{E}[Y])\prod_{g=1,\, g\neq i,k}^{n} p_{X_g}(x_g)\mathrm{dx_g}\Bigg] \\ \cdot\Bigg[\int_{\mathbb{R}^{n-2}}(\hat{f}(\mathbf{x}_{\sim ik}) - \mathbb{E}[Y])\prod_{h=1,\, h\neq i,k}^{n} p_{X_h}(x'_h)\mathrm{dx_h}\Bigg]p_{X_i,X_k}(x_{i,k})\mathrm{dx_i}\mathrm{dx_k}\end{aligned} \tag{33}$$

$$= \int_{\mathbb{R}} \int_{R} \left[\int_{\mathbb{R}^{n-2}} (f(\mathbf{x}) - \mathbb{E}[Y]) \prod_{h=1,\, h \neq i,k}^{n} p_{X_h}(x_h) \mathrm{dx_h} \right]^2 p_{X_i, X_k}(x_{i,k}) \mathrm{dx_i} \mathrm{dx_k} \tag{34}$$

$$= \mathbb{V}[\mathbb{E}[Y \mid X_i, X_k]] \tag{35}$$

Hence,

$$\mathcal{I}_{\hat{f}}(X_{ik}) = 2\mathbb{V}[Y] - 2\mathbb{V}[\mathbb{E}[Y \mid X_i, X_k]] \tag{36}$$

$$1 - \frac{\mathcal{I}_{\hat{f}}(X_{ik})}{2\mathbb{V}[Y]} = \frac{\mathbb{V}[\mathbb{E}[Y \mid X_i, X_k]]}{\mathbb{V}[Y]} \tag{37}$$

Let $V_{ik} = \mathbb{V}[\mathbb{E}[Y \mid X_i, X_k]]$ as in **§2.1**. Then, inserting identities of **Lemma 2.2** for each variable, one obtains

$$1 - \frac{\mathcal{I}_{\hat{f}}(X_{ik})}{2\mathbb{V}[Y]} + S_i - S_k + S_k - S_i = \frac{V_{ik}}{\mathbb{V}[Y]} \tag{38}$$

$$\implies 1 - \frac{\mathcal{I}_{\hat{f}}(X_{ik})}{2\mathbb{V}[Y]} - S_i - S_k = \frac{V_{ik}}{\mathbb{V}[Y]} - S_i - S_k \tag{39}$$

$$\implies 1 - \frac{\mathcal{I}_{\hat{f}}(X_{ik})}{2\mathbb{V}[Y]} - S_i - S_k = S_{ik} \tag{40}$$

as desired. ■

D Code Schematics

Algorithm 1 First-Order PVi(data: $\mathbb{R}^{m \times n}$, y_pred: $\mathbb{R}^m$, f: callable) $\to \mathbb{R}^n$

Require: data is a matrix of shape $m \times n$, y_pred is a vector of length m, f is a callable function

Ensure: Returns vector S of length n

$rows, columns \leftarrow$ shape of data

$V_y \leftarrow$ variance of y_pred

$S \leftarrow$ zeros vector of length $columns$

for i **in** $[0, 1, \ldots,$ columns $- 1]$ **do**

    $d \leftarrow$ copy of data

    $k \leftarrow d[:, i]$

    $d \leftarrow \text{roll}(d, 1, \text{axis} = 0)$

    $d[:, i] \leftarrow k$

    $S[i] \leftarrow 1 - \frac{\text{mean}(\text{square}(y\_pred - f(d)))}{2 \times V_y}$

end for

return S

Algorithm 2 Second-Order PVi2(data: $\mathbb{R}^{m \times n}$, y_pred: $\mathbb{R}^m$, f: callable, *, S_I: $\mathbb{R}^n$, n_boots: $\mathbb{N}$, alpha: $\mathbb{R}$) $\rightarrow$ (tuple[$\mathbb{R}^{n \times n}$, DataFrame])

Require: data is a matrix of shape $m \times n$, y_pred is a vector of length m, f is a callable function, S_I is a vector of length n, n_boots is an integer, $alpha$ is a real number

Ensure: Returns a tuple containing matrix S of shape $n \times n$ and a Dataframe CI

$rows, columns \leftarrow$ shape of data
$n \leftarrow n\_boots$
$CI \leftarrow$ empty dictionary
$S \leftarrow$ zeros matrix of shape $n \times n$
$V_y \leftarrow$ variance of y_pred
$S\_CI \leftarrow$ empty dictionary
for i **in** $[0, 1, \ldots,$ columns $- 1]$ **do**
    for j **in** $[0, 1, \ldots,$ columns $- 1]$ **do**
        $d \leftarrow$ copy of data
        $k \leftarrow d[:, [i, j]]$
        $d \leftarrow \text{roll}(d, 1, \text{axis} = 0)$
        $d[:, [i, j]] \leftarrow k$
        $S[i, j] \leftarrow 1 - \frac{\text{mean}(\text{square}(y\_pred - f(d)))}{2 \times V_y} - S_I[i] - S_I[j]$
        $S\_CI[(i, j)] \leftarrow 1 - \text{bootstrapp}(y\_pred, d, n\_boots)/(2 \times V_y) - S_I[i] - S_I[j]$
        $p0, p1 \leftarrow \text{quantile}(S\_CI[(i, j)], [\alpha, 1 - \alpha])$
        $CI[(i, j)] \leftarrow [p0, p1, p1 - p0]$
    end for
end for
$CI \leftarrow \text{DataFrame}(CI.values(), \text{index} = \text{CI.keys()}, \text{columns} = ['5th', '95th', 'QuantileDifference'])$
CI.index.name $\leftarrow' Interactions'$
CI.columns.name $\leftarrow' ConfidenceIntervalDifference'$
return S, CI

References

[1] Anestis Antoniadis Lambert-Lacroix, *Random forests for global sensitivity analysis: A selective review*, Reliability Engineering & System Safety **206** (2021), 107312, DOI 10.1016/j.ress.2020.107312 (en).

[2] Nazih Benoumechiara, *Treatment of dependency in sensitivity analysis for industrial reliability*, 2019 (en).

[3] ———, *Treatment of dependency in sensitivity analysis for industrial reliability*, 2019 (en).

[4] ———, *Treatment of dependency in sensitivity analysis for industrial reliability*, 2019 (en).

[5] Leo and Friedman Breiman Jerome H. and Olshen, *Classification and Regression Trees*, Chapman and Hall/CRC, 1984.

[6] ———, *Classification and Regression Trees*, Chapman and Hall/CRC, 1984.

[7] ———, *Classification and Regression Trees*, Chapman and Hall/CRC, 1984.

[8] Leo Breiman, *Random Forests*, Machine Learning **45** (July 26, 2023), no. 1, 5–32, DOI 10.1023/A:1010933404324 (en).

[9] ———, *Random Forests*, Machine Learning **45** (July 26, 2023), no. 1, 11–12, DOI 10.1023/A:1010933404324 (en).

[10] ———, *Random Forests*, Machine Learning **45** (July 26, 2023), no. 1, 23–24, DOI 10.1023/A:1010933404324 (en).

[11] *The California housing dataset — Scikit-learn course.*

[12] John A. and White Church Neil J., *Sea-Level Rise from the Late 19th to the Early 21st Century*, Surveys in Geophysics **32** (2011), no. 4, 585–602, DOI 10.1007/s10712-011-9119-1 (en).

[13] Alex and Kapelner Goldstein Adam and Bleich, *Peeking Inside the Black Box: Visualizing Statistical Learning with Plots of Individual Conditional Expectation*, posted on 2014, DOI 10.48550/arXiv.1309.6392.

[14] Brandon M. and Boehmke Greenwell Bradley C. and McCarthy, *A Simple and Effective Model-Based Variable Importance Measure*, posted on 2018, DOI 10.48550/arXiv.1805.04755.

[15] Baptiste and Michel Gregorutti Bertrand and Saint-Pierre, *Grouped variable importance with random forests and application to multiple functional data analysis*, Computational Statistics & Data Analysis **90** (2015), DOI 10.1016/j.csda.2015.04.002 (en).

[16] ———, *Correlation and variable importance in random forests*, Statistics and Computing **27** (2017), 659–678, DOI 10.1007/s11222-016-9646-1 (en).

[17] Fabrice and Janon Gamboa Alexandre and Klein, *Sensitivity analysis for multidimensional and functional outputs*, Vol. 8, Institute of Mathematical Statistics and Bernoulli Society, 2014.

[18] Alana and Wong Hough Tony E., *Analysis of the evolution of parametric drivers of high-end sea-level hazards*, Advances in Statistical Climatology, Meteorology and Oceanography **8** (2022), no. 1, 117–134, DOI 10.5194/ascmo-8-117-2022 (en).

[19] Intergovernmental Panel On Climate Change (Ipcc), *Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*, 1st ed., Cambridge University Press, 2023 (en).

[20] T. and Homma Ishigami T., *An importance quantification technique in uncertainty analysis for computer models*, posted on 1991, 398–403, DOI 10.1109/ISUMA.1990.151285 (en).

[21] Hemant Ishwaran, *Variable importance in binary regression trees and forests*, Electronic Journal of Statistics **1** (2007), 519–537, DOI 10.1214/07-EJS039.

[22] Marc C. and O'Hagan Kennedy Anthony, *Bayesian calibration of computer models*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **63** (2001), 425–464, DOI 10.1111/1467-9868.00294 (en).

[23] ———, *Bayesian calibration of computer models*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) **63** (2001), 427–428, DOI 10.1111/1467-9868.00294 (en).

[24] Dong and Jiang Li Penghao and Hu, *Comparison of local and global sensitivity analysis methods and application to thermal hydraulic phenomena*, Progress in Nuclear Energy **158** (2023), 104612, DOI 10.1016/j.pnucene.2023.104612.

[25] Christoph Molnar, *8.1 Partial Dependence Plot (PDP) | Interpretable Machine Learning.*

[26] Richard H. and Edmonds Moss Jae A. and Hibbard, *The next generation of scenarios for climate change research and assessment*, Nature **463** (2010), no. 7282, 747–756, DOI 10.1038/nature08823 (en).

[27] ———, *The next generation of scenarios for climate change research and assessment*, Nature **463** (2010), no. 7282, 748–749, DOI 10.1038/nature08823 (en).

[28] ———, *The next generation of scenarios for climate change research and assessment*, Nature **463** (2010), no. 7282, 751–752, DOI 10.1038/nature08823 (en).

[29] NASA, *Taking a Global Perspective on Earth's Climate*, Climate Change: Vital Signs of the Planet (en).

[30] Clémentine and Tarantola Prieur Stefano, *Variance-Based Sensitivity Analysis: Theory and Estimation Algorithms* (Roger and Higdon Ghanem David and Owhadi, ed.), Springer International Publishing, 2017 (en).

[31] Stefan Rahmstorf, *A semi-empirical approach to projecting future sea-level rise*, Science (New York, N.Y.) **315** (2007), no. 5810, 368–370, DOI 10.1126/science.1135456 (eng).

[32] Patrick M. and Hadjimichael Reed Antonia and Malek, *Addressing Uncertainty in MultiSector Dynamics Research*, July 18, 2023 (en).

[33] ———, *Addressing Uncertainty in MultiSector Dynamics Research*, July 18, 2023 (en).

[34] ———, *Addressing Uncertainty in MultiSector Dynamics Research*, July 18, 2023 (en).

[35] Jeremy and Lincke Rohmer Daniel and Hinkel, *Water | Free Full-Text | Unravelling the Importance of Uncertainties in Global-Scale Coastal Flood Risk Assessments under Sea Level Rise*, MDPI, 18.

[36] ———, *Water | Free Full-Text | Unravelling the Importance of Uncertainties in Global-Scale Coastal Flood Risk Assessments under Sea Level Rise*, MDPI, 11–12.

[37] Rafael and Gupta Rosolem Hoshin V. and Shuttleworth, *A fully multiple-criteria implementation of the Sobol method for parameter sensitivity analysis*, Journal of Geophysical Research: Atmospheres **117** (2012), no. D7, DOI 10.1029/2011JD016355 (en).

[38] Andrea and Ratto Saltelli Marco and Andres, *Global Sensitivity Analysis: The Primer*, Wiley, 2008.

[39] ———, *Global Sensitivity Analysis: The Primer*, Wiley, 2008.

[40] ———, *Global Sensitivity Analysis: The Primer*, Wiley, 2008.

[41] Erwan Scornet, *Trees, forests, and impurity-based variable importance in regression* (2010).

[42] Carolin and Boulesteix Strobl Anne-Laure and Kneib, *Conditional variable importance for random forests*, BMC Bioinformatics **9** (2008), no. 1, 1–11, DOI 10.1186/1471-2105-9-307 (en).

[43] ———, *Conditional variable importance for random forests*, BMC Bioinformatics **9** (2008), no. 1, 8–9, DOI 10.1186/1471-2105-9-307 (en).

[44] Carolin and Malley Strobl James and Tutz, *An introduction to recursive partitioning: Rationale, application, and characteristics of classification and regression trees, bagging, and random forests.*, Psychological Methods **14** (2009), no. 4, 323–348, DOI 10.1037/a0016973 (en).

[45] Tony E. and Bakker Wong Alexander M. R. and Ruckert, *BRICK v0.2, a simple, accessible, and transparent model framework for climate and regional sea-level projections*, Geoscientific Model Development **10** (July 17, 2017), no. 7, 2741–2760, DOI 10.5194/gmd-10-2741-2017.

[46] T. E. and Ledna Wong C. and Rennels, *Sea Level and Socioeconomic Uncertainty Drives High-End Coastal Adaptation Costs*, Earth's Future **10** (2022), no. 12, 1–24, DOI 10.1029/2022EF003061 (en).

[47] Sinan and Lu Xiao Zhenzhou, *Global sensitivity analysis based on Gini's mean difference*, Structural and Multidisciplinary Optimization ***58*** (2018), DOI 10.1007/s00158-018-1982-7.

[48] Sinan and Lu Xiao Zhenzhou and Wang, *Global sensitivity analysis based on distance correlation for structural systems with multivariate output*, Engineering Structures **167** (2018), DOI 10.1016/j.engstruct.2018.04.027.

[49] Ruoqing and Zeng Zhu Donglin and Kosorok, *Reinforcement Learning Trees*, Journal of the American Statistical Association **110** (2015), no. 512, 1770–1784, DOI 10.1080/01621459.2015.1036994.

Approval Signatures

Dr. Mihail Barbosu: Committee Member **Date**

Dr. Lucia Carichino: Committee Member **Date**

Dr. Tony Wong: Committee Member/Thesis Advisor **Date**