

# Multivariate sensitivity analysis for a large-scale climate impact and adaptation model

O. K. Oyebamiji

*Flood and Water Management, HR Wallingford, Howbery Park, Wallingford, UK.*

E-mail: o.oyebamiji@hrwallingford.com

C. Nemeth

*Department of Mathematics & Statistics, Lancaster University, United Kingdom.*

P. A. Harrison

*UK Centre for Ecology & Hydrology, Lancaster, United Kingdom.*

R. W. Dunford

*UK Centre for Ecology & Hydrology, Wallingford, United Kingdom.*

G. Cojocaru

*TIAMASG Foundation, Romania.*

**Summary.** We apply a new efficient methodology for Bayesian global sensitivity analysis for large-scale multivariate data. A multivariate Gaussian process is used as a surrogate model to replace the expensive computer model. To improve the computational efficiency and performance of the model, compactly supported correlation functions are used. The goal is to generate sparse matrices, which give crucial advantages when dealing with large datasets. The method was applied to multivariate data from the IMPRESSIONS Integrated Assessment Platform (IAP2). Our empirical results on IAP2 data show that the proposed methods are efficient and accurate for global sensitivity analysis of complex models.

*Keywords:* Bayesian methods, Gaussian process, Sensitivity analysis, Compactly supported correlation function, Robust adaptive MCMC.

## 1. Introduction

There is a growing need in the environmental science community to provide robust decision-making tools for policymakers. By linking together process-based models across sectors, environmental scientists can accurately project future climate change impacts based on climate, social, economic and political factors. Such integrated and cross-sectoral climate change impact models are essential in developing successful evidence-based adaptation strategies (Harrison et al., 2016). Full uncertainty quantification for these models is critical to ensure that policymakers develop robust strategies to adapt to climate change. Climate change impact and adaptation models, such as the IMPRESSIONS Integrated Assessment Platform (IAP2) models, are high-dimensional and computationally intensive to evaluate over the full parameter space. To develop a deeper understanding of the key processes which drive these models, it is necessary to identify

the relative contributions of the input variables and how they affect the outputs. Many input variables have no significant effect on the output variance, meaning that we can reduce the dimension of the problem by identifying and removing the uncertainties related to these variables, which have little to no impact on the output.

Using sensitivity analysis [Gamboa et al. \(2020\)](#); [Saltelli et al. \(2019\)](#), we can attribute the uncertainty in the outputs of a model to the uncertainty in the model inputs. The principal goal of sensitivity analysis is to identify the crucial parameters that are responsible for most of the variation in the output. Sensitivity analysis can also help identify regions of the input variability space that produce large and extreme output values. It can also be used to identify interactions between variables and highlight where a model can be improved to obtain better input values ([Farah and Kottas, 2014](#); [Santner et al., 2013](#)). There are two standard approaches for performing sensitivity analysis, namely *local* and *global* sensitivity. Local sensitivity quantifies the variation of the model outputs due to small changes in the uncertain model inputs, where methods such as finite differences and partial derivatives are used as sensitivity indices. We focus our attention on global sensitivity analysis, which shows how the output changes as all the input factors vary continuously over their entire ranges. It uses the full ranges of the uncertainty of the inputs, and these are usually characterised through their joint probability density functions. Global sensitivity requires more computational work than local methods but is capable of revealing hidden relationships between multiple parameters. Computing global sensitivity indices require the evaluation of a large-dimensional integral over the entire input space. The more parameters we have, the more computationally demanding. Standard numerical integration ([Robert et al., 1999](#); [South et al., 2020](#)), such as Monte Carlo integration, is usually infeasible when the computer models generating the required model outputs are computationally expensive ([Farah and Kottas, 2014](#)).

Most of the existing global sensitivity analysis techniques assume a univariate output. There exists a large literature that performs global sensitivity analysis of scalar outputs ([Farah and Kottas, 2014](#); [Svenson et al., 2014](#)). The standard approach used for multivariate outputs is to perform the global sensitivity analysis for each output. The main limitation of this approach, apart from neglecting the output correlation when performing the analysis, is that it produces many sensitivity indices that are difficult to interpret. In contrast to these methods, [Gamboa et al. \(2013\)](#); [Lamboni et al. \(2011\)](#) and recently [Xiao et al. \(2018\)](#); [Cheng et al. \(2019\)](#) generalise the Sobol sensitivity indices ([Sobol, 1993](#)) to the case of multivariate outputs. Further overview of the global sensitivity analysis methods and algorithms and their theoretical properties are provided in [Da Veiga et al. \(2021\)](#).

Interpretation of sensitivity indices for models with multivariate outputs is complicated because existing techniques usually ignore the effect of the dimension and correlation between multiple outputs. Dependencies between input variables also complicate this problem if it exists. This paper addresses these shortcomings. Firstly, we use a multivariate sparse Gaussian process (MSGP) as an efficient, alternative model to replace the expensive computer models. This combines the methods proposed in [Svenson et al. \(2014\)](#) and [Farah and Kottas \(2014\)](#). The MSGP model incorporates compact approximations of the dense covariance matrix to reduce the computational burden. The incorporation of sparse matrices offers great savings in storage and computation by in-

**Table 1.** List of user-specific input variables from 2020s to 2080s and their definitions. See Table 6 in Appendix A for the parameter ranges, where \* denotes discrete variables.

<b>X</b>	Parameters	Definitions
X <sub>1</sub> *	Emission	Emission scenarios
X <sub>2</sub> *	Model	Climate model
X <sub>3</sub> *	SSPs	Socio-economic scenarios
X <sub>4</sub>	Population	Population change
X <sub>5</sub>	GreenRed	Preference for living close to urban or rural areas
X <sub>6</sub>	RumLivDem	Change in dietary preference for beef/lamb
X <sub>7</sub>	NonRumLivDem	Change in dietary preference for chicken/pork
X <sub>8</sub>	StructChange	Water savings due to behavioural change
X <sub>9</sub>	TechFactor	Change in agricultural mechanisation
X <sub>10</sub>	TechChange	Water savings due to technological change
X <sub>11</sub>	YieldFactor	Change in agricultural yields
X <sub>12</sub>	IrrigEffFactor	Change in irrigation efficiency
X <sub>13</sub>	GDP	GDP change
X <sub>14</sub>	CostsFactor	Change in oil price
X <sub>15</sub>	ImportFactor	Change in food imports
X <sub>16</sub>	BioEneCropDemand	Change in bioenergy production
X <sub>17</sub>	ArableConservLand	Set-aside
X <sub>18</sub>	CropInpFactor	Reducing diffuse pollution from agriculture
X <sub>19</sub>	Coastal	Coastal flood event
X <sub>20</sub>	Fluvial	Fluvial flood event
X <sub>21</sub>	HabitatRecreation	Protected area changed
X <sub>22</sub>	PA Forestry	Change in protected area forest
X <sub>23</sub>	PA Agriculture	Change in protected area agriculture
X <sub>24</sub> *	FloodProtection	Level of flood protection

roducing some zeros in the matrix of the linear systems. Secondly, we parallelised the posterior computation by distributing the likelihood estimation across multiple computing nodes using a multicore computing environment. We also exploit high-performance computing libraries such as **Armadillo** (Sanderson and Curtin, 2016), which is a templated  $C^{++}$  linear algebra library to facilitate expensive matrix-matrix multiplication and inversion (Eddelbuettel and Sanderson, 2014). This procedure is combined with a robust adaptive Metropolis-within-Gibbs algorithm (Vihola, 2012) to efficiently sample from the posterior target distribution. More details of the parallelisation process is given in the pseudo-code for the MCMC algorithm in the Supplementary Material. Thirdly, we extended the computation of sensitivity indices to the case of multivariate spatially distributed outputs for computationally demanding models. This integrated approach produces a reduced-form representation of the model that enabled a substantial reduction in computation time.

In summary, we have used a multivariate sparse Gaussian process as an efficient approach for performing a multivariate sensitivity analysis, and then applied it to a large-scale climate impact and adaptation model. However, this is not the first time GPs have been used in this context. The focus of our paper is to address the challenge found in applications where it will take far too long to run the simulator and generate thousands of simulations. Large-scale simulations are required to account for different parameter combinations and to perform effective sensitivity analysis, which we can achieve in a computationally efficient manner, without sacrificing statistical accuracy.

The paper is organised as follows. Section 2 provides a brief description of the IMPRESSIONS IAP2 model and parameter sampling. The multivariate sparse Gaussian processes (MSGP) theory, its parameter estimation, and compact correlation functions are given in Section 3. Section 4 focuses on the proposed new sensitivity indices, while the analysis and results from the synthetic and IAP2 data are presented in Section 5. The paper closes with a conclusion in Section 6.

## 2. Model and simulation data

### 2.1. *IMPRESSIONS Integrated Assessment Platform version 2 (IAP2) model*

The IMPRESSIONS IAP2 model was developed to explore the impacts of climatic and socio-economic change on the interactions between different landscape sectors (agriculture, forestry, water resources, biodiversity, coastal and fluvial flooding, and urban development) Harrison et al. (2019). It provides an integrated approach to support climate change adaptation decision-making by helping decision-makers understand the complex interactions between sectors and how these are affected by changing climatic and socio-economic conditions. The latest scenarios from the Intergovernmental Panel on Climate Change research community are incorporated into the IAP2 for climate (based on Representative Concentration Pathways) and socio-economics (based on the Shared Socio-economic Pathways). The model produces outputs for a wide range of both sector-based impact and vulnerability indicators and ecosystem service indicators to link impacts directly to human well-being (Harrison et al., 2016).

The IMPRESSIONS IAP2 operates at a spatial resolution of  $10 \times 10$  arcminutes across Europe. The models within the IAP2 are hard-linked, with flows of data passed

between models as part of a hierarchical model chain. Each of the ten models within the IAP2 is represented by a simplified meta-model to enable the full model chain to run with a web-based interactive interface (Harrison et al., 2015; Kebede et al., 2015). Nevertheless, a single run of the model takes some minutes, and multiple runs can not be made parallel because it is a graphical-based model. The IMPRESSIONS IAP2 model has a large number of input parameters, namely climate variables, such as precipitation and temperature, and socio-economic variables, such as the change in the population, GDP, agricultural mechanisation and diets. Although the IAP2 adopts a meta-modelling approach for its ten component models, the expense of running such a model means that it is infeasible to run all possible parameter value combinations.

The model produces over 100 outputs, which can be categorised into land use-related indicators, such as the area of urban land, managed forests and arable land, and other indicators related to food production, carbon storage, species suitability, water exploitation and flood risk. In addition to the baseline, it simulates impacts for three timeslices (the 2020s, 2050s and 2080s), three emissions scenarios (RCPs 2.6, 4.5 and 8.5) and four socio-economic scenarios (SSPs 1, 3, 4 and 5). The ten climate models used under various emissions scenarios are given in Appendix A. Further details about the model and scenarios are given in Harrison et al. (2019).

## 2.2. Parameter sampling

Sampling the input parameter space is a critical part of both sensitivity analysis and building a useful surrogate model. The goal is to explore the entire input space with a reasonable sample size  $n$ . There exist a large number of various sampling schemes (Santner et al., 2013; Burhenne et al., 2011).

A proper sample generation technique is crucial for efficient surrogate modelling for computationally-expensive problems. It ensures an accurate estimation of model parameters. This enables large-scale sensitivity analyses and scenario exploration without the computational burden. Moreover, most of the work on sensitivity and uncertainty analyses from large-scale models assumes that the input factors in computer models are continuous variables, and most of the existing methods are not directly applicable to applications that contain both continuous and categorical variables. The presence of categorical variables makes the design of computer experiments a more laborious task because the qualitative variables do not allow for an infinite number of different points or levels.

The input parameters for running a set of runs from the IMPRESSIONS IAP2 model were generated using Latin Hypercube Sampling (LHS) (Santner et al., 2013). The procedure provides data for training a robust surrogate model to approximate the outputs from the IMPRESSIONS IAP2 model. The LHS technique is popular and provides better coverage of the input space than standard random sampling techniques. We used the optimum LHS technique to avoid clustering of the sample points and to ensure a good estimation of the statistical moments of response functions. Optimum LHS also maximises the mean distance from each design point to all the other points such that points are as spread out as possible (Carnell and Carnell, 2019; Stocki, 2005).

We selected 24 input parameters to evaluate the sensitivity of the IMPRESSIONS IAP2 model outputs; 20 of these variables are continuous parameters (see Table 1 of

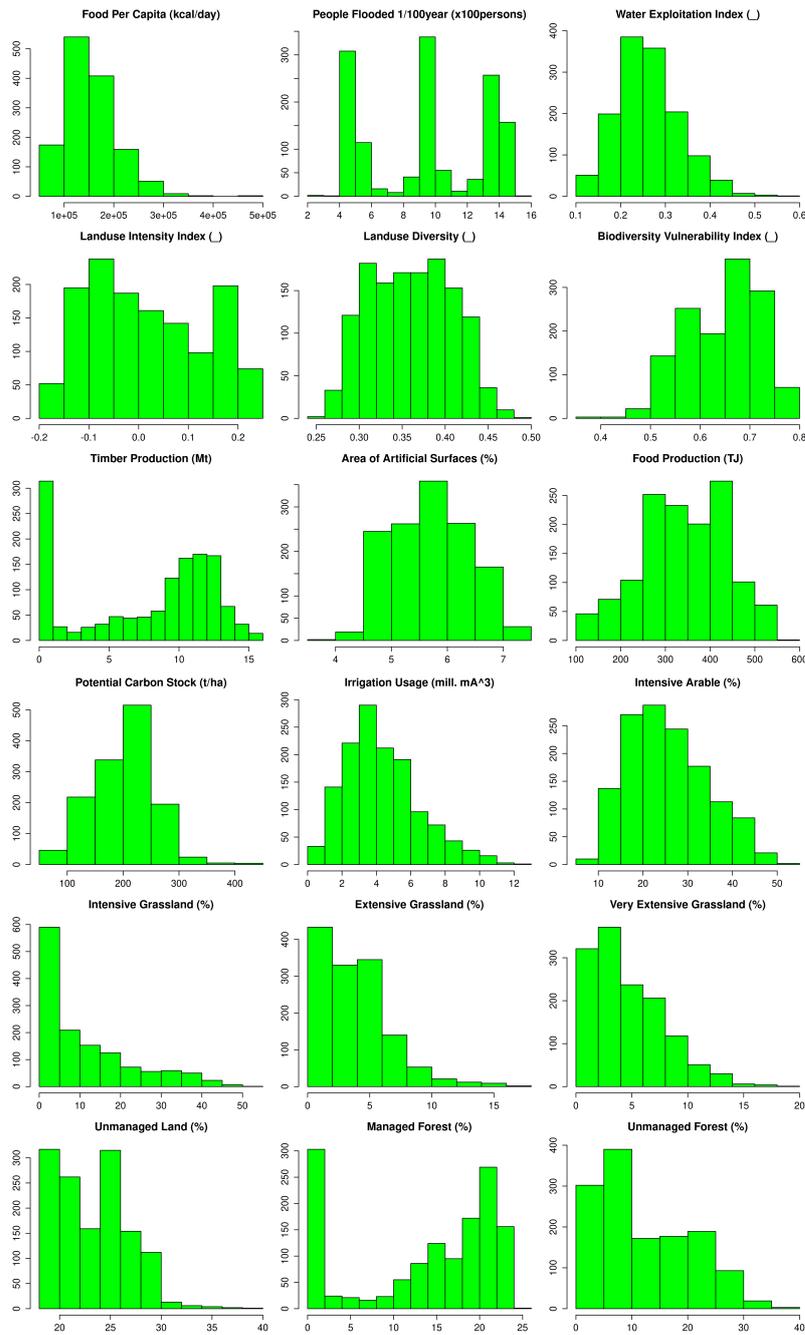
**Table 2.** Selected outputs.

Variables	
Food per capita	People flooded in a 1 in 100-year event
Water exploitation index	Landuse intensity index
Land use diversity	Biodiversity vulnerability index
Timber production	Area of artificial surfaces
Food production	Potential carbon stock
Irrigation usage	Arable land
Intensive grassland	Extensive grassland
Very extensive grassland	Unmanaged land
Managed forest	Unmanaged forest

Appendix A). The four categorical variables are the level of flood protection, emission, climate model and socio-economic scenario (denoted as \* in Table 1). We used a two-stage sampling technique. In the first stage, we use the LHS to sample the continuous variables, and in the second stage, we formed every possible combination of categorical variables according to the range of values given in Table 6. We limit our samples to just 2,000 design points. This gives  $\sim 30,000$  design points in total (Note: 30,000 simulations take  $\sim 42$  days of computer time), and each sampling point produces over a hundred spatially-distributed outputs.

### 2.3. Output data

We used the LHS algorithm to produce ensembles of simulations from the IMPRESSIONS IAP2 model. We have selected 18 important variables that reflect impacts across sectors (see Table 2). The data are on a  $10^\circ \times 10^\circ$  (10 degrees by 10 degrees) resolution grid  $\sim 16km^2$  and each has 23,871 grid cells. Figure 1 shows the histograms of selected IAP2 simulation outputs under the SSP1 scenario (which is characterized by a sustainable future with global cooperation and less inequality), averaged over all emissions scenarios, climate models and timeslices. The histograms of *food per capita* and *water exploitation index* are relatively similar and symmetric, while that of *people flooded in a 1 in 100-year event* has three major peaks. The histogram of *timber production* and *managed forest* also have a single major peak and a minor peak, making them bimodal distributions. The remaining variables such as *irrigation usage*, *intensive grassland arable*, *intensive grassland dairy*, *extensive grassland* and *very extensive grassland* are right-skewed. Apart from the nonsymmetric nature of most of the variables with no distinct shape, we note that the range of values of these variables varies widely. The diverse structural patterns clearly show the complexity of the effects we are modelling. Our model formulation needs to capture these divergent behaviours to be useful for performing sensitivity analyses. The histograms of the remaining simulated IAP2 outputs for SSP4 are given in Figure 14 in Appendix A.



**Fig. 1.** Histograms of spatially aggregated IAP2 simulation for the 18 selected outputs, averaged across the three emission scenarios and ten climate models for SSP1. Note: (L) denotes unitless.

### 3. Surrogate Modelling

#### 3.1. Multivariate sparse Gaussian processes (MSGP)

We introduce the overall statistical model structure in this section and describe each component of the model. We formulate a probabilistic model for multivariate data using a multivariate Gaussian process and describe a Markov chain Monte Carlo (MCMC) sampling algorithm for inference. We also outline how to extend this idea to sparse approximation models for efficient sampling and achieve dimension reduction for large multi-dimensional data. The main procedure involves incorporating a sparse approximation to achieve efficient matrix operations within a Bayesian framework. This multilevel model is estimated using an adaptive MCMC technique within a parallel computing environment. We implement an adaptive Metropolis algorithm [Vihola \(2012\)](#) within a Gibbs sampler to facilitate highly efficient posterior sampling.

Let  $\mathbf{x} = (x_1, \dots, x_p)^T \in \mathcal{X} \subset \mathbb{R}^p$  be the vector of  $p$  input variables with outputs from the simulator  $g : \mathbf{x} \rightarrow \mathbf{y} \in \mathcal{Y} \subset \mathbb{R}^m$  denoted as  $\mathbf{y}$ , i.e.

$$g(\mathbf{x}) = \{g_1(\mathbf{x}), \dots, g_m(\mathbf{x})\}^T$$

is the  $m \times 1$  output vector from the simulator at input combination  $\mathbf{x}$ . Also, let the matrix of input combinations be  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ , with  $\mathbf{x}_i = (x_{i1}, \dots, x_{ip})^T$  representing the  $i^{\text{th}}$  column. The multivariate outputs from the simulator, with  $n$  sample points, are

$$\text{combined into an } n \times m \text{ output matrix } \mathbf{Y} = \begin{pmatrix} g(\mathbf{x}_1)^T \\ \vdots \\ g(\mathbf{x}_n)^T \end{pmatrix}.$$

We can decompose  $\mathbf{Y}$  into a linear trend and nonlinear random process,

$$\mathbf{Y} = \mathbf{H}(\mathbf{x})\mathbf{B} + \boldsymbol{\varepsilon}(\mathbf{x}), \quad (1)$$

where  $\mathbf{B}$  is a  $p \times m$  matrix of regression coefficients capturing most of the variation in  $g(\cdot)$  and  $\mathbf{H} = [h(\mathbf{x}_1) \cdots h(\mathbf{x}_n)]^T$  is a  $n \times p$  regression matrix of explanatory variables  $h(\cdot)$ , common to all the outputs. In this paper the components of  $\mathbf{H}$ , i.e.  $\mathbf{H} = [h(\mathbf{x}_1), \dots, h(\mathbf{x}_n)]^T$ , are simple linear functions of the inputs, i.e.  $h(\mathbf{x}_1) = \mathbf{x}_1$  and  $\boldsymbol{\varepsilon}(\cdot)$  is a vector random field, modelled as a multivariate Gaussian process. The notation  $h(\cdot)$  could be simplified here as the functions are linear, however, we keep the notation  $h$  to illustrate the generality of these functions, e.g. to include nonlinear terms. We use a separable covariance function in the matrix-normal distribution  $\boldsymbol{\varepsilon} \sim \text{MN}(\mathbf{0}, \boldsymbol{\Sigma}, \mathfrak{R})$  so that we capture the remaining unexplained small scale structure. The  $n \times n$  input correlation matrix  $\mathfrak{R}$ , i.e.

$\text{Cor}(\boldsymbol{\varepsilon}(\mathbf{x}), \boldsymbol{\varepsilon}(\mathbf{x}')) = R(\mathbf{x}, \mathbf{x}')$ , is chosen by the user. A common choice for  $R(\cdot)$  is the exponential power correlation function ([Kaufman et al., 2011](#)), given in Equation (5). Further details regarding the correlation functions used in this analysis are given in the next section. The  $m \times m$  system scale matrix  $\boldsymbol{\Sigma}$  denotes the cross-covariance matrix between the  $m$  multivariate outputs.

To predict at  $n^*$  new test values  $\mathbf{x}^*$  with given estimated parameters  $\boldsymbol{\Theta}$ , the conditional distribution of  $\mathbf{Y}^*$  given training data  $\mathbf{D}$ , is a matrix-normal with mean and covariance given respectively as

$$\mathbf{Y}^* | \mathbf{x}^*, \mathbf{D}, \mathbf{B}, \boldsymbol{\Sigma} \sim \text{MN}(\mathbf{Q}, \boldsymbol{\Sigma}, \mathbf{r}^* - \mathbf{r}^T \mathfrak{R}^{-1} \mathbf{r}), \quad (2)$$

where  $\mathbf{Q} = \mathbf{H}^* \mathbf{B} + \mathbf{r}^{*T} \mathfrak{R}^{-1} (\mathbf{D} - \mathbf{H} \mathbf{B})$  and  $\mathbf{r}$  is the  $n \times n^*$  matrix of cross-correlations between the observed responses  $\mathbf{X}$  and test points  $\mathbf{x}^*$ , while  $\mathbf{r}^*$  is the  $n^* \times n^*$  correlation matrix between  $\mathbf{x}^*$ . To achieve further efficiency, we can integrate out  $\mathbf{B}$  and  $\Sigma$  (see Section 3.3 for details) to give

$$\mathbf{Y}^* | \mathbf{x}^*, \mathbf{D} \sim MT(\mathbf{Q}, \mathbf{S}, \mathfrak{D}, \delta), \quad (3)$$

where  $MT(\cdot)$  is the matrix t-distribution with location parameter  $\mathbf{Q}$  and a row-scale matrix

$$\mathfrak{D} = \mathbf{r}^* - \mathbf{r}^T \mathfrak{R}^{-1} \mathbf{r} + (\mathbf{H}^* - \mathbf{r}^T \mathfrak{R}^{-1} \mathbf{H}) \Lambda (\mathbf{H}^* - \mathbf{r}^T \mathfrak{R}^{-1} \mathbf{H})^T \quad (4)$$

while  $\mathbf{S}$  and  $\delta$  are column-scale matrices and degrees of freedom, respectively.

### 3.2. Compactly supported correlation functions

We note that making inferences for the model parameters, and creating predictions from the model, is computationally expensive. The main bottleneck is in the computation of the inverse and determinant of an  $n \times n$  matrix  $\mathfrak{R}$  and solving for  $\mathfrak{R}^{-1}(\mathbf{Y} - \mathbf{H}\mathbf{B})$ . These operations scale with an order of  $\mathcal{O}(n^3)$  and are thus intractable for large  $n$ , where  $n$  is the sample size. To overcome this problem, we introduce compactly supported correlation functions to multivariate outputs in this study. These correlation functions have been used in earlier works of Kaufman et al. (2011) and Svenson et al. (2014), which focus only on univariate applications.

The concept of compactly supported correlation functions was proposed by Gneiting (2002) to parameterise the smoothness of the associated stationary and isotropic random field. A stationary process that employs standard covariance functions for its predictions produces dense covariance matrices, and this is problematic, especially when  $n$  is large. Compactly supported correlation functions yield sparse covariance matrices. Having many zero entries in the covariance matrices can both greatly reduce computer storage requirements and the number of floating-point operations needed in computation. These types of functions with compact support are used to facilitate efficient computation in spatial predictions. A comprehensive review of compactly supported covariance functions is given in Moreaux (2008); Gneiting (2002).

Suppose  $R(\mathbf{x}, \mathbf{x}'; \phi)$  is a power exponential correlation function which is a function of the Euclidean distance between  $\mathbf{x}$  and  $\mathbf{x}'$ , we have

$$R(\mathbf{x}, \mathbf{x}') = \prod_{k=1}^p \exp \left\{ -\phi_k |x_k - x'_k|^2 \right\} \quad (\text{Power exponential}), \quad (5)$$

where  $\phi$  is a correlation range parameter for a finite set of points. The idea is that for many data sets, the correlations vanish beyond a certain cut-off distance. We have used two functions, given below, that can be directly used to approximate the power exponential function. Therefore,  $\phi_k$  is no longer used as a parameter in the model and instead we have a cut-off parameter  $\tau_k \geq 0$  and  $t = |x - x'|$ .

$$\text{Bohman:} \quad R(t; \tau) = (1 - t/\tau) \cos(\pi t/\tau) + \sin(\pi t/\tau)/\pi, 0 \leq t/\tau \leq 1,$$

$$\text{Truncated power:} \quad R(t; \tau, \alpha, \nu) = [1 - (t/\tau)^\alpha]^\nu, 0 < \alpha < 2, \nu \geq \nu_p(\alpha).$$

To ensure positive-definiteness for the truncated power function, there is a constraint on the parameter  $\nu_p(\alpha)$  so that it will produce a valid correlation matrix,  $\lim_{\alpha \rightarrow 2} \nu_p(\alpha) = \infty$ . The range of upper bounds for several values of  $\alpha$  between 1.5 and 1.955 were given in [Gneiting \(2001\)](#) because it is difficult to evaluate  $\nu_p(\alpha)$ . For instance, the upper bounds when  $\alpha = \{3/2, 5/3\}$  are 2 and 3, respectively.

We also need to restrict  $\tau$  to fall within the p-dimensional triangle to impose sparsity in the correlation functions above, i.e.

$$\tau_c = \left\{ \tau \in \mathbb{R}^p : \tau_k \geq 0, \sum_{k=1}^p \tau_k \leq c \right\}, \quad \forall k, c > 0.$$

To achieve some levels of sparsity, the parameter  $c$  is fixed to a reasonable value such that a given percentage of the  $\omega$  off-diagonal elements of  $\mathfrak{R}$  are zeroes. The procedure for determining an appropriate  $c$  to obtain a required  $\omega\%$  of zeroes among the off-diagonal elements is given below. Compute a pair distance

$$d_{i,j} = \sum_{k=1}^p |\mathbf{x}_{ik} - \mathbf{x}_{jk}| \quad \text{for } 1 \leq i < j \leq n.$$

Let  $c$  be the  $\left\lfloor \binom{n}{2} \times \omega\% \right\rfloor^{th}$  smallest value among the  $d_{i,j}$  's, where  $\lfloor \cdot \rfloor$  is the integer part of  $\binom{n}{2} \times \omega\%$ . We note that the cutoff  $\tau_c$  controls both the correlation range and degree of sparsity. See [Svenson et al. \(2014\)](#) and [Kaufman et al. \(2011\)](#) for further details.

We have allowed for various degrees of smoothness in our analysis. For instance, the Matérn correlation corresponds to a process which is  $r$  times mean-squared differentiable if  $\nu > r$ . Unlike the Matérn correlation, the truncated power function is not differentiable at the origin, even with  $\alpha < 2$ , corresponding to a process that is not mean-squared continuous. The Bohman function, on the other hand, is twice differentiable at the origin when  $\alpha < 2$  and leads to a process which is mean-square differentiable. The power exponential function, eq.(5), is smooth and infinitely differentiable at the origin when  $\alpha = 2$ .

We note that the compact correlation function preserves both compact support and the local behaviour of the correlation function at the origin. Therefore, compactly supported correlation functions can be directly used to model dependencies within random processes ([Kaufman et al., 2011](#); [Svenson et al., 2014](#); [Gneiting, 2002](#)), which is the approach we follow in this paper. They can also be used by multiplying with another strictly positive correlation function, often called tapering, for estimation and prediction in spatial statistics ([Zhang et al., 2015](#); [Kaufman et al., 2008](#)). Further details are given in Appendix D and a summary of the implementation procedure is given in the MCMC algorithm in Appendix B.

Other popular alternatives to reduce the computational burden are the earlier works of [Vecchia \(1988\)](#); [Stein et al. \(2004\)](#) where the joint likelihood is partitioned and written as a product of conditional densities. The smaller sets of conditional sets consisting

of nearest neighbours are then chosen. This approach was recently developed into the nearest neighbour Gaussian process using parallel computation in the works of [Finley et al. \(2017\)](#); [Datta et al. \(2016b\)](#) and [Zhang et al. \(2019\)](#).

Various variants of the nearest neighbourhood method have been developed which have attractive computational-efficiency features when applied to large data sets ([Finley et al., 2017](#); [Datta et al., 2016b,a](#)). Alternatively, the works of [Overstall and Woods \(2016\)](#); [Alvarez and Lawrence \(2011\)](#); [Moreno-Muñoz et al. \(2018\)](#); [Taylor-Rodriguez et al. \(2018\)](#) also cover multivariate processes but require an additional computational expense to handle large dimensional data. For instance, [Maddox et al. \(2021\)](#) exploit the use of Kronecker structures in the covariance matrices for high dimensional outputs. A sound overview of several modern methods for analyzing large spatial data, comparing their performances and identifying their strengths and weaknesses is given in ([Heaton et al., 2019](#)). This approach uses a conditional independence assumption to learn a sparse inverse covariance.

### 3.3. Posterior sampling

Let  $\Theta = \{\mathbf{B}, \Sigma, \boldsymbol{\tau}\}$  be the set of unknown parameters and  $p(\Theta)$  the prior distributions for these unknown parameters. Using the Bayesian approach, we can estimate these unknown parameters by assigning a conditionally conjugate prior,  $p(\mathbf{B}|\Sigma, \boldsymbol{\tau})p(\Sigma|\boldsymbol{\tau})p(\boldsymbol{\tau})$ . We use a matrix-normal inverse-Wishart (MNIW) prior for  $\{\mathbf{B}, \Sigma\}$ , where we assign a normal prior to the matrix of regression coefficients  $\mathbf{B}|\Sigma, \boldsymbol{\tau} \sim N(\mathbf{B}_0, \Sigma \otimes \Lambda_0)$  and cross-covariance  $\Sigma|\boldsymbol{\tau} \sim \mathcal{W}^{-1}(\mathbf{S}_0, \delta_0)$ , where  $\mathbf{B}_0, \Lambda_0, \mathbf{S}_0$  and  $\delta_0$  are prior hyperparameters. The sparsity of a given matrix depends on the particular distribution of sampling points, and we have specified the prior for  $\boldsymbol{\tau}$  to be uniformly distributed over the space. The input is scaled to  $[-1,1]$  and accommodates different levels of sparsity. A high degree of sparsity will be imposed so that the proportion of off-diagonal elements in the correlation matrix that are nonzero is small.

We can use MCMC to sample from the posterior  $p(\Theta|\mathbf{Y})$  using a Metropolis-within-Gibbs algorithm. This algorithm combines Gibbs sampling with an adaptive random-walk algorithm. We use a Gibbs sampler to sample the parameters  $\mathbf{B}$  and  $\Sigma$  using their conditional posterior distributions,

$$\mathbf{B}|\mathbf{Y}, \Sigma, \boldsymbol{\tau} \sim \mathcal{MN}_{n,m}(\mathbf{B}, \Sigma, \Lambda) \quad \text{and} \quad \Sigma|\mathbf{Y}, \mathbf{B}, \boldsymbol{\tau} \sim \mathcal{W}^{-1}(\mathbf{S}, \delta), \quad (6)$$

where  $\widehat{\mathbf{B}} = \widehat{\Lambda}(\mathbf{H}^T \mathfrak{R}^{-1} \mathbf{Y} + \Lambda_0 \mathbf{B}_0)$ ,  $\widehat{\Lambda} = (\mathbf{H}^T \mathfrak{R}^{-1} \mathbf{H} + \Lambda_0)^{-1}$ ,  $\widehat{\mathbf{S}} = \mathbf{S}_0 + \mathbf{Y}^T \mathfrak{R}^{-1} \mathbf{Y} + \mathbf{B}_0^T \Lambda_0 \widehat{\mathbf{B}}_0 - \widehat{\mathbf{B}}^T \widehat{\Lambda} \widehat{\mathbf{B}}$  and  $\widehat{\delta} = \delta_0 + n_0$ .

We sample the parameter  $\boldsymbol{\tau}$  using a robust adaptive Metropolis-Hastings algorithm ([Vihola, 2012](#)). We can simplify the posterior distribution for  $\boldsymbol{\tau}$  by integrating out the matrices  $\mathbf{B}$  and  $\Sigma$ ,

$$p(\boldsymbol{\tau}|\mathbf{Y}) \propto |\mathfrak{R}|^{-\frac{m}{2}} \times |\widehat{\Lambda}|^{m/2} |\mathbf{S}|^{-(\delta+m-1)/2} \times p(\boldsymbol{\tau}). \quad (7)$$

See [Rougier \(2007, 2008\)](#) and [Overstall and Woods \(2016\)](#) for further details. The robust adaptive Metropolis algorithm is highly efficient as it automates the tuning of the random-walk proposal. Further details on the full MCMC procedure, including pseudo-code, are given in Appendix B.

## 4. Performing sensitivity analyses

### 4.1. Evaluating global sensitivity indices using MSGP

The IAP2 model we are using in this paper consists of  $p = 24$  input variables, which produce multivariate output data, from which we have carefully selected  $m = 18$  responses for this analysis. This would require the estimation of  $p \times m$  sensitivity indices. Using the compactly supported correlation functions to model small-scale variability reduces the computational expense as operations involving zero elements need not be stored or performed. We used a two-stage approach to compute all the sensitivity indices. The first step is to sample from the posterior distribution of  $\hat{\Theta}$  using an adaptive Metropolis within Gibbs sampler as described in the previous section. As noted earlier, our choice of prior distribution allows us to work with the marginal distribution. The second step is to compute predictive means and variances for new inputs using the posterior samples  $\hat{\Theta}$  and then compute the sensitivity indices.

The variance-based approach uses the popular Hoeffding decomposition method of [Van der Vaart \(2000\)](#) to separate the total variance into different partial variances for a univariate output  $Y$ . These partial variances measure the associated output uncertainty induced by the corresponding input variables. By considering the ratio of each partial variance to the total variance, we can obtain a measure of importance for each input variable called the sensitivity index. Suppose  $E(Y|x_j) = \int_{\mathbf{x}_{\sim j}} f(\mathbf{x}) dG_{\sim j|j}(\mathbf{x}_{\sim j}|x_j)$ , where  $\mathbf{x}_{\sim j}$  denotes input vector  $\mathbf{x}$  excluding element  $x_j$ , then the variance-based method of decomposition into partial variances gives the portion of variation  $V$  that is explained by the model inputs either independently or in combination with other factors. For a  $p$ -dimensional input factor, we have

$$Y = f(\mathbf{x}) = f_0 + \sum_{j=1}^p f_j(x_j) + \sum_{j_1=1}^{p-1} \sum_{j_2=j_1+1}^p f_{j_1, j_2}(x_{j_1}, x_{j_2}) + \cdots + f_{j_1, j_2, \dots, j_p}(x_{j_1}, \dots, x_{j_p}), \quad (8)$$

where  $f_0 = E(Y)$ ,  $f_j = E(Y | x_j) - f_0$ ,  $f_{j_1, j_2} = E(Y | x_{j_1}, x_{j_2}) - f_{j_1} - f_{j_2} - f_0 \cdots$

$$\begin{aligned} f_{j_1, j_2, \dots, j_p}(x_1, \dots, x_p) = & E(Y | \mathbf{X}) - \sum_{j=1}^p f_j(x_j) - \sum_{j_1=1}^{p-1} \sum_{j_2=1+j_1}^p f_{j_1, j_2}(x_{j_1}, x_{j_2}) - \cdots \\ & - \sum_{j_1=1}^2 \cdots \sum_{j_{p-1}=1+j_{p-2}}^p f_{j_1, j_2, \dots, j_{p-1}}(x_{j_1}, \dots, x_{j_{p-1}}), \end{aligned} \quad (9)$$

and  $f_0 = E(Y) = \int_{\mathbf{x}} f(\mathbf{x}) dG(\mathbf{x})$  denotes the global mean and  $G(\mathbf{x}) = \prod_{j=1}^p G_j(x_j)$  is the probability distribution of the input factors by assuming that each factor is independent with components  $G_j(x_j)$ . Further details is given in [Xu et al. \(2019b\)](#) and [Farah and Kottas \(2014\)](#). Similarly, the next  $p$  terms  $f_j(x_j)$  are the main effect of input  $x_j$ , providing a measure of the influence of input  $x_j$  on the model output.

Using the independence assumption between the input factors, then the expression  $G(\mathbf{x}_{\sim j}|x_j)$  can be rewritten as  $G(\mathbf{x}_{\sim j})$ . The remaining terms of the decomposition are

the interactions, which quantify the combined influence on the output of two or more inputs taken together. For instance, the first-order interaction is given as

$$f_{j_1, j_2}(x_{j_1}, x_{j_2}) = E(Y|x_{j_1}, x_{j_2}) - f_{j_1}(x_{j_1}) - f_{j_2}(x_{j_2}) - E(Y).$$

Homma and Saltelli (1996) and Sobol (1993) generalise this output decomposition into a summation of partial variances, since each component of eq. (8) is orthogonal such that

$$V = \sum_{j=1}^p V_j + \sum_{1 \leq j_1 < j_2 \leq p} V_{j_1, j_2} + \dots + V_{j_1, j_2, \dots, j_p}, \quad (10)$$

where  $V = \text{var}(Y)$ ,  $V_j = \text{var}(f_j(x_j)) = \text{var}(E(Y|x_j))$ ,  $V_{j_1, j_2} = \text{var}(f_{j_1, j_2}(x_{j_1}, x_{j_2}))$  and  $V_{j_1, j_2, \dots, j_p}$  are higher order terms.

Standardizing equation (10), produces

$$S_j = \frac{V_j}{V}, \quad S_{j_1, j_2} = \frac{V_{j_1, j_2}}{V}, \dots, S_{j_1, j_2, \dots, j_p} = \frac{V_{j_1, j_2, \dots, j_p}}{V},$$

where  $S_j, S_{j_1, j_2}$ , for  $j_1 \neq j_2$ , are denoted as the first and second order sensitivity indices for input  $x_j$ , and interactive factors  $\mathbf{x}_{j_1, j_2}$  respectively. Each index measures the fractional contribution of input  $x_j$  and interactions due to inputs  $x_{j_1}$  and  $x_{j_2}$  to the total output variance, respectively. We also define the total sensitivity index  $S_j^\top$ , an additional related measure given by

$$S_j^\top = 1 - \frac{V_{\sim j}}{V}, \quad j = 1, \dots, p,$$

where  $V_{\sim j} = \text{var}(E(Y | \mathbf{x}_{\sim j}))$  is the total variance of  $\text{var}(f_j(\mathbf{x}))$  due to all inputs except  $x_j$ .

Recall that the posterior predictive density for the multivariate outputs from the computationally demanding model is approximated by the posterior predictive distribution using the relation  $\mathbf{Y}^* = f(\mathbf{x}^*)$  then

$$p(\mathbf{Y}^*|\mathbf{D}) = \int \int p(\mathbf{Y}^*|\Theta) \times p(\Theta|\mathbf{D}) d\Theta,$$

where  $\mathbf{D} = \{\mathbf{Y}, \mathbf{x}\}$  is the training data and  $\Theta = (\mathbf{B}, \Sigma, \tau)$  are the model parameters. Recalling the following identity:

$$E(Y|x_j) = \int_{\{x_{j'}: \{j' \neq j\}\}} f(x_1, \dots, x_j, \dots, x_p) \prod_{\{j': \{j' \neq j\}\}} dG_{j'}(x_{j'}), \quad (11)$$

for each specified value  $x_j$  and

$$E(Y) = \int_{\mathbf{x}} f(\mathbf{x}) \prod_{j=1}^p dG_j(x_j) \quad (12)$$

$$\text{var}(Y) = \int_{\mathbf{x}} f^2(\mathbf{x}) \prod_{j=1}^p dG_j(x_j) - (E(Y))^2. \quad (13)$$

It is possible to numerically calculate all the integrals associated with the following expectations  $(E(Y))^2$ ,  $E\left((E(Y|x_j))^2\right)$  and  $E\left((E(Y|\mathbf{x}_{\sim j}))^2\right)$ . This can be obtained from the posterior distribution of MSGP (see Appendix C for full details). For each MCMC iteration of the MSGP, the posterior distributions for the first-order sensitivity index,  $S_j$ , and the total sensitivity index,  $S_j^\top$ , are computed using the equations below

$$S_j = \frac{\text{var}(E(Y|x_j))}{\text{var}(Y)} = \frac{E\left((E(Y|x_j))^2\right) - (E(Y))^2}{\text{var}(Y)}$$

$$S_j^\top = \frac{\text{var}(Y) - \text{var}(E(Y|\mathbf{x}_{\sim j}))}{\text{var}(Y)} = 1 - \frac{E\left((E(Y|\mathbf{x}_{\sim j}))^2\right) - (E(Y))^2}{\text{var}(Y)}. \quad (14)$$

#### 4.2. Point estimates of the main effects

We describe briefly the Bayesian point estimates for the main effects using the approach of [Oakley and O'Hagan \(2004\)](#); [Farah and Kottas \(2014\)](#). To conserve space, let  $\mathbf{Y} = (Y^{(1)}, \dots, Y^{(m)})$  and  $\mathbf{f} = (f^{(1)}, \dots, f^{(m)})$ . Given any input  $\{\mathbf{x} = (x_1, \dots, x_p), \hat{Y} = f(\mathbf{x})\}$ , and as a follow up from eq. (8), we can determine the main effect of input  $x_j$ , but we need to calculate the following expectations  $E\{E(Y) | D\}$  and  $E\{E(Y | x_j) | D\}$ . This can be obtained by  $E\{E(Y) | D\} = \int_{\mathbf{f}(\mathbf{x})} E(Y)p(\mathbf{f}(x) | D)d\mathbf{f}(\mathbf{x})$  and for each specified value  $x'_j$  of  $j^{\text{th}}$  input

$$E(Y | x_{j'}) = \int_{\{x_j: j' \neq j\}} f(x_1, \dots, x_{j'}, \dots, x_p) \prod_{\{j: j' \neq j\}} dG_{j'}(x_{j'})$$

and similarly,

$$E\{E(Y | x_{j'}) | D\} = \int_{f(x_1, \dots, x_{j'}, \dots, x_p)} E(Y | x_{j'}) p(f(x_1, \dots, x_{j'}, \dots, x_p) | D) \times df(x_1, \dots, x_{j'}, \dots, x_p).$$

See further details of implementing this procedure in [Farah and Kottas \(2014\)](#).

#### 4.3. Simulation approach for computing sensitivity indices

To compute the posterior distributions of sensitivity indices  $S_j$  and  $S_j^\top$ , we shall sample from the posterior distribution of the MSGP emulator parameters. Using the procedure given in [Saltelli \(2002\)](#), generate input sample matrix  $\mathbf{A}_0$  of size  $\mathbf{s} \times p$  such that each row of  $\mathbf{A}_0$  is drawn independently from the input Uniform(-1,1) distribution of the simulator,  $G(x) = \prod_{j=1}^p G_j(x_j)$ . In our example, we have fixed  $\mathbf{s} = 5,000$  and recall that  $p = 24$ . We also generate  $p$  input sample matrices,  $\mathbf{A}_j$ , for  $j = 1, \dots, p$  of size  $\mathbf{s} \times p$  each, such that the  $j^{\text{th}}$  column of matrix  $\mathbf{A}_j$  equals the  $j^{\text{th}}$  column of matrix  $\mathbf{A}_0$ . Similarly, another  $p$  input sample matrices,  $\mathbf{A}_{\sim j}$ , for  $j = 1, \dots, p$  of size  $\mathbf{s} \times p$  each, where matrices  $\mathbf{A}_j$  and  $\mathbf{A}_0$  have all columns in common except at the  $j^{\text{th}}$  column can also be generated.

Using the MSGP emulator, and for each MCMC posterior sample  $\Theta = \{\mathbf{B}, \Sigma, \tau\}$ , we compute the following posterior predictive samples and for each row  $s$  of  $\mathbf{A}_0$ , we calculate both  $\hat{Y}_s$  and  $\hat{Y}_s^2$ .

Then for each row  $s$  of  $\mathbf{A}_j$  and  $s$  of  $\mathbf{A}_{\sim j}$ , we can sample both  $\hat{Y}'_{s,j}$  and  $\hat{Y}'_{s,\sim j}$ , respectively. For  $j = 1, \dots, p$ , estimate the following expectations using the posterior samples:

$$\begin{aligned} E(Y) &\approx \mathbf{s}^{-1} \sum_{s=1}^{\mathbf{s}} \hat{Y}_s, & E\left((E(Y | x_j))^2\right) &\approx \mathbf{s}^{-1} \sum_{s=1}^{\mathbf{s}} \hat{Y}_s \hat{Y}'_{s,j}, \\ E(Y^2) &\approx \mathbf{s}^{-1} \sum_{s=1}^{\mathbf{s}} \hat{Y}_s^2, & E\left((E(Y | x_{\sim j}))^2\right) &\approx \mathbf{s}^{-1} \sum_{s=1}^{\mathbf{s}} \hat{Y}_s \hat{Y}'_{s,\sim j}. \end{aligned}$$

By using the posterior samples above for the required expectations, we can calculate the posterior distributions for the first-order sensitivity indices,  $S_j$  and the total sensitivity indices,  $S_j^\top$  using eq. (14). This procedure allows estimation of the entire distribution for each sensitivity index thereby enabling us to quantify the uncertainty of sensitivity indices.

#### 4.4. Extension to multivariate outputs

In the previous section, we have assumed a univariate output and that input variables are independent which ensures the orthogonality property. Suppose that an input-output system structure is given

$$\mathbf{Y}^{(k)} = f^k = f_0^{(k)} + f_j^{(k)} + f_{\sim j}^{(k)} + f_{j,\sim j}^{(k)} \quad (15)$$

and  $\mathbf{Y} = \{Y^{(1)}, \dots, Y^{(m)}\}$  is a  $m$ -dimensional model output vector. The Sobol decomposition for the  $k^{\text{th}}$  scalar output is  $\mathbf{Y}^{(k)} = f^{(k)}(x_1, x_2, \dots, x_p)$ . [Gamboa et al. \(2013\)](#) and [Lamboni et al. \(2011\)](#) generalise the Sobol sensitivity indices to the case of multivariate outputs which is analogous to the univariate case by decomposing the covariance matrix  $\Omega$  of  $\mathbf{Y}$  into the sum of partial covariance matrices  $\Omega_j$ ,  $\Omega_{j,\sim j}$  and  $\Omega_{\sim j}$  respectively. This corresponds to the main contribution of  $\mathbf{X}_j$ , the interactive contribution of  $\mathbf{x}_j$  and  $\mathbf{x}_{\sim j}$ , and the main contribution of  $\mathbf{x}_{\sim j}$ . Recalling that:  $\mathbf{x}_{\sim j} = \{x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_p\}$ . By taking the covariance of both sides of eq. (15), the total unconditional covariance of  $\mathbf{Y}^{(k)}$  is given as

$$\begin{aligned} \text{Cov}(\mathbf{Y}^{(k)}) &= \text{Cov}\left(f_j^{(k)}, f_j^{(k)} + f_{j,\sim j}^{(k)} + f_{\sim j}^{(k)}\right) + \\ &\quad \text{Cov}\left(f_{j,\sim j}^{(k)}, f_j^{(k)} + f_{j,\sim j}^{(k)} + f_{\sim j}^{(k)}\right) + \text{Cov}\left(f_{\sim j}^{(k)}, f_j^{(k)} + f_{j,\sim j}^{(k)} + f_{\sim j}^{(k)}\right), \quad (16) \end{aligned}$$

where  $(k = 1, \dots, m; j = 1, \dots, p)$ . Expanding the above expression gives

$$\begin{aligned}
\boldsymbol{\Omega} &= \text{Cov}(\mathbf{Y}) = \text{Cov}\left(f^{(1)}, \dots, f^{(m)}\right) \\
&= \text{Cov}_j\left(f_j^{(1)}, \dots, f_j^{(m)}, \mathbf{f}\right) + \text{Cov}_{j,\sim j}\left(f_{j,\sim j}^{(1)}, \dots, f_{j,\sim j}^{(m)}, \mathbf{f}\right) \\
&\quad + \text{Cov}_{\sim j}\left(f_{\sim j}^{(1)}, \dots, f_{\sim j}^{(m)}, \mathbf{f}\right) \\
&= \boldsymbol{\Omega}_j + \boldsymbol{\Omega}_{j,\sim j} + \boldsymbol{\Omega}_{\sim j}
\end{aligned}$$

and  $\mathbf{f} = \left(f_j^{(1)}, \dots, f_j^{(m)}\right) + \left(f_{j,\sim j}^{(1)}, \dots, f_{j,\sim j}^{(m)}\right) + \left(f_{\sim j}^{(1)}, \dots, f_{\sim j}^{(m)}\right)$ . Taking the trace we can calculate the total sensitivity indices

$$S_j = \frac{\text{tr}(\boldsymbol{\Omega}_j)}{\text{tr}(\boldsymbol{\Omega})}, \quad S_j^\top = \frac{\text{tr}(\boldsymbol{\Omega}_j) + \text{tr}(\boldsymbol{\Omega}_{j,\sim j})}{\text{tr}(\boldsymbol{\Omega})}, \quad (17)$$

where  $\text{tr}(\boldsymbol{\Omega}) = \sum_{k=1}^m \mathbf{V}^{(k)}$  and  $\text{tr}(\boldsymbol{\Omega}_j) = \sum_{k=1}^m \text{Cov}\left(f_j^{(k)}, f_j^{(i)}\right) = \sum_{k=1}^m V_j^{(k)}$ . Therefore,  $S_j = \frac{\sum_{k=1}^m V_j^{(k)}}{\sum_{k=1}^m \mathbf{V}^{(k)}}$ . See [Xu et al. \(2019a\)](#); [Xiao and Duan \(2016\)](#) for further details.

#### 4.5. Generalised sensitivity indices for multivariate output using vector projection

We note that the generalised Sobol sensitivity indices are based on a decomposition of the covariance matrix of multiple outputs defined by [Gamboa et al. \(2013\)](#) and [Lamboni et al. \(2011\)](#) above in eq. (17) does not take into consideration the correlation between multiple outputs. This makes it difficult to interpret the sensitivity results. To address this problem, [Xu et al. \(2019b\)](#) developed new generalised indices based on the vector projection of multivariate outputs for measuring the comprehensive effects of the inputs. The method is based on vector projection and dimension normalisation. The approach involves an affine coordinate transformation of the normalised outputs based on the correlations among different variables. Firstly, the multivariate outputs are normalised into a dimensionless unit to remove the interference of different dimensions. Secondly, an affine coordinate system is computed by using the correlations among the different outputs. To obtain new sensitivity indices, we can project the variance contribution vector composed by the individual variance contribution of the input to each output on the total variance vector. The new sensitivity indices are defined as the ratio of the projection of the variance contribution vector to the norm of the total variance vector.

To obtain dimensionless outputs, the output  $\mathbf{Y}^{(k)}$  can be normalized as

$$R(\mathbf{Y}') = \begin{pmatrix} r_{1,1} & r_{1,2} & \cdots & r_{1,m} \\ r_{2,1} & r_{2,2} & \cdots & r_{2,m} \\ \vdots & \vdots & \ddots & \vdots \\ r_{m,1} & r_{m,2} & \cdots & r_{m,m} \end{pmatrix}$$

be the correlation coefficient matrix of normalized multivariate output data  $\mathbf{Y}'$ , where  $r_{i,k} = \frac{\text{Cov}(\mathbf{Y}^{(i)}, \mathbf{Y}^{(k)})}{\sqrt{\text{var}(\mathbf{Y}^{(i)}) \times \text{var}(\mathbf{Y}^{(k)})}}$ ,  $(i, k : 1, \dots, m)$ . Let  $\theta^{(i,k)} = \arccos(r_{(i,k)})$  be the angle ex-

panded by  $\mathbf{Y}^{(i)}$  and  $\mathbf{Y}^{(k)}$ . Then,  $m$ -dimensional outputs  $\mathbf{Y}^{(i)}$  ( $i = 1, \dots, m$ ) expand an affine coordinate system, respectively, with the unit directional vector  $\mathbf{e}_i$  ( $i = 1, \dots, m$ ), and  $\mathbf{e}_i^T \mathbf{e}_k = \cos(\theta^{(i),(k)}) = r_{i,k}$  ( $i, k = 1, \dots, m$ ). From the definition of the angle  $\theta^{(i),(k)}$  expanded by  $\mathbf{Y}^{(i)}$  and  $\mathbf{Y}^{(k)}$  as  $\theta^{(i),(k)} = \arccos(r_{i,k})$  the  $m$ -dimensional affine coordinate system can be determined completely in which the correlations of the multivariate outputs are incorporated. Having constructed an affine coordinate system, the individual variances of each output can be viewed as vectors  $\boldsymbol{\Omega}_j^{\mathbf{Y}^{(i)}} \mathbf{e}_i$  ( $i = 1, \dots, m$ ), and the total variance vector of  $m$ -dimensional outputs can be obtained as  $\boldsymbol{\Omega}_j = \boldsymbol{\Omega}^{\mathbf{Y}^{(1)}} \mathbf{e}_1 + \boldsymbol{\Omega}^{\mathbf{Y}^{(2)}} \mathbf{e}_2 + \dots + \boldsymbol{\Omega}^{\mathbf{Y}^{(m)}} \mathbf{e}_m$ , for ( $j = 1, \dots, p$ )

We define the matrix  $\mathbf{Q}_j$  as

$$\mathbf{Q}_j = \|\boldsymbol{\Omega}_j\| \cos \theta_j = \frac{\langle \boldsymbol{\Omega}_j, \boldsymbol{\Omega} \rangle}{\|\boldsymbol{\Omega}\|} = \frac{\left[ \boldsymbol{\Omega}_j^{(1)}, \dots, \boldsymbol{\Omega}_j^{(m)} \right] R(\mathbf{Y}') \left[ \boldsymbol{\Omega}^{(1)}, \dots, \boldsymbol{\Omega}^{(m)} \right]^T}{\sqrt{\left[ \boldsymbol{\Omega}^{(1)}, \dots, \boldsymbol{\Omega}^{(m)} \right] R(\mathbf{Y}') \left[ \boldsymbol{\Omega}^{(1)}, \dots, \boldsymbol{\Omega}^{(m)} \right]^T}},$$

and  $P_j$  is the projected first-order variance contribution which is analogous to the  $S_j$  defined in eq. (17) and corresponding expression for the total indices  $S_j^T$  is given in Appendix C.

$$\mathbf{P}_j = \frac{\mathbf{Q}_j}{\|\boldsymbol{\Omega}\|} = \frac{\langle \boldsymbol{\Omega}_j, \boldsymbol{\Omega} \rangle}{\|\boldsymbol{\Omega}\|^2}, \quad (18)$$

where  $\theta_j$  is the angle between between  $\boldsymbol{\Omega}$  and  $\boldsymbol{\Omega}_j$ ,  $\langle \cdot, \cdot \rangle$  denotes the inner product of two vectors and  $\|\cdot\|$  is the norm of a vector. The notable properties of the vector projection-based sensitivity indices are given below:

- $|\theta_j| \leq \frac{\pi}{2}, 0 \leq \hat{\mathbf{P}}_j \leq 1$ .
- Suppose there is an input  $X_i$  which has no effect on  $\mathbf{Y}$  but an input  $X_j$  has an effect on  $\mathbf{Y}$ , then  $\hat{\mathbf{P}}_{i,j} = \hat{\mathbf{P}}_j$ .
- Similarly, if  $X_i$  has no effect on  $\mathbf{Y}$ , then the total projected indices  $\hat{\mathbf{P}}_i^T = 0$ .
- $\frac{\boldsymbol{\Omega}_j}{\|\boldsymbol{\Omega}\|} = \frac{\boldsymbol{\Omega}^{\mathbf{Y}^{(1)}}}{\boldsymbol{\Omega}} \mathbf{e}_1 + \frac{\boldsymbol{\Omega}^{\mathbf{Y}^{(2)}}}{\boldsymbol{\Omega}} \mathbf{e}_2 + \dots + \frac{\boldsymbol{\Omega}^{\mathbf{Y}^{(m)}}}{\boldsymbol{\Omega}} \mathbf{e}_m$ .

See further details in [Xu et al. \(2019a\)](#) and [Xu et al. \(2019b\)](#).

It is worth noting that variance-based decomposition is not the only approach for performing sensitivity indices. Alternative approaches exist for identifying essential variables using Gaussian processes. For multivariate Gaussian process emulators, input variables can impact the response through both their mean functions and the correlation structure. Hence, in our model description above, the relative importance of the input variables can also be determined by the relative magnitude of the corresponding correlation parameters  $\boldsymbol{\tau}$ . A large value for  $\boldsymbol{\tau}_k$  indicates a weak correlation. This approach was used by [Liu et al. \(2009\)](#) and is very similar to automatic relevance determination. In this approach, the inverse length scale parameter of each input variable could be used as a proxy for variable relevance [Williams and Rasmussen \(1996\)](#); [Paananen et al. \(2019\)](#). This method is also called sparse Bayesian learning in some literature because of its capability to learn sparse

solutions to linear problems, or it can also be viewed as an in-built mechanism for feature selection. Another related approach to automatic relevance determination for active variable selection is through the sparsifying spike-and-slab priors [Savitsky et al. \(2011\)](#) and in the work of [Crawford et al. \(2019\)](#) where Kullback-Leibler divergence between the marginal distribution of the rest of the variables to their conditional distribution is fixed to a nominal distribution.

## 5. Results

### 5.1. Test cases

We illustrate our sensitivity algorithm with three test cases:

(i) An analytic function called the Sobol  $\mathbf{g}$ -function (non-correlated input) is a common example used in global sensitivity analysis ([Saltelli, 2002](#); [Saltelli et al., 2010](#)). The  $g$  function is a strongly non-monotonic, non-additive function of independent factors  $x_i$ , assumed identically and uniformly distributed in the unit cube  $I^p = \{x \mid 0 \leq x_i \leq 1; i = 1, 2, \dots, p\}$

$$\mathbf{g}_{Sobol}(x_1, x_2, \dots, x_p) = \prod_{i=1}^p \mathbf{g}_i(x_i),$$

where  $\mathbf{g}_i(x_i) = \frac{|4x_i - 2| + a_i}{1 + a_i}$ . The sensitivity indices for function  $\mathbf{g}$  can be obtained analytically from

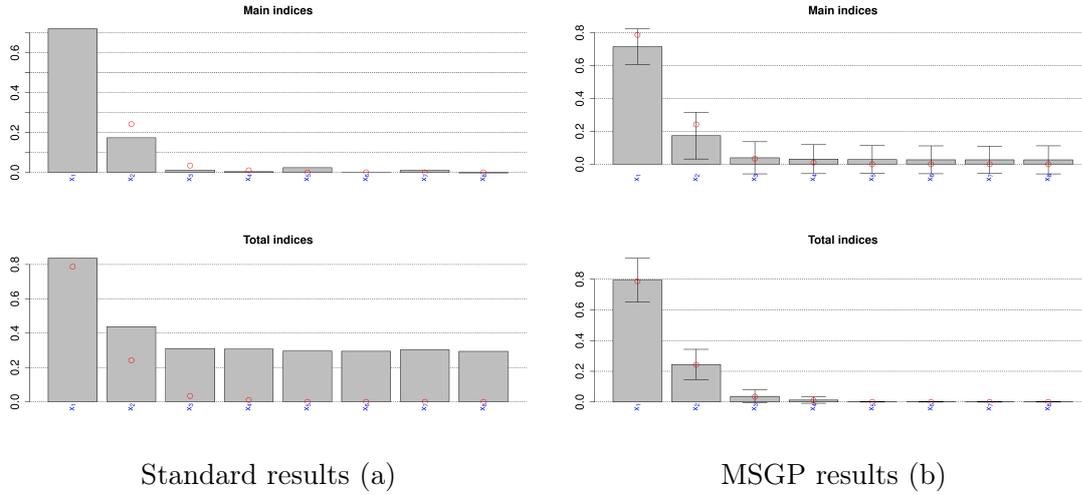
$$V_i = V_{X_i}(E_{X_{\sim i}}(Y \mid X_i)) = \frac{1/3}{(1 + a_i)^2}$$

$$V_{Ti} = V_i \prod_{j \neq i} (1 + V_j)$$

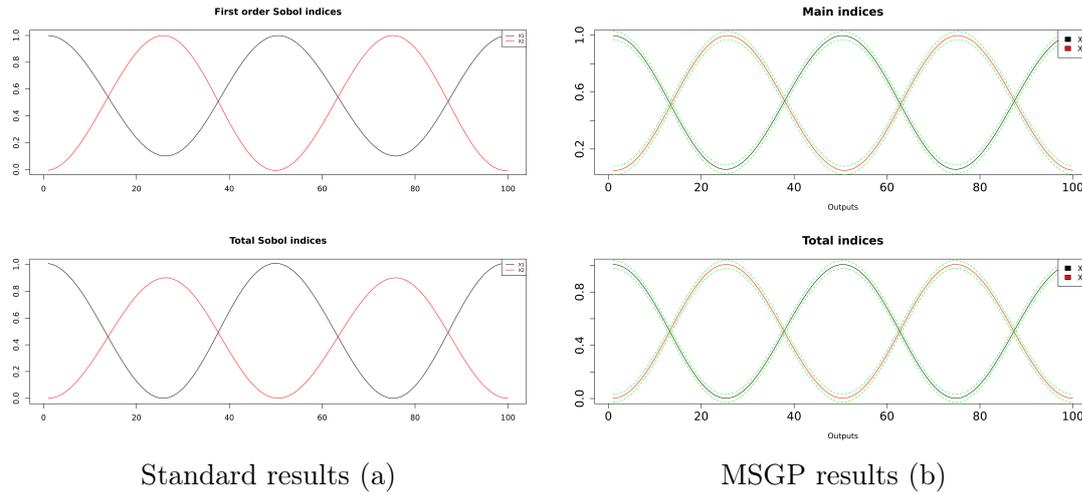
$$V = \prod_{i=1}^k (1 + V_i) - 1.$$

The Sobol  $\mathbf{g}$ -function was also used in the works of [Gamboa et al. \(2020\)](#) to perform the numerical comparisons between the standard Pick-Freeze approach and Chatterjee's estimators. In this case, we consider  $p = 8$  factors, with  $a = (0, 1, 4.5, 9, 99, 99, 99, 99)$ . We can see from Figure 2(a) that the lower the coefficient  $a$ , the larger the sensitivity indices. This result is comparable to the analytical values (given by the red dots) and the approach implemented in the R package *sensitivity* using the function *sobolshap-knn* which computes both the Sobol indices and Shapley effects ([Broto et al., 2020](#)). Both the standard approach using the *R* package and the MSGP methods slightly under-predicted the analytical values for the main indices. In contrast, the standard approach over-predicted the indices for the total sensitivity.

(ii) The second example is a multivariate case study using the arctangent temporal function with two independent variables and 100 outputs. The functional support is on  $[0, 2\pi]$ , where  $q$  is the number of discretization steps of the functional outputs ([Auder, 2011](#)). There are two factors, all following the uniform distribution between  $[-7, 7]$ . The result is given in Figure 3 below showing the comparison between the standard approach and MSGP for both main and total indices. We could see that both results are similar.



**Fig. 2.** Main and total indices for the Sobol  $g$ -function using our algorithm. The red dots denote the analytical values.



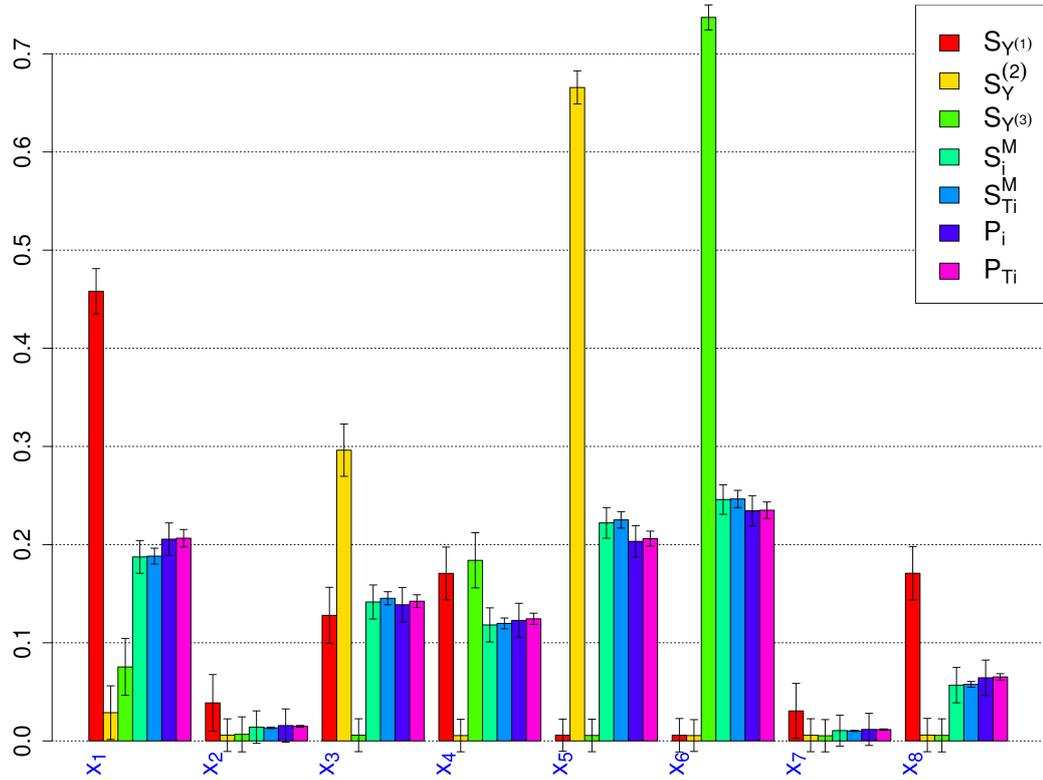
**Fig. 3.** Main and total indices for the multivariate test cases using arctangent temporal function which has 100 functional outputs with two input variables from a uniform distribution.

(iii) This third example is a nonlinear model used in [Xu et al. \(2019b\)](#) paper. The function is given below.

$$\begin{aligned}
 Y^{(1)} &= 0.03 - \frac{1.905x_1x_2^2}{x_3x_7} - \frac{0.565x_1x_2^2}{x_4x_8} \\
 Y^{(2)} &= x_5x_3 - 1.185x_1x_2 \\
 Y^{(3)} &= x_6x_4 - 0.75x_1x_2
 \end{aligned}$$

The inputs are independent and follow normal distributions. The correlation coefficients

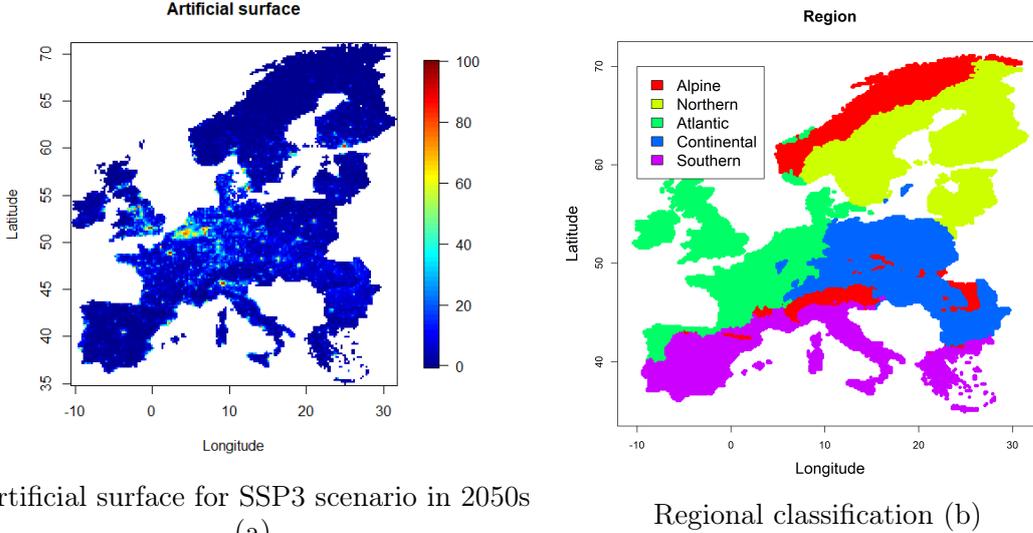
between the multivariate outputs and the distribution parameters are given in Table 8 and Table 9 of Appendix F. The results of the main and total sensitivity indices using the covariance decomposition method, and the vector projection-based method, are given in Figure 4. The results are comparable to Figure 4 in Xu et al. (2019b).



**Fig. 4.** Main and total indices for the nonlinear model used in Xu et al. (2019b) using MSGP algorithm.

### 5.2. Application to IAP2 simulation

We first describe the procedure for building the MSGP then apply the MSGP method to compute sensitivity indices of the IAP2 simulation models under various scenarios to explore a wide range of uncertainty. The effect of model inputs was evaluated on all 18 spatially-distributed model outputs using MSGP. We note that it is practically infeasible to perform a fully spatiotemporal sensitivity analysis for the IAP2 model. We, therefore, follow the works of Kebede et al. (2015) and Savall et al. (2019), where we aggregated the output data into six spatial extents (see Figure 5(b)), namely Europe-wide and five



Artificial surface for SSP3 scenario in 2050s

(a)

Regional classification (b)

**Fig. 5.** Sample output data from the IMPRESSIONS IAP2 simulation under RCP4.5 and HadGEM2-ES-RCA4 climate model and the classification of the five European regions used in the analyses.

European regions to reduce computational expense. We subsampled 600 data points from each of the five scenarios (baseline, SSP1, SSP3, SSP4, SSP5) and three timeslices (2020s, 2050s, 2080s) giving 9,000 ( $5 \times 3 \times 600$ ) samples per each output. The training data  $\mathbf{D} = \{\mathbf{X}_{n \times p}, \mathbf{Y}_{n \times m}\}$ , where  $n = 9,000$ ,  $p = 24$  and  $m = 18$  for different cases was considered in this paper. We used the multivariate sparse Gaussian process (MSGP) described in Section 3.1 to fit the data and compute the new sensitivity indices using the procedure highlighted in Section 4.1.

### 5.2.1. A procedure for building the MSGP

We fit the MSGP first by spatially averaging the entire output data (called EU-wide average datasets) and rescaling the input data to the range  $[-1, 1]$ . The input data in this analysis represent the 24 model parameters. We scaled the data by centring the column data with their respective minimum values and divided by their range, such that  $\mathbf{x}' = \left(2 \times \frac{(\mathbf{x} - x_{\min})}{(x_{\max} - x_{\min})}\right) - 1$ . The output data was transformed by normalising it to have a zero mean and unit variance. Standardising these data around the centre 0 with a standard deviation of 1 is essential because the simulated outputs have different units. This is a necessary step in data preprocessing to meet model assumptions and reduce bias in the estimated sensitivity results. This procedure is essential to reduce the effect of the magnitude/unit of the response variables on the computed sensitivity indices. We note that variables with larger magnitude would make larger contributions to the overall output variances.

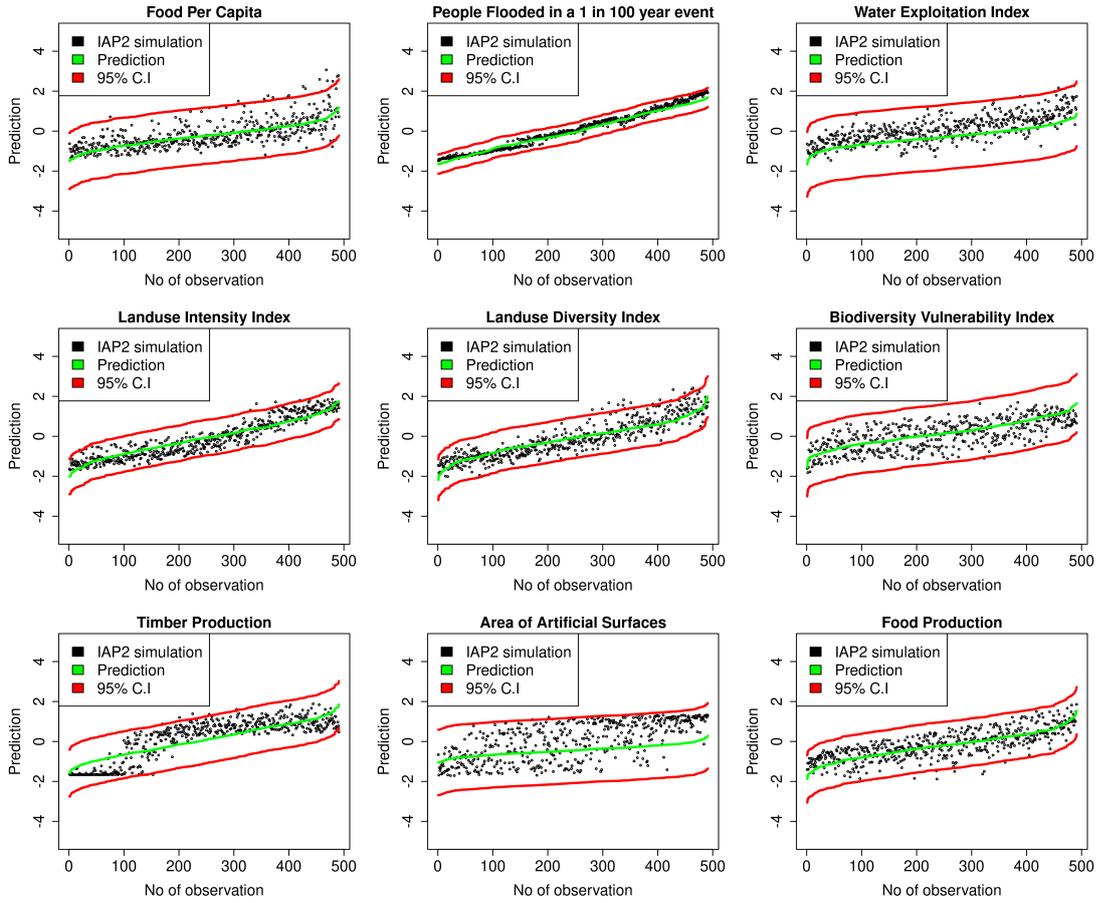
The model parameters are estimated using an MCMC procedure (see Section 3.3)

**Table 3.** Model summary: cross-validated proportion of variance  $P$ , root mean squared error  $\rho$  showing the performance of the MSGP emulators for Matérn<sub>5/2</sub> with Wendland correlation functions; and with Truncated power correlation for four different levels of sparsity for the EU-wide average data.

Models	1		2		3		4		5	
$\omega$ (%)	Matérn Wendland		80		90		95		99	
Output	<b>P</b>	$\rho$	P	$\rho$	P	$\rho$	P	$\rho$	P	$\rho$
Food Capita	<b>0.54</b>	0.68	0.52	0.70	0.52	<b>0.46</b>	0.52	0.46	0.51	0.70
People Flooded	<b>0.96</b>	0.21	0.95	0.23	0.95	<b>0.17</b>	0.95	0.17	0.95	0.23
Water Expl Ind	<b>0.40</b>	0.78	0.34	0.81	0.34	<b>0.59</b>	0.34	0.59	0.34	0.81
Landuse Int Ind	<b>0.82</b>	0.42	0.81	0.44	0.81	<b>0.33</b>	0.81	0.33	0.81	0.44
Landuse Diversity	<b>0.76</b>	0.49	0.74	0.51	0.74	<b>0.38</b>	0.74	0.38	0.74	0.51
Biodiversity	<b>0.55</b>	0.68	0.47	0.73	0.47	<b>0.58</b>	0.47	0.58	0.47	0.74
Timber Prod	<b>0.69</b>	0.56	0.66	0.58	0.66	<b>0.48</b>	0.66	0.48	0.66	0.59
Artificial Surfaces	<b>0.47</b>	0.73	0.33	0.82	0.33	<b>0.69</b>	0.33	0.69	0.34	0.82
Food Prod	<b>0.73</b>	0.52	0.66	0.59	0.66	<b>0.59</b>	0.66	0.59	0.66	0.58
Carbon Stock	<b>0.61</b>	0.63	0.56	0.66	0.56	<b>0.66</b>	0.56	0.66	0.57	0.66
Irrigation Usage	<b>0.57</b>	0.65	0.51	0.70	0.51	<b>0.70</b>	0.51	0.70	0.52	0.68
Intensive Arable	<b>0.68</b>	0.57	0.65	0.59	0.65	<b>0.59</b>	0.65	0.59	0.65	0.59
Intensive Grass	<b>0.74</b>	0.51	0.69	0.56	0.69	<b>0.56</b>	0.69	0.56	0.69	0.56
Extensive Grass	<b>0.54</b>	0.68	0.52	0.70	0.52	<b>0.70</b>	0.52	0.70	0.53	0.69
Very Ext Grass	<b>0.39</b>	0.78	0.32	0.83	0.32	<b>0.83</b>	0.32	0.83	0.32	0.82
Unmanaged Land	<b>0.55</b>	0.67	0.52	0.69	0.52	<b>0.69</b>	0.52	0.69	0.53	0.70
Managed Forest	<b>0.65</b>	0.59	0.61	0.63	0.61	<b>0.63</b>	0.61	0.63	0.61	0.63
Unmanaged Forest	<b>0.77</b>	0.48	0.74	0.51	0.74	<b>0.51</b>	0.74	0.51	0.75	0.50
Overall	<b>0.63</b>	0.59	0.59	0.63	0.59	<b>0.56</b>	0.59	0.56	0.59	0.62

to approximate the posterior distributions of the unknown parameters  $\hat{\Theta} = \{\hat{\mathbf{B}}, \hat{\Sigma}, \hat{\tau}\}$  using the MNIW prior. The parameter  $\tau$  is sampled using an adaptive Metropolis-Hastings algorithm with a multivariate normal random walk proposal distribution centred around the most current estimates of  $\tau$ . We fixed the target acceptance rate to the recommended 0.234. See [Vihola \(2012\)](#) for full details.

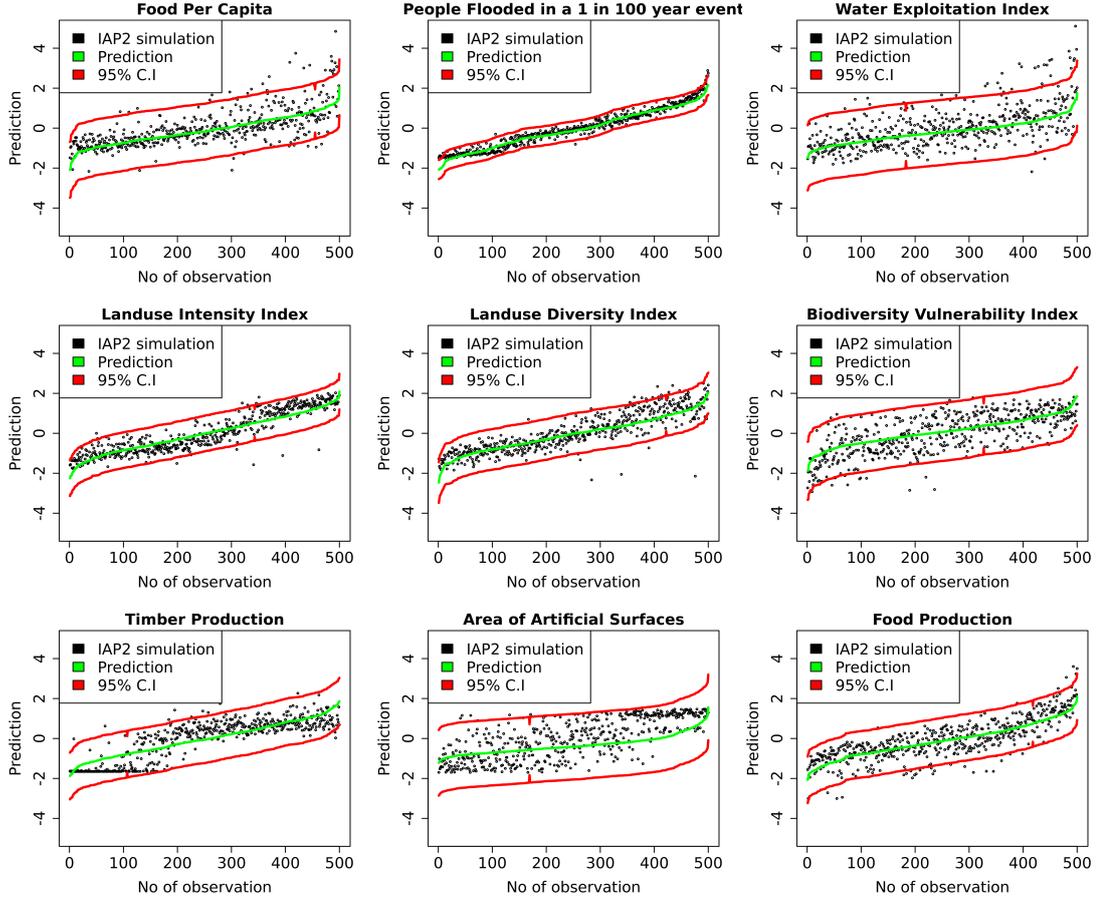
We assess the convergence of the MCMC procedure using standard diagnostic tests (see Appendix E). For example, Table 7 in Appendix E gives the potential scale reduction factor (PSRF), which is an estimated factor by which the scale of the current distribution for the target distribution can be reduced if the simulations are continued for a large number of samples. It can also be used as an estimate for the R-hat metric ([Vehtari et al., 2019](#)). To compute these values, we ran three parallel chains with different starting values. We sampled 50,000 MCMC iterations with a thinning frequency of 25 using the algorithm described in the previous section. We discarded the first 1,000 samples as burn-in. We see that the PSRF values for most of the parameters are very close to 1, which indicates they are likely to converge to the correct target distribution. According to the literature, a recommended approach is to achieve  $\text{PSRF} < 1.1$ , although this depends on the nature of the problem. A large  $\text{PSRF} > 1.2$  indicates no convergence and the possible presence of a multimodal marginal posterior distribution in which different chains may have converged to different modes. See [Gelman et al. \(1992\)](#); [Brooks and Gelman \(1998\)](#) for further details. We also included both the density and trace plots of the estimated parameter  $\tau$  in Figures 16 and 17 respectively in Appendix E as a further visual diagnostic. We note that the magnitude of the estimated scale  $\tau$  parameters is



**Fig. 6.** Comparison of the multivariate GP emulator performance with IAP2 simulation for 500 randomly sample points under the SSP4 scenario in the 2050s for EU-wide aggregated data; simulation (black) and emulator predictions (green) with 95% C.I. (red). Note: the outputs have been normalised to have a zero mean and unit variance.

also a proxy for sensitivity indices.

To implement the above procedure on the EU-wide average data, we replaced the dense  $n \times n$  matrix  $\mathfrak{R}$  with a sparse approximation using compactly supported correlation functions such as the Bohman and truncated power functions and products of Matérn $_{5/2}$  and Wendland functions, as described in Appendix B. It is infeasible to jointly estimate the levels of sparsity  $\omega$  (percentage of off-diagonal nonzero elements), the cut-off  $c$  and other parameters. However, it is possible to fix these at reasonable values with relatively little effect on predictions and model performance. Here, we used  $K$ -fold cross-validation to determine the appropriate level of sparsity to be used for the sensitivity analyses in the next section. We used root mean square errors ( $\rho$ ), and the proportion of variance ( $P$ ) explained, to compare the efficiency at fixed sample sizes. Table 3 compares the results.



**Fig. 7.** Comparison of the multivariate GP emulator performance with IAP2 simulation for 500 randomly sample points under the SSP1 scenario in the 2050s for EU-wide aggregated data; simulation (black) and emulator predictions (green) with 95% C.I. (red). Note: the outputs have been normalised to have a zero mean and unit variance.

We observe that all the fitted models performed relatively well with the proportion of variance explained  $P$  ranging from 0.34 – 0.96 and  $\rho$  0.21 – 0.83 for various sparsity levels. We note that the higher the  $P$  values, the better the model and vice-versa for the  $\rho$ . The *People flooded in a 1 in 100-year event* variable has the largest  $P$  and smallest  $\rho$  values respectively for various degrees of sparsity and is thus more predictable than the other variables. On the other hand, the *very extensive grassland* variable seems less predictable compared to the others. Generally, the models made substantial savings by using sparse models, with the time taken to fit each model decreasing with the degree of sparsity.

The results are generally similar across different levels of sparsity. We can deduce that sparsity does not necessarily degrade the performance of the models. The results are relatively similar, except that the models under the truncated power correlation

function are much faster to train. We have fixed the value of  $\omega$  as 90%, which minimised the RMSE  $\rho$  value for subsequent analyses in this paper.

Figures 6 and 7 are the graphical comparisons between the multivariate MSGP emulator with the IAP2 simulation for different outputs for model 3 in Table 3. The results are shown for the SSP4 and SSP1 scenarios in the 2050s for EU-wide aggregated data. The emulators reproduce the patterns well in most of the variables, with predictions falling within the 95% C.I. For instance, the variable *people flooded in a 1 in 100-year event* has a narrow confidence band. The predictions are relatively precise in most of the points. As suggested by Table 3, and seen again here, this variable is well-predicted compared to other variables. The *food per capita* and *water exploitation index* are relatively better predicted compared to the *land use diversity index* with several points falling outside the confidence bands in Figures 6. There are some mismatches between the emulator predictions and simulations in *timber production*. The results are also similar in Figure 7.

We now consider the regional analysis. Five different MSGP models were also fitted, each for the five European regional aggregates. This analysis allows us to compute regional sensitivity indices. Table 4 shows the model comparison. The results are generally similar across the regions. The *people flooded in a 1 in 100-year event*, *water exploitation index* and *artificial surfaces* are still well-predicted compared to the other variables, with high  $P$  and low  $\rho$  values, respectively. The corresponding regional sensitivity indices are provided in Figure 15 in Appendix A.

### 5.3. Comparison of MSGP with linear regression models.

Here, we have tested a small sample of our data based on the EU-wide average dataset and applied a linear regression to each output dataset. We also computed the relative performance metric to compare the compactly supported model when we fixed the value of  $\omega$  as 90% (see Table 4) to the linear regression model in Table 10. The index is defined below as

$$\alpha = \frac{R_l - R_m}{R_l} \times 100,$$

where  $R_l$  and  $R_m$  denote the RMSE values for the regression and MSGP model, respectively, if  $\alpha > 0$ , then the MSGP model is far better than the linear model, and vice versa. The results are very similar when  $\alpha = 0$ . We note that the performance of the MSGP model on the same dataset is better than that of linear regression in 13 out of the 18 tested outputs while linear regression does better for 5/18 outputs. See Table 10 in Appendix F for details.

### 5.4. Sensitivity results

For the practical issue of computationally demanding computer models, we show that the substitution of the original model by the MGSP models facilitates the rapid estimation of these global sensitivity indices with precision at a reasonable computational cost.

Figure 8 shows the main effects for all the input variables for some of the outputs. The input value is normalized between  $[-1, 1]$ . A large variation in the main effect plot is an indication of the greater influence of such input on the output. We see that population has a greater influence on *food per capita*, *people flooded in 1 in 100-year land use intensity*

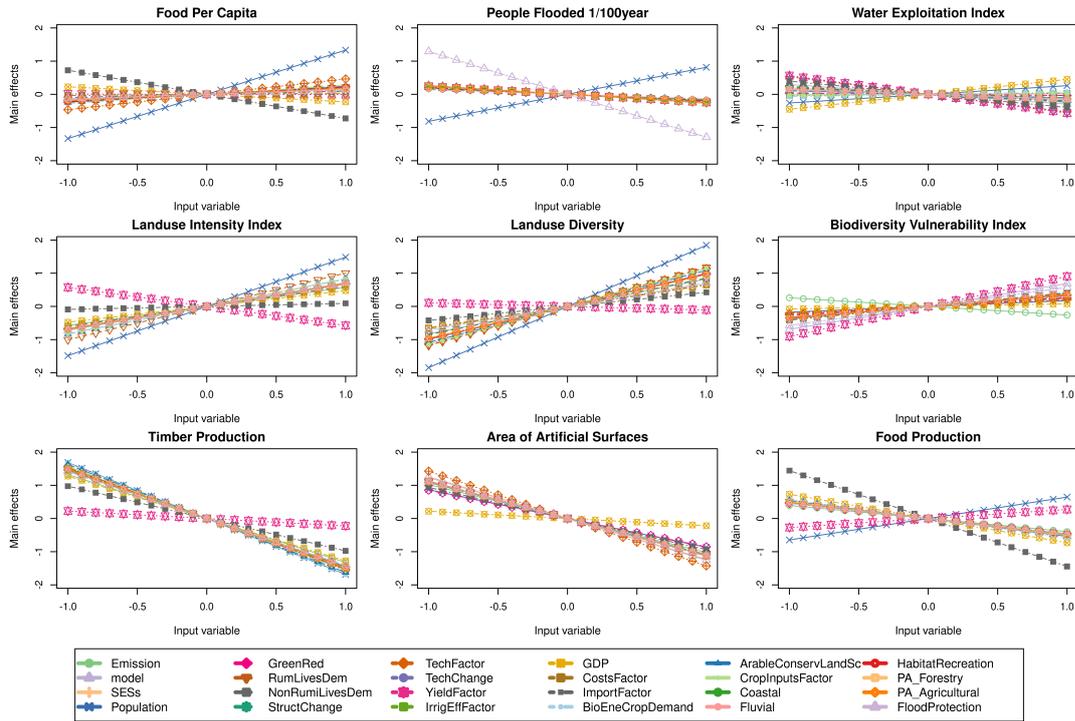
**Table 4.** Model summary: cross-validated proportion of variance  $P$ , root mean squared error  $\rho$  showing the performance of MSGP emulators for 5 regional aggregations for all the scenarios and timeslices. AL=Alpine, NO=Northern; AT=Atlantic; CO=Continental; SO=Southern.

	AL		NO		AT		CO		SO	
	P	$\rho$	P	$\rho$	P	$\rho$	P	$\rho$	P	$\rho$
Food Capita	0.41	0.78	0.53	0.69	0.41	0.77	<b>0.61</b>	<b>0.62</b>	0.79	0.46
People Flooded	0.76	0.49	0.79	0.46	0.88	0.35	<b>0.75</b>	<b>0.50</b>	0.77	0.48
Water Exploit	0.35	0.81	0.25	0.87	0.29	0.84	<b>0.35</b>	<b>0.80</b>	0.32	0.83
Landuse Inten	0.70	0.55	0.70	0.55	0.83	0.41	<b>0.78</b>	<b>0.47</b>	0.56	0.67
Landuse Diver	0.64	0.60	0.69	0.56	0.30	0.84	<b>0.57</b>	<b>0.66</b>	0.58	0.65
Biodiversity Vul	0.54	0.68	0.40	0.78	0.46	0.74	<b>0.47</b>	<b>0.73</b>	0.61	0.63
Timber Prod	0.54	0.68	0.55	0.67	0.55	0.67	<b>0.72</b>	<b>0.54</b>	0.41	0.76
Artificial Surfaces	0.32	0.82	0.32	0.83	0.32	0.82	<b>0.35</b>	<b>0.81</b>	0.33	0.82
Food Prod	0.49	0.72	0.57	0.66	0.49	0.71	<b>0.70</b>	<b>0.54</b>	0.72	0.53
Carbon Stock	0.34	0.82	0.49	0.72	0.51	0.71	<b>0.57</b>	<b>0.66</b>	0.29	0.85
Irrigation	0.41	0.77	0.27	0.86	0.49	0.72	<b>0.50</b>	<b>0.71</b>	0.45	0.74
Intensive Arable	0.46	0.73	0.63	0.61	0.40	0.77	<b>0.59</b>	<b>0.64</b>	0.45	0.74
Intensive Grass	0.60	0.64	0.42	0.77	0.76	0.50	<b>0.54</b>	<b>0.68</b>	0.70	0.55
Extensive Grass	0.44	0.75	0.32	0.83	0.44	0.74	<b>0.51</b>	<b>0.70</b>	0.38	0.79
Very Ext. Grass	0.30	0.84	0.24	0.87	0.34	0.82	<b>0.20</b>	<b>0.90</b>	0.40	0.77
Unmanaged Land	0.62	0.62	0.57	0.66	0.44	0.74	<b>0.37</b>	<b>0.79</b>	0.30	0.83
Managed For	0.54	0.68	0.55	0.67	0.53	0.68	<b>0.69</b>	<b>0.56</b>	0.42	0.76
Unmanaged For	0.67	0.58	0.61	0.63	0.71	0.54	<b>0.71</b>	<b>0.54</b>	0.61	0.63
Overall	0.51	0.70	0.49	0.70	0.51	0.69	<b>0.55</b>	<b>0.66</b>	0.50	0.69

and *landuse diversity* indices. An increase in population causes these indices to increase. Import factor has a negative effect on *food per capita*, *timber* and *food production* while the influence of yield factor also reduces with an increase in *landuse intensity index*. The plots for the remaining outputs are given in Figure 13 of Appendix A.

Figures 9 and 10 are bar plots showing the posterior estimates for the first-order and total sensitivity indices for the EU-wide aggregated data for some selected outputs. The relative contribution of each input  $X_j$  can be either positive or very small negative values close to zero. This usually occurs when the posterior samples for the variances are computed using differences in expectations, especially when the Monte Carlo sample size is not large enough. Therefore, the absolute magnitudes of the sensitivity indices can be used to rank the inputs. We focus on the four largest values for the first-order indices using their posterior mean estimates. We see that the *food per capita* is highly sensitive to changes in food imports, population, agricultural mechanisation and agricultural yields.

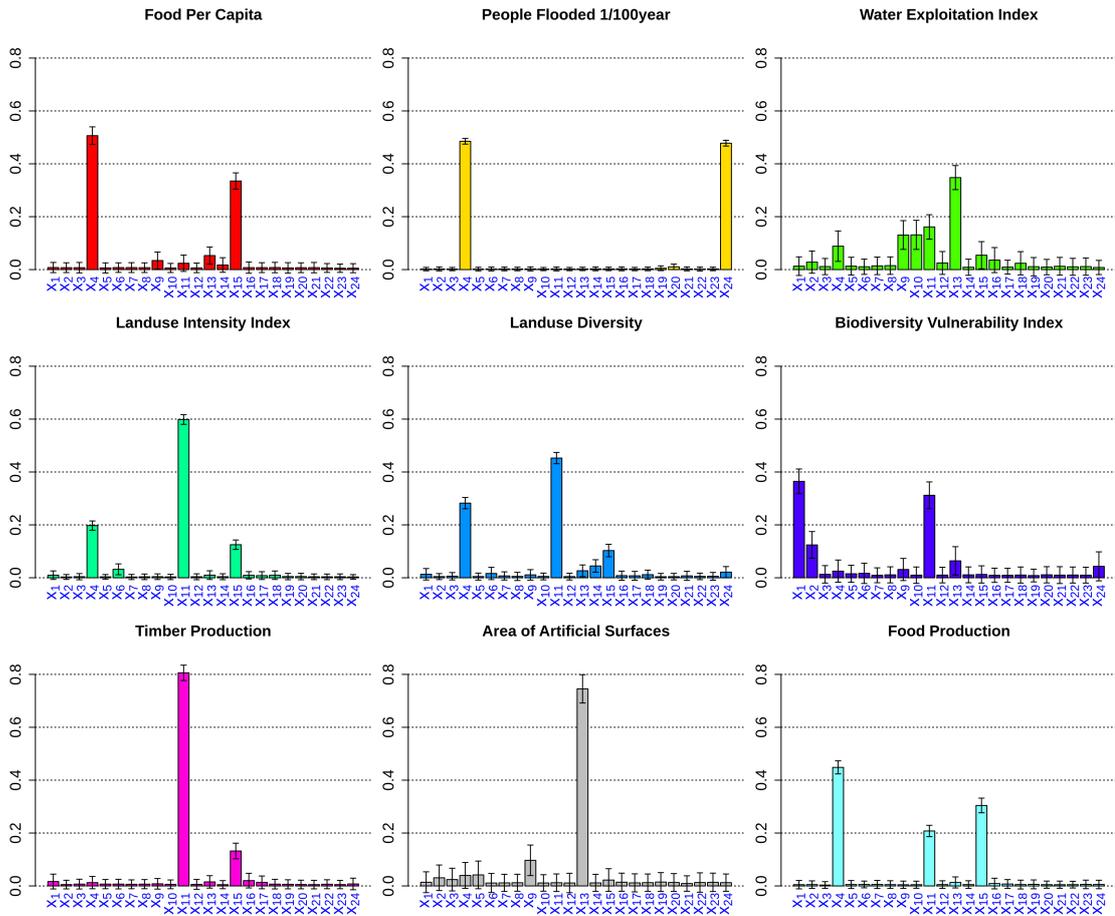
The *people flooded in a 1 in 100-year event* is mostly affected by the level of flood protection and population change. It is also completely nonsensitive to the other remaining variables. The *water exploitation index* is affected mostly by GDP, change in agricultural yields, water-saving due to technological change and food imports. We note that change in agricultural yields, population and change in food imports are the most important variables affecting the *land use intensity index*, *land use diversity index* and *timber production*. The emissions scenario, climate model, and socio-economic scenarios are mostly non-significant factors, except for the *biodiversity vulnerability index*, which shows some degree of sensitivity to many of the inputs. This reflects its position at the bottom of the modelling chain, where it is influenced by most of the other sectoral models within the IAP2 model.



**Fig. 8.** Main effect plots for EU-wide aggregated data; simulation (black) and emulator predictions (green) with 95% C.I. (red). Note: the outputs have been normalised to have a zero mean and unit variance.

For the total sensitivity indices in Figure 10, there are significant differences between the main and total indices emphasising the contribution of interactions among the inputs. We noted that the values of the main indices are different from the corresponding total effect indices. Some of the sectoral outputs are still sensitive to the variables identified under the main indices as highly significant. For example, the *food per capita* output is still most affected by changes in food imports and population, but it also shows sensitivity to fluvial flood events, preferences for living close to urban or rural areas, and changes in agricultural inputs to reduce diffuse pollution.

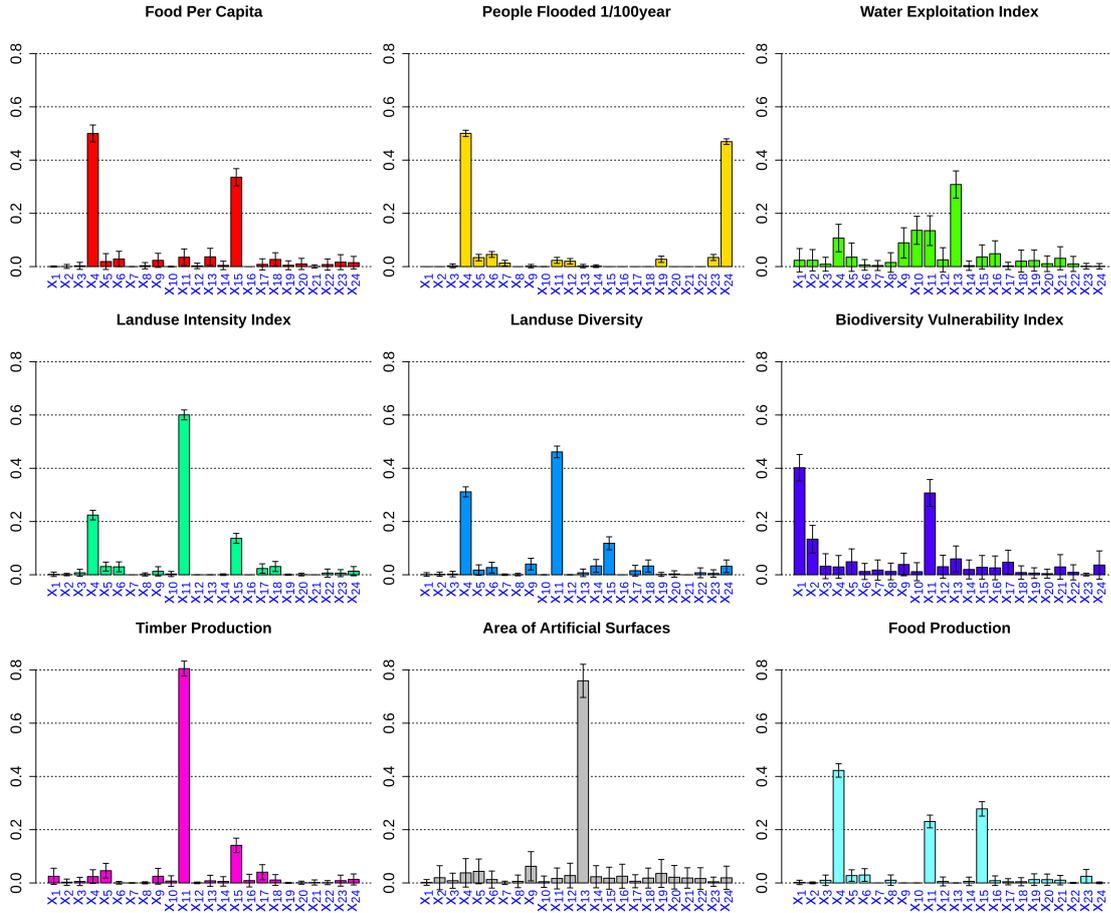
Similar to the main indices, changes in agricultural yields have the predominant influence on the *land use intensity index* and *land use diversity index* and *timber production*, but these also show sensitivity to many other inputs, such as bioenergy crop demand and land set aside for conservation, in the total effect indices. Other variables also show greater sensitivity to other parameters due to multiple interactions between the variables. For instance, the *biodiversity vulnerability index* is now sensitive to other parameters such as changes in food imports and bioenergy production, set-aside, change in irrigation efficiency and fluvial flood events. It is interesting that it shows sensitivity to lots of parameters but that there is a large uncertainty associated with the interaction between factors. This is an indication that interaction between factors played a major



**Fig. 9.** Barplots of the posterior mean estimates of the first-order indices from the multivariate sensitivity analyses for some selected IAP2 outputs for EU-wide average data.

role in these sets of data.

The generalised sensitivity indices based on the vector projection for multivariate outputs, which measures the comprehensive effect of the inputs on the multiple outputs, are given in Figure 11. This takes into consideration the correlation between multiple outputs when computing the indices. The results are similar to those reported in Figures 9 and 10. Clearly, the change in agricultural yields, population, GDP and food import variables are significant for both main and total indices. The less significant variables are the level of flood protection, preference for living close to urban or rural areas and the emissions scenario. We note that the change in agricultural yields dominates the given results. We can see from Figure 11 that the standard case (without output correlation) has identified variable  $X_{11}$  (Yield factor) as the most important parameter whereas the vector projected approach (with output correlation) recognized  $X_4$  (Change in population) as the most important variable. This shows that the correlation between the outputs needs to be considered to have a complete effect of the inputs on the uncertainty

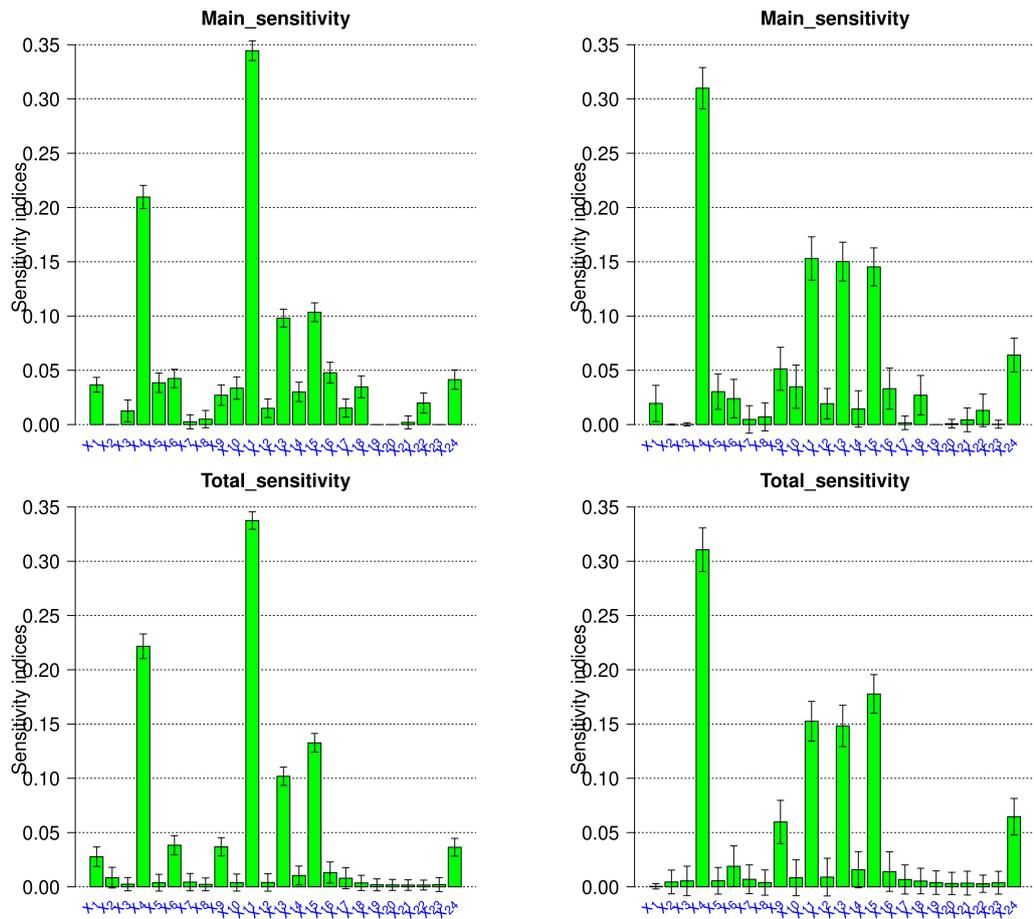


**Fig. 10.** Barplots of the posterior mean estimates of the total indices from the multivariate sensitivity analyses for some selected IAP2 outputs for EU-wide average data.

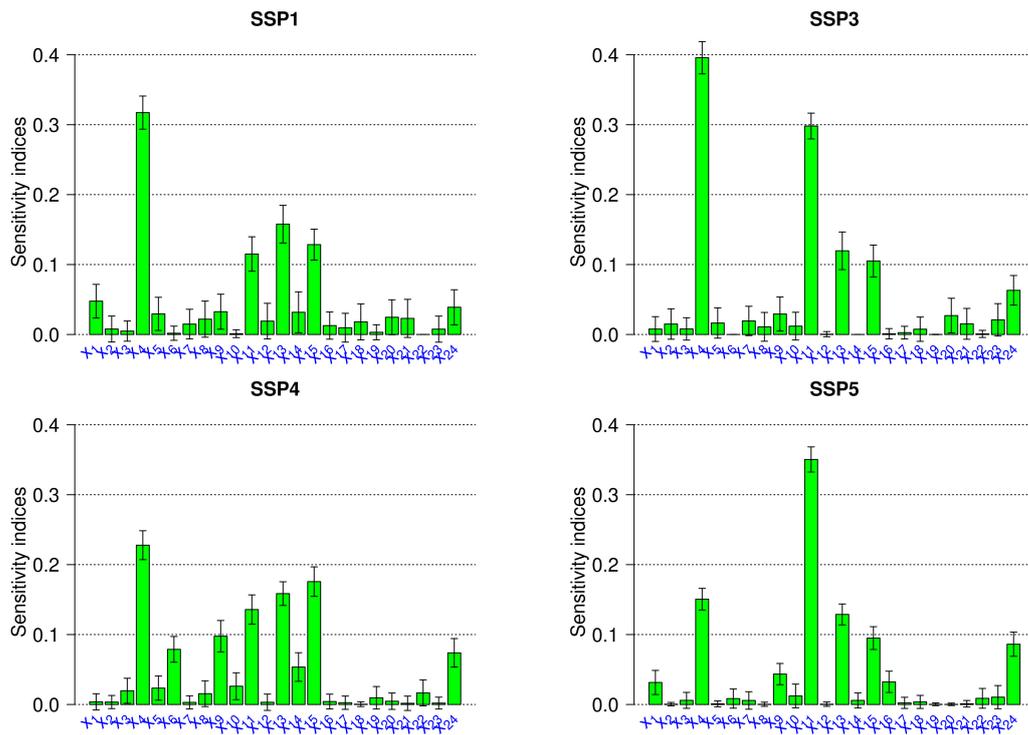
of the multivariate outputs.

Next, we study the implications of the socio-economic scenarios on the sensitivity indices. Figure 12 shows the barplot for the posterior distribution of the main and total effects. There is a divergence in patterns across the scenarios. Furthermore, a different combination of variables is selected in each scenario as the essential parameters.

We also performed the sensitivity analysis for the regional data. The corresponding regional sensitivity indices are provided in Figure 15 in Appendix A. The results are similar across the regions to those reported in Figure 11. The first four most sensitive parameters are the change in agricultural yields, GDP, food imports and population. Some variables such as water savings due to behavioural change, reducing diffuse source pollution from agriculture and protected area change are generally nonsensitive to most of the variables and thus may require less attention when parameterising the IAP2 model. This set of variables could be fixed to reduce the dimension of the problem.



**Fig. 11.** Boxplots of the posterior mean for the multivariate generalised indices for EU-wide average data. Note: The left plots correspond to the standard approach while the right plots are computed from the vector projected method using affine transformation.



**Fig. 12.** Boxplots of the posterior mean and their uncertainty estimates for the aggregated main indices for SSP1, SSP3, SSP4 and SSP5 respectively.

## 6. Conclusion

We proposed a new methodology for Bayesian global sensitivity analysis for large multivariate output data with a focus on computationally demanding models. The method was applied to a large dataset with a large number of correlated variables from the IMPRESSIONS IAP2 model. The goal was to estimate the expected outputs of such models given the distribution of inputs. These models are usually heavily parameterised, and the relative contribution of these parameters is often ignored in uncertainty quantification. The main objective of this paper was to investigate how changes in each parameter affect the sectoral and cross-sectoral indicators of the IAP2 model.

The computational cost of performing sensitivity analysis for complex models depends on the number of input variables making the approach intractable for high dimensional input data. To address the problem, we have developed an efficient strategy for computing sensitivity indices for large dimensional and multivariate data. By exploiting the sparsity of covariance matrices, using compactly supported correlation functions, we have shown how the parameters from the multivariate sparse Gaussian process model could be estimated. The procedure requires neither storing nor inverting the prohibitive full  $n \times n$  covariance matrix, thus significantly reducing the computational and memory complexity. The MSGP emulator provided a faster and cheaper approximation to the real data.

Farah and Kottas (2014); Kaufman et al. (2011) have earlier used this approach, but we have extended it to the case of multivariate outputs and adapted it for computing sensitivity indices for a high-dimensional problem which has not been done before. Our proposed sensitivity method that utilized cross-correlation between the outputs provides improved estimates of sensitivity indices. We also explored parallelisation in posterior computation by distributing the likelihood estimation on different nodes using the multicore computing environment. For instance, a comparison between a standard GP using the 'DiceKriging' library in R was applied to the same sample of the data to test the performance of the MSGP model. It takes around  $\sim 2.5$  hours to train a sample of 4500 data for a single output. The MSGP model was able to produce similar results within  $\sim 20$  minutes for 9000 training samples with 18 outputs using 1000 MCMC iterations. This combined effort enabled us to achieve massive scalability in the calculation of sensitivity indices, as seen in Table 3.

We applied the proposed methods on three different test cases the non-monotonic Sobol g-function Saltelli (2002), the arctangent temporal function for multivariate applications Auder (2011) and a nonlinear model used in Xu et al. (2019b) paper. Our method provides a good approximation of the sensitivity results. Next, the method was applied to the IAP2 model for the computation of main and total sensitivity indices whether or not the input is correlated. We observed that changes in agricultural yields, food imports, GDP and population have large indices and are highly significant. However, variables such as water savings due to behavioural change, reducing diffuse source pollution from agriculture, and protected area change have negligible effects and could be fixed.

We have applied spatial aggregation as a strategy to reduce the data dimension. The effect of this on the MSGP and its implications for the outcomes of the sensitivity analysis is not clear. Similarly, it would be interesting to demonstrate this proposed method for computing sensitivity indices for spatiotemporal data without aggregation.

The analysis provides a thorough assessment of the effects of input uncertainty on the IAP2 model predictions. This enables a better understanding and interpretation of the complex relationships between the input drivers and the sectoral output indicators under the different socio-economic scenarios for the five European regions. Such information is essential for further refinement and application of the IAP2 model for use in climate change adaptation planning across Europe.

## Acknowledgements

The authors are grateful to the Associate Editor and the reviewers for their useful comments which significantly improved the quality of the paper. The authors also gratefully acknowledge the support of EPSRC and NERC grants EP/R01860X/1, EP/S00159X/1, EP/V022636/1 and NE/T004002/1.

## Data availability

The data and code that support the findings of this study are available from the following GitHub repository: [https://github.com/wolemi2/JRSSC\\_paper\\_code](https://github.com/wolemi2/JRSSC_paper_code).

## References

- Alvarez, M. A. and Lawrence, N. D. (2011) Computationally efficient convolved multiple output Gaussian processes. *The Journal of Machine Learning Research*, **12**, 1459–1500.
- Auder, B. (2011) *Classification et modélisation de sorties fonctionnelles de codes de calcul: application aux calculs thermo-hydrauliques accidentels dans les réacteurs à eau pressurisés (REP)*. Ph.D. thesis, Paris 6.
- Brooks, S. P. and Gelman, A. (1998) General methods for monitoring convergence of iterative simulations. *Journal of computational and graphical statistics*, **7**, 434–455.
- Broto, B., Bachoc, F. and Depecker, M. (2020) Variance reduction for estimation of shapley effects and adaptation to unknown input distribution. *SIAM/ASA Journal on Uncertainty Quantification*, **8**, 693–716.
- Burhenne, S., Jacob, D. and Henze, G. P. (2011) Sampling based on sobol sequences for monte carlo techniques applied to building simulations. In *Building Simulation 2011: 12th Conference of International Building Performance Simulation Association, Sydney, Australia, Nov*, 14–16.
- Carnell, R. and Carnell, M. R. (2019) Package lhs. URL <http://cran.stat.auckland.ac.nz/web/packages/lhs/lhs.pdf>, **780**.
- Cheng, K., Lu, Z. and Zhang, K. (2019) Multivariate output global sensitivity analysis using multi-output support vector regression. *Structural and Multidisciplinary Optimization*, **59**, 2177–2187.

- Crawford, L., Flaxman, S. R., Runcie, D. E. and West, M. (2019) Variable prioritization in nonlinear black box methods: A genetic association case study. *The annals of applied statistics*, **13**, 958.
- Da Veiga, S., Gamboa, F., Iooss, B. and Prieur, C. (2021) *Basics and Trends in Sensitivity Analysis: Theory and Practice in R*. SIAM.
- Datta, A., Banerjee, S., Finley, A. O. and Gelfand, A. E. (2016a) On nearest-neighbor Gaussian process models for massive spatial data. *Wiley Interdisciplinary Reviews: Computational Statistics*, **8**, 162–171.
- Datta, A., Banerjee, S., Finley, A. O., Hamm, N. A. and Schaap, M. (2016b) Nonseparable dynamic nearest neighbor Gaussian process models for large spatio-temporal data with an application to particulate matter analysis. *The annals of applied statistics*, **10**, 1286.
- Eddelbuettel, D. and Sanderson, C. (2014) Rcpparmadillo: Accelerating r with high-performance C++ linear algebra. *Computational Statistics & Data Analysis*, **71**, 1054–1063.
- Farah, M. and Kottas, A. (2014) Bayesian inference for sensitivity analysis of computer simulators, with an application to radiative transfer models. *Technometrics*, **56**, 159–173.
- Finley, A. O., Datta, A., Cook, B. C., Morton, D. C., Andersen, H. E. and Banerjee, S. (2017) Applying nearest neighbor Gaussian processes to massive spatial data sets forest canopy height prediction across tanana valley alaska. *arXiv preprint arXiv:1702.00434*.
- Furrer, R., Genton, M. G. and Nychka, D. (2006) Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, **15**, 502–523.
- Gamboa, F., Gremaud, P., Klein, T. and Lagnoux, A. (2020) Global sensitivity analysis: a new generation of mighty estimators based on rank statistics. *arXiv preprint arXiv:2003.01772*.
- Gamboa, F., Janon, A., Klein, T. and Lagnoux, A. (2013) Sensitivity indices for multivariate outputs. *arXiv preprint arXiv:1303.3574*.
- Gelman, A., Rubin, D. B. et al. (1992) Inference from iterative simulation using multiple sequences. *Statistical science*, **7**, 457–472.
- Gneiting, T. (2001) Criteria of p̄glya type for radial positive definite functions. *Proceedings of the American Mathematical Society*, **129**, 2309–2318.
- Gneiting, T. (2002) Compactly supported correlation functions. *Journal of Multivariate Analysis*, **83**, 493–508.
- Harrison, P. A., Dunford, R. W., Holman, I. P., Cojocar, G., Madsen, M. S., Chen, P.-Y., Pedde, S. and Sandars, D. (2019) Differences between low-end and high-end climate change impacts in europe across multiple sectors. *Regional environmental change*, **19**, 695–709.

- Harrison, P. A., Dunford, R. W., Holman, I. P. and Rounsevell, M. D. (2016) Climate change impact modelling needs to include cross-sectoral interactions. *Nature climate change*, **6**, 885.
- Harrison, P. A., Holman, I. P. and Berry, P. (2015) Assessing cross-sectoral climate change impacts, vulnerability and adaptation: an introduction to the climsave project. *Climatic Change*, **128**, 153–167.
- Heaton, M. J., Datta, A., Finley, A. O., Furrer, R., Guinness, J., Guhaniyogi, R., Gerber, F., Gramacy, R. B., Hammerling, D., Katzfuss, M. et al. (2019) A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, **24**, 398–425.
- Homma, T. and Saltelli, A. (1996) Importance measures in global sensitivity analysis of nonlinear models. *Reliability Engineering & System Safety*, **52**, 1–17.
- Horn, R. A. and Johnson, C. R. (2012) *Matrix analysis*. Cambridge university press.
- Kaufman, C. G., Bingham, D., Habib, S., Heitmann, K., Frieman, J. A. et al. (2011) Efficient emulators of computer experiments using compactly supported correlation functions, with an application to cosmology. *The Annals of Applied Statistics*, **5**, 2470–2492.
- Kaufman, C. G., Schervish, M. J. and Nychka, D. W. (2008) Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association*, **103**, 1545–1555.
- Kebede, A., Dunford, R., Mokrech, M., Audsley, E., Harrison, P., Holman, I., Nicholls, R., Rickebusch, S., Rounsevell, M., Sabaté, S. et al. (2015) Direct and indirect impacts of climate and socio-economic change in europe: a sensitivity analysis for key land-and water-based sectors. *Climatic change*, **128**, 261–277.
- Lamboni, M., Monod, H. and Makowski, D. (2011) Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models. *Reliability Engineering & System Safety*, **96**, 450–459.
- Liu, F., West, M. et al. (2009) A dynamic modelling strategy for Bayesian computer model emulation. *Bayesian Analysis*, **4**, 393–411.
- Maddox, W. J., Balandat, M., Wilson, A. G. and Bakshy, E. (2021) Bayesian optimization with high-dimensional outputs. *Advances in Neural Information Processing Systems*, **34**, 19274–19287.
- Moreaux, G. (2008) Compactly supported radial covariance functions. *Journal of Geodesy*, **82**, 431–443.
- Moreno-Muñoz, P., Artés, A. and Alvarez, M. (2018) Heterogeneous multi-output Gaussian process prediction. *Advances in neural information processing systems*, **31**.
- Oakley, J. E. and O’Hagan, A. (2004) Probabilistic sensitivity analysis of complex models: a Bayesian approach. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **66**, 751–769.

- Overstall, A. M. and Woods, D. C. (2016) Multivariate emulation of computer simulators: model selection and diagnostics with application to a humanitarian relief model. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **65**, 483–505.
- Paananen, T., Piironen, J., Andersen, M. R. and Vehtari, A. (2019) Variable selection for Gaussian processes via sensitivity analysis of the posterior predictive distribution. In *The 22nd International Conference on Artificial Intelligence and Statistics*, 1743–1752. PMLR.
- Robert, C. P., Casella, G. and Casella, G. (1999) *Monte Carlo statistical methods*, vol. 2. New York, Springer.
- Rougier, J. (2007) Lightweight emulators for multivariate deterministic functions. *Unpublished, available at <http://www.maths.bris.ac.uk/~mazjcr/lightweight1.pdf>*.
- Rougier, J. (2008) Efficient emulators for multivariate deterministic functions. *Journal of Computational and Graphical Statistics*, **17**, 827–843.
- Saltelli, A. (2002) Making best use of model evaluations to compute sensitivity indices. *Computer Physics Communications*, **145**, 280–297.
- Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S. and Wu, Q. (2019) Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices. *Environmental modelling & software*, **114**, 29–39.
- Saltelli, A., Aleksankina, K., Becker, W., Fennell, P., Ferretti, F., Holst, N., Li, S. and Wu, Q. (2019) Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. *Computer physics communications*, **181**, 259–270.
- Sanderson, C. and Curtin, R. (2016) Armadillo: a template-based C++ library for linear algebra. *Journal of Open Source Software*, **1**, 26.
- Santner, T. J., Williams, B. J. and Notz, W. I. (2013) *The design and analysis of computer experiments*. Springer Science & Business Media.
- Savall, J. F., Franqueville, D., Barbillon, P., Benhamou, C., Durand, P., Taupin, M.-L., Monod, H. and Drouet, J.-L. (2019) Sensitivity analysis of spatio-temporal models describing nitrogen transfers, transformations and losses at the landscape scale. *Environmental Modelling & Software*, **111**, 356–367.
- Savitsky, T., Vannucci, M. and Sha, N. (2011) Variable selection for nonparametric Gaussian process priors: Models and computational strategies. *Statistical science: a review journal of the Institute of Mathematical Statistics*, **26**, 130.
- Sobol, I. M. (1993) Sensitivity estimates for nonlinear mathematical models. *Mathematical modelling and computational experiments*, **1**, 407–414.
- South, L. F., Karvonen, T., Nemeth, C., Girolami, M., Oates, C. et al. (2020) Semi-exact control functionals from sard’s method. *arXiv preprint arXiv:2002.00033*.

- Stein, M. L., Chi, Z. and Welty, L. J. (2004) Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **66**, 275–296.
- Stocki, R. (2005) A method to improve design reliability using optimal latin hypercube sampling. *Computer Assisted Mechanics and Engineering Sciences*, **12**, 393.
- Svenson, J., Santner, T., Dean, A. and Moon, H. (2014) Estimating sensitivity indices based on Gaussian process metamodels with compactly supported correlation functions. *Journal of Statistical Planning and Inference*, **144**, 160–172.
- Taylor-Rodriguez, D., Finley, A. O., Datta, A., Babcock, C., Andersen, H.-E., Cook, B. D., Morton, D. C. and Banerjee, S. (2018) Spatial factor models for high-dimensional and large spatial data: An application in forest variable mapping. *arXiv preprint arXiv:1801.02078*.
- Van der Vaart, A. W. (2000) *Asymptotic statistics*, vol. 3. Cambridge university press.
- Vecchia, A. V. (1988) Estimation and model identification for continuous spatial processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, **50**, 297–312.
- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B. and Bürkner, P.-C. (2019) Rank-normalization, folding, and localization: An improved R for assessing convergence of MCMC. *arXiv preprint arXiv:1903.08008*.
- Vihola, M. (2012) Robust adaptive Metropolis algorithm with coerced acceptance rate. *Statistics and Computing*, **22**, 997–1008.
- Williams, C. and Rasmussen, C. (1996) Gaussian processes for regression. *Advances in Neural Information Processing Systems*.
- Xiao, H. and Duan, Y. (2016) Sensitivity analysis of correlated inputs: Application to a riveting process model. *Applied Mathematical Modelling*, **40**, 6622–6638.
- Xiao, S., Lu, Z. and Wang, P. (2018) Multivariate global sensitivity analysis based on distance components decomposition. *Risk Analysis*, **38**, 2703–2721.
- Xu, L., Lu, Z., Li, L. and Shi, Y. (2019a) Sensitivity analysis method for model with correlated inputs and multivariate output and its application to aircraft structure. *Computer Methods in Applied Mechanics and Engineering*, **355**, 373–404.
- Xu, L., Lu, Z. and Xiao, S. (2019b) Generalized sensitivity indices based on vector projection for multivariate output. *Applied Mathematical Modelling*, **66**, 592–610.
- Zhang, B., Konomi, B. A., Sang, H., Karagiannis, G. and Lin, G. (2015) Full scale multi-output Gaussian process emulator with nonseparable auto-covariance functions. *Journal of Computational Physics*, **300**, 623–642.
- Zhang, L., Datta, A. and Banerjee, S. (2019) Practical Bayesian modeling and inference for massive spatial data sets on modest computing environments. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, **12**, 197–209.

# Appendices

## A. LHS sampling details and additional Figures

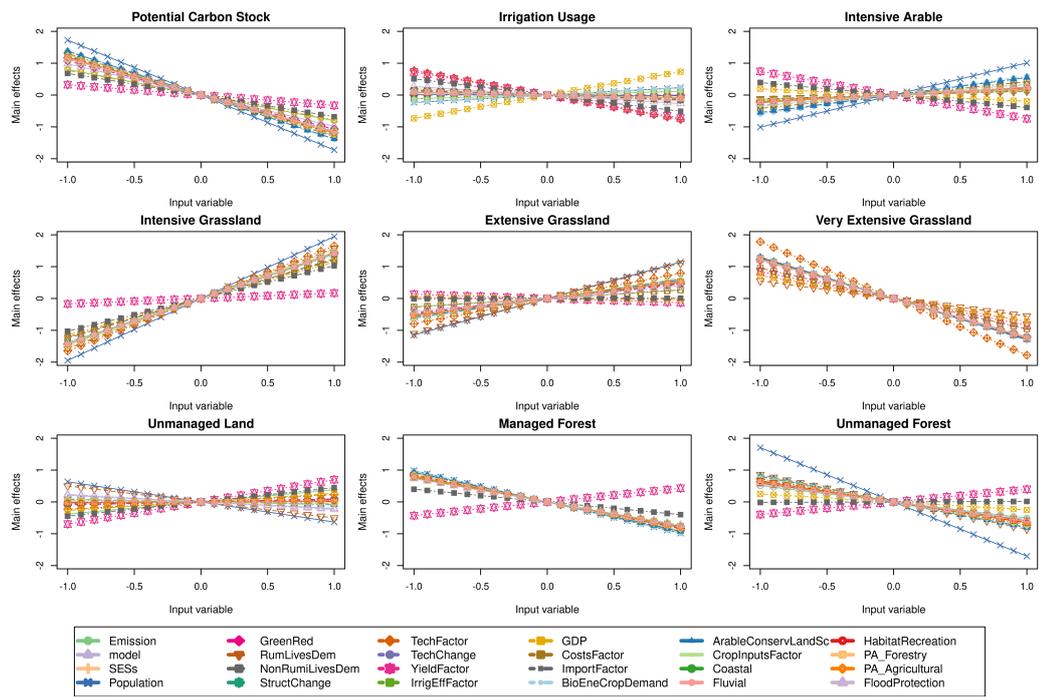
The three emission levels used in this analysis are denoted as 0 = *RCP2.6*; 1 = *RCP4.5*; 2 = *RCP8.5*. For all three emissions, we have a total of ten climate models (ie, RCP 2.6, 4.5 and 8.5 have 3, 3 and 4 climate models, respectively.) The following ten climate/e-emission combinations used for the analyses are given in Table 5 below. We also have five socio-economic scenarios (SSPs) denoted as 0 = "SSP1 / We are the world"; 1 = "SSP3 / Icarus"; 2 = "SSP5 / Should I Stay Or Should I Go"; 3 = "SSP4 / Riders on the Storm"; 4 = "Baseline" and three time slices. Together, we have performed 30,000 simulation runs ( $2000 \times 3 \times 5$ ). To train the MSGP emulators, we subsample 600 data points randomly from 2000 emission/climate combinations for each time slice and SSPs, making 9000 samples ( $600 \times 3 \times 5$ ). Each sample represents a matrix of  $\mathbf{y}_{n0,m}$  spatial outputs with  $n0 = 23871$  grid points and  $m = 18$  which is the dimension of output data. To further pre-process the data, we have averaged the data EU-wide and regionally. The output data for emulation is denoted as  $\mathbf{Y}_{9000 \times 18}$  for each of the six averages (EU-wide, five regions). This produces an array of  $9000 \times 18 \times 6$  output data.

**Table 5.** The following are the ten emission/climate combinations used for analyses.

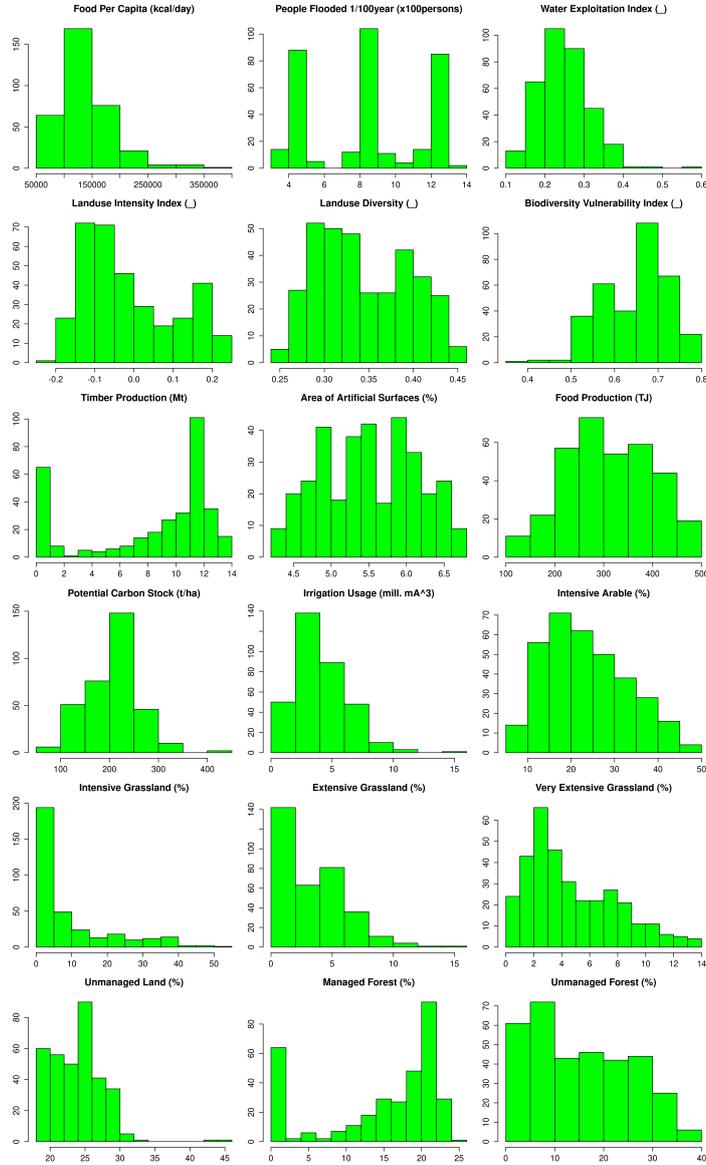
RCPs	
2.6	0 = EC-EARTH RCA4; 1 = MPIESMLR-REMO; 2 = NorESM1-M RCA4
4.5	0 = HadGEM2-ES RCA4; 1 = MPI-ESM-LR CCLM4; 2 = GFDL-ESM2M RCA4
8.5	0=HadGEM2-ESRCA4; 1=CanESM2-CanRCM4; 2=IPSL-CM5A-MRWRF; 3=GFDL-ESM2

**Table 6.** IMPRESSIONS IAP2 parameter ranges for the four socio-economic scenarios from 2020s to 2080s. The 20 continuous variables were sampled from a uniform distribution.

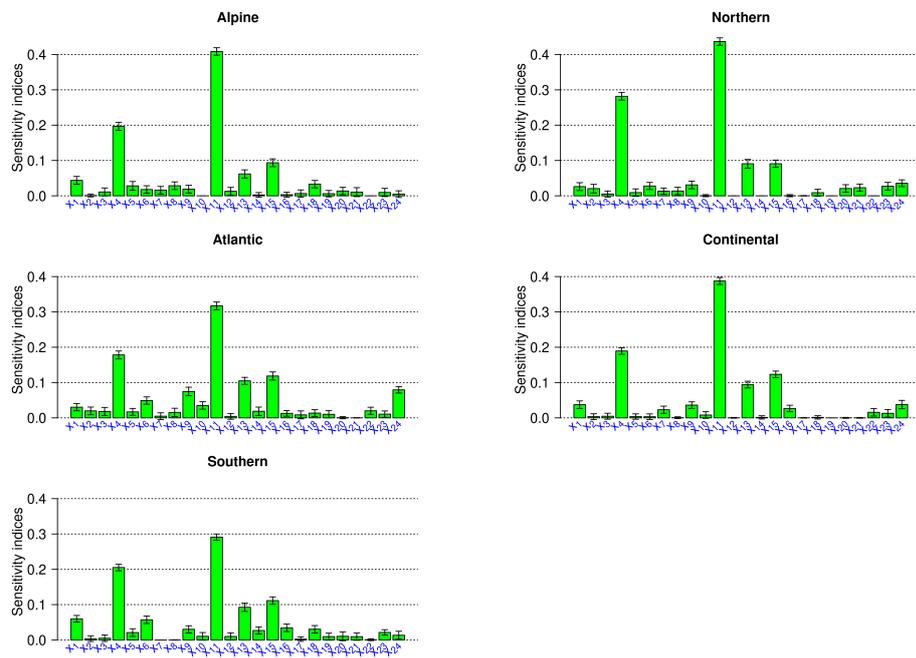
	Variables	SSP1	SSP3	SSP5	SSP4
X <sub>1</sub> *	Emission	{0, 1, 2}	{0, 1, 2}	{0, 1, 2}	{0, 1, 2}
X <sub>2</sub> *	Model	{0,1, 2, 3}	{0, 1, 2, 3}	{0, 1, 2, 3}	{0, 1, 2, 3}
X <sub>3</sub> *	SSPs	{0}	{1}	{2}	{3}
X <sub>4</sub>	Population	[3, 16]	[-37, 1]	[5, 69]	[-18, 3]
X <sub>5</sub>	GreenRed	[3, 5]	[2, 5]	[3, 5]	[1, 5]
X <sub>6</sub>	RumLivesDem	[-81, -5]	[-27, 38]	[-10, 94]	[-25, 39]
X <sub>7</sub>	NonRumiLivesDem	[-33, 5]	[2, 54]	[29, 97]	[4, 54]
X <sub>8</sub>	StructChange	[-5, 55]	[-12, 13]	[-24, 14]	[-15, 24]
X <sub>9</sub>	TechFactor	[5, 125]	[-30, 1]	[5, 125]	[4, 125]
X <sub>10</sub>	TechChange	[-10, 77]	[-10, 65]	[2, 87]	[-10, 65]
X <sub>11</sub>	YieldFactor	[-22, 0.7]	[-34, -4]	[7, 105]	[5, 106]
X <sub>12</sub>	IrrigEffFactor	[4, 66]	[-37, -3]	[4, 66]	[4, 66]
X <sub>13</sub>	GDP	[9, 427]	[17, 66]	[21, 1312]	[19, 309]
X <sub>14</sub>	CostsFactor	[90, 253]	[108, 467]	[59, 110]	[143, 452]
X <sub>15</sub>	ImportFactor	[-8, 3]	[-6, 1]	[.4, 15]	[1.1, 8.2]
X <sub>16</sub>	BioEneCropDemand	[.7, 18]	[1.8, 24]	[1, 21]	[1, 18]
X <sub>17</sub>	ArableConservLand	[3, 6.3]	[2.2, 3]	[0.0 ,2.2]	[3, 6.3]
X <sub>18</sub>	CropInputsFactor	[1, 2]	[.93, 1]	[.93, 1.2]	[.93, 1.2]
X <sub>19</sub>	Coastal	[11, 500]	[13, 500]	[12, 500]	[14, 500]
X <sub>20</sub>	Fluvial	[11, 500]	[13, 500]	[11, 500]	[9, 500]
X <sub>21</sub>	HabitatRecreation	[-100,100]	[-100,100]	[-100,100]	[-100,100]
X <sub>22</sub>	PA Forestry	[0,100]	[0,100]	[0,100]	[0,100]
X <sub>23</sub>	PA Agriculture	[0,100]	[0,100]	[0,100]	[0,100]
X <sub>24</sub> *	FloodProtection	{0,1}	{0,1}	{0,1}	{0,1}



**Fig. 13.** Main effect plots for EU-wide aggregated data; simulation (black) and emulator predictions (green) with 95% C.I. (red). Note: the outputs have been normalised to have a zero mean and unit variance.



**Fig. 14.** Histograms of spatially aggregated IAP2 simulation for the 18 selected outputs, averaged across the three emission scenarios and ten climate models for SSP4. Note: (□) denotes unitless.



**Fig. 15.** Boxplots of the posterior mean and their uncertainty estimates for the aggregated main indices for Alpine, Northern, Atlantic, Continental and Southern regional averaged data.

## B. MCMC sampling details

We follow the procedure below to generate samples from the posterior distribution  $p(\boldsymbol{\tau}|\mathbf{Y})$  and make a prediction at a new location. To sample  $\boldsymbol{\tau}$ , we implement the adaptive Metropolis algorithm of [Vihola \(2012\)](#) to speed up the mixing of the target distribution. Suppose the value of tuning parameter  $\gamma \in (0, 1]$  and given a fixed target acceptance rate  $\alpha^*$  and symmetric proposal density  $w$  such that  $w(\mathbf{x}) = \hat{w}(\|\mathbf{x}\|)$  for all  $\mathbf{x} \in \mathbb{R}^d$  and  $\Upsilon_1 \in \mathbb{R}^{d \times d}$  be a lower-diagonal matrix with positive diagonal elements and  $\{\omega_n\}_{n \geq 1} \subset (0, 1]$ . Then at  $j^{\text{th}}$  iteration of the MCMC:

- Compute  $\boldsymbol{\tau}^{(j)} = \boldsymbol{\tau}^{(j-1)} + \Upsilon_{j-1} \mathbf{u}_j$ , where  $\mathbf{u}_j \sim w$  is an independent random vector
- Accept the proposal with probability  $\alpha_j := \min \left\{ 1, \frac{p(\boldsymbol{\tau}')p(\mathbf{Y}|\boldsymbol{\tau}')}{p(\boldsymbol{\tau}_{j-1})p(\mathbf{Y}|\boldsymbol{\tau}_{j-1})} \right\}$ .
- If the proposal  $\boldsymbol{\tau}'$  is accepted, set  $\boldsymbol{\tau}_j = \boldsymbol{\tau}'$ . Otherwise, set  $\boldsymbol{\tau}_j = \boldsymbol{\tau}_{j-1}$
- Compute the Cholesky factor  $\Upsilon_j$  that satisfy  $\Upsilon_j \Upsilon_j^T = \Upsilon_{j-1} \left( I + \omega_j (\alpha_j - \alpha^*) \frac{\mathbf{u}_j \mathbf{u}_j^T}{\|\mathbf{u}_j\|^2} \right) \Upsilon_{j-1}^T$ , where  $\Upsilon_j$  is the lower-diagonal matrix with positive diagonal elements and  $I \in \mathbb{R}^{d \times d}$  stands for an identity matrix.

Once the posterior samples from  $\boldsymbol{\tau}^{(j)}$  have been drawn. We can easily obtain the predictive distribution of  $Y^{(j)}$  given  $\mathbf{h}(\mathbf{x}_0)$  and  $\mathbf{H}(\mathbf{x})$ . Recall that the joint conditional distribution of  $\mathbf{Y}(\mathbf{x}_0)$  at a new location  $\mathbf{x}_0$  and  $\mathbf{Y}(\mathbf{x})$  is given as a matrix  $t$ -distribution  $p(\mathbf{Y}_0|\mathbf{Y}, \boldsymbol{\tau}) \sim MT(\mathbf{Q}, \mathbf{S}, \mathfrak{D}, \delta)$  with location matrix  $\mathbf{Q}$  and scale matrix  $\mathfrak{D}$ . Suppose further that  $\mathbf{q}_i$  is the  $i^{\text{th}}$  row of  $\mathbf{Q}$  and  $q_{is}$  is the  $is^{\text{th}}$  element of  $\mathbf{Q}$ ,  $\mathfrak{S}_{ii}$  is the  $i^{\text{th}}$  diagonal element of  $\mathfrak{R}$  and  $\mathbf{S}_{ss}$  is the  $s^{\text{th}}$  diagonal element of  $\mathbf{S}$ , then the marginal posterior predictive distributions at the  $i^{\text{th}}$  simulator point is given as  $y_{0i} = f(\mathbf{x}_{0i})$ , and the  $s^{\text{th}}$  output,  $y_{0, is} = f_s(\mathbf{x}_{0i})$  are multivariate and univariate  $t$  distributions, respectively:

$$y_{0i}|\mathbf{Y}, \boldsymbol{\tau} \quad \sim \quad t_q \left( \mathbf{q}_i^T, \frac{\mathfrak{S}_{ii}\mathbf{S}}{\hat{\delta}}, \hat{\delta} \right), \quad y_{0, is}|\mathbf{Y}, \boldsymbol{\tau} \quad \sim \quad t \left( q_{is}, \frac{\mathfrak{S}_{ii}\mathbf{S}_{ss}}{\hat{\delta}}, \hat{\delta} \right) \quad (19)$$

### B.1. Pseudo-code for MCMC Algorithm

1 Substitute  $R(x, x') = \prod_{k=1}^p \exp \left\{ -\phi_k \left| \mathbf{x}_{\cdot, k} - \mathbf{x}'_{\cdot, k} \right|^2 \right\}$  (Power exponential) with either truncated power or Bohman correlation function  $R(t; \boldsymbol{\tau})$  given above

- (a) Identify two points  $x_i$  and  $x_j$  in sample space such that  $t = |\mathbf{x}_{ik} - \mathbf{x}_{jk}| < \boldsymbol{\tau}_k$ ,  $\forall k$ .
- (b) Create a sparse representation of the  $n \times n$  correlation matrix  $\mathfrak{R}(\boldsymbol{\tau})$  with only these pairs while the remaining pair will contribute zero to the correlation matrix to complete the model specification
- (c) Store only the non-zero off-diagonal elements with their row and column indices

- (d) Parallelise steps (a-c) above using the Rcpp and Armadillo linear Algebra packages
- 2 Assign a matrix-normally inverse-Wishart (MNIW) prior distribution to  $\mathbf{B}$  as  $p(\mathbf{B}|\boldsymbol{\Sigma}, \phi) \sim N(\mathbf{B}_0, \boldsymbol{\Sigma} \otimes \boldsymbol{\Lambda}_0)$  and  $\boldsymbol{\Sigma}$  as  $p(\boldsymbol{\Sigma}|\phi) \sim \mathcal{W}^{-1}(\mathbf{S}_0, \delta_0)$  where  $\mathbf{B}_0, \boldsymbol{\Lambda}_0, \mathbf{S}_0$  and  $\delta_0$  are given prior hyperparameters
  - 3 Obtain the marginal posterior distributions for  $\mathbf{B}$  and  $\boldsymbol{\Sigma}$  using Gibbs sampling
  - 4 Generate a sample of the posterior  $p(\boldsymbol{\tau}|\mathbf{D})$  using a robust adaptive Metropolis-within-Gibbs step, for each  $\mathbf{B}$  and  $\boldsymbol{\Sigma}$
  - 5 Solve the sparse linear systems, such as  $\mathfrak{R}^{-1}(\mathbf{D} - \mathbf{H}\mathbf{B})$ , that occur in the posterior density
  - 6 Repeat steps (1-5)  $N$  times to obtain  $N$  iterations of MCMC samples for  $\hat{\boldsymbol{\Theta}} = \{\mathbf{B}, \boldsymbol{\Sigma}, \boldsymbol{\tau}\}$ .

## B.2. Model complexity

To derive an approximate likelihood that facilitates efficient computation, we replace the dense correlation function of the spatial process with the sparse approximation using a compact function. Following up from Section 3.1, and given that  $\mathbf{Y}$  is matrix-normally distributed as  $MN(\mathbf{H}\mathbf{B}, \boldsymbol{\Sigma}, \mathfrak{R})$ , the posterior distribution for  $\boldsymbol{\tau}$  can be obtained by integrating out the matrices  $\mathbf{B}$  and  $\boldsymbol{\Sigma}$  to give the approximate likelihood below,

$$p(\boldsymbol{\tau}|\mathbf{Y}) \propto |\mathfrak{R}|^{-\frac{m}{2}} \times |\hat{\boldsymbol{\Lambda}}|^{m/2} |\mathbf{S}|^{-(\delta+m-1)/2} \times p(\boldsymbol{\tau}). \quad (20)$$

We see that  $\mathfrak{R}$  is an  $n \times n$  sparse matrix and the approximate likelihood only contains inversion, determinant and multiplication terms with sparse  $n \times n$  matrices as well as the inversion and determinant of  $m \times m$  matrices. Therefore, the computational complexity of the log-likelihood calculation is of the order  $O(nk^2m)$ , where  $k$  is the average number of nonzero entries per row of the correlation matrix  $\mathfrak{R}$ ,  $m$  is the output dimension and  $n$  training samples, respectively. By manipulating the parameter  $\boldsymbol{\tau}$  we can increase the degree of sparsity in matrix  $\mathfrak{R}$  and then the computational cost of fitting the spatial MSGP model can be greatly reduced compared to the high computational cost of using the standard covariance function, where the computational complexity is typically of the order  $\mathcal{O}(n^3m^3)$ .

## C. Posterior expectations for computing the sensitivity indices

The formula required for computing all the expectations for  $S_j$  and  $S_j^\top$ , that is,  $\text{cov}(Y) = E(Y^2) - (E(Y))^2$ ,  $E^* \left( (E(Y | x_j))^2 \right)$ , and  $E \left( (E(Y | x_{\sim j}))^2 \right)$  are given below: The expectation and covariance of  $Y$  are given by

$$E(Y) = \int_{\mathbf{x}} f(\mathbf{x}) \prod_{j=1}^p dG_j(x_j)$$

$$\text{var}(Y) = \int_{\mathbf{x}} f^2(\mathbf{x}) \prod_{j=1}^p dG_j(x_j) - (E(Y))^2.$$

Define

$$E(Y | x_j) = \int_{\{x_{j'}:j' \neq j\}} f(x_1, \dots, x_j, \dots, x_p) \prod_{\{j':j' \neq j\}} dG_{j'}(x_{j'}). \quad (21)$$

For each specified value  $x_j$  of the  $j^{\text{th}}$  input, squaring the expression for  $E(Y|x_j)$  in (21) and taking its expectation, we have

$$\begin{aligned} E^* \left( (E(Y | \mathbf{x}_j))^2 \right) &= \\ &= \int \left\{ \int_{\{x_{j'}:j' \neq j\}} f(x_1, \dots, x_j, \dots, x_p) \prod_{\{j':j' \neq j\}} dG_{j'}(x_{j'}) \right\}^2 \times dG_j(x_j) \\ &= \int \left\{ \int_{\{x_{j'}:j' \neq j\}} \int_{\{x'_{j'}:j' \neq j\}} f(x_1, \dots, x_j, \dots, x_p) f(x'_1, \dots, x_j, \dots, x'_p) \prod_{\{j':j' \neq j\}} dG_{j'}(x_{j'}) \prod_{\{j':j' \neq j\}} dG_{j'}(x'_{j'}) \right\} \\ &\quad \times dG_j(x_j) \\ &= \int_{\mathbf{x}} \int_{\{x'_{j'}:j' \neq j\}} f(\mathbf{x}) f(x'_1, \dots, x_j, \dots, x'_p) \prod_{j'=1}^p dG_{j'}(x_{j'}) \prod_{\{j':j' \neq j\}} dG_{j'}(x'_{j'}). \quad (22) \end{aligned}$$

Similarly, for the expectations required for the total sensitivity indices, let

$$\mathbf{x}_{\sim j} = (x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_p),$$

then

$$E(Y | \mathbf{x}_{\sim j}) = \int f(x_j, \mathbf{x}_{\sim j}) dG_j(x_j), \quad (23)$$

and, analogously to the derivation above. Similarly, we can write

$$E^* \left( (E(Y | \mathbf{x}_{\sim j}))^2 \right) = \iint f(x_j, \mathbf{x}_{\sim j}) f(x'_j, \mathbf{x}_{\sim j}) dG_j(x_j) \prod_{j'=1}^p dG_{j'}(x_{j'}). \quad (24)$$

### C.1. Algorithm summary 2

We have summarised the algorithm to compute the main and total sensitivity indices for the multivariate outputs:

- 1 Normalize  $\mathbf{Y} = (Y^{(1)}, \dots, Y^{(m)})$  to common scale by dividing the centered column of  $\mathbf{Y}$  by their standard deviations to obtain  $\mathbf{Y}' = \frac{(\mathbf{Y} - \bar{\mathbf{Y}})}{\sigma(\mathbf{Y})}$ , where  $\sigma$  is the standard deviation.

- 2 Develop a fast approximation of a continuous function  $\hat{\mathbf{Y}} = f(\mathbf{x})$  using the multivariate sparse Gaussian process (MSGP) model based on the marginal samples of the inputs and corresponding set of simulated multivariate outputs.
- 3 Compute the following functions  $\left\{ f_j^{(k)}, f_{\sim j}^{(k)}, f_{j,\sim j}^{(k)} \right\}_{k=1}^m$  of multivariate outputs using the posterior predictive predictions given in the equation (3).
- 4 Estimate main and total sensitivity indices of the input variables for noncorrelated multivariate outputs given in equation (17) from their full marginal contributions

$$\left\{ \Omega_j^{(k)}, \Omega_{\sim j}^{(k)}, \Omega_{j,\sim j}^{(k)} \right\}_{k=1}^m$$

- 5 Obtain the empirical estimate of the correlation coefficient matrix  $R_{m,m}(\mathbf{Y})$  from the training data  $\mathbf{D}$ . Calculate the projected main and total sensitivity indices of the input variables for correlated multivariate outputs given in equations (18 and 27) from their full marginal contributions

$$\left\{ \Omega_j^{(k)}, \Omega_{\sim j}^{(k)}, \Omega_{j,\sim j}^{(k)} \right\}_{k=1}^m.$$

$$\Omega_j = \begin{bmatrix} \text{Cov} \left( f_j^{(1)}, \mathbf{f}^{(1)} \right) & \text{Cov} \left( f_j^{(1)}, \mathbf{f}^{(2)} \right) & \cdots & \text{Cov} \left( f_j^{(1)}, \mathbf{f}^{(m)} \right) \\ \text{Cov} \left( f_j^{(2)}, \mathbf{f}^{(1)} \right) & \text{Cov} \left( f_j^{(2)}, \mathbf{f}^{(2)} \right) & \cdots & \text{Cov} \left( f_j^{(2)}, \mathbf{f}^{(m)} \right) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov} \left( f_j^{(m)}, \mathbf{f}^{(1)} \right) & \text{Cov} \left( f_j^{(m)}, \mathbf{f}^{(2)} \right) & \cdots & \text{Cov} \left( f_j^{(m)}, \mathbf{f}^{(m)} \right) \end{bmatrix} \quad (25)$$

Note:  $\Omega_{\sim j}$  are given analogously with "j" replaced with " $\sim j$ ".

$$\Omega_{j,\sim j} = \begin{bmatrix} \text{Cov} \left( f_{j,\sim j}^{(1)}, \mathbf{f}^{(1)} \right) & \cdots & \text{Cov} \left( f_{j,\sim j}^{(1)}, \mathbf{f}^{(m)} \right) \\ \vdots & \ddots & \vdots \\ \text{Cov} \left( f_{j,\sim j}^{(m)}, \mathbf{f}^{(1)} \right) & \cdots & \text{Cov} \left( f_{j,\sim j}^{(m)}, \mathbf{f}^{(m)} \right) \end{bmatrix}. \quad (26)$$

The interaction sensitivity index  $\mathbf{P}_{j,\sim j}$  can be defined by normalizing  $\mathbf{Q}_{j,\sim j}$  with  $\|\Omega\|$  as

$$\mathbf{P}_{j,\sim j} = \frac{\mathbf{Q}_{j,\sim j}}{\|\Omega\|} = \frac{\langle \Omega_{j,\sim j}, \Omega(\mathbf{Y}) \rangle}{\|\Omega\|^2}.$$

The total sensitivity index  $\mathbf{P}^T$  based on the vector projection can be defined by

$$\mathbf{P}^T = \mathbf{P}_{j,\sim j} + \mathbf{P}_j = \frac{\langle (\Omega_{j,\sim j} + \Omega_j), \Omega(\mathbf{Y}) \rangle}{\|\Omega(\mathbf{Y})\|^2}. \quad (27)$$

## D. Covariance tapering and Schur product theorem

We also consider the product of Matérn and Wendland correlation functions given below. Wendland's compactly supported functions belong to the family of compactly supported radial functions. It can be constructed by a repeated operation on the truncated power function (which we know to be strictly positive definite and radial on  $\mathbb{R}^s$  for  $s \leq 2\ell - 1$ ).

$$\varphi_{s,2}(t) \doteq (1-t)_+^{\ell+2} [(\ell^2 + 4\ell + 3)t^2 + (3\ell + 6)t + 3] \quad (28)$$

where  $\ell = \lfloor s/2 \rfloor + k + 1$ ,  $t = \|\mathbf{x} - \mathbf{x}'\|$  and the symbol  $\doteq$  denotes equality up to a multiplicative positive constant.

$$R(\mathbf{x}_i, \mathbf{x}_j) = C_{\phi,\nu}(\|\mathbf{x}_i - \mathbf{x}_j\|) = \frac{1}{2^{\nu-1}\Gamma(\nu)} (\|\mathbf{x}_i - \mathbf{x}_j\| \phi)^\nu \mathcal{K}_\nu(\|\mathbf{x}_i - \mathbf{x}_j\| \phi); \phi > 0, \nu > 0. \quad (29)$$

Here  $\Gamma$  is the Gamma function, and  $\mathcal{K}_\nu$  is the modified Bessel function of the second kind of order  $\nu$ . The parameter  $\phi$  is related to the effective range. For certain values of  $\nu$ , the Matérn covariance function has appealing forms. For example, if  $\nu = 0.5$ ,  $C_{\phi,\nu}$  becomes an exponential covariance, if  $\nu = n_0 + 0.5$  with  $n_0$  an integer,  $C_{\phi,\nu}$  becomes the product of an exponential covariance and a polynomial of order  $n$ .

Tapering involves multiplying the data covariance matrix element-wise by a sparse correlation matrix. In both cases, one must maintain the positive definiteness of any sparse matrices. Let  $C_\tau$  be a correlation function that is identically zero outside a particular range described by  $\tau$ . Now consider a tapered correlation that is the Schur or Hadamard product of  $C_\tau$  and  $C$

$$C_{\text{tap}}(\mathbf{x}, \mathbf{x}'; \tau) = C(\mathbf{x}, \mathbf{x}') C_\tau(\mathbf{x}, \mathbf{x}'). \quad (30)$$

An approximation to the original process is obtained by replacing the correlation matrices in 30 based on  $C$  with those defined by  $C_{\text{tap}}$  to model dependence. The assumption is that the product  $C_{\text{tap}}$  of correlation matrices preserves some of the shapes of original correlation  $C$  which is identically zero outside the chosen range  $\tau$  thus making  $C_{\text{tap}}$  a valid correlation, by using the Schur product theorem the tapered covariance function is positive definite and thus a valid covariance function (Furrer et al., 2006; Horn and Johnson, 2012). See the proof of the Schur product theorem below.

### D.1. Schur theorem

If  $A$  and  $B$  are positive semidefinite  $n \times n$  matrices, then so is  $A \circ B$ . If, in addition,  $B$  is positive definite, and  $A$  has no diagonal entry equal to 0, then  $A \circ B$  is positive definite. In particular, if both  $A$  and  $B$  are positive definite, then so is  $A \circ B$ .

Definition: Two  $m \times n$  matrices,  $A$  and  $B$ , have Schur product  $A \circ B = \{a_{ij}b_{ij}\}$ .

Proof:

Consider an integral operator  $C(f) = \int_a^b C(\mathbf{x}, \mathbf{x}') f(\mathbf{x}) d\mathbf{x}$  such that  $f \in [a, b]^R$  and covariance function  $H(\mathbf{x}, \mathbf{x}')$  having the same condition, and consider the pointwise product kernel such that  $L(\mathbf{x}, \mathbf{x}') = C(\mathbf{x}, \mathbf{x}')H(\mathbf{x}, \mathbf{x}')$ , and the associated integral operator then,

$$L(f) = \int_a^b L(\mathbf{x}, \mathbf{x}') f(\mathbf{x}) d\mathbf{x} = \int_a^b C(\mathbf{x}, \mathbf{x}') H(\mathbf{x}, \mathbf{x}') f(\mathbf{x}) d\mathbf{x}. \quad (31)$$

The linear mapping  $f \rightarrow C(f)$  is a limit of matrix-vector multiplications (approximate the integral as a finite Riemann sum), and many properties of integral operators can be obtained by taking appropriate limits of results known for matrices. Therefore, the pointwise product of integral kernels produces an integral operator that is a continuous analogue of the Hadamard product of matrices.

Suppose, an integral kernel  $C(\mathbf{x}, \mathbf{x}')$  has the property that

$$\int_a^b \int_a^b C(\mathbf{x}, \mathbf{x}') f(\mathbf{x}) \bar{f}(\mathbf{x}') d\mathbf{x} d\mathbf{x}' \geq 0,$$

for all  $f \in [a, b]^R$ , then  $C(\mathbf{x}, \mathbf{x}')$  is said to be a positive semidefinite kernel. This is a standard result from (Mercer's theorem) that if  $C(\mathbf{x}, \mathbf{x}')$  is a continuous positive semidefinite kernel on a finite interval  $[a, b]$ , then there exist positive real numbers  $\lambda_1, \lambda_2, \dots$  (eigenvalues) and continuous functions  $\zeta_1(\mathbf{x}), \zeta_2(\mathbf{x}), \dots$  (eigenfunctions) such that

$$C(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{\infty} \frac{\zeta_i(\mathbf{x}) \bar{\zeta}_i(\mathbf{x}')}{\lambda_i},$$

on  $[a, b] \times [a, b]$  and the series converges absolutely and uniformly. Therefore, if  $C(\mathbf{x}, \mathbf{x}')$  and  $H(\mathbf{x}, \mathbf{x}')$  are both continuous positive semidefinite kernel s on  $[a, b]$ , then  $L(\mathbf{x}, \mathbf{x}')$  also has an absolute and uniform convergence such that

$$L(\mathbf{x}, \mathbf{x}') = \sum_{i=1}^{\infty} \frac{\psi_i(\mathbf{x}) \bar{\psi}_i(\mathbf{x}')}{\mu_i},$$

on  $[a, b] \times [a, b]$  with all  $\mu_i > 0$ . Therefore, pointwise product kernel  $L(\mathbf{x}, \mathbf{x}') = C(\mathbf{x}, \mathbf{x}')H(\mathbf{x}, \mathbf{x}')$  also has the representation

$$L(\mathbf{x}, \mathbf{x}') = \sum_{i,j=1}^{\infty} \frac{\zeta_i(\mathbf{x}) \psi_j(\mathbf{x}) \bar{\zeta}_i(\mathbf{x}') \bar{\psi}_j(\mathbf{x}')}{\lambda_i \mu_j},$$

on  $[a, b] \times [a, b]$  which also converges absolutely and uniformly. Then

$$\int_a^b \int_a^b C(\mathbf{x}, \mathbf{x}') f(\mathbf{x}) \bar{f}(\mathbf{x}') d\mathbf{x} d\mathbf{x}' = \sum_{i,j=1}^{\infty} \frac{1}{\lambda_i \mu_j} \left| \int_a^b \zeta_i(\mathbf{x}) \psi_j(\mathbf{x}) f(\mathbf{x}) d\mathbf{x} \right|^2 \geq 0. \quad (32)$$

Therefore,  $L(\mathbf{x}, \mathbf{x}')$  is also positive semidefinite.

Theorem: Let  $A$  and  $B \in R^n$  be positive semidefinite.

- $A \circ B$  is positive semidefinite.
- if  $A$  is positive definite and every main diagonal entry of  $B$  is positive, then  $A \circ B$  is positive definite.
- if both  $A$  and  $B$  are positive definite, then  $A \circ B$  is positive definite.
- Commutativity. Unlike the standard matrix product,  $A \circ B = B \circ A$ .
- Trace. For square matrices  $A, B$ , and  $C$ , with  $B$  symmetric,  $tr(A \circ B)C = trA(B \circ C)$ .

E. MCMC diagnostics

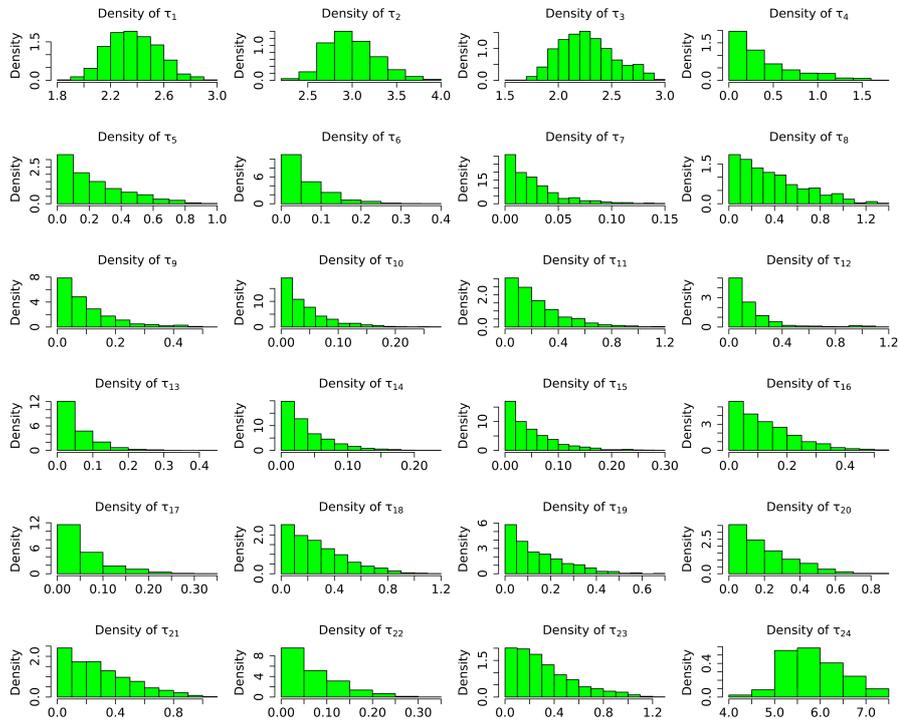


Fig. 16. Posterior summary showing the densities of each parameter for EU-wide data.

**Table 7.** Potential scale reduction factor for each MCMC chain, together with upper and lower confidence limits. Approximate convergence is diagnosed when the upper limit is close to 1.

	Point est.	Upper C.I.
1	1.01	1.01
2	1.05	1.09
3	1.00	1.01
4	1.82	3.49
5	1.04	1.17
6	1.02	1.06
7	1.00	1.01
8	1.04	1.18
9	1.01	1.01
10	1.00	1.01
11	1.33	2.04
12	1.17	1.58
13	1.00	1.00
14	1.02	1.02
15	1.00	1.01
16	1.02	1.09
17	1.00	1.02
18	1.20	1.67
19	1.00	1.02
20	1.06	1.23
21	1.16	1.57
22	1.01	1.05
23	1.06	1.15
24	1.10	1.39

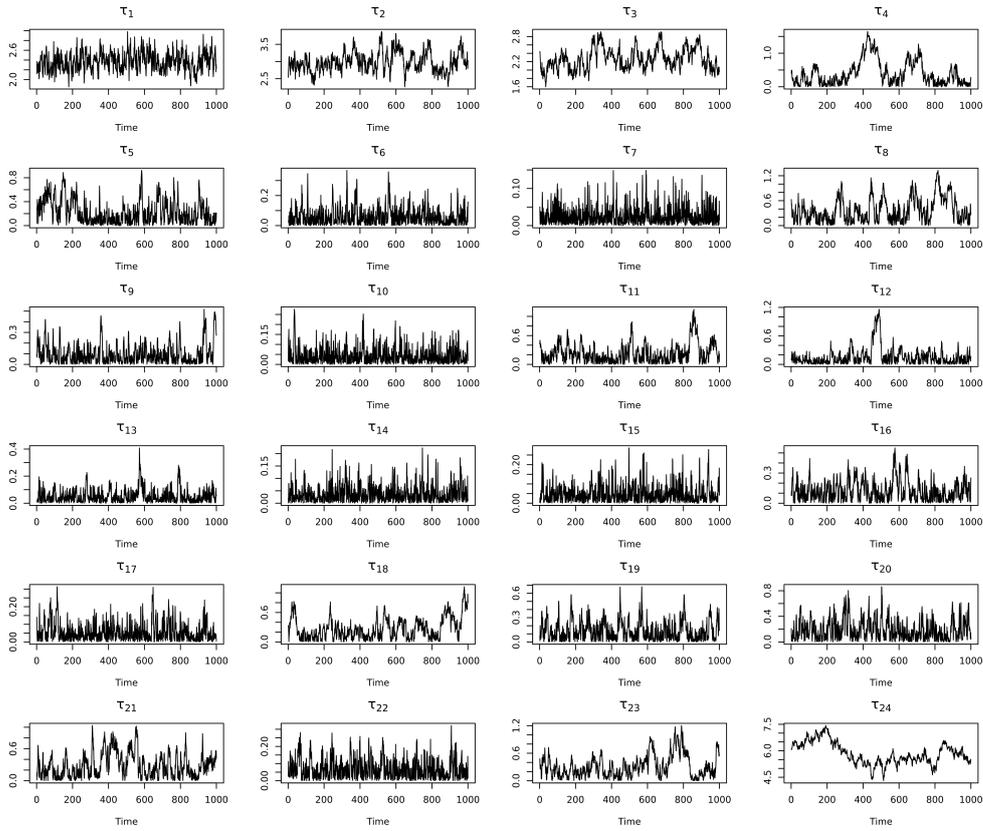


Fig. 17. Posterior summary showing the trace plots of each parameter for EU-wide data.

## F. Further results

This section contains miscellaneous extra results for test case 3 from Section 5.1. The results included Tables 8 and 9 for the correlation coefficients  $\rho$  between multivariate outputs and distribution parameters respectively.

Similarly, following the discussion from Section 5.3 in the main paper, Table 10 compared the performance of linear regression using the DiceKriging library in R and the MSGP models.

**Table 8.** Correlation coefficients  $\rho$  between outputs for test case 3.

$\rho$	$Y^{(1)}$	$Y^{(2)}$	$Y^{(3)}$
$Y^{(1)}$	1.000	0.311	0.369
$Y^{(2)}$	0.311	1.000	0.046
$Y^{(3)}$	0.369	0.046	1.000

**Table 9.** Distribution parameters for test case 3.

Variables	$x_1$	$x_2$	$x_3$	$x_4$	$x_5$	$x_6$	$x_7$	$x_8$
Mean	$2 \times 10^4$	12	0.04	$9.82 \times 10^{-4}$	$1.34 \times 10^7$	$3.35 \times 10^8$	$2 \times 10^{10}$	$1 \times 10^{11}$
SD	$1.4 \times 3$	0.12	0.004	$5.89 \times 10^{-5}$	$2.412 \times 10^5$	$4.02 \times 10^7$	$1.2 \times 10^9$	$6 \times 10^9$

**Table 10.** Relative performance of linear regression and compactly supported model when fixed the value of  $\omega$  as 90% in Table 3 using their RMSE values. Note: if  $\alpha > 0$ , then the MSGP model is far better than the linear regression model, and vice versa. The results are very similar when  $\alpha = 0$ .

	Output	$\alpha(\%)$
1	Food Per Capita	31.30
2	People Flooded 1/100year	19.00
3	Water Exploitation Index	21.30
4	Landuse Intensity Index	26.70
5	Landuse Diversity	25.50
6	Biodiversity Vulnerability Index	18.30
7	Timber Production	21.30
8	Area of Artificial Surfaces	-3.00
9	Food Production	-1.70
10	Potential Carbon Stock	0.00
11	Irrigation Usage	-9.40
12	Intensive Arable	-1.70
13	Intensive Grassland	3.40
14	Extensive Grassland	-1.40
15	Very Extensive Grassland	1.20
16	Unmanaged Land	0.00
17	Managed Forest	0.00
18	Unmanaged Forest	1.90

## Author notes

**Conflict of interest** The authors have no conflict of interest to declare.

## Figure legends

Fig. 1. Histograms of spatially aggregated IAP2 simulation for the 18 selected outputs, averaged across the three emission scenarios and ten climate models for SSP1. Note: (—) denotes unitless.

Fig. 2. Main and total indices for the Sobol  $g$ - function using our algorithm. The red dots denote the analytical values.

Fig. 3. Main and total indices for the multivariate test cases using arctangent temporal function which has 100 functional outputs with two input variables from a uniform distribution.

Fig. 4. Main and total indices for the nonlinear model used in Xu et al. (2019b) using MSGP algorithm.

Fig. 5. Sample output data from the IMPRESSIONS IAP2 simulation under RCP4.5 and HadGEM2-ES-RCA4 climate model and the classification of the five European regions used in the analyses.

Fig. 6. Comparison of the multivariate GP emulator performance with IAP2 simulation for 500 randomly sample points under the SSP4 scenario in the 2050s for EU-wide aggregated data; simulation (black) and emulator predictions (green) with 95% C.I. (red). Note: the outputs have been normalised to have a zero mean and unit variance.

Fig. 7. Comparison of the multivariate GP emulator performance with IAP2 simulation for 500 randomly sample points under the SSP1 scenario in the 2050s for EU-wide aggregated data; simulation (black) and emulator predictions (green) with 95% C.I. (red). Note: the outputs have been normalised to have a zero mean and unit variance.

Fig. 8. Main effect plots for EU-wide aggregated data; simulation (black) and emulator predictions (green) with 95% C.I. (red). Note: the outputs have been normalised to have a zero mean and unit variance.

Fig. 9. Barplots of the posterior mean estimates of the first-order indices from the multivariate sensitivity analyses for some selected IAP2 outputs for EU-wide average data.

Fig. 10. Barplots of the posterior mean estimates of the total indices from the multivariate sensitivity analyses for some selected IAP2 outputs for EU-wide average data.

Fig. 11. Boxplots of the posterior mean for the multivariate generalised indices for EU-wide average data. Note: The left plots correspond to the standard approach while the right plots are computed from the vector projected method using affine transformation.

Fig. 12. Boxplots of the posterior mean and their uncertainty estimates for the aggregated main indices for SSP1, SSP3, SSP4 and SSP5 respectively.

Fig. 13. Main effect plots for EU-wide aggregated data; simulation (black) and emulator predictions (green) with 95% C.I. (red). Note: the outputs have been normalised to have a zero mean and unit variance.

Fig. 14. Histograms of spatially aggregated IAP2 simulation for the 18 selected outputs, averaged across the three emission scenarios and ten climate models for SSP4.

Note: ( ) denotes unitless.

Fig. 15. Boxplots of the posterior mean and their uncertainty estimates for the aggregated main indices for Alpine, Northern, Atlantic, Continental and Southern regional averaged data.

Fig. 16. Posterior summary showing the densities of each parameter for EU-wide data.

Fig. 17. Posterior summary showing the trace plots of each parameter for EU-wide data.