Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

Geoscientific Model Development Discussions

Fast sensitivity analysis methods for computationally expensive models with multi-dimensional output

Edmund Ryan[1]*, Oliver Wild[1], Fiona O'Connor[2], Apostolos Voulgarakis[3], and Lindsay Lee[4]

[1]Lancaster Environment Centre, Lancaster University, Lancaster. UK

[2]UK Met Office Hadley Centre, Exeter. UK

[3]Department of Physics, Imperial College London, London. UK

[4]School of Earth and Environment, University of Leeds. UK

*Corresponding author:

Lancaster Environment Centre, Lancaster University,

Bailrigg, Lancaster.

LA1 4YQ.

UK

Tel: +44 (0)1524 594009

edmund.ryan@lancaster.ac.uk

Keywords: *global sensitivity analysis, emulator, meta-model, multi-dimensional output, principal component analysis, FAST, Sobol, partial least squares, generalized additive model, atmospheric chemical transport models, tropospheric methane lifetime.*

For submission to: *Geoscientific Model Development*

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

Geoscientific Model Development Discussions

Abstract

Global sensitivity analysis (GSA) is a critical approach in identifying which inputs or parameters of a model most affect model output. This determines which inputs to include when performing model calibration or uncertainty analysis. GSA allows quantification of the sensitivity index (SI) of a particular input – the percentage of the total variability in the output attributed to the changes in that input – by averaging over the other inputs rather than fixing them at specific values. Traditional methods of computing the SIs (e.g. Sobol) involve running a model thousands of times, but this may not be feasible for computationally expensive earth system models. GSA methods that use a statistical emulator in place of the expensive model are popular as they require far fewer model runs. Here, we perform an eight-input GSA on two computationally expensive atmospheric chemistry transport models using emulators that were trained with 80 runs of the models. We consider two methods to further reduce the computational cost of GSA: (1) a dimension reduction approach and (2) an emulator-free approach. When the output of a model is multi-dimensional, it is common practice to build a separate emulator for each dimension of the output space. Here, we use principal component analysis (PCA) to reduce the output dimension and build an emulator for each of the transformed outputs. We consider the global distribution of the annual column mean lifetime of atmospheric methane, which requires ~2000 emulators without PCA, but only 5-40 emulators with PCA. As an alternative, we apply an emulator-free method using a generalised additive model (GAM) to estimate the SIs using only the training runs. Compared to the emulator-only method, the hybrid PCA-emulator and GAM methods are 6 and 30 times quicker, respectively, at computing the SIs for the ~2000 methane lifetime outputs. The SIs computed using the two computationally faster methods are almost identical to those computed using the standard emulator-only method.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

Geoscientific Model Development Discussions

Introduction

Sensitivity analysis is a powerful tool for understanding the behaviour of a numerical model. It allows quantification of the sensitivity in the model outputs to changes in each of the model inputs. If the inputs are fixed values such as model parameters, then sensitivity analysis allows study of how the uncertainty in the model outputs can be attributed to the uncertainty in these inputs. Sensitivity analysis is important for a number of reasons: (i) to identify which parameters contribute the largest uncertainty to the model outputs; (ii) to prioritise estimation of model parameters from observational data, and (iii) to understand the potential of observations as a model constraint, and (iv) to diagnose differences in behaviour between different models;

Different approaches for sensitivity analysis

By far, the most common types of sensitivity analysis are those performed one-at-a-time (OAT) and locally. OAT sensitivity analysis involves running a model a number of times, varying each input in turn whilst fixing other inputs at their nominal values. For example, Wild (2007) showed that the tropospheric ozone budget was highly sensitive to differences in global NOx emissions from lightning. The observation-based range of 3-8 TgN/yr in the magnitude of these emissions could result in a 10% difference in predicted tropospheric ozone burden. OAT sensitivity analysis is used in a variety of research fields including environmental science (Bailis et al., 2005; Campbell et al., 2008; de Gee et al., 2008; Saltelli and Annoni, 2010), medicine (Coggan et al., 2005; Stites et al., 2007; Wu et al., 2013), economics (Ahtikoski et al., 2008) and physics (Hill et al., 2012). While the ease of implementing OAT sensitivity analysis is appealing, a major drawback of this approach is that it assumes that the model response to different inputs is independent, which in most cases is unjustified (Saltelli and Annoni, 2010) and can result in biased results (Carslaw et al., 2013). Local sensitivity analysis, another popular

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

method, involves mathematical differentiation of model output with respect to each input of interest (Hakami et al., 2004). By 'local' we refer to small perturbations in the inputs at a point where the differentiated model is applied. The linearity of the model is assumed but only in the locality of a point in the input space, thus it is more mathematically robust than OAT sensitivity analysis (Saltelli et al., 2008). A deficiency with local sensitivity analysis is its inability to explore the effects on the output of changes in the inputs over the whole input space.

Global sensitivity analysis (GSA) is an approach which can be used to explore the whole input space, whilst still maintaining mathematical rigour. However, the number of sensitivity analysis studies using this global method has been very small. Ferretti et al. (2016) found that out of around 1.75 million research articles surveyed up to 2014, only 1 in 20 of studies mentioning 'sensitivity analysis' also use or refer to 'global sensitivity analysis'. A common type of GSA is the variance based method, which operates by apportioning the variance of the model's output into different sources of variation in the inputs. More specifically, it quantifies the sensitivity of a particular input – the percentage of the total variability in the output attributed to the changes in that input – by averaging over the other inputs rather than fixing them at specific values. The Fourier Amplitude Sensitivity Test (FAST) was one of the first of these variance based methods (Cukier et al., 1973). The classical FAST method uses spectral analysis to apportion the variance, after first exploring the input space using sinusoidal functions of different frequencies for each input factor or dimension (Saltelli et al., 2012). Modified versions of FAST include the extended FAST method which improves its computational efficiency (Saltelli et al., 1999) and the random-based-design (RBD) FAST method which samples from the input space more efficiently (Tarantola et al., 2006). Another widely used GSA method is the Sobol method (Sobol, 1990; Homma and Saltelli, 1996; Saltelli, 2002), which has been found to

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

outperform FAST (Saltelli, 2002). Most applications of the Sobol and FAST methods involve a small number of input factors. However, Mara and Tarantola (2008) carried out a 100-input sensitivity analysis using the RBD version of FAST and a modified version of the Sobol method and found that both methods gave estimates of the SIs that were close to the known analytical solutions. A downside to the Sobol method is that a large number of runs of the model typically need to be carried out. For the model used in Mara and Tarantola (2008), 10,000 runs were required for the Sobol method but only 1000 were needed for FAST.

Emulators and meta-models

If a model is computationally expensive, carrying out 1000 simulations may not be feasible. A solution is to use a surrogate function for the model called a meta-model that maps the same set of inputs to the same set of outputs, but is computationally much faster. Thus, much less time is required to perform GSA using the meta-model than using the slow-running model. A meta-model can be any function that maps the inputs of a model to its outputs, e.g. linear or quadratic functions, splines, neural networks, etc. Here, we use a statistical emulator because it has two useful properties. First, an emulator is an interpolating function which means that at inputs of the model that are used to train the emulator, the resulting outputs of the emulator must exactly match those of the model (Iooss and Lemaître, 2015). Secondly, for inputs that the emulator is not trained at, the probability distribution of the outputs represents their uncertainty (O'Hagan, 2006). The vast majority of emulators are based on Gaussian process (GP) theory due to its attractive properties (Kennedy and O'Hagan, 2000; Oakley and O'Hagan, 2004; O'Hagan, 2006), which make GP emulators easy to implement while providing accurate representations of the computationally-expensive model (e.g. Kennedy et al., 2008; Lee et al., 2013; Chang et al., 2015; Gómez-Dans et al., 2016). A GP is a multivariate Normal distribution applied to a function

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

rather than a set of variables. The original GP emulator in a Bayesian setting was developed by Currin et al. (1991) (for basic overview see also O'Hagan, 2006) and is mathematically equivalent to the Kriging interpolation methods used in geostatistics (E.g. Cressie, 1990; Ripley, 2005). Kriging regression has been used as an emulator method since the 1990s (Welch et al., 1992; Koehler and Owen, 1996). More recently there has been considerable interest in using this Kriging emulator approach for practical purposes such as GSA or inverse modelling (Marrel et al., 2009; Roustant et al., 2012). Examples of its application can be found in atmospheric modelling (Carslaw et al., 2013; Lee et al., 2013), medicine (Degroote et al., 2012) and electrical engineering (Pistone and Vicario, 2013).

For GSA studies involving multi-dimensional output, a traditional approach is to apply a separate GP emulator for each dimension of the output space. However, if the output consists of many thousands of points on a spatial map or time-series (Lee et al., 2013) then the need to use thousands of emulators can impose substantial computational constraints even using the FAST methods. A solution is to adopt a GSA method that does not rely on an emulator, but is based on generalized additive modelling (Mara and Tarantola, 2008; Strong et al., 2014; Strong et al., 2015b) or on a partial least squares approach (Sobie, 2009; Chang et al., 2015). A separate generalized additive model (GAM) can be built for each input against the output of the expensive model, and the sensitivity of the output to changes in each input are then computed using these individual GAM models. Partial least squares (PLS) is an extension of the more traditional multivariate linear regression where the number of samples (i.e. model runs in this context) can be small, and may even be less that the number of inputs (Sobie, 2009).

An alternative way of reducing the computational constraints is to use principal component analysis (PCA) to reduce the dimensionality of the output. This means that we

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

require far fewer emulators to represent the outputs, reducing the GSA calculations by a large margin, although there is some loss of detail. This PCA-emulator hybrid approach has been successfully used in radiative transfer models (Gómez-Dans et al., 2016), a very simple chemical reaction model (Saltelli et al., 2012) and general circulation models (Sexton et al., 2012). While we hypothesize that both emulator-free and PCA-based methods are suited to large-scale GSA problems (e.g. those involving more than 20 input factors), a focus of our work is to determine the accuracy of these methods for a smaller scale GSA study.

Aims of this study

Recent research comparing different GSA methods based on Gaussian Process emulators has been limited in application to relatively simple models and low-dimensional output (Mara and Tarantola, 2008). Using two computationally expensive models of global atmospheric chemistry and transport we compare the accuracy and efficiency of global sensitivity analysis using emulators and emulator-free methods, and we investigate the benefits of using PCA to reduce the number of emulators needed. We compare and contrast a number of ways of computing the first order sensitivity indices for the expensive atmospheric models: (i) the Sobol method using an emulator; (ii) the extended FAST method using an emulator; (iii) generalised additive modelling; (iv) a partial least squares approach; (v) an emulator-PCA hybrid approach. Hereafter, we refer to (i) and (ii) as emulator-based GSA methods and (iii) and (iv) as emulator-free GSA methods.

Materials and methods

Atmospheric chemistry models

Global atmospheric chemistry and transport models simulate the composition of trace gases in the atmosphere (e.g. O_3, CH_4, CO, SO_x) at a given spatial resolution (latitude × longitude ×

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

altitude). The evolution in atmospheric composition over time is controlled by a range of different dynamical and chemical processes, our understanding of which remains incomplete. Trace gases are emitted from anthropogenic sources (e.g., NO from traffic and industry) and from natural sources (e.g. isoprene from vegetation, NO from lightning), they may undergo chemical transformation (e.g., formation of O_3) and transport (e.g., convection or boundary layer mixing), and may be removed through wet or dry deposition. Global sensitivity analysis is needed to understand the sensitivity of our simulations of atmospheric composition and its evolution to assumptions about these governing processes.

In this study, we performed global sensitivity analysis (GSA) on two such atmospheric models. We used the Frontier Research System for Global Change version of the University of California, Irvine chemistry transport model, the FRSGC/UCI CTM (Wild and Prather, 2000; Wild et al., 2004), and the Goddard Institute for Space Studies general circulation model, the GISS GCM (Shindell et al., 2006; Schmidt et al., 2014). We used results from 104 model runs carried out with both of these models from a comparative GSA study (Wild et al., in prep.). This involved varying eight inputs or parameters over specified ranges using a maximin Latin hypercube design: global surface NOx emissions (30-50 TgN/year), global lightning NOx emissions (2-8 TgN/year), global isoprene emissions (200-800 TgC/year), dry deposition rates (model value ± 80%), wet deposition rates (model value ± 80%), humidity (model value ± 50%), cloud optical depth (model value × 0.1–10) and boundary layer mixing (model value × 0.01–100). For this study, we focus on a single model output, the global distribution of annual tropospheric column mean lifetime of methane (CH_4). The CH_4 lifetime is an important indicator of the amount of highly reactive hydroxyl radical in the troposphere (Voulgarakis et al., 2013), and we choose this output because of its contrasting behaviour in the two models. The native

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

spatial resolution of the models is 2.8°×2.8° for FRSGC and 2.5°×2.0° for GISS, but we combine neighbouring grid points so that both models have a comparable resolution of 5-6°, giving a total of 2048 grid points for FRSGC/UCI and 2160 grid points for GISS.

Global sensitivity analysis using the Sobol and extended FAST methods

A common way of conducting global sensitivity analysis is to compute the first order sensitivity indices (SIs) using variance based decomposition; this apportions the variance in output of the chemistry model to different sources of variation in the different model inputs. Assuming that the inputs are independent of one another, the ith first-order SI, corresponding to the ith input (i=1, 2, …, p), is given by:

$$S_i = \frac{Var[E(Y|X_i)]}{Var(Y)} \quad (1)$$

The simplest way of computing S_i is by brute force, but this is also the most computationally intensive (Saltelli et al., 2008).

The Sobol method, developed in the 1990s, is much faster than brute force at computing the terms in equation (1), in part because it requires fewer executions of the model (Sobol, 1990; Homma and Saltelli, 1996; Saltelli, 2002; Saltelli et al., 2008). The method operates by first generating a N×2p matrix of random numbers from a space filling sampling design (e.g. a Latin hypercube design), where N is the number of samples and p is the number of input factors. The matrix is split in half to form two new matrices, $\boldsymbol{A}$ and $\boldsymbol{B}$, each of size N×p. To compute the ith SI ($1 \le i \le p$), we define two new matrices $\boldsymbol{C_i}$, and $\boldsymbol{D_i}$, where $\boldsymbol{C_i}$ is formed by taking the ith column from $\boldsymbol{A}$ and the remaining columns from $\boldsymbol{B}$, and $\boldsymbol{D_i}$ is formed by taking the ith column from $\boldsymbol{B}$ and the remaining columns from $\boldsymbol{A}$. We then apply model $\boldsymbol{f}$ to each set of inputs given by the rows of matrices $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C_i}$ and $\boldsymbol{D_i}$. This gives vectors $\mathbf{y_A}=\boldsymbol{f}(\boldsymbol{A})$, $\mathbf{y_B}=\boldsymbol{f}(\boldsymbol{B})$, $\mathbf{y_{Ci}}=\boldsymbol{f}(\boldsymbol{C_i})$ and $\mathbf{y_{Di}}=\boldsymbol{f}(\boldsymbol{D_i})$. Vectors $\mathbf{y_A}$ and $\mathbf{y_{Ci}}$ are then substituted into eqn (2):

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

Geoscientific Model Development Discussions

$$\widehat{S}_i = \frac{\widehat{Var}\left[\hat{E}(Y|X_i)\right]}{\widehat{Var}(Y)} = \frac{\mathbf{y_A} \cdot \mathbf{y_{C_i}} - \left(\frac{1}{N}\sum_{j=1}^{N} y_A^{(j)}\right)^2}{\mathbf{y_A} \cdot \mathbf{y_A} - \left(\frac{1}{N}\sum_{j=1}^{N} y_A^{(j)}\right)^2} \quad (2)$$

where $\mathbf{y_A} \cdot \mathbf{y_{C_i}} = \left(\frac{1}{N}\sum_{j=1}^{N} y_A^{(j)} y_{C_i}^{(j)}\right)$. Saltelli (2002) and Tarantola et al. (2006) suggested using eight variants of equation (2), using different combinations of $\mathbf{y_A}$, $\mathbf{y_B}$, $\mathbf{y_{Ci}}$ and $\mathbf{y_{Di}}$ (Appendix A). Lilburne et al. (2009) proposed using the average of these eight SI estimates as they deemed this to be more accurate than a single estimate.

An alternative and even faster way of estimating the terms in equation (1) is to use the extended-FAST method, first developed by Saltelli *et al.* (1999) and widely used since (Koehler and Owen, 1996; Queipo et al., 2005; Saltelli et al., 2008; Carslaw et al., 2013; Vanuytrecht et al., 2014; Vu-Bac et al., 2015). Defining f to represent the model, so that $\mathbf{y} = f(\mathbf{x})$ where $\mathbf{x}=(x_1, \ldots, x_p)$ and $\mathbf{y}$ is the scalar output, a multi-dimensional Fourier transformation of f allows a variance-based decomposition that samples the input space along a curve defined by:

$$x_i(s) = G_i\left(sin(\omega_i s)\right), \quad (3)$$

where $s \in \mathbb{R}$ is a variable over the range (-∞,∞), G_i is the ith transformation function (Appendix A), and ω_i is the ith user-specified frequency corresponding to each input. Varying s allows a multi-dimensional exploration of the input space due to x_is being simultaneously varied. After applying $\boldsymbol{f}$, the resulting output $\boldsymbol{y}$ produces different periodic functions based on different ω_i. If the output is sensitive to changes in the ith input factor, the periodic function of $\boldsymbol{y}$ corresponding to frequency ω_i will have a high amplitude. Further details of extended-FAST including the formulae for $Var[E(Y|X_i)]$ and $Var(Y)$ are given in Saltelli et al. (1999). The difference between the original and the extended versions of the FAST method are given in Appendix A.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

Gaussian Process Emulators

When the model is expensive to run – like the chemistry transport models used here – we substitute it with an emulator, an accurate surrogate of the expensive model but much faster to run. If we are confident that the emulator is accurate, then we can compute the first order SIs using the outputs of the emulator rather than the expensive model. Mathematically, an emulator is a statistical model that mimics the input-output relationship of a computationally expensive model. As stated in the *Introduction*, an emulator is an interpolating function at model outputs it is trained at and gives a probability distribution and other outputs (O'Hagan, 2006).

An emulator is trained using a set of inputs denoted by $\mathbf{x_1}$, $\mathbf{x_2}$, .., $\mathbf{x_n}$, and scalar outputs represented by $\mathbf{y_1}=f(\mathbf{x_1})$, $\mathbf{y_2}=f(\mathbf{x_2})$,…, $\mathbf{y_n}=f(\mathbf{x_n})$, where f represents the expensive model. The most common form of an emulator is a Gaussian Process (GP) since it has attractive mathematical properties that allow an analytical derivation of the mean and variance of the emulated output (given by $\hat{f}(\mathbf{x})$ for a general input $\mathbf{x}$). A notable exception is Goldstein and Rougier (2006) who used a non-GP emulator based on a Bayes linear approach. More formally, a GP is an extension of the multivariate Gaussian distribution to infinitely many variables (Rasmussen, 2006). The multivariate Gaussian distribution is specified by a mean vector μ and covariance matrix $\mathbf{\Sigma}$. A GP has a mean function which is typically given by $m(\mathbf{x}) = E(f(\mathbf{x}))$ and covariance function given by c(x,x') = cov($f(\mathbf{x})$, $f(\mathbf{x'})$). For the latter we used a Matern(5/2) function (Roustant et al., 2012), which is given by:

$$c(\mathbf{x}, \mathbf{x}') = s^2 + \left(1 + \sqrt{5}\left(\frac{|\mathbf{x}-\mathbf{x}'|}{\theta}\right) + \frac{5}{3}\left(\frac{|\mathbf{x}-\mathbf{x}'|}{\theta}\right)^2\right) \times exp\left(-\sqrt{5}\left(\frac{|\mathbf{x}-\mathbf{x}'|}{\theta}\right)\right), \quad (4)$$

where s denotes the standard deviation and θ is the vector of range parameters (sometimes called *length-scales*). In addition, a GP has the property that $f(\mathbf{x_1})$, $f(\mathbf{x_2})$, …, $f(\mathbf{x_n})$ are jointly Normally distributed for any set of inputs $\mathbf{x_1}$, $\mathbf{x_2}$, …, $\mathbf{x_n}$. GP emulators for uncertainty quantification were

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

originally developed within a Bayesian framework (Currin et al., 1991; Kennedy and O'Hagan, 2000; Oakley and O'Hagan, 2004; O'Hagan, 2006).

Developed around the same time, the Kriging interpolation methods used in geostatistics are mathematically equivalent to the GP methods developed by Currin et al. (E.g. Cressie, 1990; Ripley, 2005). Kriging based emulators have been used for 25 years (Welch et al., 1992; Koehler and Owen, 1996), with recent implementations including the DICE-Kriging R packages used for GSA and inverse modelling (Marrel et al., 2009; Roustant et al., 2012). Since the latter approach is computationally faster, we adopted the DICE-Kriging version of the GP emulator for this study. For the statistical theory behind both emulator versions and descriptions of related R packages, see Hankin (2005) and Roustant et al. (2012).

Emulator-free global sensitivity analysis

For GSA studies involving highly multi-dimensional output, the time to compute the SIs can be significantly reduced by employing an emulator-free GSA approach. In this study, we consider two such methods using: (i) generalised additive modelling and (ii) a partial least squares regression approach.

A generalized additive model (GAM) is a generalized linear model where the predictor variables are represented by smooth functions (Wood, 2017). The general form of a GAM is:

$$\mathbf{Y} = g(\mathbf{X}) + \varepsilon \tag{5a}$$

$$g(\mathbf{X}) = s(\mathbf{x_1}) + s(\mathbf{x_2}) + \cdots + s(\mathbf{x_p}) \tag{5b}$$

where $\mathbf{Y}$ is an $n \times 1$ vector of model outputs, $\mathbf{X}=[\mathbf{x_1},\mathbf{x_2},\ldots,\mathbf{x_p}]$ is an $n \times p$ matrix which stores n samples of the p-dimensional inputs, $s(.)$ is the smoothing function such as a cubic spline, and ε is a zero-mean Normally distributed error term with constant variance. If we wish to include second order terms in $g(\mathbf{X})$, we would add $s(\mathrm{x}_1, \mathrm{x}_2) + s(\mathrm{x}_1, \mathrm{x}_3) + \cdots + s(\mathrm{x}_{p-1}, \mathrm{x}_p)$ to the right

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

hand size of equation (5b). A GAM it is not an emulator as defined by O'Hagan (2006) because the fitted values of the GAM are not exactly equal to the outputs of the training data (Wood, S.N, personal communication). It is still a meta-model and we can use it as a surrogate of the expensive model in order to perform variance based sensitivity analysis using for example the Sobol or extended FAST method. However, we have found that we require far more runs of the expensive chemistry model, compared to a GP emulator, for it to be an accurate surrogate for the model. Instead, it is possible to obtain accurate estimates of the first order SIs by using a GAM to estimate the components of equation (1) directly (Strong et al., 2014; Stanfill et al., 2015; Strong et al., 2015b). To compute the ith first order SI ($1 \leq i \leq p$), we first recognise that taking the expectation of equation (5a) leads to $E(\mathbf{Y}) = g(\mathbf{X})$. The expression for $E(\mathbf{Y}|\mathbf{X_i})$ is thus the marginal distribution of $E(\mathbf{Y})$. We could fit the full model and then compute this marginal distribution following Stanfill et al. (2014). However, an easier and quicker way is to fit a GAM to the ($\mathbf{X_i}$, $\mathbf{y}$) "data" where $\mathbf{X_i}$ represents the ith column of input matrix $\mathbf{X}$. Then, $E(\mathbf{Y}|\mathbf{X_i})$ consists of the fitted values of this reduced model (Strong et al., 2015b). Thus, $Var[E(\mathbf{Y}|\mathbf{X_i})]$ is determined by computing the variance of the n points from this fitted model. Finally, $Var(\mathbf{Y})$ is computed by taking the variance of the n samples of the training output data stored in $\mathbf{Y}$. The ith first order SI is then easily found by substituting the values of $Var[E(\mathbf{Y}|\mathbf{X_i})]$ and $Var(\mathbf{Y})$ into equation 1.

The partial least squares (PLS) method is the only one of the four GSA methods considered here that is not variance-based (Chang et al., 2015). Multivariate linear regression (MLR) is a commonly used tool to represent a set of outputs or response variables ($\mathbf{Y}$) based on a set of inputs or predictor variables ($\mathbf{X}$), where $\mathbf{X}$ and $\mathbf{Y}$ are matrices. MLR is only appropriate to use when the different inputs (columns in $\mathbf{X}$) are independent and not excessive in number. In

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

many situations, such as GSA studies, there can be a large number of inputs and they can be highly correlated (Sobie, 2009). PLS is an extension of MLR which is able to deal with these more challenging multivariate modelling problems (Wold et al., 2001). However, we note that for our experimental setup the inputs are in fact independent since this is an assumption of the variance based GSA which the three previously described methods are based on. The main reason for choosing PLS over other applicable regression approaches is that it has been shown to give similar estimates of the sensitivity indices to a variance based GSA approach (Chang et al., 2015). Thus, for sensitivity analysis problems when the inputs are correlated, this PLS method could be considered an alternative to the variance based GAM method which assumes that the inputs are independent. The mathematical details behind the PLS method can be found in the supplementary material (supplement S1).

Principal Component Analysis

As an alternative approach for speeding up the sensitivity analysis calculations, we computed the SIs from the Sobol GSA method using a hybrid approach involving principal component analysis (PCA) to reduce the dimensionality of the output space, and then use separate Gaussian Process emulators for each of the transformed outputs (Saltelli et al., 2012; Sexton et al., 2012; Gómez-Dans et al., 2016). PCA transforms the outputs onto a projected space with maximal variance. Mathematically, we obtain the matrix of transformed outputs $\mathbf{Z}$ – whose columns are orthogonal to one another – by multiplying the transposed matrix of outputs $\mathbf{Y}'$ by a matrix $\mathbf{A}$. The first column of $\mathbf{A}$ ($\mathbf{a_1}$) is chosen such that var($\mathbf{Y}'\mathbf{a_1}$) is maximised subject to the constraint $\mathbf{a_1}'\mathbf{a_1} = 1$. The vector $\mathbf{a_1}$ is called the first principal component (PC1), and we define λ_1 to be the principle eigenvalue of $\mathbf{S}$=Var($\mathbf{Y}'$) which is the largest variance of the outputs Y with respect to PC1. The ith (i>2) column of $\mathbf{A}$ ($\mathbf{a_1}$) is referred as the ith principal component, with λ_2, λ_3, λ_4, …

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

representing the second, third, fourth, … largest variance of **Y** with respect to PC2, PC3, PC4, …, respectively. PC1 contains the most information in the output, followed by PC2, then PC3, etc. The number of principal components required is determined by plotting the following points: $(1, \lambda_1)$, $(2, \lambda_1+\lambda_2)$, $(3, \lambda_1+\lambda_2+\lambda_3)$, …, and identifying the point where the line begins to flatten out. This is equivalent to choosing a cut off when most of the variance is explained.

Experimental setup

The sequence of tasks to complete when performing global sensitivity analysis is shown schematically in figure 1. The choice of inputs (e.g. parameters) to include in the sensitivity analysis will depend upon which have the greatest effects, based on expert knowledge of the model and field of study. Expert judgement is also needed to define the ranges of these inputs. A space-filling design such as maximin Latin hypercube sampling (see Appendix B for R code) or sliced Latin hypercube sampling (Ba et al., 2015) is required in order to sample from the input space with the minimum sufficient number of model runs. The third stage is to run the model at the set of input points specified by the space-filling sampling design.

If we are employing an emulator, the next stage is build the emulator using the training runs. We also need to perform runs of the computationally expensive model to validate the emulators. For this study, we ran the models with an additional set of inputs for validation. A simple comparison like this is usually sufficient, but more sophisticated diagnostics can also be carried out if needed (Bastos and O'Hagan, 2009). If employing the emulator-free approach, validation is also needed to do because we are using a statistical model to infer the SIs. Such a validation is not a central part of our results but is included in the supplemental material (Fig. S2). For the emulator-PCA hybrid approach (Figure 1), we found that the first 5 (for FRSGC) and 40 (for GISS) principal components were required to account for 99% of the variance. This

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

means that only 5-40 emulators are required to generate a global map in place of ~2000 needed if each grid point is emulated separately, which provides a large computational saving.

The final stage is to compute the first-order SIs for all the inputs; these quantify the sensitivity of the output to changes in each input. The SIs are also known as the main effects. The eFAST, Sobol and GAM approaches can also be used to compute the total effects, defined as the sum of the sensitivities of the output to changes in input *i* on its own and interacting with other inputs. For this study, we do not consider total effects as the sum of the main effects was close to 100% in each case.

Results

Validation of the emulators

Since the emulators we employed are based on a scalar output, we built a separate emulator for each of the ~2000 model grid points to represent the spatial distribution of the CH_4 lifetimes. At the 24 sets of inputs set aside for emulator validation, the predicted outputs from the emulators compared extremely well with the corresponding outputs from both chemistry models (Figure 2a,b, R^2=0.9996-0.9999, median absolute difference = 0.1-0.18 years). When PCA is used to reduce the output dimension from ~2000 to 5-40 (depending on the chemistry model), the accuracy of the predicted outputs was not as good (Figure 2c,d, R^2=0.9759-0.9991, median absolute difference = 0.94-3.44) but was still sufficient for this study.

Comparison of sensitivity indices

As expected, the two emulator-based global sensitivity analysis approaches (eFAST and Sobol) produced almost identical global maps of first order sensitivity indices (SIs, %) of CH_4 lifetime, see Figures 3 and 4. The statistics (mean, 95th percentile and 99th percentile) of the differences in

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

SIs between the two GSA methods over all 8 inputs at 2000 output points for the FRSGC and GISS models are shown in Figure 5, M1 vs M2.

Our results show that the GAM emulator-free GSA method produces very similar estimates of the SIs to the emulator-based methods (Figures 3-4; (a) vs (c)). The 95th and 99th percentiles of differences of the emulator-based method (eFAST or Sobol) versus GAM are 5% and 9% for FRSGC, and 7% and 10% for GISS (Figure 5; M1 vs M3). For both models, the PLS non-emulator-based method produced SIs that were significantly different from those using the eFAST and Sobol methods (Figures 3-4; (a) vs (d)). For FRSGC, the mean and 95th percentile of the differences in SIs for the emulator based method versus PLS was around 21% and 31%, while for GISS the corresponding values were around 14% and 23% (Figure 5; M1 vs M4). Thus, our results indicate that the PLS method is not suitable for use as an emulator-free approach to estimating the SIs.

The global map of SIs using the emulator-PCA hybrid approach compared well to those from the emulator-only approach (Figures 3-4; (a) vs (e)). The 95th and 99th percentiles of differences between the two approaches were 6% and 10%, respectively for FRSGC (Figure 5a, M1 vs M5) and 3% and 5%, respectively for GISS (Figure 5b, M1 vs M5). These are both higher than the corresponding values for the emulator-only methods (Figure 5, M1 vs M2; <2% and <3%, respectively). These higher values for the emulator-PCA hybrid approach is also reflected in the poorer estimates of the validation outputs using this approach versus the emulator-only approach (Figure 2). Such poorer estimates are expected because the PCA transformed outputs only explain 99% of the variance of the untransformed outputs used in the emulator-only approach.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

Discussion

Comparison of sensitivity indices

Our results align with the consensus that the eFAST method or other modified versions of the FAST method (e.g. RBD-FAST) produce very similar SIs to the Sobol method. Mathematically, the two methods are equivalent (Saltelli et al., 2012) and when the analytical (true) values of the SIs can be computed, both methods are able to accurately estimate these values (Mara and Tarantola, 2008; Iooss and Lemaître, 2015). However, many studies have noted that the Sobol method requires more model (or emulator) runs to compute the SIs. Saltelli et al. (2012) states that $\frac{2}{k} \times 100$ (%) more model runs are required for the Sobol method compared to eFAST, where k is the number of input factors (e.g. if $k=8$, then 25% more runs are needed for Sobol). Mara and Tarantola (2008) found that the Sobol method required ~10,000 runs of their model to achieve the same level of aggregated absolute error to that of FAST, which only needed 1000 runs. This is comparable to our analysis where the Sobol method required 18,000 runs of the emulator but only 1000 runs were needed for the eFAST method.

Given recent interest in applying generalized additive models (GAMs) to perform GSA (Strong et al., 2014, 2015a; Strong et al., 2015b), only Stanfill et al. (2015) has compared how they perform against other variance based approaches. The authors found that first order SIs estimated from the original FAST method were very close to the true values using 600 executions of the model, whereas the GAM approach only required 90-150 model runs. This is roughly consistent with our results, as we estimated the SIs using 80 runs of the chemistry models for GAM and 1000 runs of the emulator for the eFAST method.

There are a limited number of studies comparing the accuracy of the SIs of the GAM method amongst different models, as in our study. Stanfill et al. (2015) found that the GAM

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

method was accurate at estimating SIs based on a simple model (3-4 parameters) as well as a more complex one (10 parameters). However, if more models of varying complexity and type (e.g. process versus empirical) were to apply the GAM approach, we expect that while GAM would work well for some models, but for others the resulting SIs may be substantially different to that produced using the more traditional Sobol or eFAST methods. Saltelli et al. (1993) suggests that the performance of a GSA method can be model dependent, especially when the model is linear versus non-linear, monotonic versus non-monotonic, or if transformations are applied on the output (e.g. logarithms) or not. This is particularly true for GSA methods based on correlation or regression coefficients (Saltelli et al., 1999), which might explain why the SIs calculated from the PLS method in our analysis also disagreed with those of the eFAST/Sobol methods for the FRSGC versus GISS models. Not all GSA methods are model dependent; for example the eFAST method is not (Saltelli et al., 1999).

Principal Component Analysis

For both models, using principal component analysis (PCA) to significantly reduce the number of emulators needed resulted in SIs very similar to those calculated using an emulator-only approach. For the GISS model, this was encouraging given that the spread of points and their bias in the emulator against the model was noticeably larger than those of the FRSGC model (Figure 2c,d). If we had increased the number of principle components so that 99.9% of the variance in the output was captured rather than 99% , following Verrelst et al. (2016), then we would expect less bias in the validation plot for GISS. However, the poor validation plots did not translate into poorly estimated SIs for the emulator-PCA approach. On the contrary, the estimated SIs for GISS are consistent with the estimated SIs using the emulator-only approach (Fig. 5).

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

Implications for large scale sensitivity analysis studies

GSA studies for expensive models involving a small number of inputs (e.g. <10) are useful and straightforward to implement (Lee et al., 2012). However, the inferences made are limited due to the large number of parameters on which these models depend and the number of processes that they simulate. Hence, interest is growing in carrying out large scale GSA studies involving a high number of inputs to improve understanding of an individual model (e.g. Lee et al., 2013) or to diagnose differences between models (Wild et al., in prep.). For GSA studies when the number of inputs is small, our study has demonstrated that the GAM approach is a good candidate for carrying out emulator-free GSA since it calculates very similar SIs without the computational demands of emulation. A caveat is that the performance of GAM may depend on the behaviour of the model; although we have found it is a good GSA method for our models (FRSGC and GISS) and output (CH_4 lifetimes) its suitability may not be as good in all situations.

5. Conclusion

Global sensitivity analysis (GSA) is a powerful tool for understanding model behaviour, for diagnosing differences between models and for determining which parameters to choose for model calibration. In this study, we compared different methods for computing first order sensitivity indices for computationally expensive models based on modelled spatial distributions of CH_4 lifetimes. We have demonstrated that the more established emulator-based methods (eFAST and Sobol) can be used to efficiently derive meaningful sensitivity indices for multi-dimensional output from atmospheric chemistry transport models. We have shown that an emulator-free method based on a generalised additive model (GAM) and an emulator-PCA hybrid method produce first order sensitivity indices that are consistent with the emulator-only

methods. For a reasonably smooth system with few parameters, as investigated here, the GAM and PCA methods are viable and effective options for GSA, and are robust over models that exhibit distinctly different responses. Moreover, the computational benefits of these alternative methods is apparent, with the PCA and GAM approaches allowing calculation of variance based sensitivity indices about 6 and 30 times faster than traditional emulator-only methods. Finally, we have provided guidance on how to implement these methods in a reproducible way.

Code Availability

The R code to carry out global sensitivity analysis using the methods described in this paper are available in appendices B-E following the *Data Availability* section, and in the appendices S2-S4 of the supplemental material. This R code as well as the R code used to validate the emulators is also be found via http://doi.org/10.5281/zenodo.1038667.

Data Availability

The inputs and outputs of the FRSGC chemistry model that was used to train the emulators in this paper can be found via http://doi.org/10.5281/zenodo.1038670.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

Geoscientific Model Development Discussions

Appendix A: Further details of the Sobol and eFAST global sensitivity analysis methods

Sobol method: Saltelli (2002) and Tarantola et al. (2006) suggest using eight variants of equation (2), using different combinations of $\mathbf{y_A}$, $\mathbf{y_B}$, $\mathbf{y_{Ci}}$ and $\mathbf{y_{Di}}$:

$$\widehat{S}_i^{\,I} = \frac{\mathbf{y_A}\cdot\mathbf{y_{C_i}} - \left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{A}}^{(j)}\right)\left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{B}}^{(j)}\right)}{\mathbf{y_A}\cdot\mathbf{y_A} - \left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{A}}^{(j)}\right)\left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{B}}^{(j)}\right)} \qquad \widehat{S}_i^{\,II} = \frac{\mathbf{y_A}\cdot\mathbf{y_{D_i}} - \left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{A}}^{(j)}\right)\left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{B}}^{(j)}\right)}{\mathbf{y_B}\cdot\mathbf{y_B} - \left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{A}}^{(j)}\right)\left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{B}}^{(j)}\right)}$$

$$\widehat{S}_i^{\,III} = \frac{\mathbf{y_A}\cdot\mathbf{y_{C_i}} - \left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{A}}^{(j)}\right)\left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{B}}^{(j)}\right)}{\mathbf{y_B}\cdot\mathbf{y_B} - \left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{A}}^{(j)}\right)\left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{B}}^{(j)}\right)} \qquad \widehat{S}_i^{\,IV} = \frac{\mathbf{y_A}\cdot\mathbf{y_{C_i}} - \left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{C_i}}^{(j)}\right)\left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{D_i}}^{(j)}\right)}{\mathbf{y_{C_i}}\cdot\mathbf{y_{C_i}} - \left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{C_i}}^{(j)}\right)\left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{D_i}}^{(j)}\right)}$$

$$\widehat{S}_i^{\,V} = \frac{\mathbf{y_A}\cdot\mathbf{y_{C_i}} - \left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{C_i}}^{(j)}\right)\left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{D_i}}^{(j)}\right)}{\mathbf{y_{D_i}}\cdot\mathbf{y_{D_i}} - \left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{C_i}}^{(j)}\right)\left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{D_i}}^{(j)}\right)} \qquad \widehat{S}_i^{\,VI} = \frac{\mathbf{y_B}\cdot\mathbf{y_{D_i}} - \left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{A}}^{(j)}\right)\left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{B}}^{(j)}\right)}{\mathbf{y_A}\cdot\mathbf{y_A} - \left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{A}}^{(j)}\right)\left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{B}}^{(j)}\right)}$$

$$\widehat{S}_i^{\,VII} = \frac{\mathbf{y_B}\cdot\mathbf{y_{D_i}} - \left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{C_i}}^{(j)}\right)\left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{D_i}}^{(j)}\right)}{\mathbf{y_{C_i}}\cdot\mathbf{y_{C_i}} - \left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{C_i}}^{(j)}\right)\left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{D_i}}^{(j)}\right)} \qquad \widehat{S}_i^{\,VIII} = \frac{\mathbf{y_B}\cdot\mathbf{y_{D_i}} - \left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{C_i}}^{(j)}\right)\left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{D_i}}^{(j)}\right)}{\mathbf{y_{D_i}}\cdot\mathbf{y_{D_i}} - \left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{C_i}}^{(j)}\right)\left(\frac{1}{N}\sum_{j=1}^{N} y_{\mathbf{D_i}}^{(j)}\right)}$$

Thus, the *i*th first order Sobol SI estimate is:

$$\widehat{S}_i = \frac{1}{8}\left(\widehat{S}_i^{\,I} + \widehat{S}_i^{\,II} + \widehat{S}_i^{\,III} + \widehat{S}_i^{\,IV} + \widehat{S}_i^{\,V} + \widehat{S}_i^{\,VI} + \widehat{S}_i^{\,VII} + \widehat{S}_i^{\,VIII}\right)$$

The extended FAST (eFAST) method: The main difference between classical FAST (Cukier et al., 1973), and extended FAST (Saltelli et al., 1999) when computing first order SIs is the choice of transformation function G_i:

Classical FAST: $G_i(z) = \bar{x}_i e^{\bar{v}_s z}$, ($\bar{x}_i$, $\bar{v}_s$ are user-specified) (A1a)

Extended FAST: $G_i(z) = \frac{1}{2} + \frac{1}{\pi}\arcsin(z)$ (A1b)

Using equation (A1b), equation (3) now becomes a straight line equation:

$$x_i(s) = \frac{1}{2} + \frac{1}{\pi}\omega_i s$$

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

Appendix B: Setting up the global sensitivity analysis experiment

Prior to running the scripts accompanying this paper, the following packages are required:

install.packages("lhs"); install.packages("emulator"); install.packages("mvtnorm");

install.packages("mgcv"); install.packages("sensitivity"); install.packages("DiceKriging");

install.packages("DiceOptim")

(For first time users of the open source programming language R, go to https://www.r-project.org/ to download it for free).

We decided on the model inputs/parameters, the ranges of the inputs, and the outputs. For applications to other problems, note that the outputs could be a spatial map, a time-series or a combination of both. We created the design matrix by running the following as a new script file, specifying the number of input factor and their ranges:

```
#Things to specify by the user:
setwd("C:/Users/....")  #Location of folder where files are stored.
Np = ??    #No. of input factors
mink = c(p1min,p2min,p3min,....)  #Min values of the inputs/parameters.
maxk = c(p1max,p2max,p3max,...) #Max values of the inputs/parameters.
library(lhs); library(emulator)
inputs_norm = maximinLHS(Np*10,Np)
write.table(inputs_norm,"InputsNorm_TrainingData.csv",row.names=F,col.names=F,sep=",")
inputs = matrix(-9999,nrow=k*10,ncol=k)
for (i in 1:k){inputs[,i] = (Inputs_norm[,i]*(maxk[i]-mink[i])) + mink[i]}c
write.table(inputs,"Inputs_TrainingData.csv", row.names=F, col.names=F, sep=",")
```

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

For input ranges that were on the log-scale (e.g. the min/max of input/parameter *p* is 0.01*pCtrl and 100*pCtrl, where pCtrl is the control run value of the input/parameter *p*), then we first transformed to a linear scale before running the script.

We ran the chemistry model for each of the rows of the *Inputs_TrainingData.csv* file and stored the outputs in a csv file *Outputs_TrainingData.csv*. The outputs in this csv file consisted of *Nx* rows and *Ny* columns, where *Nx*= *Np**10 is the number of runs of the model, and *Ny* is the length of the row vector storing the output for a given input.

Appendix C: Calculating first order and total sensitivity indices using the extended FAST (eFAST) method

We ran the following as a new R script file:

```
library(sensitivity); library(DiceKriging); library(DiceOptim)
X =read.csv('InputsNorm_TrainingData.csv', header=FALSE)
yALL =read.csv('Outputs_TrainingData.csv', header=FALSE)
SI = matrix(-9999,nrow=dim(X)[2],ncol= dim(yALL)[2])
SI.total = matrix(-9999,nrow=dim(X)[2],ncol= dim(yALL)[2])
for (j in 1: dim(yALL)[2]){y = as.matrix(yALL[,j],rownames.force=NA)
m = km(~ ., design = X, response = y, covtype = "matern3_2")
kriging.mean = function(Xnew, m){predict.km(m, Xnew, "UK", se.compute = FALSE,
checkNames = FALSE)$mean}
temp = fast99(model = kriging.mean, factors = dim(X)[2], n = 1000, q = "qunif", q.arg =
list(min = 0, max = 1), m = m)
SI[,j] = as.matrix(temp$D1/temp$V)
SI.total[,j] = as.matrix((temp$V- temp$Dt)/temp$V)}
```

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017


```
write.table(SI*100," SIs_eFAST_EmulatorOnly.csv", row.names=F, col.names=F, sep=",")
write.table(SI.total*100, "TotalSIs_eFAST_EmulatorOnly.csv", row.names=F, col.names=F,
sep=",")
```

Appendix D: Calculating first and higher order sensitivity indices using the Generalized Additive Model (GAM) method.

We ran the following as a new R script file:

```
library(mgcv);
X =read.csv('InputsNorm_TrainingData.csv', header=FALSE)
yALL =read.csv('OutputsNorm_TrainingData.csv', header=FALSE)
SI = matrix(-9999,nrow=dim(inputs)[2],ncol= dim(outputs)[2])
SI.total = matrix(-9999,nrow=dim(inputs)[2],ncol= dim(outputs)[2])
for (j in 1:dim(outputs)[2]){y = as.matrix(yALL[,j],rownames.force=NA)
SI = matrix(-9999,nrow=dim(inputs)[2],ncol=dim(outputs)[2])
for (j in 1:dim(outputs)[2]){y=yALL[,j]; vary=var(Y); v=rep(-9999,8);
for (i in 1:dim(inputs)[2]){gam.model = gam(Y ~ te(X[,i])); v[i]=var(gam.model$fitted)}-
SI[,j]=(v/varY)*100}
write.table(SI, "SIs_GAM.csv", row.names=F, col.names=F, sep=",")
```

Note that the second order SIs of the *i*th and *k*th inputs can be computed by replacing *gam(Y ~ te(X[,i]))* with *gam.model = gam(Y ~ ti(X[,i], X[,k]))*.

Author contributions

ER and OW designed the study. ER conducted the analysis and wrote the manuscript and OW gave feedback during the analysis and writing up phases. OW, FO and AW provided output from the global atmospheric model runs needed to carry out the analysis. LL advised on statistical aspects of the analysis. All coauthors gave feedback on drafts of the manuscript.

Acknowledgements

This work was supported by the Natural Environment Research Council [grant number NE/N003411/1].

References

Ahtikoski, A., Heikkilä, J., Alenius, V., and Siren, M.: Economic viability of utilizing biomass energy from young stands—the case of Finland, Biomass and Bioenergy, 32, 988-996, 2008.

Ba, S., Myers, W. R., and Brenneman, W. A.: Optimal sliced Latin hypercube designs, Technometrics, 57, 479-487, 2015.

Bailis, R., Ezzati, M., and Kammen, D. M.: Mortality and greenhouse gas impacts of biomass and petroleum energy futures in Africa, Science, 308, 98-103, 2005.

Bastos, L. S. and O'Hagan, A.: Diagnostics for Gaussian process emulators, Technometrics, 51, 425-438, 2009.

Campbell, J. E., Carmichael, G. R., Chai, T., Mena-Carrasco, M., Tang, Y., Blake, D., Blake, N., Vay, S. A., Collatz, G. J., and Baker, I.: Photosynthetic control of atmospheric carbonyl sulfide during the growing season, Science, 322, 1085-1088, 2008.

Carslaw, K., Lee, L., Reddington, C., Pringle, K., Rap, A., Forster, P., Mann, G., Spracklen, D., Woodhouse, M., and Regayre, L.: Large contribution of natural aerosols to uncertainty in indirect forcing, Nature, 503, 67-71, 2013.

Chang, E. T., Strong, M., and Clayton, R. H.: Bayesian sensitivity analysis of a cardiac cell model using a Gaussian process emulator, PloS one, 10, e0130252, 2015.

Coggan, J. S., Bartol, T. M., Esquenazi, E., Stiles, J. R., Lamont, S., Martone, M. E., Berg, D. K., Ellisman, M. H., and Sejnowski, T. J.: Evidence for ectopic neurotransmission at a neuronal synapse, Science, 309, 446-451, 2005.

Cressie, N.: The origins of kriging, Mathematical geology, 22, 239-252, 1990.

Cukier, R., Fortuin, C., Shuler, K. E., Petschek, A., and Schaibly, J.: Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory, The Journal of chemical physics, 59, 3873-3878, 1973.

Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D.: Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments, Journal of the American Statistical Association, 86, 953-963, 1991.

de Gee, M., Lof, M. E., and Hemerik, L.: The effect of chemical information on the spatial distribution of fruit flies: II parameterization, calibration, and sensitivity, Bulletin of mathematical biology, 70, 1850, 2008.

Degroote, J., Couckuyt, I., Vierendeels, J., Segers, P., and Dhaene, T.: Inverse modelling of an aneurysm's stiffness using surrogate-based optimization and fluid-structure interaction simulations, Structural and Multidisciplinary Optimization, 46, 457-469, 2012.

Ferretti, F., Saltelli, A., and Tarantola, S.: Trends in sensitivity analysis practice in the last decade, Science of The Total Environment, 2016.

Goldstein, M. and Rougier, J.: Bayes linear calibrated prediction for complex systems, Journal of the American Statistical Association, 101, 1132-1143, 2006.

Gómez-Dans, J. L., Lewis, P. E., and Disney, M.: Efficient Emulation of Radiative Transfer Codes Using Gaussian Processes and Application to Land Surface Parameter Inferences, Remote Sensing, 8, 119, 2016.

Hakami, A., Odman, M. T., and Russell, A. G.: Nonlinearity in atmospheric response: A direct sensitivity analysis approach, Journal of Geophysical Research: Atmospheres, 109, 2004.

Hankin, R. K.: Introducing BACCO, an R package for Bayesian analysis of computer code output, Journal of Statistical Software, 14, 1-21, 2005.

Hill, T. C., Ryan, E., and Williams, M.: The use of CO2 flux time series for parameter and carbon stock estimation in carbon cycle research, Global Change Biology, 18, 179-193, 2012.

Homma, T. and Saltelli, A.: Importance measures in global sensitivity analysis of nonlinear models, Reliability Engineering & System Safety, 52, 1-17, 1996.

Iooss, B. and Lemaître, P.: A review on global sensitivity analysis methods. In: Uncertainty Management in Simulation-Optimization of Complex Systems, Springer, 2015.

Kennedy, M., Anderson, C., O'Hagan, A., Lomas, M., Woodward, I., Gosling, J. P., and Heinemeyer, A.: Quantifying uncertainty in the biospheric carbon flux for England and Wales, Journal of the Royal Statistical Society: Series A (Statistics in Society), 171, 109-135, 2008.

Kennedy, M. C. and O'Hagan, A.: Predicting the output from a complex computer code when fast approximations are available, Biometrika, 87, 1-13, 2000.

Koehler, J. and Owen, A.: 9 Computer experiments, Handbook of statistics, 13, 261-308, 1996.

Lee, L., Carslaw, K., Pringle, K., and Mann, G.: Mapping the uncertainty in global CCN using emulation, Atmospheric Chemistry and Physics, 12, 9739-9751, 2012.

Lee, L., Pringle, K., Reddington, C., Mann, G., Stier, P., Spracklen, D., Pierce, J., and Carslaw, K.: The magnitude and causes of uncertainty in global model simulations of cloud condensation nuclei, Atmos. Chem. Phys, 13, 8879-8914, 2013.

Lilburne, L. and Tarantola, S.: Sensitivity analysis of spatial models, International Journal of Geographical Information Science, 23, 151-168, 2009.

Mara, T. A. and Tarantola, S.: Application of global sensitivity analysis of model output to building thermal simulations, 2008, 290-302.

Marrel, A., Iooss, B., Laurent, B., and Roustant, O.: Calculations of sobol indices for the gaussian process metamodel, Reliability Engineering & System Safety, 94, 742-751, 2009.

O'Hagan, A.: Bayesian analysis of computer code outputs: a tutorial, Reliability Engineering & System Safety, 91, 1290-1300, 2006.

Oakley, J. E. and O'Hagan, A.: Probabilistic sensitivity analysis of complex models: a Bayesian approach, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 66, 751-769, 2004.

Pistone, G. and Vicario, G.: Kriging prediction from a circular grid: application to wafer diffusion, Applied Stochastic Models in Business and Industry, 29, 350-361, 2013.

Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., and Tucker, P. K.: Surrogate-based analysis and optimization, Progress in aerospace sciences, 41, 1-28, 2005.

Rasmussen, C. E.: Gaussian processes for machine learning, 2006.

Ripley, B. D.: Spatial statistics, John Wiley & Sons, 2005.

Roustant, O., Ginsbourger, D., and Deville, Y.: DiceKriging, DiceOptim: Two R packages for the analysis of computer experiments by kriging-based metamodeling and optimization, 2012.

Saltelli, A.: Making best use of model evaluations to compute sensitivity indices, Computer Physics Communications, 145, 280-297, 2002.

Saltelli, A., Andres, T., and Homma, T.: Sensitivity analysis of model output: an investigation of new techniques, Computational statistics & data analysis, 15, 211-238, 1993.

Saltelli, A. and Annoni, P.: How to avoid a perfunctory sensitivity analysis, Environmental Modelling & Software, 25, 1508-1517, 2010.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S.: Global sensitivity analysis: the primer, John Wiley & Sons, 2008.

Saltelli, A., Ratto, M., Tarantola, S., and Campolongo, F.: Update 1 of: Sensitivity analysis for chemical models, Chemical reviews, 112, PR1-PR21, 2012.

Saltelli, A., Tarantola, S., and Chan, K.-S.: A quantitative model-independent method for global sensitivity analysis of model output, Technometrics, 41, 39-56, 1999.

Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., Bauer, M., Bauer, S. E., Bhat, M. K., and Bleck, R.: Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive, Journal of Advances in Modeling Earth Systems, 6, 141-184, 2014.

Sexton, D. M., Murphy, J. M., Collins, M., and Webb, M. J.: Multivariate probabilistic projections using imperfect climate models part I: outline of methodology, Climate dynamics, 38, 2513-2542, 2012.

Shindell, D., Faluvegi, G., Unger, N., Aguilar, E., Schmidt, G., Koch, D., Bauer, S. E., and Miller, R. L.: Simulations of preindustrial, present-day, and 2100 conditions in the NASA GISS composition and climate model G-PUCCINI, Atmospheric Chemistry and Physics, 6, 4427-4459, 2006.

Sobie, E. A.: Parameter sensitivity analysis in electrophysiological models using multivariable regression, Biophysical journal, 96, 1264-1274, 2009.

Sobol, I. y. M.: On sensitivity estimation for nonlinear mathematical models, Matematicheskoe Modelirovanie, 2, 112-118, 1990.

Stanfill, B., Mielenz, H., Clifford, D., and Thorburn, P.: Simple approach to emulating complex computer models for global sensitivity analysis, Environmental Modelling & Software, 74, 140-155, 2015.

Stites, E. C., Trampont, P. C., Ma, Z., and Ravichandran, K. S.: Network analysis of oncogenic Ras activation in cancer, Science, 318, 463-467, 2007.

Strong, M., Oakley, J. E., and Brennan, A.: An efficient method for computing the Expected Value of Sample Information. A non-parametric regression approach. 2015a.

Strong, M., Oakley, J. E., and Brennan, A.: Estimating multiparameter partial expected value of perfect information from a probabilistic sensitivity analysis sample a nonparametric regression approach, Medical Decision Making, 34, 311-326, 2014.

Strong, M., Oakley, J. E., Brennan, A., and Breeze, P.: Estimating the expected value of sample information using the probabilistic sensitivity analysis sample a fast nonparametric regression-based method, Medical Decision Making, 0272989X15575286, 2015b.

Tarantola, S., Gatelli, D., and Mara, T. A.: Random balance designs for the estimation of first order global sensitivity indices, Reliability Engineering & System Safety, 91, 717-727, 2006.

Vanuytrecht, E., Raes, D., and Willems, P.: Global sensitivity analysis of yield output from the water productivity model, Environmental Modelling & Software, 51, 323-332, 2014.

Verrelst, J., Sabater, N., Rivera, J. P., Muñoz-Marí, J., Vicent, J., Camps-Valls, G., and Moreno, J.: Emulation of Leaf, Canopy and Atmosphere Radiative Transfer Models for Fast Global Sensitivity Analysis, Remote Sensing, 8, 673, 2016.

Voulgarakis, A., Naik, V., Lamarque, J.-F., Shindell, D. T., Young, P., Prather, M. J., Wild, O., Field, R., Bergmann, D., and Cameron-Smith, P.: Analysis of present day and future OH and methane lifetime in the ACCMIP simulations, Atmospheric Chemistry and Physics, 13, 2563-2587, 2013.

Vu-Bac, N., Rafiee, R., Zhuang, X., Lahmer, T., and Rabczuk, T.: Uncertainty quantification for multiscale modeling of polymer nanocomposites with correlated parameters, Composites Part B: Engineering, 68, 446-464, 2015.

Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J., and Morris, M. D.: Screening, predicting, and computer experiments, Technometrics, 34, 15-25, 1992.

Wild, O.: Modelling the global tropospheric ozone budget: exploring the variability in current models, Atmospheric Chemistry and Physics, 7, 2643-2660, 2007.

Wild, O., Pochanart, P., and Akimoto, H.: Trans-Eurasian transport of ozone and its precursors, Journal of Geophysical Research: Atmospheres, 109, 2004.

Wild, O. and Prather, M. J.: Excitation of the primary tropospheric chemical mode in a global three-dimensional model, Journal of geophysical research, 105, 2000.

Wild, O., Ryan, E., O'Connor, F., Vougarakis, A., and Lee, L.: Reducing Uncertainty in Model Budgets of Tropospheric Ozone and OH., Intended for submission to Atmospheric Chemistry and Physics, in prep.

Wold, S., Sjöström, M., and Eriksson, L.: PLS-regression: a basic tool of chemometrics, Chemometrics and intelligent laboratory systems, 58, 109-130, 2001.

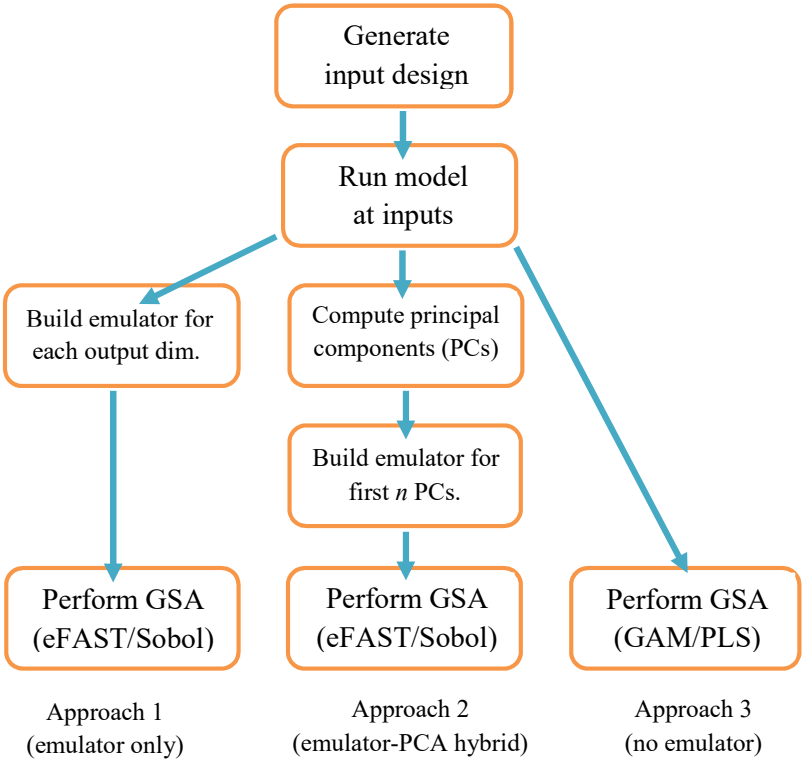
Wood, S. N.: Generalized additive models: an introduction with R, CRC press, 2017.

Wu, J., Dhingra, R., Gambhir, M., and Remais, J. V.: Sensitivity analysis of infectious disease models: methods, advances and their application, Journal of The Royal Society Interface, 10, 20121018, 2013.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

Geoscientific Model Development Discussions

1 **Figure and Tables**



2

3 **Figure 1.** Flow-chart for order of tasks to complete in order to perform global sensitivity analysis
4 (GSA) on a computationally expensive model. The ranges on the inputs, from which its design
5 is based, are determined by expert elicitation. For approach 1, the dimensions of the output
6 consist of different spatial or temporal points of the same output variable (CH_4 lifetime for this
7 study). For approach 2, a principal component (PCs) is a linear combination of the different
8 dimensions of the output, where *n* is chosen such that the first *n* PCs explain 99% of the variance
9 of the output.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

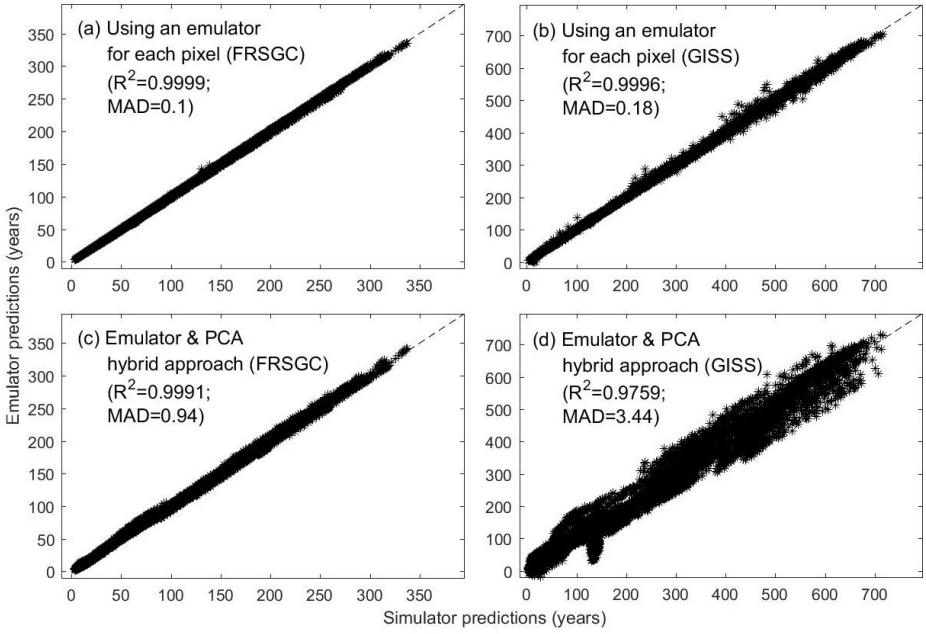


Figure 2. Annual column mean CH_4 lifetime calculated by the FRSGC and GISS chemistry models from each of 24 validation runs (x-axis) versus that predicted by the emulator (y-axis). In each plot, the R^2 and median absolute difference (MAD) are given as metrics for the accuracy of the emulator predictions. Each validation run contains ~2000 different output values, corresponding to different latitude-longitude grid squares.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

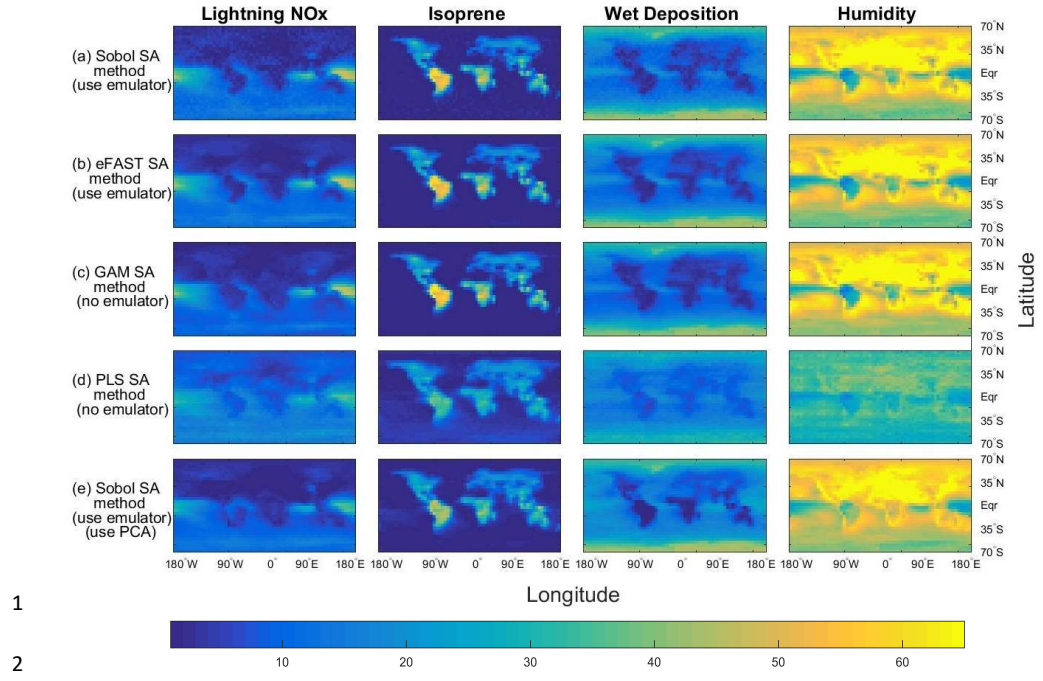


Figure 3. The sensitivity indices (percentage of the total variance in a given output) for the four dominant inputs, for annual column mean CH_4 lifetime in the FRSGC chemistry transport model. The rows show the results from five different methods for performing sensitivity analysis (SA).

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

Geoscientific Model Development Discussions
Open Access
EGU

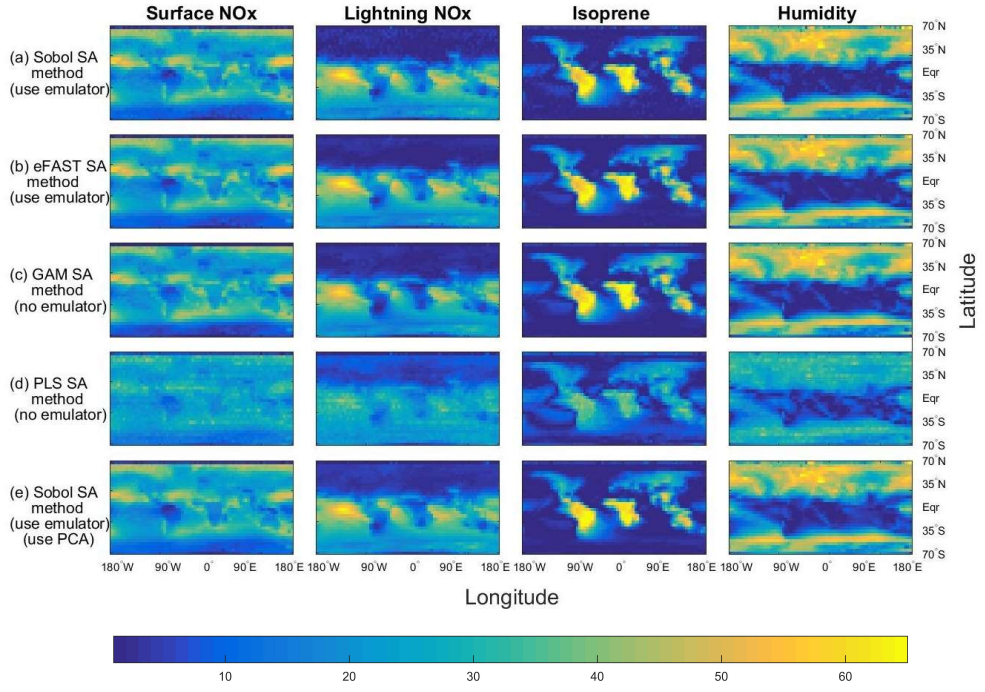


Figure 4. As figure 3, showing sensitivity indices for CH_4 lifetime from the GISS chemistry transport model. Note that some of the four dominant inputs differ from those in Fig 3.

Geosci. Model Dev. Discuss., https://doi.org/10.5194/gmd-2017-271
Manuscript under review for journal Geosci. Model Dev.
Discussion started: 13 November 2017

Geoscientific Model Development Discussions
Open Access
EGU

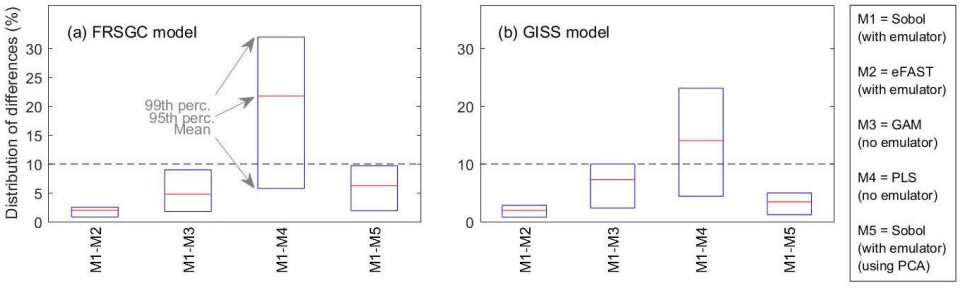


Figure 5. Statistics (mean, 95% percentile and maximum) of the distribution of differences in sensitivity indices (SIs) between pairs of methods. For each comparison, the 16,000 pairs of SIs are made up of ~2000 pairs of SIs for each of the 8 inputs.