Generalized Sobol' indices for multi-output regression models

Robert A. Milton, Solomon F. Brown, Aaron S. Yeardley

Department of Chemical and Biological Engineering, University of Sheffield, Sheffield, S1 3JD, United Kingdom

Abstract

Variance based global sensitivity usually measures the relevance of inputs to a single output using Sobol' indices. This paper extends the definition in a natural way to multiple outputs, directly measuring the relevance of inputs to the linkages between outputs in a correlation-like matrix of indices. The usual Sobol' indices constitute the diagonal of this matrix. Existence, uniqueness and uncertainty quantification are established by developing the indices from a putative regression model. Analytic expressions for generalized Sobol' indices and their standard errors are computed for Gaussian Process regression with an anisotropic radial basis function kernel. The formulae allow for rotation of the inputs to facilitate locating an active subspace.

Keywords: Global Sensitivity Analysis, Sobol' Index, Surrogate Model, Multi-Output, Gaussian Process, Uncertainty Quantification

1. Introduction

This paper is concerned with analysing the results of experiments or computer simulations in a design matrix of $M \geq 1$ input and $L \geq 1$ output columns, over N rows (datapoints). Global Sensitivity Analysis (GSA) [1] examines the relevance of the various inputs to the various outputs. When pursued via ANOVA decomposition, this leads naturally to the well known

Email addresses: `r.a.milton@sheffield.ac.uk` (Robert A. Milton), `s.f.brown@sheffield.ac.uk` (Solomon F. Brown), `asyeardley1@sheffield.ac.uk` (Aaron S. Yeardley)

Sobol' indices, which have by now been applied across most fields of science and engineering [2, 3].

The Sobol' decomposition apportions the variance of an output to sets of one or more inputs [4]. We shall use ordinals of inputs $\mathbf{m} := (0, \ldots, m-1) \subseteq \mathbf{M}$, tuples which are conveniently also naive sets. The maximal ordinal $\mathbf{M}$ of all M inputs explains everything explicable, so its Sobol' index is 1 by definition. The void ordinal $\mathbf{0}$ explains nothing, so its Sobol' index is 0 by definition. The influence of an isolated ordinal of inputs $\mathbf{m}$ is measured by its closed Sobol' index $S_{\mathbf{m}} \in [0, 1]$. A first-order Sobol' index $S_{m'}$ is simply the closed Sobol' index of a single input m'. Because inputs in an isolated ordinal may act in concert with each other, the influence of the ordinal often exceeds the sum of first-order contributions from its members, always obeying

$$S_{\mathbf{m}} \geq \sum_{m' \in \mathbf{m}} S_{m'}$$

The total Sobol index $S^T_{\mathbf{M}-\mathbf{m}} \geq 0$ of the set theoretic complement $\mathbf{M} - \mathbf{m}$ is $1 - S_{\mathbf{m}}$, which expresses the influence of non-isolated inputs $\mathbf{M} - \mathbf{m}$ allowed to act in concert with each other *and* isolated inputs $\mathbf{m}$. When speaking of irrelevant inputs $\mathbf{M} - \mathbf{m}$, we mean that $S^T_{\mathbf{M}-\mathbf{m}} \approx 0$. This is synonymous with the isolated ordinal of inputs $\mathbf{m}$ explaining everything explicable $S_{\mathbf{m}} \approx 1$. It is apparent that we can readily obtain any Sobol' index of interest by ordering input dimensions appropriately and calculating the closed index $S_{\mathbf{m}}$ of some ordinal set $\mathbf{m}$.

Perhaps the most significant use of closed Sobol' indices is to identify a representative reduced model of $m \leq M$ inputs within the full model $\mathbf{M}$. Apportioning variance is mathematically equivalent to squaring a correlation coefficient to produce a coefficient of determination R^2 [5]. A closed Sobol' index is thus a coefficient of determination between the predictions from the reduced model $\mathbf{m}$ and predictions from the full model $\mathbf{M}$. A closed Sobol' index close to 1 confirms that the two models make nearly identical predictions. Simplicity and economy (not least of calculation) motivate the adoption of a reduced model, a closed Sobol' index close to 1 permits it.

The discussion thus far, and almost all prior GSA, has dealt with a single (i.e scalar) output. With multiple (i.e vector) outputs, the Sobol' decomposition apportions the covariance matrix of outputs rather than the variance of a single output. With L outputs, the closed Sobol' index $S_{\mathbf{m}}$ is generally a symmetric $\mathbf{L} \times \mathbf{L}$ matrix. The diagonal elements express the relevance of inputs to the output variables themselves. The off-diagonal elements express relevance to the linkages between outputs. This may be of considerable interest when outputs are, for example, yield and purity of a product, or perhaps a single output measured at various times. The Sobol indices reveal (amongst other things) which inputs it is worthwhile varying in an effort to alter the linkages between outputs. Prior work on Sobol' indices with multiple outputs [6, 7, 8] has settled ultimately on just the diagonal elements of the covariance matrix, so this linkage remains unexamined. Although output covariance has been incoroporated indirectly in prior studies by performing principal component analysis (PCA) on ouputs prior to GSA on the (diagonal) variances of the resulting output basis [9]. This has been used in particular to study synthetic "multi-outputs" which are actually the dynamic response of a single output over time [10, 11].

Accurate calculation of Sobol' indices even for a single output is computationally expensive and requires 10,000+ datapoints [12]. A (sometimes) more efficient approach is calculation via a surrogate model, such as Polynomial Chaos Expansion (PCE) [13, 14, 15], low-rank tensor approximation [16, 17], and support vector regression [18]. As well as being efficient, surrogate models also smooth out noise in the output, which is often highly desirable in practice. This paper employs one of the most popular surrogates, the Gaussian Processes (GP) [19, 20] as it is highly tractable. We shall follow the multi-output form (MOGP) described in [21], in order to examine the linkages between outputs. This paper deals exclusively with the anisotropic Radial Basis Function kernel, known as RBF/ARD, which is widely accepted as the kernel of choice for smooth outputs [22]. This uses the classic Gaussian bell curve to express the proximity of two input points, described in detail in Sections 4.1 and 4.2.

Semi-analytic expressions for Sobol' indices are available for scalar PCEs [23], and the diagonal elements of multi-output PCEs [8]. Semi-analytic expressions for Sobol' indices of GPs have been provided in integral form by [24] and alternatively by [25]. These approaches are implemented, examined and compared in [26, 27]. Both [24, 26] estimate the errors on Sobol' indices in semi-analytic, integral form. Fully analytic, closed form expressions have been derived without error estimates for uniformly distributed inputs [28] with an RBF kernel. There are currently no closed form expressions for MOGPs, or the errors on Sobol' indices, or any GPs for which inputs are not uniformly distributed.

In this paper we provide explicit, closed-form analytic formulae for the $\mathbf{L}\times$

$\mathbf{L}$ matrices of closed Sobol' indices and their errors, for a class of MOGP with an RBF/ARD kernel applicable to smoothly varying outputs. We transform uniformly distributed inputs u to normally distributed inputs z prior to fitting a GP and performing analytic calculation of closed Sobol' indices. This leads to relatively concise expressions in terms of exponentials, and enables ready calculation of the errors (variances) of these expressions. It also allows for an arbitrary rotation Θ of inputs, as normal variables are additive, whereas summing uniform inputs does not produce uniform inputs. If the goal is reducing inputs, rotating their basis first boosts the possibilities immensely [29]. It presents the possibilty of choosing Θ to maximise the closed Sobol' index of the first few inputs.

The quantities to be calculated and their formal context are introduced in Section 2, assuming only that the output is an integrable function of the input. Our approach effectively regards a regression model which quantifies uncertainty with each prediction as just another name for a stochastic process. A great deal of progress is made in Section 3 using general stochastic (not necessarily Gaussian) processes. This approach is analytically cleaner, as it is not obfuscated by the GP details. Furthermore, it turns out that the desirable properties of the Gaussian (lack of skew, simple kurtosis) are not actually helpful, as these terms cancel of their own accord. This development leaves just two terms to be calculated, which require the stochastic process to be specified. MOGPs with an RBF/ARD kernel are tersely developed and described in Section 4, then used to calculate the two unknown terms in Sections 5 and 6. Methods to reduce computational complexity are discussed in Section 7.Conclusions are drawn in Section 8.

2. Generalized Sobol' indices

Apply a constant offset to a Lebesgue integrable model so that

$$y\colon [0,1]^{M+1} \to \mathbb{R}^L \quad \text{obeys} \quad \int y(u)\,\mathrm{d}u = [0]_{\mathbf{L}} \tag{1}$$

taking as input a uniformly distributed random variable (RV)

$$\mathsf{u} \sim \mathsf{U}\big([0]_{\mathbf{M+1}}\,, [1]_{\mathbf{M+1}}\big) := \mathsf{U}(0,1)^{M+1} \tag{2}$$

Throughout this paper exponentiation is categorical – repeated cartesian $\times$ or tensor $\otimes$ – unless otherwise specified. Square bracketed quantities are tensors, carrying their axes as a subscript tuple. In this case the subscript tuple is the von Neumann ordinal

$$\mathbf{M}+\mathbf{1} := (0,\ldots,M) \supset \mathbf{m} := (0,\ldots,m-1 \leq M-1)$$

with void $\mathbf{0} := ()$ voiding any tensor it subscripts. Ordinals are concatenated into tuples by Cartesian $\times$ and will be subtracted like sets, as in $\mathbf{M}-\mathbf{m} := (m,\ldots,M-1)$. Subscripts label the tensor prior to any superscript operation, so $[y(\mathsf{u})]^2_{\mathbf{L}}$ is an $\mathbf{L}^2 := \mathbf{L}\times\mathbf{L}$ tensor, for example. The preference throughout this work is for uppercase constants and lowercase variables, in case of ordinals the lowercase ranging over the uppercase. We prefer o for an unbounded positive integer, avoiding O.

Expectations and variances will be subscripted by the dimensions of u marginalized. Conditioning on the remaining inputs is left implicit after Eq. (3), to lighten notation. Now, construct $M+1$ stochastic processes (SPs)

$$[\mathsf{y}_{\mathsf{m}}]_{\mathbf{L}} := \mathbb{E}_{\mathbf{M}-\mathbf{m}}[y(\mathsf{u})] := \mathbb{E}_{\mathbf{M}-\mathbf{m}}\big[y(\mathsf{u})\big|\,[u]_{\mathbf{m}}\big] \tag{3}$$

ranging from $[\mathsf{y}_0]_{\mathbf{L}}$ to $[\mathsf{y}_{\mathsf{M}}]_{\mathbf{L}}$. Every SP depends stochastically on the ungoverned noise dimension $[\mathsf{u}]_M \perp [\mathsf{u}]_{\mathbf{M}}$ and deterministically on the first m governed inputs $[u]_{\mathbf{m}}$, marginalizing the remaining inputs $[\mathsf{u}]_{\mathbf{M}-\mathbf{m}}$. Sans serif symbols such as u,y generally refer to RVs and SPs, italic u, y being reserved for (tensor) functions and variables. Each SP is simply a regression model for y on the first m dimensions of u.

Following the Kolmogorov extension theorem [30, pp.124] we may regard an SP as a random function, from which we shall freely extract finite dimensional distributions generated by a design matrix $[u]_{\mathbf{M}\times\mathbf{o}}$ of $o \in \mathbb{Z}^+$ input samples. The Kolmogorov extension theorem incidentally secures u. Because y is (Lebesgue) integrable it must be measurable, guaranteeing $[\mathsf{y}_0]_{\mathbf{L}}$. Because all probability measures are finite, integrability of y implies integrability of y^n for all $n \in \mathbb{Z}^+$ [31]. So Fubini's theorem [32, pp.77] allows all expectations to be taken in any order. These observations suffice to secure every object appearing in this paper. Changing the order of expectations, as permitted by Fubini's theorem, is the vital tool used throughout to construct this work.

Our aim is to compare predictions from a reduced regression model y_{m} with those from the full regression model y_{M}. Correlation between these predictions is squared – using element-wise (Hadamard) multiplication $\circ$ and division / – to form an RV called the coefficient of determination

$$\left[\mathsf{R}^2_{\mathsf{mM}}\right]_{\mathbf{L}^2} := \frac{\mathbb{V}_{\mathbf{M}}[\mathsf{y}_{\mathsf{m}}, \mathsf{y}_{\mathsf{M}}] \circ \mathbb{V}_{\mathbf{M}}[\mathsf{y}_{\mathsf{m}}, \mathsf{y}_{\mathsf{M}}]}{\mathbb{V}_{\mathbf{m}}[\mathsf{y}_{\mathsf{m}}] \circ \mathbb{V}_{\mathbf{M}}[\mathsf{y}_{\mathsf{M}}]} = \frac{\mathbb{V}_{\mathbf{m}}[\mathsf{y}_{\mathsf{m}}]}{\mathbb{V}_{\mathbf{M}}[\mathsf{y}_{\mathsf{M}}]} \tag{4}$$

However, this is undefined whenever $\mathbb{V}_{\mathbf{M}}[\mathsf{y}_{\mathsf{M}}]_{l\times l'} = 0$, obscuring potentially useful information about $\mathbb{V}_{\mathbf{m}}[\mathsf{y}_{\mathsf{m}}]_{l\times l'}$. Introducing 1-tensors representing the square root diagonal of a covariance matrix

$$\left[\sqrt{\mathbb{V}[\cdot,\cdot]_{\mathbf{L}^2}}\right]_l := \mathbb{V}[\cdot,\cdot]^{1/2}_{l^2} \tag{5}$$

the correlation coefficient between output dimensions is

$$[\mathsf{R}_{\mathsf{m}}]_{\mathbf{L}\times\mathbf{L}'} := \frac{\mathbb{V}_{\mathbf{m}}[\mathsf{y}_{\mathsf{m}}]_{\mathbf{L}\times\mathbf{L}'}}{\sqrt{\mathbb{V}_{\mathbf{m}}[\mathsf{y}_{\mathsf{m}}]_{\mathbf{L}^2}} \otimes \sqrt{\mathbb{V}_{\mathbf{m}}[\mathsf{y}_{\mathsf{m}}]_{\mathbf{L}'^2}}} \quad \forall \mathbf{m} \subseteq \mathbf{M} \tag{6}$$

Let us define the multi-output closed Sobol' index as the product of the full correlation between output dimensions and the coefficient of determination

$$[\mathsf{S}_{\mathsf{m}}]_{\mathbf{L}\times\mathbf{L}'} := [\mathsf{R}_{\mathsf{M}}]_{\mathbf{L}\times\mathbf{L}'} \circ \left[\mathsf{R}^2_{\mathsf{mM}}\right]_{\mathbf{L}\times\mathbf{L}'} \tag{7}$$

and the multi-output total Sobol' index as its complement

$$\left[\mathsf{S}^{\mathsf{T}}_{\mathsf{M-m}}\right]_{\mathbf{L}\times\mathbf{L}'} := [\mathsf{S}_{\mathsf{M}}]_{\mathbf{L}\times\mathbf{L}'} - [\mathsf{S}_{\mathsf{m}}]_{\mathbf{L}\times\mathbf{L}'} \tag{8}$$

These definitions coincide precisely with the traditional Sobol' index along the diagonal $\sqrt{[\mathsf{S}_{\mathsf{m}}]_{\mathbf{L}^2}} \circ \sqrt{[\mathsf{S}_{\mathsf{m}}]_{\mathbf{L}^2}}$, which has been very much the focus of prior literature [6, 7, 8]. The off-diagonal elements are bound by the diagonal as

$$-\,[\mathsf{S}_{\mathsf{m}}]^{1/2}_{l^2}\,[\mathsf{S}_{\mathsf{m}}]^{1/2}_{l'^2} \leq [\mathsf{S}_{\mathsf{m}}]_{l\times l'} = [\mathsf{R}_{\mathsf{m}}]_{l\times l'}\,[\mathsf{S}_{\mathsf{m}}]^{1/2}_{l^2}\,[\mathsf{S}_{\mathsf{m}}]^{1/2}_{l'^2} \leq [\mathsf{S}_{\mathsf{m}}]^{1/2}_{l^2}\,[\mathsf{S}_{\mathsf{m}}]^{1/2}_{l'^2} \tag{9}$$

To calculate moments over ungoverned noise we use the Taylor series method [33, pp.353], which is valid provided $\mathbb{V}_{\mathbf{M}}[\mathsf{y}_{\mathsf{M}}]_{l^2}$ is well approximated by its mean

$$[V_{\mathbf{M}}]_{l^2} := \mathbb{E}_M \mathbb{V}_{\mathbf{M}}[\mathsf{y}_{\mathsf{M}}]_l \gg \left|\mathbb{V}_{\mathbf{M}}[\mathsf{y}_{\mathsf{M}}]_{l^2} - [V_{\mathbf{M}}]_{l^2}\right| \tag{10}$$

This provides the mean Sobol' index

$$[S_{\mathbf{m}}]_{\mathbf{L}\times\mathbf{L}'} := \mathbb{E}_M[\mathsf{S}_{\mathsf{m}}]_{\mathbf{L}\times\mathbf{L}'} = \frac{[V_{\mathbf{m}}]_{\mathbf{L}\times\mathbf{L}'}}{\sqrt{[V_{\mathbf{M}}]_{\mathbf{L}^2}} \otimes \sqrt{[V_{\mathbf{M}}]_{\mathbf{L}'^2}}} \tag{11}$$

$$\text{where } [V_{\mathbf{m}}]_{\mathbf{L}^2} := \mathbb{E}_M\,\mathbb{V}_{\mathbf{m}}[\mathsf{y}_{\mathsf{m}}] \quad \forall \mathbf{m} \subseteq \mathbf{M} \tag{12}$$

with variance due to ungoverned noise of

$$[T_{\mathbf{m}} \circ T_{\mathbf{m}}]_{(\mathbf{L}\times\mathbf{L}')^2} := \mathbb{V}_M[\mathsf{S}_{\mathbf{m}}]_{(\mathbf{L}\times\mathbf{L}')^2} = \frac{[Q_{\mathbf{m}}]_{(\mathbf{L}\times\mathbf{L}')^2}}{[V_{\mathbf{M}}]^{2/2}_{\mathbf{L}^2} \otimes [V_{\mathbf{M}}]^{2/2}_{\mathbf{L}'^2}} \tag{13}$$

where improper fractions exponentiate a square root diagonal of $V_{\mathbf{M}}$, and

$$\begin{aligned}[Q_{\mathbf{m}}]_{(\mathbf{L}\times\mathbf{L}')^2} := [W_{\mathbf{mm}}]_{(\mathbf{L}\times\mathbf{L}')^2} - [V_{\mathbf{m}}]_{\mathbf{L}\times\mathbf{L}'} \circ \sum_{\mathbf{L}^\circ\in\{\mathbf{L},\mathbf{L}'\}} \frac{[W_{\mathbf{Mm}}]_{\mathbf{L}^{\circ 2}\times\mathbf{L}\times\mathbf{L}'}}{[V_{\mathbf{M}}]^{2/2}_{\mathbf{L}^{\circ 2}}} \\ + \frac{[V_{\mathbf{m}}]^2_{\mathbf{L}\times\mathbf{L}'}}{4} \circ \sum_{\mathbf{L}^\circ\in\{\mathbf{L},\mathbf{L}'\}} \frac{[W_{\mathbf{MM}}]_{\mathbf{L}^{\circ 2}\times\mathbf{L}^2}}{[V_{\mathbf{M}}]^{2/2}_{\mathbf{L}^{\circ 2}} \otimes [V_{\mathbf{M}}]^{2/2}_{\mathbf{L}^2}} + \frac{[W_{\mathbf{MM}}]_{\mathbf{L}^{\circ 2}\times\mathbf{L}'^2}}{[V_{\mathbf{M}}]^{2/2}_{\mathbf{L}^{\circ 2}} \otimes [V_{\mathbf{M}}]^{2/2}_{\mathbf{L}'^2}}\end{aligned} \tag{14}$$

$$[W_{\mathbf{mm}'}]_{\mathbf{L}^4} := \mathbb{V}_M[\mathbb{V}_{\mathbf{m}}[\mathsf{y}_{\mathsf{m}}]\,, \mathbb{V}_{\mathbf{m}'}[\mathsf{y}_{\mathsf{m}'}]] \tag{15}$$

It is satisfying to note that these Equations enforce

$$[T_{\mathbf{M}}]_{l^4} = 0 \quad \text{on the diagonal} \quad [\mathsf{S}_{\mathsf{M}}]_{l^2} = 1 \tag{16}$$

In practice it may be best to retain only the term in $W_{\mathbf{mm}}$, ignoring the uncertainty in $V_{\mathbf{M}}$ conveyed by $W_{\mathbf{Mm}}, W_{\mathbf{MM}}$, because these may drastically reduce uncertainty whenever $V_{\mathbf{m}} \approx V_{\mathbf{M}}$, which is the circumstance of greatest interest. In a similar vein, the total index is defined as the difference between the closed index and $[\mathsf{S}_{\mathsf{M}}]_{\mathbf{L}\times\mathbf{L}'}$, which is exactly 1 on the diagonal and usually highly correlated with $[\mathsf{S}_{\mathsf{m}}]_{\mathbf{L}\times\mathbf{L}'}$ off the diagonal. Tiny differences between terms in the correlated case are swamped by numerical errors in practice. The conservative standard error

$$\begin{aligned}\left[T^T_{\mathbf{M}-\mathbf{m}}\right]_{(l\times l')^2} &:= \sqrt{[T_{\mathbf{M}} \circ T_{\mathbf{M}}]_{(l\times l')^2}} + \sqrt{[T_{\mathbf{m}} \circ T_{\mathbf{m}}]_{(l\times l')^2}} \\ &\geq \sqrt{\mathbb{V}_M \left[\mathsf{S}^T_{\mathbf{M}-\mathbf{m}}\right]_{(l\times l')^2}}\end{aligned} \tag{17}$$

achieves equality on the diagonal, and is robust and sufficiently precise for most practical purposes.

The remainder of this paper is devoted to calculating these two quantities – the generalized closed Sobol' Index $S_{\mathbf{m}}$ and its standard error due to ungoverned noise $T_{\mathbf{m}}$.

3. Stochastic Process estimates

The central problem in calculating errors on Sobol' indices is that they involve ineluctable covariances between differently marginalized SPs, via their moments over ungoverned noise. But marginalization and moment determination are both a matter of taking expectations. So the ineluctable can be avoided by reversing the order of expectations – taking moments over ungoverned noise, then marginalizing. To this end, adopt as design matrix a triad of inputs to condition $[\mathsf{u}]_{(\mathbf{M}+\mathbf{1})\times\mathbf{3}}$, eliciting the response

$$[\mathsf{y}]_{\mathbf{L}\times\mathbf{3}} := \mathbb{E}_{\mathbf{M}}\,\mathbb{E}_{\mathbf{M}'-\mathbf{m}'}\,\mathbb{E}_{\mathbf{0}''}\left[y([\mathsf{u}]_{(\mathbf{M}+\mathbf{1})\times\mathbf{3}})\middle|\,[[u]_{\mathbf{0}}\,,[u]_{\mathbf{m}'}\,,[u]_{\mathbf{M}''}]\right] \tag{18}$$

Primes mark independent inputs, otherwise expectations are shared by all three members of the triad. It is not always obvious whether inputs are independent or shared by the triad, but this can be mechanically checked against the measure of integration behind an expectation. Repeated expectations over the same axis are rare here, usually indicating that apparent repetitions must be "primed". The purpose of the triad is to interrogate its response for moments in respect of ungoverned noise (which is shared by the triad members)

$$[\mu_n]_{(\mathbf{L}\times\mathbf{3})^n} := \mathbb{E}_M\left[[\mathsf{y}]^n_{\mathbf{L}\times\mathbf{3}}\right] \quad \forall n \in \mathbb{Z}^+ \tag{19}$$

for these embody

$$[\mu_{\mathbf{m}'\ldots\mathbf{m}^{n\prime}}]_{\mathbf{L}^n} := [\mu_n]_{\prod_{j=1}^{n}(\mathbf{L}\times i_j)} = \mathbb{E}_M[[\mathsf{y}_{\mathbf{m}'}]_{\mathbf{L}} \otimes \cdots \otimes [\mathsf{y}_{\mathbf{m}^{n\prime}}]_{\mathbf{L}}]$$

where $i_j \in \mathbf{3}$ corresponds to $\mathbf{m}^{j\prime} \in \{\mathbf{0}, \mathbf{m}, \mathbf{M}\}$. This expression underpins the quantities we seek. The reduction which follows repeatedly realizes the iterated expectation law

$$[\mu_{\mathbf{0}\ldots\mathbf{0}\mathbf{m}^{j\prime}\ldots\mathbf{m}^{n\prime}}]_{\mathbf{L}^n} := \mathbb{E}_{\mathbf{M}}[\mu_{\mathbf{M}\ldots\mathbf{M}\mathbf{m}^{j\prime}\ldots\mathbf{m}^{n\prime}}]_{\mathbf{L}^n} = \mathbb{E}_{\mathbf{m}}[\mu_{\mathbf{m}\ldots\mathbf{m}\mathbf{m}^{j\prime}\ldots\mathbf{m}^{n\prime}}]_{\mathbf{L}^n} \tag{20}$$

and that y was offset in Eq. (1) to obey

$$[\mu_{\mathbf{0}}]_{\mathbf{L}} = [0]_{\mathbf{L}} \tag{21}$$

and in any case

$$[V_{\mathbf{0}}]_{\mathbf{L}^2} = [\mu_{\mathbf{00}}]_{\mathbf{L}^2} - [\mu_{\mathbf{0}}]^2_{\mathbf{L}} = [0]_{\mathbf{L}^2} \tag{22}$$

Defining

$$[\mathsf{e}]_{\mathbf{L}\times\mathbf{3}} := \mathsf{y} - \mu_1 \tag{23}$$

the expected conditional variance in Eq. (11) amounts to

$$\begin{aligned}
[V_{\mathbf{m}}]_{\mathbf{L}^2} &= \mathbb{E}_{\mathbf{m}}\mathbb{E}_M\left[[\mathsf{e}_\mathsf{m}+\mu_{\mathbf{m}}]^2_{\mathbf{L}}\right] - \mathbb{E}_M\left[[\mathsf{e}_0+\mu_{\mathbf{0}}]^2_{\mathbf{L}}\right] \\
&= \mathbb{E}_{\mathbf{m}}\left[[\mu_{\mathbf{m}}]^2_{\mathbf{L}}\right] - [\mu_{\mathbf{0}}]^2_{\mathbf{L}} + \mathbb{E}_{\mathbf{m}}[\mu_{\mathbf{mm}}]_{\mathbf{L}^2} - [\mu_{\mathbf{00}}]_{\mathbf{L}^2} \\
&= \mathbb{E}_{\mathbf{m}}\left[[\mu_{\mathbf{m}}]^2_{\mathbf{L}}\right]
\end{aligned} \tag{24}$$

and the covariance between conditional variances in Eq. (15) is

$$\begin{aligned}
[W_{\mathbf{mm}'}]_{\mathbf{L}^4} &:= \mathbb{V}_M[\mathbb{V}_{\mathbf{m}}[\mathsf{y}_\mathsf{m}]\,,\mathbb{V}_{\mathbf{m}'}[\mathsf{y}_{\mathsf{m}'}]] \\
&= \mathbb{V}_M\left[\mathbb{E}_{\mathbf{m}}\left[[\mathsf{y}_\mathsf{m}]^2_{\mathbf{L}} - [\mathsf{y}_0]^2_{\mathbf{L}}\right], \mathbb{E}_{\mathbf{m}'}\left[[\mathsf{y}_{\mathsf{m}'}]^2_{\mathbf{L}} - [\mathsf{y}_0]^2_{\mathbf{L}}\right]\right] \\
&= \mathbb{E}_M\left[\mathbb{E}_{\mathbf{m}}\left[[\mathsf{y}_\mathsf{m}]^2_{\mathbf{L}} - [\mathsf{y}_0]^2_{\mathbf{L}}\right] \otimes \mathbb{E}_{\mathbf{m}'}\left[[\mathsf{y}_{\mathsf{m}'}]^2_{\mathbf{L}} - [\mathsf{y}_0]^2_{\mathbf{L}}\right]\right] \\
&\quad - [V_{\mathbf{m}}]_{\mathbf{L}^2} \otimes [V_{\mathbf{m}'}]_{\mathbf{L}^2} \\
&= [A_{\mathbf{mm}'} - A_{\mathbf{0m}'} - A_{\mathbf{m0}} + A_{\mathbf{00}}]_{\mathbf{L}^4}
\end{aligned}$$

Here, the inputs within any $\mathbf{m},\mathbf{m}' \subseteq \mathbf{M}$ clearly vary independently, and

$$\begin{aligned}
[A_{\mathbf{mm}'}]_{\mathbf{L}^4} &:= \mathbb{E}_M\mathbb{E}_{\mathbf{m}'}\mathbb{E}_{\mathbf{m}}\left[[\mathsf{y}_\mathsf{m}]^2_{\mathbf{L}} \otimes [\mathsf{y}_{\mathsf{m}'}]^2_{\mathbf{L}}\right] - [V_{\mathbf{m}}]_{\mathbf{L}^2} \otimes [V_{\mathbf{m}'}]_{\mathbf{L}^2} \\
&= \mathbb{E}_{\mathbf{m}}\mathbb{E}_{\mathbf{m}'}\mathbb{E}_M\left[[\mathsf{e}_\mathsf{m}+\mu_{\mathbf{m}}]^2_{\mathbf{L}} \otimes [\mathsf{e}_{\mathsf{m}'}+\mu_{\mathbf{m}'}]^2_{\mathbf{L}} - [\mu_{\mathbf{m}}]^2_{\mathbf{L}} \otimes [\mu_{\mathbf{m}'}]^2_{\mathbf{L}}\right]
\end{aligned}$$

exploiting the fact that $V_{\mathbf{0}} = [0]_{\mathbf{L}^2}$. Equation (20) cancels all terms beginning with $[\mathsf{e}_\mathsf{m}]^2_{\mathbf{L}}$, first across $A_{\mathbf{mm}'} - A_{\mathbf{0m}'}$ then across $A_{\mathbf{m0}} - A_{\mathbf{00}}$. All remaining terms ending in $[\mu_{\mathbf{m}'}]^2_{\mathbf{L}}$ are eliminated by centralization $\mathbb{E}_M[\mathsf{e}_\mathsf{m}] = 0$. Similar arguments eliminate $[\mathsf{e}_{\mathsf{m}'}]^2_{\mathbf{L}}$ and $[\mu_{\mathbf{m}}]^2_{\mathbf{L}}$. Effectively then

$$[A_{\mathbf{mm}'}]_{\mathbf{L}^4} = \sum_{\pi(\mathbf{L}^2)}\sum_{\pi(\mathbf{L}'^2)} \mathbb{E}_{\mathbf{m}}\mathbb{E}_{\mathbf{m}'}[\mu_{\mathbf{m}} \otimes \mu_{\mathbf{mm}'} \otimes \mu_{\mathbf{m}'}]_{\mathbf{L}^2\times\mathbf{L}'^2}$$

so Eq. (21) entails

$$[W_{\mathbf{mm}'}]_{\mathbf{L}^4} = \sum_{\pi(\mathbf{L}^2)}\sum_{\pi(\mathbf{L}'^2)} \mathbb{E}_{\mathbf{m}}\mathbb{E}_{\mathbf{m}'}[\mu_{\mathbf{m}} \otimes \mu_{\mathbf{mm}'} \otimes \mu_{\mathbf{m}'}]_{\mathbf{L}^2\times\mathbf{L}'^2} \tag{25}$$

where each summation is over permutations of tensor axes

$$\pi(\mathbf{L}^2) := \{(\mathbf{L}\times\mathbf{L}''), (\mathbf{L}''\times\mathbf{L})\} \quad ; \quad \pi(\mathbf{L}'^2) := \{(\mathbf{L}'\times\mathbf{L}'''), (\mathbf{L}'''\times\mathbf{L}')\}$$

Primes on constants are for bookeeping purposes only ($\mathbf{L}^{j\prime} = \mathbf{L}$ always), they do not change the value of the constant – unlike primes on variables ($\mathbf{m}^{j\prime}$

need not equal $\mathbf{m}$ in general). One is mainly interested in variances (errors), constituted by the diagonal $\mathbf{L}'^2 = \mathbf{L}^2$, for which the summation in Eq. (25) is over a pair of transposed pairs.

In order to further elucidate these estimates, we must fill in the details of the underlying SPs, sufficiently identifying the regression y by its first two moments μ_1, μ_2. Then the Sobol' indices are given by Eqs. (11) and (24), and their standard error by Eqs. (13), (14) and (25).

4. Interlude: Gaussian Process regression

The development in this Section is based on [21], with slightly different notation. A GP over x is formally defined and specified by

$$[\mathsf{y}_\mathsf{M}]_\mathbf{L} \,\big|\, [x]_{\mathbf{M}\times\mathbf{o}} \sim \mathsf{N}^\dagger\Big([\bar{y}(x)]_{\mathbf{L}\times\mathbf{o}}\,, [k_\mathsf{y}(x,x)]_{(\mathbf{L}\times\mathbf{o})^2}\Big) \quad \forall o \in \mathbb{Z}^+$$

where tensor ranks concatenate into a multivariate normal distribution

$$[]_{\mathbf{L}\times\mathbf{o}} \sim \mathsf{N}^\dagger\Big([]_{\mathbf{L}\times\mathbf{o}}\,, []_{(\mathbf{L}\times\mathbf{o})^2}\Big) \iff []^\dagger_{\mathbf{L}\times\mathbf{o}} \sim \mathsf{N}\Big([]^\dagger_{\mathbf{L}\times\mathbf{o}}\,, []^\dagger_{(\mathbf{L}\times\mathbf{o})^2}\Big)$$
$$\Big[[]^\dagger_{\mathbf{L}\times\mathbf{o}}\Big]_{\mathbf{lo}-(\mathbf{l}-\mathbf{1})\mathbf{o}} := []_{(l-1)\times\mathbf{o}}$$
$$\Big[[]^\dagger_{(\mathbf{L}\times\mathbf{o})^2}\Big]_{(\mathbf{lo}-(\mathbf{l}-\mathbf{1})\mathbf{o})\times(\mathbf{l'o}-(\mathbf{l'}-\mathbf{1})\mathbf{o})} := []_{(l-1)\times\mathbf{o}\times(l'-1)\times\mathbf{o}}$$

supporting the fundamental definition of the GP kernel, as a covariance (over ungoverned noise) between responses

$$[k_\mathsf{y}(x,x)]_{l\times o\times l'\times o'} := \mathbb{V}_M\big[[\mathsf{y}_\mathsf{M}|x]_{l\times o}\,, [\mathsf{y}_\mathsf{M}|x]_{l'\times o'}\big]$$

4.1. Tensor Gaussians

Henceforth, tensors will be broadcast when necessary, as described in [34, 35]. This means that ranks and dimensions are implicitly expanded as necessary to perform an algebraic operation between tensors of differing signature. A tensor Gaussian like $\mathrm{p}\big([x]_{\mathbf{m}\times\mathbf{o}}\big|\,[x']_{\mathbf{m}\times\mathbf{o}'}\,, [\Sigma]_{\mathbf{L}^2\times\mathbf{m}^2}\big)$ is defined element-wise, using broadcasting

$$\big[\mathrm{p}\big([x]_{\mathbf{m}\times\mathbf{o}}\big|\,[x']_{\mathbf{m}\times\mathbf{o}'}\,, [\Sigma]_{\mathbf{L}^2\times\mathbf{m}^2}\big)\big]_{l\times o\times l'\times o'} := (2\pi)^{-M/2}\,\big|[\Sigma]_{l\times l'}\big|^{-1/2}$$
$$\exp\left(-\frac{[x-x']^\intercal_{\mathbf{m}\times l\times o\times l'\times o'}\,[\Sigma]^{-1}_{l\times l'\times\mathbf{m}\times\mathbf{m}'}\,[x-x']_{\mathbf{m}'\times l\times o\times l'\times o'}}{2}\right) \tag{26}$$

for $\mathbf{m}' = \mathbf{m}$ and transposition $^\intercal$ moving first rank to last.

Remarkably, the algebraic development in the remainder of this paper relies almost exclusively on an invaluable product formula reported in [22]:

$$\begin{aligned}\mathrm{p}(z|\,a,A)\circ\mathrm{p}(\Theta^\intercal z|\,b,B) &= \mathrm{p}(0|\,(b-\Theta^\intercal a),(B+\Theta^\intercal A\Theta))\\ &\circ\,\mathrm{p}\big(z|\,(A^{-1}+\Theta B^{-1}\Theta^\intercal)^{-1}(A^{-1}a+\Theta B^{-1}b),(A^{-1}+\Theta B^{-1}\Theta^\intercal)^{-1}\big)\end{aligned} \tag{27}$$

This formula and the tensor Gaussians behind it will appear in a variety of guises.

4.2. Prior GP

GP regression decomposes output $[\mathsf{y_M}]_{\mathbf{L}}$ into signal GP $[\mathsf{f_M}]_{\mathbf{L}}$, and independent noise GP $[\mathsf{e}_M]_{\mathbf{L}}$ with homoskedastic noise (also known as likelihood) covariance $[E]_{\mathbf{L}^2}$

$$\begin{aligned}[\mathsf{y_M}|E]_{\mathbf{L}} &= [\mathsf{f_M}]_{\mathbf{L}} + [\mathsf{e}_M|E]_{\mathbf{L}}\\ [\mathsf{e}_M|E]_{\mathbf{L}}\,\big|\,[x]_{\mathbf{M}\times\mathbf{o}} &\sim \mathsf{N}^\dagger\Big([0]_{\mathbf{L}\times\mathbf{o}}\,,[E]_{(\mathbf{L}\times 1)^2}\circ\langle 1\rangle_{(1\times\mathbf{o})^2}\Big)\end{aligned}$$

Angle brackets denote a (perhaps broadcast) diagonal tensor, such as the identity matrix $\langle 1\rangle_{(1\times\mathbf{o})^2} =: \big\langle [1]_{(1\times\mathbf{o})^2}\big\rangle$.

The RBF kernel is hyperparametrized by signal covariance $[F]_{\mathbf{L}^2}$ and the tensor $[\Lambda]_{\mathbf{L}^2\times\mathbf{M}}$ of characteristic lengthscales, which must be symmetric $[\Lambda]_{l\times l'\times\mathbf{M}} = [\Lambda]_{l'\times l\times\mathbf{M}}$. Now use

$$\begin{aligned}\big\langle\Lambda^2\pm I\big\rangle_{l\times l'\times\mathbf{M}^2} &:= \big\langle[\Lambda]_{l\times\mathbf{M}}\circ[\Lambda]_{l'\times\mathbf{M}}\pm[I]_{\mathbf{M}}\big\rangle \qquad I\in\{0\}\cup\mathbb{Z}^+\\ \big\langle\Lambda^2\big\rangle_{l\times l'\times\mathbf{M}^2} &:= \big\langle\Lambda^2\pm 0\big\rangle_{l\times l'}\\ [\pm F]_{l\times l'} &:= (2\pi)^{M/2}\left|\big\langle\Lambda^2\big\rangle_{l\times l'}\right|^{1/2}[F]_{l\times l'}\end{aligned}$$

to implement the non-informative RBF prior according to Eq. (26)

$$[\mathsf{f_M}|F,\Lambda]_{\mathbf{L}}\,\big|\,[x]_{\mathbf{M}\times\mathbf{o}}\sim\mathsf{N}^\dagger\Big([0]_{\mathbf{L}\times\mathbf{o}}\,,[\pm F]_{(\mathbf{L}\times 1)^2}\circ\mathrm{p}\Big([x]_{\mathbf{M}\times\mathbf{o}}\big|\,[x]_{\mathbf{M}\times\mathbf{o}}\,,\big\langle\Lambda^2\big\rangle_{\mathbf{L}^2\times\mathbf{M}^2}\Big)\Big)$$

4.3. Predictive GP

Bayesian inference for GP regression further conditions the hyper-parametrized GP $\mathsf{y}|E,F,\Lambda$ on the observed realization (over ungoverned noise) of the random variable $[\mathsf{y}|X]$

$$[Y]^\dagger_{\mathbf{L}\times\mathbf{N}} := \big[[\mathsf{y_M}|E,F,\Lambda]_{\mathbf{L}}\,\big|\,[X]_{\mathbf{M}\times\mathbf{N}}\big]^\dagger(\omega)\in\mathbb{R}^{LN}$$

To this end we define

$$
\begin{aligned}
[K_{\mathsf{e}}]_{\mathbf{Lo}\times\mathbf{Lo}} &:= \mathbb{V}_M\Big[\big[[\mathsf{e}_M|E]_{\mathbf{L}}\,\big|\,[x]_{\mathbf{M}\times\mathbf{o}}\big]^{\dagger}\Big] \\
&= \Big[[E]_{(\mathbf{L}\times 1)^2}\circ\langle 1\rangle_{(1\times\mathbf{o})^2}\Big]^{\dagger} \\
[k(x,x')]_{\mathbf{Lo}\times\mathbf{Lo'}} &:= \mathbb{V}_M\Big[\big[[\mathsf{f}_{\mathsf{M}}|F,\Lambda]_{\mathbf{L}}\,\big|\,[x]_{\mathbf{M}\times\mathbf{o}}\big]^{\dagger},\big[[\mathsf{f}_{\mathsf{M}}|F,\Lambda]_{\mathbf{L}}\,\big|\,[x']_{\mathbf{M}\times\mathbf{o'}}\big]^{\dagger}\Big] \\
&= \Big[[\pm F]_{\mathbf{L}^2}\circ \mathrm{p}\Big([x]_{\mathbf{M}\times\mathbf{o}}\Big|\,[x']_{\mathbf{M}\times\mathbf{o'}}\,,\big\langle\Lambda^2\big\rangle_{\mathbf{L}^2\times\mathbf{M}^2}\Big)\Big]^{\dagger} \\
[K_Y]_{\mathbf{LN}\times\mathbf{LN}} &:= \mathbb{V}_M\Big[\big[[\mathsf{y}|E,F,\Lambda]_{\mathbf{L}}\,\big|\,[X]_{\mathbf{M}\times\mathbf{N}}\big]^{\dagger}\Big] \\
&= k([X]_{\mathbf{M}\times\mathbf{N}}\,,[X]_{\mathbf{M}\times\mathbf{N}}) + [K_{\mathsf{e}}]_{\mathbf{LN}\times\mathbf{LN}}
\end{aligned}
\tag{28}
$$

Applying Bayes' rule

$$
\begin{aligned}
p(\mathsf{f}_{\mathsf{M}}|Y)p(Y) = p(Y|\mathsf{f}_{\mathsf{M}})p(\mathsf{f}_{\mathsf{M}}) &= \mathrm{p}\big(Y^{\dagger}\big|\,{\mathsf{f}_{\mathsf{M}}}^{\dagger},K_{\mathsf{e}}\big)\,\mathrm{p}\big({\mathsf{f}_{\mathsf{M}}}^{\dagger}\big|\,[0]_{\mathbf{LN}}\,,k(X,X)\big) \\
&= \mathrm{p}\big({\mathsf{f}_{\mathsf{M}}}^{\dagger}\big|\,Y^{\dagger},K_{\mathsf{e}}\big)\,\mathrm{p}\big({\mathsf{f}_{\mathsf{M}}}^{\dagger}\big|\,[0]_{\mathbf{LN}}\,,k(X,X)\big)
\end{aligned}
$$

Product formula Eq. (27) immediately reveals the marginal likelihood

$$
p\big([Y|E,F,\Lambda]\,\big|X\big) = \mathrm{p}\Big([Y]^{\dagger}_{\mathbf{L}\times\mathbf{N}}\Big|\,[0]_{\mathbf{LN}}\,,K_Y\Big) \tag{29}
$$

and the posterior distribution

$$
\big[[\mathsf{f}_{\mathsf{M}}|Y|E,F,\Lambda]\,\big|X\big]^{\dagger}_{\mathbf{L}\times\mathbf{N}} \sim \mathsf{N}\big(k(X,X)K_Y^{-1}Y^{\dagger},\ k(X,X)-k(X,X)K_Y^{-1}k(X,X)\big)
$$

The ultimate goal is the posterior predictive GP which extends the posterior distribution to arbitrary – usually unobserved – $[x]_{\mathbf{M}\times\mathbf{o}}$. This is formally derived from the definition of conditional probability, but this seems unnecessary, for the extension must recover the posterior distribution when $x = X$. There is but one way of selectively replacing X with x in the posterior formula which preserves the coherence of tensor ranks:

$$
\big[[\mathsf{f}_{\mathsf{M}}|Y|E,F,\Lambda]\,\big|x\big]^{\dagger}_{\mathbf{L}\times\mathbf{o}} \sim \mathsf{N}\big(k(x,X)K_Y^{-1}Y^{\dagger},\ k(x,x)-k(x,X)K_Y^{-1}k(X,x)\big) \tag{30}
$$

In order to calculate the last term, the Cholesky decomposition $K_Y^{1/2}$ is used to write (slightly abusing notation)

$$
\big[k(x,X)K_Y^{-1}k(X,x)\big]_{\mathbf{Lo}\times\mathbf{Lo}} = [K_Y^{-1/2}k(X,x)]^2_{\mathbf{Lo}}
$$

4.4. GP Optimization

Henceforth we implicitly condition on optimal hyperparameters, which maximise the marginal likelihood Eq. (29).

$$[E]_{\mathbf{L}^2}, [F]_{\mathbf{L}^2}, [\Lambda]_{\mathbf{L}^2\times\mathbf{M}} := \operatorname{argmax} \mathrm{p}\Big([Y]^{\dagger}_{\mathbf{L}\times\mathbf{N}} \Big| [0]_{\mathbf{LN}}, K_Y \Big) \tag{31}$$

5. Gaussian Process moments

This Section calculates the SP moments of GP Regression, absorbing Section 4 into the perspective of Section 3. Let $c\colon \mathbb{R} \to [0,1]$ be the (bijective) CDF of the standard, univariate normal distribution, and define the triads

$$\begin{aligned} [\mathsf{z}]_{\mathbf{M}\times\mathbf{3}} &:= c^{-1}\big([\mathsf{u}]_{\mathbf{M}\times\mathbf{3}}\big) \sim \mathsf{N}\big([0]_{\mathbf{M}\times\mathbf{3}}, \langle 1\rangle_{\mathbf{M}^2}\big) \\ [\mathsf{x}]_{\mathbf{M}'\times\mathbf{3}} &:= [\Theta]^{\intercal}_{\mathbf{M}\times\mathbf{M}'} [\mathsf{z}]_{\mathbf{M}\times\mathbf{3}} \end{aligned}$$

Here, the rotation matrix $[\Theta]^{\intercal}_{\mathbf{M}\times\mathbf{M}'} = [\Theta]^{-1}_{\mathbf{M}\times\mathbf{M}'}$ is broadcast to multiply the triad $[\mathsf{z}]_{\mathbf{M}\times\mathbf{3}}$. The purpose of ths arbitrary rotation is to allow GPs whose input basis x is not aligned with the fundamental basis u of the coefficient of determination. The latter is aligned with z which is the input we must condition. This generalization is cheap, given product formula Eq. (27), and of great potential benefit. One could, for example, imagine optimizing Θ to maximize some seminorm on $S_{\mathbf{m}}$.

Throughout the remainder of this paper, primed ordinal subscripts are used to specify Einstein summation contraction of tensors, the multiplication and summation of elements over a matching index which underpins matrix multiplication. In this work, whenever a subscript primed in a specific fashion appears in adjacent tensors (those not separated by algebraic operations $+, -, \circ, \otimes$) and does not subscript the result, it is contracted over, according to the Einstein convention. Implementation examples of the convention are given under `einsum` in [34].

Adding shared Gaussian noise $[\mathsf{e}_M|E]_{\mathbf{L}}$ to Eq. (30) yields

$$\begin{aligned} \big[y([\mathsf{u}]_{\mathbf{M+1}\times\mathbf{3}}) \big| [u]_{\mathbf{M}\times\mathbf{3}}\big]^{\dagger}_{\mathbf{L}\times\mathbf{3}} &= \big[[\mathsf{y}_{\mathsf{M}}|Y|E,F,\Lambda] \,\big|[z]_{\mathbf{M}\times\mathbf{3}}\big]^{\dagger}_{\mathbf{L}\times\mathbf{3}} \sim \\ &\mathsf{N}\Big(k(x,X)K_Y^{-1}Y^{\dagger},\ k(x,x) - [K_Y^{-1/2}k(X,x)]^{2}_{\mathbf{Lo}} + E^{\dagger}\Big) \end{aligned} \tag{32}$$

using broadcast $[E^{\dagger}]_{\mathbf{L3}\times\mathbf{L3}} := [[E]_{(\mathbf{L}\times 1)^2} \circ [1]_{(1\times\mathbf{3})^2}]^{\dagger}_{(\mathbf{L}\times\mathbf{3})^2}$. To bring the GP estimate fully under the umbrella of the SP estimate we should identify its ungoverned noise, and ascribe it to $[\mathsf{u}]_M$ of the SP. Let $d\colon (0,1) \to (0,1)^L$ concatenate every L^{th} decimal place starting at l, for each output dimension $l \leq L$ of $(0,1)^L$, then Eq. (32) can be written as

$$\begin{aligned}\big[y([\mathsf{u}]_{\mathbf{M+1\times 3}})\big|\,[u]_{\mathbf{M\times 3}}\big]^{\dagger}_{\mathbf{L\times 3}} \\ = [\mu_1]^{\dagger}_{\mathbf{L\times 3}} + [\mu_2]^{\dagger/2}_{\mathbf{L\times 3\times L'\times 3'}}\Big[\big[c^{-1}(d\,([\mathsf{u}]_M))\big]_{\mathbf{L}\times 1}\circ[1]_{1\times\mathbf{3}}\Big]^{\dagger}_{\mathbf{L'\times 3'}}\end{aligned} \tag{33}$$

where $[\mu_2]^{\dagger/2}_{(\mathbf{L\times 3})^2}$ denotes the lower triangular Cholesky decomposition of the matrix $[\mu_2]^{\dagger}_{(\mathbf{L\times 3})^2}$. From the development in Section 3, the first two moments μ_1, μ_2 are sufficient to compute the coefficient of determination and its variance.

The crucial moments μ_1, μ_2 are simply read from Eqs. (32) and (33), but still need conditioning. This is a process of marginalizing predictive content out of μ_1 to increase the conditioned uncertainty μ_2 in Eq. (33). Effectively, variation in the marginalized inputs $[\mathsf{u}]_{\mathbf{M-m}}$ is transferred to the ungovernable noise dimension $[\mathsf{u}]_M$. The calculation is entirely a matter of repeatedly applying product formula Eq. (27), together with the familiar Gaussian identities

$$\begin{aligned}[\mathsf{z}]_{\mathbf{M}} \sim \mathsf{N}\big([Z]_{\mathbf{M}}\,, [\Sigma]_{\mathbf{M\times M}}\big) &\Rightarrow [\mathsf{z}]_{\mathbf{m}} \sim \mathsf{N}\big([Z]_{\mathbf{m}}\,, [\Sigma]_{\mathbf{m\times m}}\big) \\ [\mathsf{z}]_{\mathbf{m}} \sim \mathsf{N}\big([Z]_{\mathbf{m}}\,, [\Sigma]_{\mathbf{m\times m}}\big) &\Rightarrow [\Theta]^{\intercal}_{\mathbf{m\times m}}\,[\mathsf{z}]_{\mathbf{m}} \sim |\Theta|^{-1}\,\mathsf{N}(\Theta^{\intercal}Z, \Theta^{\intercal}\Sigma\Theta)\end{aligned}$$

Henceforth the ordinal set $\mathbf{m}''$, whether or not decorated with a further *even* number of primes, should be taken as equal to $\mathbf{m}$. Likewise the ordinal set $\mathbf{m}'''$, whether or not decorated with a further *even* number of primes, should be taken as equal to $\mathbf{m}'$. Superscript * will stand for four consecutive primes $''''$. So $\mathbf{m}, \mathbf{m}'$ are identified by the parity (even or odd) of the primes adorning $\mathbf{m}$. Such explicit notation is required to maintain the integrity of einstein summation. This only applies to ordinal sets, not singleton values, so the many different prime decorations of l *always* indicate potentially different values.

5.1. First Moments

The first moment of the GP for any $\mathbf{m} \subseteq \mathbf{M}$ is given by

$$[\mu_{\mathbf{m}}]_{\mathbf{L}} = \mathbb{E}_{\mathbf{M-m}}\big[k([\mathsf{x}]_{\mathbf{M}}\,, X)\,K_Y^{-1}Y^{\dagger}\big|\,[z]_{\mathbf{m}}\big] = [g_{\mathbf{m}}]^{\dagger}_{\mathbf{L\times L''\times N''}}\,\big[K_Y^{-1}Y^{\dagger}\big]_{\mathbf{L''N''}}$$

where

$$\frac{[g_{\mathbf{m}}]_{l\times l''\times\mathbf{N}''}}{[g_{\mathbf{0}}]_{l\times l''\times\mathbf{N}''}} := \frac{\mathrm{p}\big([z]_{\mathbf{m}}\big|\,[G]_{\mathbf{m}\times l\times l''\times\mathbf{N}''}\,,[\Gamma]_{l\times l''}\big)}{\mathrm{p}([z]_{\mathbf{m}}|\,[0]_{\mathbf{m}}\,,\langle 1\rangle_{\mathbf{m}^2})}$$
$$= \frac{\mathrm{p}\big([\Phi]_{l\times l''\times\mathbf{m}''\times\mathbf{m}}\,[z]_{\mathbf{m}}\big|\,[G]_{\mathbf{m}\times l\times l''\times\mathbf{N}''}\,,[\Gamma]_{l\times l''\times\mathbf{m}^*\times\mathbf{m}''}\,[\Phi]_{l\times l''\times\mathbf{m}''\times\mathbf{m}}\big)}{\mathrm{p}\big([0]_{\mathbf{m}}\big|\,[G]_{\mathbf{m}\times l\times l''\times\mathbf{N}''}\,,[\Phi]_{l\times l''}\big)}$$

and

$$\begin{aligned}
[g_{\mathbf{0}}]_{l\times l''\times\mathbf{N}''} &:= [\pm F]_{l\times l''}\,\mathrm{p}\Big([0]_{\mathbf{M}}\big|\,[X]_{\mathbf{M}\times\mathbf{N}''}\,,\big\langle\Lambda^2+1\big\rangle_{l\times l''}\Big)\\
[G]_{\mathbf{m}\times l\times l''\times\mathbf{N}''} &:= [\Theta]_{\mathbf{m}\times\mathbf{M}}\,\big\langle\Lambda^2+1\big\rangle^{-1}_{l\times l''\times\mathbf{M}\times\mathbf{M}''}\,[X]_{\mathbf{M}''\times\mathbf{N}''}\\
[\Phi]_{l\times l''\times\mathbf{m}''\times\mathbf{m}} &:= [\Theta]_{\mathbf{m}''\times\mathbf{M}}\,\big\langle\Lambda^2+1\big\rangle^{-1}_{l\times l''\times\mathbf{M}\times\mathbf{M}''}\,[\Theta]^{\intercal}_{\mathbf{m}\times\mathbf{M}''}\\
[\Gamma]_{l\times l''\times\mathbf{m}^2} &:= \langle 1\rangle_{\mathbf{m}^2}-[\Phi]_{l\times l''\times\mathbf{m}^2}
\end{aligned}$$

Note that when $\mathbf{m}=\mathbf{M}$, Θ factors out entirely.

5.2. Second Moments

The second moment of the GP for any $\mathbf{m},\mathbf{m'}\subseteq\mathbf{M}$ is given by

$$[\mu_{\mathbf{mm'}}]_{\mathbf{L}^2} = [F]_{\mathbf{L}^2}\circ[\phi_{\mathbf{mm'}}]_{\mathbf{L}^2} - [\psi_{\mathbf{mm'}}]_{\mathbf{L}^2} + [E]_{\mathbf{L}^2}$$

where

$$[\phi_{\mathbf{mm'}}]_{l\times l'} := \frac{\mathbb{E}_{\mathbf{M}-\mathbf{m}}\,\mathbb{E}_{\mathbf{M'}-\mathbf{m'}}\big[k([\mathsf{x}]_{\mathbf{M}}\,,[\mathsf{x}]_{\mathbf{M'}})\,\big|\,[z]_{\mathbf{m}}\,,[z]_{\mathbf{m'}}\big]_{l\times l'}}{[F]_{l\times l'}}$$
$$= \frac{\big|\langle\Lambda^2\rangle_{l\times l'\times\mathbf{M}^2}\big|^{1/2}\,\mathrm{p}\big([\mathsf{z}]_{\mathbf{m}}\big|\,[0]_{\mathbf{m}}\,,[1-\Upsilon]_{l\times l'\times\mathbf{m}^2}\big)\,\mathrm{p}\big([\mathsf{z}]_{\mathbf{m'}}\big|\,[Z]_{l\times l'\times\mathbf{m'}}\,,[\Pi]_{l\times l'\times\mathbf{m'}^2}\big)}{\big|\langle\Lambda^2+2\rangle_{l\times l'\times\mathbf{M}^2}\big|^{1/2}\,\mathrm{p}([\mathsf{z}]_{\mathbf{m}}|\,[0]_{\mathbf{m}}\,,\langle 1\rangle_{\mathbf{m}^2})\,\mathrm{p}([\mathsf{z}]_{\mathbf{m'}}|\,[0]_{\mathbf{m'}}\,,\langle 1\rangle_{\mathbf{m'}^2})}$$

$$[\psi_{\mathbf{mm'}}]_{\mathbf{L}\times\mathbf{L'}} := \mathbb{E}_{\mathbf{M}-\mathbf{m}}\,\mathbb{E}_{\mathbf{M'}-\mathbf{m'}}\big[k([\mathsf{x}]_{\mathbf{M}}\,,X)\,K_Y^{-1}k(X,[\mathsf{x}]_{\mathbf{M'}})\,\big|\,[z]_{\mathbf{m}}\,,[z]_{\mathbf{m'}}\big]_{\mathbf{L}\times\mathbf{L'}}$$
$$= \Big([g_{\mathbf{m}}]^{\dagger}_{\mathbf{L}\times\mathbf{L''}\times\mathbf{N''}}\,[K_Y]^{-1/2}_{\mathbf{L'''}\mathbf{N'''}\times\mathbf{L''}\mathbf{N''}}\Big)\Big([g_{\mathbf{m'}}]^{\dagger}_{\mathbf{L'}\times\mathbf{L''}\times\mathbf{N''}}\,[K_Y]^{-1/2}_{\mathbf{L'''}\mathbf{N'''}\times\mathbf{L''}\mathbf{N''}}\Big)$$

using the lower triangular Cholesky decomposition $[K_Y]^{1/2}_{\mathbf{LN}\times\mathbf{LN}}$ and

$$\begin{aligned}
[\Upsilon]_{l\times l'\times\mathbf{m}\times\mathbf{m}''} &:= [\Theta]_{\mathbf{m}\times\mathbf{M}}\,\big\langle\Lambda^2+2\big\rangle^{-1}_{l\times l'\times\mathbf{M}\times\mathbf{M'}}\,[\Theta]^{\intercal}_{\mathbf{m}''\times\mathbf{M'}}\\
[\Pi]^{-1}_{l\times l'\times\mathbf{M'}\times\mathbf{M'''}} &:= \langle 1\rangle_{\mathbf{M'}\times\mathbf{M'''}} + [\Phi]_{l\times l'\times\mathbf{M'}\times\mathbf{M'''}} +\\
&\qquad [\Phi]_{l\times l'\times\mathbf{M'}\times\mathbf{m}}\,[\Gamma]^{-1}_{l\times l'\times\mathbf{m}\times\mathbf{m}''}\,[\Phi]_{l\times l'\times\mathbf{m}''\times\mathbf{M'''}}\\
[Z]_{l\times l'\times\mathbf{m'}} &:= [\Pi]_{l\times l'\times\mathbf{m'}\times\mathbf{M}}\,[\Phi]_{l\times l'\times\mathbf{M}\times\mathbf{m}''}\,[\Gamma]^{-1}_{l\times l'\times\mathbf{m}''\times\mathbf{m}}\,[\mathsf{z}]_{\mathbf{m}}
\end{aligned}$$

Again, when $\mathbf{m}=\mathbf{M}$, Θ factors out entirely.

6. Gaussian Process estimates

Using the work of the last two Sections, we are finally in a position to calculate Gaussian process estimates for the Sobol' indices Eq. (11) and their standard error Eq. (13) via the two unknown quantities in Eq. (24) and Eq. (25), as described in Section 3.

6.1. Expected Value

Using the shorthand

$$[g_{\mathbf{0}}KY]^{\dagger}_{l\times\mathbf{L}''\mathbf{N}''} := [g_{\mathbf{0}}]^{\dagger}_{l\times\mathbf{L}''\times\mathbf{N}''} \circ \left[K_Y^{-1}Y^{\dagger}\right]_{\mathbf{L}''\mathbf{N}''}$$

to write

$$\mathbb{E}_{\mathbf{m}}\left[[\mu_{\mathbf{m}}]^2_{\mathbf{L}}\right]_{l\times l'} =: [g_{\mathbf{0}}KY]^{\dagger}_{l\times\mathbf{L}''\mathbf{N}''} \, [H_{\mathbf{m}}]^{\dagger}_{l\times\mathbf{L}''\times\mathbf{N}''\times l'\times\mathbf{L}'''\times\mathbf{N}'''} \, [g_{\mathbf{0}}KY]^{\dagger}_{l'\times\mathbf{L}'''\mathbf{N}'''}$$

results in

$$\begin{aligned}
&[H_{\mathbf{m}}]_{l\times\mathbf{L}''\times\mathbf{N}''\times l'\times\mathbf{L}'''\times\mathbf{N}'''} \\
&:= \mathbb{E}_{\mathbf{m}}\left[\frac{\mathrm{p}\big([\mathsf{z}]_{\mathbf{m}} \big| \, [G]_{\mathbf{m}\times l\times\mathbf{L}''\times\mathbf{N}''} \, , [\Gamma]_{l\times\mathbf{L}''}\big) \otimes \mathrm{p}\big([\mathsf{z}]_{\mathbf{m}} \big| \, [G]_{\mathbf{m}\times l'\times\mathbf{L}'''\times\mathbf{N}'''} \, , [\Gamma]_{l'\times\mathbf{L}'''}\big)}{\mathrm{p}\big([\mathsf{z}]_{\mathbf{m}} \big| \, [0]_{\mathbf{m}} \, , \langle 1\rangle_{\mathbf{m}\times\mathbf{m}}\big) \, \mathrm{p}\big([\mathsf{z}]_{\mathbf{m}} \big| \, [0]_{\mathbf{m}} \, , \langle 1\rangle_{\mathbf{m}\times\mathbf{m}}\big)}\right] \\
&= \frac{\mathrm{p}\big([\Phi]_{l\times l''} \, [G]_{\mathbf{m}\times l'\times l'''\times\mathbf{N}'''} \big| \, [G]_{\mathbf{m}\times l\times l''\times\mathbf{N}''} \, , [\Psi]_{l\times l''\times l'\times l'''} \, [\Phi]_{l\times l''\times\mathbf{m}''\times\mathbf{m}}\big)}{\mathrm{p}\big([0]_{\mathbf{m}} \big| \, [G]_{\mathbf{m}\times l\times l''\times\mathbf{N}''} \, , [\Phi]_{l\times l''}\big)}
\end{aligned}$$

where

$$\begin{aligned}
[\Psi]_{l\times l''\times l'\times l'''\times\mathbf{m}^*\times\mathbf{m}''} := {} & [\Gamma]_{l\times l''\times\mathbf{m}^*\times\mathbf{m}''} + [\Gamma]_{l'\times l'''\times\mathbf{m}^*\times\mathbf{m}''} \\
& - [\Gamma]_{l\times l''\times\mathbf{m}^*\times\mathbf{m}} \, [\Gamma]_{l'\times l'''\times\mathbf{m}\times\mathbf{m}''}
\end{aligned}$$

6.2. Variance

Recall from Eq. (25) that the inputs comprising $\mathbf{m}, \mathbf{m}'$ vary independently when calculating a covariance $W_{\mathbf{mm}'}$ via $A_{\mathbf{mm}'}$. Eqs. (20) and (21) eliminate terms containing the ungoverned noise variance $[E]_{\mathbf{L}^2}$, leaving just two terms in

$$\begin{aligned}
&\mathbb{E}_{\mathbf{m}}\mathbb{E}_{\mathbf{m}'}[\mu_{\mathbf{m}} \otimes \mu_{\mathbf{mm}'} \otimes \mu_{\mathbf{m}'}]_{\mathbf{L}^2\times\mathbf{L}'^2} \\
&\qquad = \mathbb{E}_{\mathbf{m}}\mathbb{E}_{\mathbf{m}'}[\mu_{\mathbf{m}} \otimes (F \circ \phi_{\mathbf{mm}'} - \psi_{\mathbf{mm}'} + E) \otimes \mu_{\mathbf{m}'}]_{\mathbf{L}^2\times\mathbf{L}'^2} \\
&\qquad = [F]_{\mathbf{L}\times\mathbf{L}'} \circ \mathbb{E}_{\mathbf{m}}\mathbb{E}_{\mathbf{m}'}[\mu_{\mathbf{m}} \otimes \phi_{\mathbf{mm}'} \otimes \mu_{\mathbf{m}'}]_{\mathbf{L}^2\times\mathbf{L}'^2} \\
&\qquad\qquad - \mathbb{E}_{\mathbf{m}}\mathbb{E}_{\mathbf{m}'}[\mu_{\mathbf{m}} \otimes \psi_{\mathbf{mm}'} \otimes \mu_{\mathbf{m}'}]_{\mathbf{L}^2\times\mathbf{L}'^2} \qquad (34)
\end{aligned}$$

The first term is calculated componentwise as

$$
\begin{aligned}
&[F]_{l''\times l'''} \circ \mathbb{E}_{\mathbf{m}}\,\mathbb{E}_{\mathbf{m}'}\big[[\mu_{\mathbf{m}}]_{l} \otimes [\phi_{\mathbf{m}\mathbf{m}'}]_{l''\times l'''} \otimes [\mu_{\mathbf{m}'}]_{l'}\big] = \\
&\qquad \frac{[F]_{l''\times l'''}\,\big|\langle \Lambda^2\rangle_{l''\times l'''\times \mathbf{M}^2}\big|^{1/2}}{\big|\langle \Lambda^2+2\rangle_{l''\times l'''\times \mathbf{M}^2}\big|^{1/2}}\,[g_{\mathbf{0}}KY]^{\dagger}_{l\times \mathbf{L}^{*}\mathbf{N}^{*}} \otimes [g_{\mathbf{0}}KY]^{\dagger}_{l'\times \mathbf{L}^{*\prime}\mathbf{N}^{*\prime}} \\
&\Bigg[(2\pi)^{m/2}\mathrm{p}\Big([0]_{\mathbf{m}}\big|\,[\Upsilon]^{1/2}_{l''\times l'''}\,[G]_{\mathbf{m}\times l\times \mathbf{L}^{*}\times \mathbf{N}^{*}}\,,\langle 1\rangle - [\Upsilon]^{1/2}_{l''\times l'''}\,[\Phi]_{l\times \mathbf{L}^{*}}\,[\Upsilon]^{\intercal/2}_{l''\times l'''}\Big) \\
&\quad\circ\frac{\mathrm{p}\Big([G]_{\mathbf{m}'\times l'\times \mathbf{L}^{*\prime}\times \mathbf{N}^{*\prime}}\big|\,[\Omega]\,[C]\,[\Gamma]^{-1}_{l\times \mathbf{L}^{*}}\,[G]_{\mathbf{m}\times l\times \mathbf{L}^{*}\times \mathbf{N}^{*}}\,,[B]+[\Omega]\,[C]\,[\Omega]^{\intercal}\Big)}{\mathrm{p}\Big([0]_{\mathbf{m}'}\big|\,[G]_{\mathbf{m}'\times l'\times \mathbf{L}^{*\prime}\times \mathbf{N}^{*\prime}}\,,[\Phi]_{l'\times \mathbf{L}^{*\prime}}\Big)}\Bigg]^{\dagger}
\end{aligned}
$$

using the lower triangular Cholesky decomposition

$$
[\Upsilon]_{l''\times l'''\times \mathbf{m}^2} = [\Upsilon]^{1/2}_{l''\times l'''}\,[\Upsilon]^{\intercal/2}_{l''\times l'''}
$$

and

$$
\begin{aligned}
[\Omega]_{\mathbf{m}'\times \mathbf{m}} &:= [\Phi]_{l'\times l^{*\prime}\times \mathbf{m}'\times \mathbf{m}'''}\,[\Pi]_{l''\times l'''\times \mathbf{m}'''\times \mathbf{M}}\,[\Phi]_{l''\times l'''\times \mathbf{M}\times \mathbf{m}''}\,[\Gamma]^{-1}_{l''\times l'''\times \mathbf{m}''\times \mathbf{m}} \\
[B]_{\mathbf{m}'\times \mathbf{m}'''} &:= [\Gamma]_{l'\times l^{*\prime}\times \mathbf{m}'\times \mathbf{m}^{*\prime}}\,[\Phi]_{l'\times l^{*\prime}\times \mathbf{m}^{*\prime}\times \mathbf{m}'''} + \\
&\qquad [\Phi]_{l'\times l^{*\prime}\times \mathbf{m}'\times \mathbf{m}^{*\prime\prime\prime}}\,[\Pi]_{l''\times l'''\times \mathbf{m}^{*\prime\prime\prime}\times \mathbf{m}^{*\prime}}\,[\Phi]_{l'\times l^{*\prime}\times \mathbf{m}^{*\prime}\times \mathbf{m}'''} \\
[C]_{\mathbf{m}\times \mathbf{m}''} &:= [1-\Upsilon]_{l''\times l'''\times \mathbf{m}\times \mathbf{m}^{*}} \\
&\qquad \big[\langle 1\rangle - [\Phi]_{l\times l^{*}\times \mathbf{m}^{**}\times \mathbf{m}^{**\prime\prime}}\,[\Upsilon]_{l''\times l'''\times \mathbf{m}^{**\prime\prime}\times \mathbf{m}^{***}}\big]^{-1}_{\mathbf{m}^{*}\times \mathbf{m}^{*\prime\prime}} \\
&\qquad [\Gamma]_{l\times l^{*}\times \mathbf{m}^{*\prime\prime}\times \mathbf{m}''}
\end{aligned}
$$

The second term is calculated componentwise as

$$
\mathbb{E}_{\mathbf{m}}\,\mathbb{E}_{\mathbf{m}'}\big[[\mu_{\mathbf{m}}]_{l} \otimes [\psi_{\mathbf{m}\mathbf{m}'}]_{l''\times l'''} \otimes [\mu_{\mathbf{m}'}]_{l'}\big] = [E_{\mathbf{m}}]_{l\times l''\times \mathbf{L}^{**}\mathbf{N}^{**}}\,[E_{\mathbf{m}'}]_{l'\times l'''\times \mathbf{L}^{**}\mathbf{N}^{**}}
$$

where

$$
\begin{aligned}
&[E_{\mathbf{m}'}]_{l'\times l'''\times \mathbf{L}^{**}\mathbf{N}^{**}} := \Big([K_Y]^{-1/2}_{\mathbf{L}^{**}\mathbf{N}^{**}\times \mathbf{L}^{**}\mathbf{N}^{**}} \otimes [g_{\mathbf{0}}KY]^{\dagger}_{l'\times \mathbf{L}^{*\prime}\mathbf{N}^{*\prime}}\Big) \\
&\quad\left[\frac{[g_{\mathbf{0}}]_{l'''\times \mathbf{L}^{**}\times \mathbf{N}^{**}} \circ \mathrm{p}\Big([\Phi]_{l'\times \mathbf{L}^{*\prime}}\,[G]_{\mathbf{m}'\times l'''\times \mathbf{L}^{**}\times \mathbf{N}^{**}}\big|\,[G]_{\mathbf{m}'\times l'\times \mathbf{L}^{*\prime}\times \mathbf{N}^{*\prime}}\,,[D]\Big)}{\mathrm{p}\Big([0]_{\mathbf{m}'}\big|\,[G]_{\mathbf{m}'\times l'\times \mathbf{L}^{*\prime}\times \mathbf{N}^{*\prime}}\,,[\Phi]_{l'\times \mathbf{L}^{*\prime}}\Big)}\right]^{\dagger} \\
&[D]_{l'\times l'''\times l^{*\prime}\times l^{**}\times \mathbf{m}'^{2}} := [\Phi]_{l'\times l^{*\prime}\times \mathbf{m}'\times \mathbf{m}'''} \\
&\qquad\qquad - [\Phi]_{l'\times l^{*\prime}\times \mathbf{m}'\times \mathbf{m}^{**}}\,[\Phi]_{l'''\times l^{**}\times \mathbf{m}^{**}\times \mathbf{m}^{*\prime}}\,[\Phi]_{l'\times l^{*\prime}\times \mathbf{m}^{*\prime}\times \mathbf{m}'''}
\end{aligned}
$$

and $[E_{\mathbf{m}'}]_{l\times l''\times \mathbf{L}^{**}\mathbf{N}^{**}}$ substitutes $\mathbf{m}' \mapsto \mathbf{m}$, $l' \mapsto l$, $l''' \mapsto l''$ in these definitions. In other words, drop a single prime superscript from every symbol.

This completes the calculation of all quantities of interest.

7. Complexity and simplifications

In this Section we highlight the computational cost of these calculations, assuming GP regression has already performed the Cholesky decomposition $[K_Y]^{1/2}_{(\mathbf{LN})^2}$. GP regression typically requires this to be computed several times in optimizing the GP hyperparameters E, F, Λ. This optimization will always dominate compute time, in repeatedly calculating $\exp(\ldots)$ of $O(L^2N^2M)$ and the Cholesky decomposition of $O(L^3N^3)$ [36].

We shall therefore concentrate on the memory demands of calculating $[S_{\mathbf{m}}]_{\mathbf{L}\times\mathbf{L}'}, [T_{\mathbf{m}}]_{(\mathbf{L}\times\mathbf{L}')^2}$, which are substantial. In all cases, we consider the largest two tensors occurring in an Einstein summation to be the effective memory requirement. All this assumes that the number of inputs is moderate $M \ll N$, which is required for GP regression to avoid the curse of dimensionality [37, 38] anyway.

From Section 6.1 the memory required to compute a closed index $S_{\mathbf{m}}$ is

$$\text{cost of } \mathbb{E}_{\mathbf{m}}[\mu_{\mathbf{m}}]^2 = O(L^4N^2)$$

From Section 6.2 the memory required to compute a standard error $T_{\mathbf{m}}$ is

$$\text{cost of } \mathbb{E}_{\mathbf{m}'}\mathbb{E}_{\mathbf{m}}[\mu_{\mathbf{m}} \otimes \phi_{\mathbf{mm}'} \otimes \mu_{\mathbf{m}'}] = O(L^6N^2)$$

For a medium-sized problem of 10 outputs and 1000 datapoints, this is $L^6N^2 = 10^{12}$, for larger problems this could reach 10^{20}. This will challenge the memory limitations of a CPU or GPU. However, there are two simplifications which substantially ease this burden.

7.1. Diagonal Uncertainty

The first simplification observes that assessing standar error concerns the variances of Sobol' indices, not the cross covariances between them. This costlessy reduces the tensor elements which must be calculated according to

$$l' = l \ ; \ l''' = l'' \quad \text{or} \quad l' = l'' \ ; \ l''' = l \quad \text{or} \quad l'' = l \quad \text{throughout Section 6.2.}$$

The memory required to compute a standard error $T_{\mathbf{m}}$ is reduced by $O(L)$ to

$$\text{cost of } \mathbb{E}_{\mathbf{m}'}\mathbb{E}_{\mathbf{m}}[\mu_{\mathbf{m}} \otimes \phi_{\mathbf{mm}'} \otimes \mu_{\mathbf{m}'}] = O(L^5N^2)$$

7.2. Independent Kernels

The second simplification is to restrict the GP to independent kernels, by constraining the signal covariance to be diagonal

$$[F]_{l\times l''} = 0 \text{ unless } l'' = l$$

In which case the off-diagonal elements $[\Lambda]_{l\times l''\times\mathbf{M}}$ of the lengthscales tensor are completely irrelevant and need not be specified. Note, however, that the off-diagonal elements of ungovened noise (likelihood) covariance $[E]_{l\times l''}$ are unconstrained, provided $[E]_{\mathbf{L}^2}$ is a covariance, and therefore symmetric positive definite.

Restricting $[F]_{\mathbf{L}^2}$ to be diagonal implies that

$$l'' = l \ ; \ l''' = l' \quad \text{throughout Section 6.1.}$$
$$l''' = l'' \text{ and } l^* = l \ ; \ l^{*\prime} = l' \quad \text{throughout Section 6.2.}$$

This reduces the memory required to compute a closed index $S_{\mathbf{m}}$ by $O(L^2)$ to

$$\text{cost of } \mathbb{E}_{\mathbf{m}}[\mu_{\mathbf{m}}]^2 = O(L^2N^2)$$

and the memory required to compute an standard error $T_{\mathbf{m}}$ by $O(L^3)$ to

$$\text{cost of } \mathbb{E}_{\mathbf{m}'}\mathbb{E}_{\mathbf{m}}[\mu_{\mathbf{m}} \otimes \phi_{\mathbf{mm}'} \otimes \mu_{\mathbf{m}'}] = O(L^2N^2)$$

Using these simplifications, the time and memory demands of the Sobol' calculation are no greater than those of the MOGP regression enabling it.

8. Conclusion

In this paper, we transformed uniformly distributed inputs u to normally distributed inputs z, enabling an arbitrary rotation by Θ to inputs x which are still normal. We then preformed Multi-Output Gaussian Process (MOGP) regression with an anisotropic radial basis function (RBF/ARD) kernel on x, broadly applicable to smoothly varying outputs. Using this surrogate, analytic expressions for closed Sobol' indices $S_{\mathbf{m}}$ are given by Eqs. (11) and (24) and Section 6.1. Analytic expressions for the standard error of these estimates over ungoverned noise is given by Eqs. (13) to (15) and (25) and Section 6.2. Reasonably cheap simplifications of the results are described in Section 7. In conclusion, we shall assess the utility of these results, pointing to further research directions.

The value of these novel formulae is somewhat limited by their high computational expense. Although calculations can be performed in seconds, their memory demands are large. Section 7 provides ways to ameliorate this so that the Sobol' calculation requires no more memory than MOGP regression enabling it. Furthermore, MOGP regression is far slower than the Sobol' index calculation, and may run into time constraints. Overall, the computational cost of a direct (presumably Monte Carlo) numerical evaluation of Sobol' indices is far lower.

The technique we have developed here will be preferable, even indispensable, when training data is scarce or expensive. The use of a surrogate greatly eases data requirements, preliminary tests indicating that $N = 100$ datapoints is more than sufficient for Sobol' indices within 10% accuracy, as opposed to $N \geq 10,000$ datapoints for direct calculation. Furthermore, future research could implement our Sobol' index calculations using sparse GPs [39, 40, 41], wherein N training data are replicated by $N^* \ll N$ inducing points.

Perhaps related to the remarkably low data requirements for accuracy, it is highly significant that the regression noise E cancels from the calculation of Sobol' indices, and their standard error. This means that the accuracy of Sobol' estimates is largely unaffected by noise in the training data. Put another way, the Sobol' comparison between predictions relying on M inputs and predictions only using $m \leq M$ inputs is indifferent to the absolute quality of those predictions. Two poor predictors are compared just as accurately as two good ones. Obviously this is extremely attractive whenever the output is inherently noisy, incorporating unavoidable random error. The Sobol' calculation on GPs is remarkably immune to this error.

A significant limitation is bound to be the number of inputs M. All GPs are extremely susceptible to every aspect of the curse of dimensionality [37, 38]. Every datapoint tends to the same Euclidean distance from every other, and inevitably lies adjacent to the bounding hypersurface as the dimensionality of the input hyperspace increases. We would caution against allowing more than $M = 15$ inputs in any case.

This is where arbitrary rotation Θ of inputs comes into its own. If the goal is reducing inputs, rotating their basis first boosts the possibilities immensely [29]. This presents the possibilty of choosing Θ to maximise the closed Sobol' index of the first few inputs, called the active subspace. Exciting future work could thereby build the input space from the bottom up, as follows. From a large number of inputs, take 10 likely suspects, fit a GP and rotate to an active subspace of 5 inputs. Then fit a new GP to the active subspace plus 5 inputs previously ignored, and rotate to a new active subspace, again optimizing Θ to maximize the first 5 Sobol' indices. In this manner a high fidelity surrogate may be achieved without ever confronting the curse of dimensionality. Furthermore, the Sobol' calculations at every step will be accurate and robust, because the inputs ignored will manifest as noise in the output, to which our technique is immune. We can even allow this noise to be correlated between outputs – as it no doubt will be, considering its source – because $[E]_{\mathbf{L}^2}$ can be non-diagonal in our calculations, even when independent kernels (diagonal $[F]_{\mathbf{L}^2}$) are employed. In fact, this should only affect the fidelity of GP regression, as the Sobol' calculation is immune to E.

Finally, we should emphasise that that the fidelity and robustness of everything we have suggested is easily monitored because we provide uncertainty quantification in the form of the standard error of the Sobol' indices over ungoverned noise. Because of our multi-output approach, all our results and suggestions apply not only to the outputs themselves, but equally to the linkages (covariance) between them.

References

[1] S. Razavi, A. Jakeman, A. Saltelli, C. Prieur, B. Iooss, E. Borgonovo, E. Plischke, S. L. Piano, T. Iwanaga, W. Becker, S. Tarantola, J. H. Guillaume, J. Jakeman, H. Gupta, N. Melillo, G. Rabitti, V. Chabridon, Q. Duan, X. Sun, S. Smith, R. Sheikholeslami, N. Hosseini, M. Asadzadeh, A. Puy, S. Kucherenko, H. R. Maier, The future of sensitivity analysis: An essential discipline for systems modeling and policy support, Environmental Modelling & Software 137 (2021-03) 104954. `doi:10.1016/j.envsoft.2020.104954`.

[2] A. Saltelli, K. Aleksankina, W. Becker, P. Fennell, F. Ferretti, N. Holst, S. Li, Q. Wu, Why so many published sensitivity analyses are false: A systematic review of sensitivity analysis practices, Environmental Modelling & Software 114 (2019-04) 29–39. `doi:10.1016/j.envsoft.2019.01.012`.

[3] R. Ghanem, D. Higdon, H. Owhadi, et al., Handbook of uncertainty quantification, Vol. 6, Springer, 2017.

[4] I. M. Sobol, Global sensitivity indices for nonlinear mathematical models and their monte carlo estimates, Mathematics and Computers in Simulation 55 (2001) 271–280.

[5] D. Chicco, M. J. Warrens, G. Jurman, The coefficient of determination r-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, PeerJ Computer Science 7 (1) (2021-07) e623. doi:10.7717/peerj-cs.623.

[6] F. Gamboa, A. Janon, T. Klein, A. Lagnoux, Sensitivity indices for multivariate outputs, Comptes Rendus Mathematique 351 (7-8) (2013) 307–310. doi:10.1016/j.crma.2013.04.016.

[7] S. Xiao, Z. Lu, F. Qin, Estimation of the generalized sobol's sensitivity index for multivariate output model using unscented transformation, Journal of Structural Engineering 143 (5) (may 2017). doi:10.1061/(asce)st.1943-541x.0001721.

[8] O. Garcia-Cabrejo, A. Valocchi, Global sensitivity analysis for multivariate output using polynomial chaos expansion, Reliability Engineering & System Safety 126 (6) (2014) 25–36. doi:10.1016/j.ress.2014.01.005.

[9] K. Campbell, M. D. McKay, B. J. Williams, Sensitivity analysis when model outputs are functions, Reliability Engineering & System Safety 91 (10-11) (2006) 1468–1472. doi:10.1016/j.ress.2005.11.049.

[10] M. Lamboni, H. Monod, D. Makowski, Multivariate sensitivity analysis to measure global contribution of input factors in dynamic models, Reliability Engineering & System Safety 96 (4) (2011) 450–459. doi:10.1016/j.ress.2010.12.002.

[11] K. Zhang, Z. Lu, K. Cheng, L. Wang, Y. Guo, Global sensitivity analysis for multivariate output model and dynamic models, Reliability Engineering & System Safety 204 (2020) 107195. doi:10.1016/j.ress.2020.107195.

[12] B. Lamoureux, N. Mechbal, J. R. Massé, A combined sensitivity analysis and kriging surrogate modeling for early validation of health indicators, Reliability Engineering and System Safety 130 (2014) 12–26. doi:10.1016/j.ress.2014.03.007.

[13] R. G. Ghanem, P. D. Spanos, Spectral techniques for stochastic finite elements, Archives of Computational Methods in Engineering 4 (1) (1997) 63–100. doi:10.1007/BF02818931.

[14] D. Xiu, G. E. Karniadakis, The wiener–askey polynomial chaos for stochastic differential equations, SIAM Journal on Scientific Computing 24 (2) (2002) 619–644. doi:10.1137/s1064827501387826.

[15] D. Xiu, Numerical Methods for Stochastic Computations: A Spectral Method Approach, Princeton University Press, 2010.

[16] M. Chevreuil, R. Lebrun, A. Nouy, P. Rai, A least-squares method for sparse low rank approximation of multivariate functions, SIAM/ASA Journal on Uncertainty Quantification 3 (1) (2015) 897–921. arXiv:http://arxiv.org/abs/1305.0030v2, doi:10.1137/13091899X.

[17] K. Konakli, B. Sudret, Global sensitivity analysis using low-rank tensor approximations, Reliability Engineering & System Safety 156 (2016) 64–83. doi:10.1016/j.ress.2016.07.012.

[18] C. Cortes, V. Vapnik, Support-vector networks, Machine Learning 20 (3) (1995) 273–297. doi:10.1007/bf00994018.

[19] J. Sacks, W. J. Welch, T. J. Mitchell, H. P. Wynn, Design and analysis of computer experiments, Statistical Science 4 (4) (1989) 409–423.

[20] C. E. Rasmussen, C. K. I. Williams, Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning series), The MIT Press, 2005.

[21] M. A. Alvarez, L. Rosasco, N. D. Lawrence, Kernels for vector-valued functions: a review (2011). arXiv:http://arxiv.org/abs/1106.6251v2.

[22] C. E. Rasmussen, Some useful gaussian and matrix equations (2016). URL http://mlg.eng.cam.ac.uk/teaching/4f13/1617/gaussian%20and%20matrix%20equations.pdf

[23] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, Reliability Engineering & System Safety 93 (7) (2008) 964–979. doi:10.1016/j.ress.2007.04.002.

[24] J. E. Oakley, A. O'Hagan, Probabilistic sensitivity analysis of complex models: a bayesian approach, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 66 (3) (2004) 751–769. doi:10.1111/j.1467-9868.2004.05304.x.

[25] W. Chen, R. Jin, A. Sudjianto, Analytical variance-based global sensitivity analysis in simulation-based design under uncertainty, Journal of Mechanical Design 127 (5) (2005) 875. doi:10.1115/1.1904642.

[26] A. Marrel, B. Iooss, B. Laurent, O. Roustant, Calculations of sobol indices for the gaussian process metamodel, Reliability Engineering & System Safety 94 (3) (2009) 742–751. doi:10.1016/j.ress.2008.07.008.

[27] A. Srivastava, A. K. Subramaniyan, L. Wang, Analytical global sensitivity analysis with gaussian processes, Artificial Intelligence for Engineering Design, Analysis and Manufacturing 31 (03) (2017) 235–250. doi:10.1017/s0890060417000142.

[28] Z. Wu, D. Wang, P. O. N, F. Hu, W. Zhang, Global sensitivity analysis using a gaussian radial basis function metamodel, Reliability Engineering & System Safety 154 (2016) 171–179. doi:10.1016/j.ress.2016.06.006.

[29] P. G. Constantine, Active Subspaces: Emerging Ideas for Dimension Reduction in Parameter Studies, Society for Industrial and Applied Mathematics, 2015. doi:10.1137/1.9781611973860.

[30] L. C. G. Rogers, D. Williams, Diffusions, Markov Processes, and Martingales, Cambridge University Press, 2000.

[31] A. Villani, Another note on the inclusion $\mathrm{L}_p(\mu) \subset \mathrm{L}_q(\mu)$, The American Mathematical Monthly 92 (7) (1985) 485. doi:10.2307/2322503.

[32] D. Williams, Probability with Martingales, Cambridge University Press, 1991.

[33] M. G. Kendall, Kendall's advanced theory of statistics, John Wiley & Sons, 1994.

[34] NumPy user guide: Broadcasting (2022-09-24).
URL `https://numpy.org/doc/stable/user/basics.broadcasting.html`

[35] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, T. E. Oliphant, Array programming with NumPy, Nature 585 (7825) (2020-09) 357–362. `doi:10.1038/s41586-020-2649-2`.

[36] Z. Dai, Scalability of gaussian process (2022).
URL `http://gpss.cc/gpss22/slides/zhenwen.pdf`

[37] R. Bellman, Dynamic programming, Science 153 (3731) (1966) 34–37. `doi:10.1126/science.153.3731.34`.

[38] M. Binois, N. Wycoff, A survey on high-dimensional gaussian process modeling with application to bayesian optimization (Nov. 2021). `arXiv:2111.05040`.

[39] E. Snelson, Z. Ghahramani, Sparse gaussian processes using pseudo-inputs, in: Y. Weiss, B. Schölkopf, J. C. Platt (Eds.), Advances in Neural Information Processing Systems 18, MIT Press, 2006, pp. 1257–1264.

[40] M. Titsias, Variational learning of inducing variables in sparse gaussian processes, in: D. van Dyk, M. Welling (Eds.), Proceedings of the Twelth International Conference on Artificial Intelligence and Statistics, Vol. 5 of Proceedings of Machine Learning Research, PMLR, Hilton Clearwater Beach Resort, Clearwater Beach, Florida USA, 2009, pp. 567–574.
URL `https://proceedings.mlr.press/v5/titsias09a.html`

[41] J. Hensman, N. Fusi, N. D. Lawrence, Gaussian processes for big data (Sep. 2013). `arXiv:1309.6835`.