The Coefficient of Determination of a Reduced Order Model

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

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

---

Abstract

Keywords: Global Sensitivity Analysis, Sobol' Index, Surrogate Model, Active Subspace

---

1. Introduction

This paper is concerned with analysing the results of experiments or computer simulations embodying $M \geq 1$ inputs and $L \geq 1$ outputs. Global Sensitivity Analysis [1] examines the relevance of the various inputs to the various outputs. When pursued via ANOVA decomposition, this leads naturally to the well known Sobol' indices [2]. Each closed Sobol' index is essentially a coefficient of determination (correlation squared) between the predictions of a full model with M inputs and those of a reduced model with $m \leq M$ inputs. As described in [3], with multiple outputs the ANOVA is on the covariance matrix of outputs. Thus one examines not just the relevance of inputs to the output variables themselves, but also to the linkages between these outputs. This may be of considerable interest when outputs are, for example, yield and purity of a product, or perhaps a single outut measured at various times. The Sobol indices reveal (amongst other things) which inputs are responsible for the covariance (linkages) between these outputs.

Accurate calculation of Sobol' indices even for a single output is computationally expensive and requires 10,000+ datapoints [4]. A more efficient

---

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)

approach is calculation via a surrogate model, such as Polynomial Chaos Expansions [5, 6, 7], low-rank tensor approximations [8, 9], and support vector regression [10]. This paper employs one of the most popular surrogates, the Gaussian Processes (GP) [11, 12] in its multi-output form MOGP [13], as it is highly tractable. As well as being efficient, surrogate models also smooth out noise in the outputs, which is often highly desirable in practice. Analytic expressions for Sobol' indices and their errors have been provided for integral [14, 15] and parametrized [16] GPs, and on uniformly distributed inputs [17, 18] but never for MOGPs. In this paper we provide explicit formulae for a class of MOGP with a radial basis function (RBF) kernel applicable to smoothly varying outputs.

The quantities to be calculated and their formal context are introduced in Section 2. Our approach effectively regards a regression model (with prediction errors) as just another name for a stochastic process. A great deal of progress is made in Section 3 using just 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 kernel are tersely developed and described in Section 4, then used to calculate the two unknown terms in Sections 5 and 6. Conclusions are drawn in Section 7.

2. Coefficient of Determination

Given a model

$$\text{Integrable } y\colon [0,1]^{M+1} \mapsto \mathbb{R}^L$$

take as input a uniformly distributed random variable (RV)

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

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

$$\mathbf{M+1} := (0,\ldots,M) \supseteq \mathbf{m+1} := (0,\ldots,m \le M)$$

with void $\mathbf{0} = ()$ voiding any tensor it subscripts. Ordinals are concatenated by Cartesian $\times$ and may be subtracted like sets, as in $\mathbf{M} - \mathbf{m} := (m, \ldots, M-1)$. Subscripts refer to 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 dimensions is left implicit after Eq. (1), to lighten notation. Now, construct $M+1$ stochastic processes (SPs)

$$[\mathsf{y}_\mathsf{m}]_{\mathbf{L}} := \mathbb{E}_{\mathbf{M}-\mathbf{m}}[y(\mathsf{u})] := \mathbb{E}_{\mathbf{M}-\mathbf{m}}\left[y(\mathsf{u}) \middle| [u]_{\mathbf{m}}\right] \tag{1}$$

ranging from $[\mathsf{y}_0]_{\mathbf{L}}$ to $[\mathsf{y}_\mathsf{M}]_{\mathbf{L}}$. Every SP depends stochastically on the ungovernable noise dimension $[\mathsf{u}]_M \perp [\mathsf{u}]_{\mathbf{M}}$ and deterministically on the first m governed input dimensions $[u]_{\mathbf{m}}$, while input dimensions $[\mathsf{u}]_{\mathbf{M}-\mathbf{m}}$. 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 Daniell-Kolmogorov [19] 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 $o \in \mathbb{Z}^+$ samples. Daniell-Kolmogorov 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 $n \in \mathbb{Z}^+$ [20]. So Fubini's Theorem [21] pp.77 allows all expectations to be taken in any order. These observations suffice to ensure every object appearing in this work. Changing the order of expectations, as permitted by Fubini's Theorem, is the vital tool in the construction of this work.

Our aim is to compare predictions from a reduced regression model y_m with those of 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 or closed Sobol' index

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

The closed Sobol' index is the complement of the commonplace total Sobol' index

$$[\mathsf{S}_{\mathsf{m}}]_{\mathbf{L}^2} =: [1]_{\mathbf{L}^2} - \left[\mathsf{S}^{\mathsf{T}}_{\mathsf{M}-\mathsf{m}}\right]_{\mathbf{L}^2}$$

It has mean value over the ungovernable noise dimension of

$$[S_{\mathbf{m}}]_{\mathbf{L}^2} := \mathbb{E}_M[\mathsf{S}_{\mathsf{m}}] = \frac{V_{\mathbf{m}}}{V_{\mathbf{M}}} \tag{3}$$

$$\text{where } [V_{\mathbf{m}}]_{\mathbf{L}^2} := \mathbb{E}_M \mathbb{V}_{\mathbf{m}}[\mathsf{y}_{\mathsf{m}}] \quad \forall \mathbf{m} \subseteq \mathbf{M} \tag{4}$$

and variance due to ungovernable noise of

$$[T_{\mathbf{m}}]_{\mathbf{L}^4} := \mathbb{V}_M[\mathsf{S}_{\mathsf{m}}] = \frac{V^2_{\mathbf{m}}}{V^2_{\mathbf{M}}} \circ \left(\frac{W_{\mathbf{mm}}}{V^2_{\mathbf{m}}} - 2\frac{W_{\mathbf{Mm}}}{V_{\mathbf{M}} \otimes V_{\mathbf{m}}} + \frac{W_{\mathbf{MM}}}{V^2_{\mathbf{M}}}\right) \tag{5}$$

$$\text{where } [W_{\mathbf{mm'}}]_{\mathbf{L}^4} := \mathbb{V}_M[\mathbb{V}_{\mathbf{m}}[\mathsf{y}_{\mathsf{m}}], \mathbb{V}_{\mathbf{m'}}[\mathsf{y}_{\mathsf{m'}}]] \quad \forall \mathbf{m}, \mathbf{m'} \subseteq \mathbf{M} \tag{6}$$

In practice it is best to retain only the term in $W_{\mathbf{mm}}$, ignoring the uncertainty in $V_{\mathbf{M}}$ conveyed by $W_{\mathbf{Mm}}, W_{\mathbf{MM}}$. This is because one is normally interested in adequate reduced models, for which $V_{\mathbf{m}} \approx V_{\mathbf{M}}$ implies $W_{\mathbf{mm}} - 2W_{\mathbf{Mm}} + W_{\mathbf{MM}} \approx 0$, giving a drastically tiny uncertainty in the Sobol' index.

The remainder of this paper is devoted to calculating these two quantities – the coefficient of determination and its variance over noise (measurement error, squared).

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 ungovernable 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 ungovernable noise, then marginalizing. To this end, adopt as design matrix a triad of inputs to condition $[\mathsf{u}]_{\mathbf{M+1}\times\mathbf{3}}$, eliciting the response

$$[\mathsf{y}]_{\mathbf{L}\times\mathbf{3}} := \mathbb{E}_{\mathbf{M}}\,\mathbb{E}_{\mathbf{M'}-\mathbf{m'}}\,\mathbb{E}_{\mathbf{0''}}\left[y([\mathsf{u}]_{\mathbf{M+1}\times\mathbf{3}})\middle|\,[[u]_{\mathbf{0}}, [u]_{\mathbf{m'}}, [u]_{\mathbf{M''}}]\right] \tag{7}$$

Primes mark independent input axes, otherwise expectations are shared by all three members of the triad. It is not always obvious whether axes 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 ungovernable noise (which is shared)

$$[\mu_n]_{(\mathbf{L}\times\mathbf{3})^n} := \mathbb{E}_M\left[[\mathsf{y}]^n_{\mathbf{L}\times\mathbf{3}}\right] \quad \forall n\in\mathbb{Z}^+ \tag{8}$$

for these embody

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

where $\mathbf{m}^{i\prime}\in\{\mathbf{0},\mathbf{m},\mathbf{M}\}$. This expression underpins the quantities we seek. The reduction which follows repeatedly realizes

$$[\mu_{\mathbf{0}\ldots\mathbf{0}\mathbf{m}^{j\prime}\ldots\mathbf{m}^{n\prime}}]_{\mathbf{L}^n} := \mathbb{E}_{\mathbf{M}}[\mu_{\mathbf{M}\ldots\mathbf{M}\mathbf{m}^{j\prime}\ldots\mathbf{m}^{n\prime}}]_{\mathbf{L}^n} = \mathbb{E}_{\mathbf{m}}[\mu_{\mathbf{m}\ldots\mathbf{m}\mathbf{m}^{j\prime}\ldots\mathbf{m}^{n\prime}}]_{\mathbf{L}^n} \tag{9}$$

Defining

$$[\mathsf{e}]_{\mathbf{L}\times\mathbf{3}} := \mathsf{y}-\mu_1 \tag{10}$$

the expected conditional variance in Eq. (3) amounts to

$$\begin{aligned}[V_{\mathbf{m}}]_{\mathbf{L}^2} &= \mathbb{E}_{\mathbf{m}}\mathbb{E}_M\left[[\mathsf{e}_{\mathsf{m}}+\mu_{\mathbf{m}}]^2_{\mathbf{L}}\right] - \mathbb{E}_M\left[[\mathsf{e}_{\mathsf{0}}+\mu_{\mathbf{0}}]^2_{\mathbf{L}}\right]\\ &= \mathbb{E}_{\mathbf{m}}\left[[\mu_{\mathbf{m}}]^2_{\mathbf{L}}\right] - [\mu_{\mathbf{0}}]^2_{\mathbf{L}} + \mathbb{E}_{\mathbf{m}}[\mu_{\mathbf{mm}}]_{\mathbf{L}^2} - [\mu_{\mathbf{00}}]_{\mathbf{L}^2}\\ &= \mathbb{E}_{\mathbf{m}}\left[[\mu_{\mathbf{m}}]^2_{\mathbf{L}}\right] - [\mu_{\mathbf{0}}]^2_{\mathbf{L}}\end{aligned} \tag{11}$$

and the covariance between conditional variances in Eq. (5) is

$$\begin{aligned}[W_{\mathbf{mm}'}]_{\mathbf{L}^4} &:= \mathbb{V}_M[\mathbb{V}_{\mathbf{m}}[\mathsf{y}_{\mathsf{m}}]\,,\mathbb{V}_{\mathbf{m}'}[\mathsf{y}_{\mathsf{m}'}]]\\ &= \mathbb{V}_M\left[\mathbb{E}_{\mathbf{m}}\left[[\mathsf{y}_{\mathsf{m}}]^2_{\mathbf{L}} - [\mathsf{y}_{\mathsf{0}}]^2_{\mathbf{L}}\right], \mathbb{E}_{\mathbf{m}'}\left[[\mathsf{y}_{\mathsf{m}'}]^2_{\mathbf{L}} - [\mathsf{y}_{\mathsf{0}}]^2_{\mathbf{L}}\right]\right]\\ &= \mathbb{E}_M\left[\mathbb{E}_{\mathbf{m}}\left[[\mathsf{y}_{\mathsf{m}}]^2_{\mathbf{L}} - [\mathsf{y}_{\mathsf{0}}]^2_{\mathbf{L}}\right]\otimes\mathbb{E}_{\mathbf{m}'}\left[[\mathsf{y}_{\mathsf{m}'}]^2_{\mathbf{L}} - [\mathsf{y}_{\mathsf{0}}]^2_{\mathbf{L}}\right]\right]\\ &\quad - [V_{\mathbf{m}}]_{\mathbf{L}^2}\otimes[V_{\mathbf{m}'}]_{\mathbf{L}^2}\\ &= [A_{\mathbf{mm}'} - A_{\mathbf{0m}'} - A_{\mathbf{m0}} + A_{\mathbf{00}}]_{\mathbf{L}^4}\end{aligned} \tag{12}$$

Here, for any $\mathbf{m},\mathbf{m}'\subseteq\mathbf{M}$

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

exploiting the fact that $V_{\mathbf{0}}=[0]_{\mathbf{L}^2}$. Equation (9) cancels all terms beginning with $[\mathsf{e}_{\mathsf{m}}]^2_{\mathbf{L}}$, first across $A_{\mathbf{m},\mathbf{m}'}-A_{\mathbf{0},\mathbf{m}'}$ then across $A_{\mathbf{m},\mathbf{0}}-A_{\mathbf{0},\mathbf{0}}$. All remaining terms ending in $[\mu_{\mathbf{m}'}]^2_{\mathbf{L}}$ are eliminated by centralization $\mathbb{E}_M[\mathsf{e}_{\mathsf{m}}]=0$ and

$$\begin{aligned}&\mathbb{E}_{\mathbf{m}}\mathbb{E}_{\mathbf{m}'}\left[[\mu_{\mathbf{m}}]^2_{\mathbf{L}}\otimes[\mu_{\mathbf{m}'}]^2_{\mathbf{L}}\right] - [V_{\mathbf{m}}]_{\mathbf{L}^2}\otimes[V_{\mathbf{m}'}]_{\mathbf{L}^2}\\ &\qquad = [V_{\mathbf{m}}]_{\mathbf{L}^2}\otimes[\mu_{\mathbf{0}}]^2_{\mathbf{L}} + [\mu_{\mathbf{0}}]^2_{\mathbf{L}}\otimes[V_{\mathbf{m}'}]_{\mathbf{L}^2} + [\mu_{\mathbf{0}}]^4_{\mathbf{L}}\end{aligned}$$

cancelling across $A_{\mathbf{m},\mathbf{m}'} - A_{\mathbf{0},\mathbf{m}'} - A_{\mathbf{m},\mathbf{0}} + A_{\mathbf{0},\mathbf{0}}$. Similar arguments eliminate $[\mathsf{e}_{\mathsf{m}'}]^2_{\mathbf{L}}$ and $[\mu_{\mathbf{m}}]^2_{\mathbf{L}}$. Effectively then

$$[A_{\mathbf{mm}'}]_{\mathbf{L}^4} = \sum_{\pi(\mathbf{L}^2)} \sum_{\pi(\mathbf{L}'^2)} \mathbb{E}_{\mathbf{m}} \mathbb{E}_{\mathbf{m}'} [\mu_{\mathbf{m}} \otimes \mu_{\mathbf{mm}'} \otimes \mu_{\mathbf{m}'}]_{\mathbf{L}^2 \times \mathbf{L}'^2} \tag{13}$$

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 ($\mathbf{L}^{i\prime} = \mathbf{L}$ always), they do not change the value of the constant – unlike primes on variables ($\mathbf{m}^{i\prime}$ need not equal $\mathbf{m}$ in general). One is normally only interested in variances (errors), constituted by the diagonal $\mathbf{L}'^2 = \mathbf{L}^2$, for which the summation in Eq. (13) is over a pair of identical pairs.

In order to further elucidate these estimates, we must fill in the details of the underlying stochastic processes, sufficiently identifying the regression y by its first two moments μ_1, μ_2. Then all the answers we desire are given by Eqs. (3) and (11) and Eqs. (5), (12) and (13).

4. Interlude: Gaussian Process Regression

The development in this Section is based on [13], but introduces different notation. A Gaussian Process (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

$$[]_{\mathbf{L}\times\mathbf{o}} \sim \mathsf{N}^{\dagger}\Big([]_{\mathbf{L}\times\mathbf{o}}\,, []_{(\mathbf{L}\times\mathbf{o})^2}\Big) \Longleftrightarrow []^{\dagger}_{\mathbf{L}\times\mathbf{o}} \sim \mathsf{N}\Big([]^{\dagger}_{\mathbf{L}\times\mathbf{o}}\,, []^{\dagger}_{(\mathbf{L}\times\mathbf{o})^2}\Big)$$

$$\Big[[]^{\dagger}_{\mathbf{L}\times\mathbf{o}}\Big]_{\mathbf{lo}-(\mathbf{l}-\mathbf{1})\mathbf{o}} := []_{(l-1)\times\mathbf{o}}$$

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

supporting the fundamental definition of the GP kernel, as a covariance between responses

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

4.1. Tensor Gaussians

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

$$\big[\mathrm{p}\big([x]_{\mathbf{m}\times\mathbf{o}}\big|\,[x']_{\mathbf{m}\times\mathbf{o}'}\,,[\Sigma]_{\mathbf{L}^2\times\mathbf{m}^2}\big)\big]_{l\times o\times l'\times o'} := (2\pi)^{-M/2}\left|[\Sigma]_{l\times l'}\right|^{-1/2} \exp\left(-\frac{[x-x']^{\mathsf{T}}_{\mathbf{m}\times l\times o\times l'\times o'}\,[\Sigma]^{-1}_{l\times l'}\,[x-x']_{\mathbf{m}\times l\times o\times l'\times o'}}{2}\right) \tag{14}$$

in terms of the matrix

$$[\Sigma]_{l\times l'} := [\Sigma]_{l\times l'\times\mathbf{m}^2}$$

and the transpose $^{\mathsf{T}}$ (moving first multi-index to last) of the broadcast difference between two tensors

$$[x-x']_{\mathbf{m}\times l\times o\times l'\times o'} := [x]_{\mathbf{m}\times o} - [x']_{\mathbf{m}\times o'}$$

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

$$\begin{aligned}\mathrm{p}(z|\,a,A)\circ\mathrm{p}(\Theta^{\mathsf{T}}z|\,b,B) &= \mathrm{p}(0|\,(b-\Theta^{\mathsf{T}}a),(B+\Theta^{\mathsf{T}}A\Theta))\\ &\circ\,\mathrm{p}\big(z|\,(A^{-1}+\Theta B^{-1}\Theta^{\mathsf{T}})^{-1}(A^{-1}a+\Theta B^{-1}b),(A^{-1}+\Theta B^{-1}\Theta^{\mathsf{T}})^{-1}\big)\end{aligned} \tag{15}$$

This formula and the Gaussian tensors behind it will appear in a variety of guises.

4.2. Prior GP

GP regression decomposes output $[\mathsf{y}_\mathsf{M}]_\mathbf{L}$ into signal GP $[\mathsf{f}_\mathsf{M}]_\mathbf{L}$, and independent noise GP $[\mathsf{e}_M]_\mathbf{L}$ with constant noise covariance $[E]_{\mathbf{L}^2}$

$$[\mathsf{y}_\mathsf{M}|E]_\mathbf{L} = [\mathsf{f}_\mathsf{M}]_\mathbf{L} + [\mathsf{e}_M|E]_\mathbf{L}$$

$$[\mathsf{e}_M|E]_\mathbf{L}\;\big|\;[x]_{\mathbf{M}\times\mathbf{o}} \sim \mathsf{N}^\dagger\Big([0]_{\mathbf{L}\times\mathbf{o}}\,,[E]_{(\mathbf{L}\times 1)^2}\circ\langle 1\rangle_{(1\times\mathbf{o})^2}\Big)$$

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 $[\Lambda]_{l\times l'\times\mathbf{M}} = [\Lambda]_{l'\times l\times\mathbf{M}}$. Angle brackets denoting a (perhaps broadcast) diagonal tensor, such as the identity matrix $\langle 1\rangle_{\mathbf{m}^2} =: \langle[1]_\mathbf{m}\rangle$, are used to define

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

and implement the objective RBF prior using Eq. (14)

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

4.3. Predictive GP

Bayesian inference for GP regression further conditions the hyper-parametrized GP $\mathsf{y}|E,F,\Lambda$ on the observed realization 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

$$\begin{aligned}
[K_{\mathsf{e}}]_{\mathbf{Lo}\times\mathbf{Lo}} &:= \mathbb{V}_{\mathbf{Lo}}\Big[\big[[\mathsf{e}_M|E]_{\mathbf{L}} \,\big|\, [x]_{\mathbf{M}\times\mathbf{o}}\big]^{\dagger}\Big] \\
&= \Big[[E]_{(\mathbf{L}\times 1)^2} \circ \langle 1\rangle_{(1\times\mathbf{o})^2}\Big]^{\dagger} \\
[k(x,x')]_{\mathbf{Lo}\times\mathbf{Lo}'} &:= \mathbb{V}_{\mathbf{Lo}}\Big[\big[[\mathsf{f_M}|F,\Lambda]_{\mathbf{L}} \,\big|\, [x]_{\mathbf{M}\times\mathbf{o}}\big]^{\dagger}, \big[[\mathsf{f_M}|F,\Lambda]_{\mathbf{L}} \,\big|\, [x']_{\mathbf{M}\times\mathbf{o}'}\big]^{\dagger}\Big] \\
&= \Big[[\pm F]_{\mathbf{L}^2} \circ \mathrm{p}\Big([x]_{\mathbf{M}\times\mathbf{o}} \Big|\, [x']_{\mathbf{M}\times\mathbf{o}'}\,, \langle\Lambda^2\rangle_{\mathbf{L}^2\times\mathbf{M}^2}\Big)\Big]^{\dagger} \\
[K_Y]_{\mathbf{LN}\times\mathbf{LN}} &:= \mathbb{V}_{\mathbf{Lo}}\Big[\big[[\mathsf{y}|E,F,\Lambda]_{\mathbf{L}} \,\big|\, [X]_{\mathbf{M}\times\mathbf{N}}\big]^{\dagger}\Big] \\
&= k([X]_{\mathbf{M}\times\mathbf{N}}\,, [X]_{\mathbf{M}\times\mathbf{N}}) + [K_{\mathsf{e}}]_{\mathbf{LN}\times\mathbf{LN}}
\end{aligned} \tag{16}$$

Applying Bayes' rule

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

Eq. (15) immediately reveals the marginal likelihood

$$\mathsf{p}\big([Y|E,F,\Lambda]\,\big|X\big) = \mathrm{p}\Big([Y]^{\dagger}_{\mathbf{L}\times\mathbf{N}}\Big|\, [0]_{\mathbf{LN}}\,, K_Y\Big) \tag{17}$$

and the posterior distribution

$$\begin{aligned}
&\big[[\mathsf{f_M}|Y|E,F,\Lambda]\,\big|X\big]^{\dagger}_{\mathbf{L}\times\mathbf{N}} \sim \\
&\qquad\qquad \mathsf{N}\big(k(X,X)K_Y^{-1}Y^{\dagger},\ k(X,X) - k(X,X)K_Y^{-1}k(X,X)\big)
\end{aligned}$$

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 traditionally derived from the definition of conditional probability, but this seems unnecessary, for the extension must recover the posterior distribution when $x = X$. There is only one way of selectively replacing X with x in the posterior formula which preserves the coherence of tensor ranks:

$$\left[[\mathsf{f_M}|Y|E,F,\Lambda]\,|x\right]^{\dagger}_{\mathbf{L}\times\mathbf{o}} \sim \\ \mathsf{N}\big(k(x,X)K_Y^{-1}Y^{\dagger},\ k(x,x) - k(x,X)K_Y^{-1}k(X,x)\big) \quad (18)$$

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

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

4.4. GP Optimization

Henceforth we implicitly condition on optimal hyperparameters, which maximise the marginal likelihood Eq. (17).

$$[E]_{\mathbf{L}^2}\,, [F]_{\mathbf{L}^2}\,, [\Lambda]_{\mathbf{L}^2\times\mathbf{M}} := \operatorname{argmax} \mathrm{p}\Big(\,[Y]^{\dagger}_{\mathbf{L}\times\mathbf{N}}\Big|\,[0]_{\mathbf{LN}}\,, K_Y\Big) \quad (19)$$

5. Gaussian Process Moments

This Section calculates the stochastic process 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

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

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.

Throughout the remainder of this paper, primed ordinal subscripts are used to specify Einstein sum (einsum) 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 einsummed over.

Adding shared Gaussian noise $[\mathsf{e}_M|E]_{\mathbf{L}}$ to Eq. (18) yields

$$\big[y([\mathsf{u}]_{\mathbf{M+1\times 3}})\big|\,[u]_{\mathbf{M\times 3}}\big]^{\dagger}_{\mathbf{L\times 3}} = \big[[\mathsf{y_M}|Y|E,F,\Lambda]\,\big|[z]_{\mathbf{M\times 3}}\big]^{\dagger}_{\mathbf{L\times 3}} \sim \mathsf{N}\big(k(x,X)K_Y^{-1}Y^{\dagger},\ k(x,x)-k(x,X)K_Y^{-1}k(X,x)+E^{\dagger}\big) \quad (20)$$

using broadcast $[E^{\dagger}]_{\mathbf{L3\times L3}} := [[E]_{(\mathbf{L}\times 1)^2}\circ[1]_{(1\times\mathbf{3})^2}]^{\dagger}_{(\mathbf{L\times 3})^2}$. To bring the GP estimate fully under the umbrella of the SP estimate we should identify its ungovernable 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 $l\le L$ of $(0,1)^L$, then Eq. (20) can be written as

$$\begin{aligned}&\big[y([\mathsf{u}]_{\mathbf{M+1\times 3}})\big|\,[u]_{\mathbf{M\times 3}}\big]^{\dagger}_{\mathbf{L\times 3}}\\ &\quad = [\mu_1]^{\dagger}_{\mathbf{L\times 3}} + [\mu_2]^{\dagger/2}_{\mathbf{L\times 3\times L'\times 3'}}\Big[\big[c^{-1}(d([\mathsf{u}]_M))\big]_{\mathbf{L}\times 1}\circ[1]_{1\times\mathbf{3}}\Big]^{\dagger}_{\mathbf{L'\times 3'}}\end{aligned} \quad (21)$$

where $[\mu_2]^{\dagger/2}_{(\mathbf{L\times 3})^2}$ denotes the 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 can be determined from Eqs. (20) and (21), but still need conditioning. This is entirely a matter of repeatedly applying product formula Eq. (15), together with the familiar Gaussian identities

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

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

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

and

$$
\begin{aligned}
[G]_{\mathbf{m}\times l\times l'\times\mathbf{N}'} &:= [\Theta]_{\mathbf{m}\times\mathbf{M}} \left\langle \Lambda^2+1\right\rangle^{-1}_{l\times l'\times\mathbf{M}\times\mathbf{M}'} [X]_{\mathbf{M}'\times\mathbf{N}'} \\
[\Phi]_{l\times l'\times\mathbf{m}\times\mathbf{m}'} &:= [\Theta]_{\mathbf{m}\times\mathbf{M}} \left\langle \Lambda^2+1\right\rangle^{-1}_{l\times l'\times\mathbf{M}\times\mathbf{M}'} [\Theta]^{\intercal}_{\mathbf{m}'\times\mathbf{M}'} \\
[\Gamma]_{l\times l'\times\mathbf{m}^2} &:= \langle 1\rangle_{\mathbf{m}^2} - [\Phi]_{l\times l'\times\mathbf{m}^2}
\end{aligned}
$$

In particular, when $\mathbf{m}=\mathbf{M}$, Θ factors out entirely. On the other hand

$$
[g_{\mathbf{0}}]_{\mathbf{L}^2\times\mathbf{N}'} = [\pm F]_{\mathbf{L}^2} \circ \mathrm{p}\Big([0]_{\mathbf{M}} \big|\, [X]_{\mathbf{M}\times\mathbf{N}'}\,, \left\langle\Lambda^2+1\right\rangle_{\mathbf{L}^2\times\mathbf{M}^2}\Big)
$$

Standardization of X and Y instills a totally marginal expectation of $\mu_{\mathbf{0}} \approx [0]_{\mathbf{L}}$, but this is usually inexact.

5.2. Second Moments

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

$$
[\mu_{\mathbf{mm}'}]_{\mathbf{L}^2} = [F]_{\mathbf{L}^2}\circ[\phi_{\mathbf{mm}'}]_{\mathbf{L}^2} - [\psi_{\mathbf{mm}'}]_{\mathbf{L}^2} + [E]_{\mathbf{L}^2} \tag{22}
$$

where

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

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

by applying the Cholesky decomposition $[K_Y^{-1/2}]$ and

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

6. Gaussian Process Estimates

6.1. Expected Value

Using the shorthand

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

to write

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

results in

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

where

$$\begin{aligned}
[\Sigma]_{l\times l''\times l'\times l'''\times\mathbf{m}^2} &:= [\Gamma]_{l\times l''\times\mathbf{m}^2} + [\Gamma]_{l'\times l'''\times\mathbf{m}^2}\\
[\Psi]_{l\times l''\times l'\times l'''\times\mathbf{m}\times\mathbf{m}''} &:= [\Sigma]_{l\times l''\times l'\times l'''\times\mathbf{m}\times\mathbf{m}''}\\
&\quad - [\Gamma]_{l\times l''\times\mathbf{m}\times\mathbf{m}^{*\prime}}\,[\Gamma]_{l'\times l'''\times\mathbf{m}^{*\prime}\times\mathbf{m}''}\\
\left[|\Psi|^{-1}\right]_{l\times l''\times l'\times l'''} &:= \left|[\Psi]_{l\times l''\times l'\times l'''\times\mathbf{m}^2}\right|^{-1}\\
[\Sigma G]_{\mathbf{m}\times l\times l''\times\mathbf{N}''\times l'\times l'''\times\mathbf{N}'''} &:= [\Gamma]_{l'\times l'''\times\mathbf{m}\times\mathbf{m}'}\,[G]_{\mathbf{m}'\times l\times l''\times\mathbf{N}''}\\
&\quad + [\Gamma]_{l\times l''\times\mathbf{m}\times\mathbf{m}'}\,[G]_{\mathbf{m}'\times l'\times l'''\times\mathbf{N}'''}\\
[\Sigma\Psi]_{l\times l''\times l'\times l'''\times\mathbf{m}\times\mathbf{m}'} &:= [\Sigma]_{l\times l''\times l'\times l'''\times\mathbf{m}\times\mathbf{m}''}\,[\Psi]_{l\times l''\times l'\times l'''\times\mathbf{m}''\times\mathbf{m}'}
\end{aligned}$$

6.2. Variance

In calculating

$$\mathbb{E}_{\mathbf{m}}\,\mathbb{E}_{\mathbf{m}'}[\mu_{\mathbf{m}}\otimes\mu_{\mathbf{m}\mathbf{m}'}\otimes\mu_{\mathbf{m}'}]_{\mathbf{L}^2\times\mathbf{L}'^2}$$

from Eq. (22) the terms containing E reduce to squares of $g_{\mathbf{0}}$ by iterated expectations, and these will obviously cancel across Eq. (12). We may therefore assume $E=0$ in Eq. (22). This leaves just two terms. Firstly

$$\mathbb{E}_{\mathbf{m}}\mathbb{E}_{\mathbf{m}'}\left[[\mu_{\mathbf{m}}]_{l}\otimes[\phi_{\mathbf{mm}'}]_{l''\times l'}\otimes[\mu_{\mathbf{m}'}]_{l'''}\right]= \frac{\left|\langle\Lambda^2\rangle_{l''\times l'\times\mathbf{M}^2}\right|^{1/2}(2\pi)^{m/2}}{\left|\langle\Lambda^2+2\rangle_{l''\times l'\times\mathbf{M}^2}\right|^{1/2}}[KYg_{\mathbf{0}}]_{l\times\mathbf{L}^*\mathbf{N}^*}\otimes[KYg_{\mathbf{0}}]_{l'''\times\mathbf{L}^{*\prime}\mathbf{N}^{*\prime}}$$
$$\left[\mathrm{p}\Big([0]_{\mathbf{m}}\big|\,[1-\Upsilon]^{1/2}_{l''\times l'}[G]_{\mathbf{m}\times l\times\mathbf{L}^*\times\mathbf{N}^*},\langle 1\rangle-[1-\Upsilon]^{1/2}_{l''\times l'}[\Phi]_{l\times\mathbf{L}^*}[1-\Upsilon]^{\intercal/2}_{l''\times l'}\Big)\right.$$
$$\left.\circ\frac{\mathrm{p}\Big([G]_{\mathbf{m}'\times l'''\times\mathbf{L}^{*\prime}\times\mathbf{N}^{*\prime}}\big|\,[\Omega][C][\Gamma]^{-1}_{l\times\mathbf{L}^*}[G]_{\mathbf{m}\times l\times\mathbf{L}^*\times\mathbf{N}^*},[B]+[\Omega][C][\Omega]^\intercal\Big)}{\mathrm{p}\Big([0]_{\mathbf{m}'}\big|\,[G]_{\mathbf{m}'\times l'''\times\mathbf{L}^{*\prime}\times\mathbf{N}^{*\prime}},[\Phi]_{l'''\times\mathbf{L}^{*\prime}}\Big)}\right]^\dagger$$

using the Cholesky decomposition

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

and

$$\begin{aligned}
[\Omega]_{\mathbf{m}'\times\mathbf{m}} &:= [\Phi]_{l'''\times l^{*\prime}\times\mathbf{m}'\times\mathbf{m}'''}[\Pi]_{l''\times l'\times\mathbf{m}'''\times\mathbf{M}}[\Phi]_{l''\times l'\times\mathbf{M}\times\mathbf{m}''}[\Gamma]^{-1}_{l''\times l'\times\mathbf{m}''\times\mathbf{m}}\\
[B]_{\mathbf{m}'^2} &:= [\Gamma]_{l'''\times l^{*\prime}\times\mathbf{m}'\times\mathbf{m}^{*\prime}}[\Phi]_{l'''\times l^{*\prime}\times\mathbf{m}^{*\prime}\times\mathbf{m}'''}+\\
&\quad [\Phi]_{l'''\times l^{*\prime}\times\mathbf{m}'\times\mathbf{m}^{*\prime\prime\prime}}[\Pi]_{l''\times l'\times\mathbf{m}^{*\prime\prime\prime}\times\mathbf{m}^{*\prime}}[\Phi]_{l'''\times l^{*\prime}\times\mathbf{m}^{*\prime}\times\mathbf{m}'''}\\
[C]_{\mathbf{m}^2} &:= [\Upsilon]_{l''\times l'\times\mathbf{m}\times\mathbf{m}^*}\\
&\quad\Big[\big[[\Gamma]_{l\times l^*\times\mathbf{m}^{**2}}+[\Phi]_{l\times l^*\times\mathbf{m}^{**}\times\mathbf{m}^{**\prime\prime}}[\Upsilon]_{l''\times l'\times\mathbf{m}^{**\prime\prime}\times\mathbf{m}^{***}}\big]^{-1}\Big]_{\mathbf{m}^*\times\mathbf{m}^{*\prime\prime}}\\
&\quad[\Gamma]_{l\times l^*\times\mathbf{m}^{*\prime\prime}\times\mathbf{m}''}
\end{aligned}$$

Secondly

$$\mathbb{E}_{\mathbf{m}}\mathbb{E}_{\mathbf{m}'}\left[[\mu_{\mathbf{m}}]_{l}\otimes[\psi_{\mathbf{mm}'}]_{l''\times l'}\otimes[\mu_{\mathbf{m}'}]_{l'''}\right]=[E_{\mathbf{m}}]_{l\times l''\times\mathbf{L}^{*\prime}\mathbf{N}^{*\prime}}[E_{\mathbf{m}'}]_{l'''\times l'\times\mathbf{L}^{*\prime}\mathbf{N}^{*\prime}}$$

where

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

$$\begin{aligned}
[D]_{l'''\times l^{*\prime\prime\prime}\times l'\times l^{*\prime}\times\mathbf{m}'^2} &:= [\Phi]_{l'''\times l^{*\prime\prime\prime}\times\mathbf{m}'\times\mathbf{m}^{*\prime\prime\prime}}-\\
&\quad[\Phi]_{l'''\times l^{*\prime\prime\prime}\times\mathbf{m}'\times\mathbf{m}^{*\prime\prime\prime}}[\Phi]_{l'\times l^{*\prime}\times\mathbf{m}^{*\prime\prime\prime}\times\mathbf{m}^{*\prime}}[\Phi]_{l'''\times l^{*\prime\prime\prime}\times\mathbf{m}^{*\prime}\times\mathbf{m}'''}
\end{aligned}$$

and $[E_{\mathbf{m}}]_{l\times l''\times\mathbf{L}^{*\prime}\mathbf{N}^{*\prime}}$ substitutes $\mathbf{m}'\to\mathbf{m},\ l'''\to l,\ l'\to l''$ in these definitions.

7. Conclusion

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 137 104954. `doi:10.1016/j.envsoft.2020.104954`.

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

[3] 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`.

[4] 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`.

[5] 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`.

[6] 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`.

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

[8] 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`.

[9] 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.

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

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

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

[13] M. A. Alvarez, L. Rosasco, N. D. Lawrence, Kernels for vector-valued functions: a reviewarXiv:http://arxiv.org/abs/1106.6251v2.

[14] 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.

[15] R. Jin, W. Chen, A. Sudjianto, Analytical metamodel-based global sensitivity analysis and uncertainty propagation for robust design, SAE Transactions Journal of Materials & Manufacturing (2004). doi:10.4271/2004-01-0429.

[16] 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.

[17] 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.

[18] 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.

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

[20] A. Villani, Another note on the inclusion $L_p(\mu) \subset L_q(\mu)$, The American Mathematical Monthly 92 (7) (1985) 485. doi:10.2307/2322503.

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

[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