Emulation of Computer Models with Multivariate Output

Ksenia Kyzyurova [1] with Jim Berger and Robert Wolpert

Duke University, Department of Statistical Science

April 19, 2018

[1] Current affiliation: Statistics program, KAUST

Eyjafjallajökull

credit: hiticeland.com

Uncertainty quantification

Experiments and observations are rare (e.g. volcano eruptions *(Bursik et al., 2012)*.)

Computer models are simulators based on mathematical representation of reality.

Emulators are fast approximations to computationally expensive simulators *(Sacks et al., 1989)*.

Sometimes the output of a computer model is multivariate.

Model of a volcano eruption ash column has five outputs.

Minimum and maximum of a height of the ash column are highly correlated.

Myth

Multivariate modeling of a computer model multivariate output leads to “better" emulation results than individual modeling of each output.

Outline

Gaussian stochastic process (GASP) emulator of a computer model.

Emulation of computer models with *multivariate output*.

Gaussian process emulator

Function g is a simulator of a computer model.
$\mathbf{z} = \{\mathbf{z}_1, \ldots, \mathbf{z}_m\}$ is a vector of computer model inputs.

If $\{g(\mathbf{z}_1), \ldots, g(\mathbf{z}_m)\} = \mathbf{g}(\mathbf{z})$ are the runs of the computer model g at these inputs, then with a Gaussian stochastic process $g^M(\cdot)$ prior on g *(Bayarri et al., 2007)*

$$\mathbf{g}^{\mathbf{M}}(\mathbf{z}) \sim \mathcal{N}(\boldsymbol{\mu}, \sigma_g^2 \mathbf{C}_z), \tag{1}$$

where $\boldsymbol{\mu} = (\tilde{\eta}(\mathbf{z}_1), \ldots, \tilde{\eta}(\mathbf{z}_m))$ and $\tilde{\eta}(\cdot)$ is the mean function of the process, σ_g^2 is the unknown variance and $\mathbf{C}_z$ is the correlation matrix whose (k, l) element is given by a correlation function $c(\mathbf{z}_k, \mathbf{z}_l)$.

Parameters of the GASP

$\tilde{\eta}(\cdot) = \mathbf{h}(\cdot)\boldsymbol{\beta}$ where $\mathbf{h}(\cdot)^{\mathrm{T}}$ is a vector of regression functions and $\boldsymbol{\beta}$ are unknown regression coefficients *(Sacks et al., 1989)*.

Correlation function $c(\cdot, \cdot)$ between outputs at two inputs $\mathbf{z}_k$ and $\mathbf{z}_l$ equals

$$c(\mathbf{z}_k, \mathbf{z}_l) = \prod_{j=1}^{d} c(z_{kj}, z_{lj}). \tag{2}$$

For the jth coordinate

$$c(z_{kj}, z_{lj}) = \exp\left(-\left(\frac{|z_{kj} - z_{lj}|}{\delta_j}\right)^{\alpha_j}\right) \tag{3}$$

with range $\delta_j \in (0, \infty)$ and smoothness $\alpha_j \in (0, 2]$.

GASP parameters $\boldsymbol{\theta}_g = (\boldsymbol{\beta}, \sigma^2, \{\alpha_j\}_{j=1,\ldots,m}, \{\delta_j\}_{j=1,\ldots,m})$.

GASP

Computer model evaluations $\mathbf{g}(\mathbf{z})$ at $\mathbf{z}$ are training data.

The posterior predictive GASP at a new input $\mathbf{z}'$ (given GASP parameters $\boldsymbol{\theta}_g$) is $\mathcal{N}(\mu^*(\mathbf{z}'), \sigma^{*2}(\mathbf{z}'))$

$$\mu^*(\mathbf{z}') = \mu(\mathbf{z}') + c(\mathbf{z}', \mathbf{z})\mathbf{C}_z^{-1}(\mathbf{g}(\mathbf{z}) - \boldsymbol{\mu}), \tag{4}$$

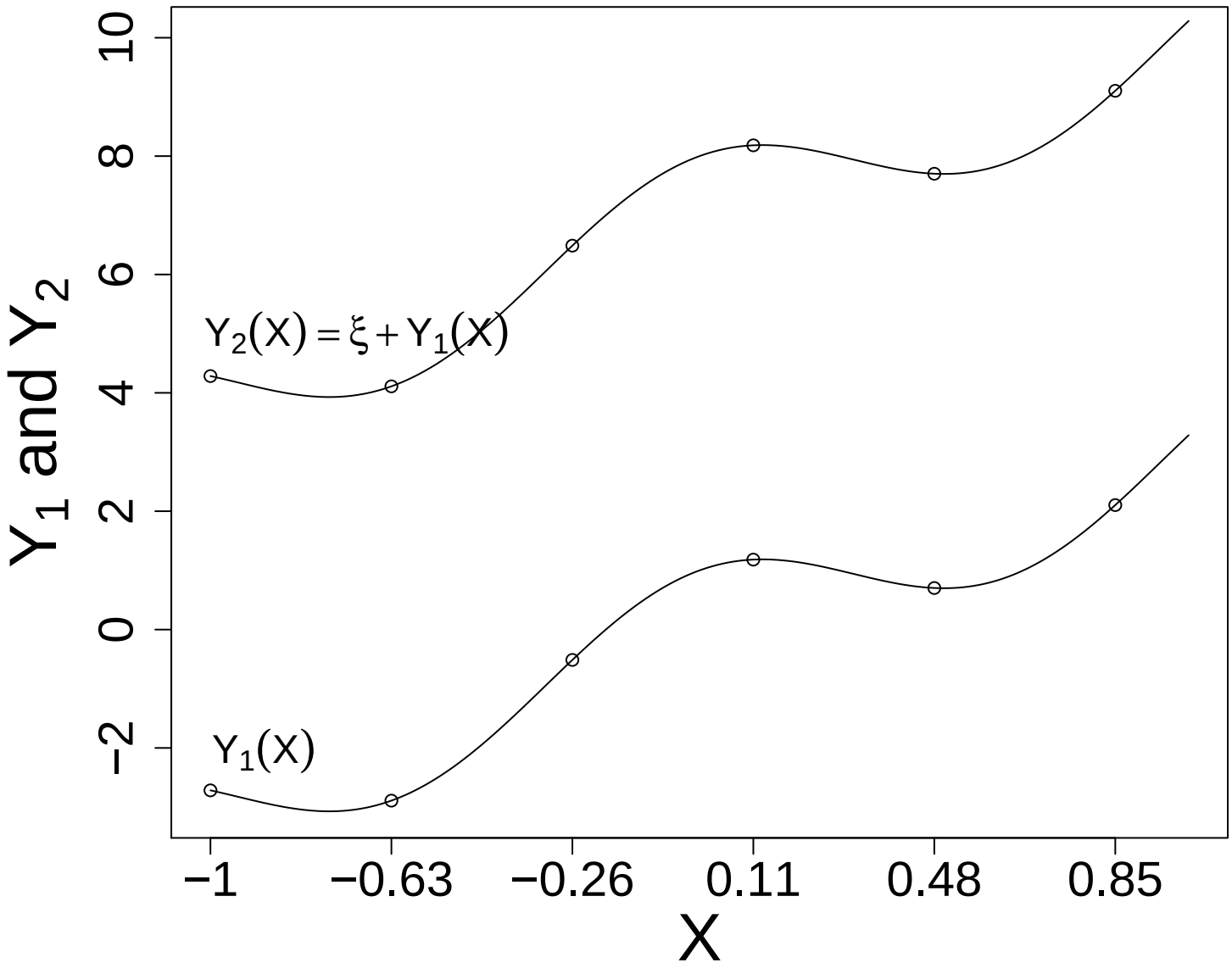
$$\sigma^{*2}(\mathbf{z}') = \sigma^2(\mathbf{C}_{z'} - c(\mathbf{z}', \mathbf{z})\mathbf{C}_z^{-1}c(\mathbf{z}, \mathbf{z}')), \tag{5}$$

where $\mathbf{C}_{z'}$ is the correlation matrix whose (k, l) element is $c(\mathbf{z}'_k, \mathbf{z}'_l)$.

Posterior GASP is

$$g^M(\cdot) \mid \mathbf{g}(\mathbf{z}), \boldsymbol{\theta}_g \sim \mathcal{GASP}(\mu^*(\cdot), \sigma^{*2}(\cdot, \cdot)). \tag{6}$$

Motivating example



Motivating example

Independent modeling

$$
\begin{aligned}
f_1^M(\cdot) \mid \mathbf{f_1}(\mathbf{z}), \boldsymbol{\theta}_{f_1} &\sim \mathcal{GASP}(\mu_1^*(\cdot), \sigma_1^{*2}(\cdot,\cdot)), \\
f_2^M(\cdot) \mid \mathbf{f_2}(\mathbf{z}), \boldsymbol{\theta}_{f_2} &\sim \mathcal{GASP}(\mu_2^*(\cdot), \sigma_2^{*2}(\cdot,\cdot)) \,.
\end{aligned} \tag{7}
$$

Multivariate modeling through conditional specification

$$
\begin{aligned}
f_1^M(\cdot) \mid \mathbf{f_1}(\mathbf{z}), \boldsymbol{\theta}_{f_1} &\sim \mathcal{GASP}(\mu_1^*(\cdot), \sigma_1^{*2}(\cdot,\cdot)), \\
f_2^M(\cdot) \mid f_1^M(\cdot) &= f_1^M(\cdot) + \xi \,.
\end{aligned} \tag{8}
$$

Perfectly correlated functions

Proposition. Let $\mathbf{y}_2 = \lambda \mathbf{y}_1 + \xi$, $\eta_1(\cdot)$ and $\eta_2(\cdot)$ have the same form of linear regression with an intercept and the same correlation function form in each process. Then marginal emulators constructed with independent modeling and multivariate, conditionally specified, coincide exactly, if MLEs of parameters are used.

At a new point $\mathbf{x}'$

$$f_1^M(\mathbf{x}') \sim \mathcal{N}(\mu_1^*(\mathbf{x}'), \sigma_1^{*2}(\mathbf{x}'))\,, \tag{9}$$

$$f_2^M(\mathbf{x}') \sim \mathcal{N}(\xi + \lambda\mu_1^*(\mathbf{x}'), \lambda^2\sigma_1^{*2}(\mathbf{x}'))\,. \tag{10}$$

We should not expect improvement from joint modeling for less correlated functions.

Perfectly correlated functions

Proposition. Let $\mathbf{y}_2 = \lambda \mathbf{y}_1 + \xi$, $\eta_1(\cdot)$ and $\eta_2(\cdot)$ have the same form of linear regression with an intercept and the same correlation function form in each process. Then marginal emulators constructed with independent modeling and multivariate, conditionally specified, coincide exactly, if MLEs of parameters are used.

At a new point $\mathbf{x}'$

$$f_1^M(\mathbf{x}') \sim \mathcal{N}(\mu_1^*(\mathbf{x}'), \sigma_1^{*2}(\mathbf{x}'))\,, \tag{9}$$

$$f_2^M(\mathbf{x}') \sim \mathcal{N}(\xi + \lambda\mu_1^*(\mathbf{x}'), \lambda^2\sigma_1^{*2}(\mathbf{x}'))\,. \tag{10}$$

We should not expect improvement from joint modeling for less correlated functions.

Linear model of coregionalization

LMC for model-based geostatistics attempts to model outputs that “covary". *(Schmidt and Gelfand, 2003)*.

If $W_i \overset{\text{ind}}{\sim} \mathcal{GP}(0, c_i(\cdot,\cdot))$, $i = 1, \ldots, p$ with unit variances. $\mathbf{A}$ is a $p \times p$ matrix. Then p outputs of a computer model $\mathbf{Y} = (Y_1, \ldots, Y_p)$ are modeled as

$$\mathbf{Y} = \boldsymbol{\eta}(\cdot) + \mathbf{AW}\,, \tag{11}$$

where $\mathbf{W} = (W_1, \ldots, W_p)$ and $\boldsymbol{\eta}(\cdot) = (\eta_1(\cdot), \ldots, \eta_p(\cdot))$ is a vector of p mean functions.

$\mathbf{AA}^{\mathrm{T}} = \Sigma$ is interpreted as covariance matrix of p outputs.

LMC likelihood

n inputs $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and outputs $\mathbf{y} = (\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(n)})$, where $\mathbf{y}^{(j)} = (y_1^{(j)}, \ldots, y_p^{(j)})$.

$$\mathbf{y} \sim \mathcal{N}\left(\boldsymbol{\mu}, \tilde{\boldsymbol{\Sigma}} = \sum_{j=1}^{p} \mathbf{C}_j \otimes \mathbf{T}_j\right), \tag{12}$$

where

$\boldsymbol{\mu} = (\boldsymbol{\mu}^{(1)}, \ldots, \boldsymbol{\mu}^{(n)})$;
$\boldsymbol{\mu}^{(j)} = (\eta_1(\mathbf{x}_j), \ldots, \eta_p(\mathbf{x}_j)) \; \forall j = 1, \ldots, n$,
$\mathbf{T}_j = \mathbf{A}_{\cdot j}\mathbf{A}_{\cdot j}^{\mathrm{T}}$ and $\mathbf{A}_{\cdot j}$ is the jth column of matrix $\mathbf{A}$,
$\mathbf{C}_j$ is the jth correlation matrix whose (k, l)th element is $c_j(\mathbf{x}_k, \mathbf{x}_l)$.

We investigate theoretical properties of the LMC model.

Crucial computational lemmas

Determinant and inverse of the LMC covariance matrix.

$$\det \tilde{\Sigma} = \det \left(\sum_{j=1}^{p} \mathbf{C}_j \otimes \mathbf{T}_j \right) = (\det \Sigma)^n \prod_{j=1}^{p} \det \mathbf{C}_j \,, \tag{13}$$

$$\tilde{\Sigma}^{-1} = \sum_{j=1}^{p} \mathbf{C}_j^{-1} \otimes \mathbf{S}_j \,, \tag{14}$$

where $\mathbf{S}_j = (\mathbf{A}^{-\mathrm{T}}{}_{\cdot j}) (\mathbf{A}^{-\mathrm{T}}{}_{\cdot j})^{\mathrm{T}}$.

LMC GASP emulator

Joint and conditional distributions.

Theorem. The LMC GASP emulator, conditional on the computer model evaluations $\mathbf{y}$, at any new point $\mathbf{x}'$ is $\mathcal{N}(\boldsymbol{\mu}', \mathbf{R}')$ with

$$\boldsymbol{\mu}' = \boldsymbol{\eta}(\mathbf{x}') + \left(\sum_{j=1}^{p} \mathbf{R}_{j_{x,x'}}^{\mathrm{T}} \left(\mathbf{R}_{j_{x,x}} \right)^{-1} \otimes \mathbf{T}_j \mathbf{S}_j \right) (\mathbf{y} - \boldsymbol{\mu}) \ ,$$

$$\mathbf{R}' = \sum_{j=1}^{p} \left(1 - \mathbf{R}_{j_{x,x'}}^{\mathrm{T}} \left(\mathbf{R}_{j_{x,x}} \right)^{-1} \mathbf{R}_{j_{x,x'}} \right) \mathbf{T}_j \ ,$$

where $\mathbf{R}_{j_{x,x'}}$ is the cross-correlation between outputs at a new input $\mathbf{x}'$ and inputs $\mathbf{x}_{1:n}$ and $\mathbf{R}_{j_{x,x}}$ is the correlation matrix between outputs of inputs $\mathbf{x}_{1:n}$.

Independent model (IND)

LMC model with diagonal $\mathbf{A}$, such that $A_{ij} = 0$ for any $i \neq j$ and $A_{jj} \neq 0$.

Then jth output of the model is independent from any other output, and

$$Y_j \sim \mathcal{GP}(\eta_j(\cdot), A_{jj}^2 c_j(\cdot, \cdot)) \,. \tag{15}$$

Each process Y_j has variance A_{jj}^2.

Separable model

$c_j(\cdot,\cdot) = c_i(\cdot,\cdot) = c(\cdot,\cdot)$ for any $i, j = 1, \ldots, p$.

Then the joint distribution of p output variables obtained from evaluation of a computer model at n input points $\mathbf{x}$ is

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \tilde{\Sigma} = \mathbf{R} \otimes \Sigma)\,, \tag{16}$$

where $\mathbf{R}$ is correlation matrix given by correlation function $c(\cdot,\cdot)$.
Separable model does not depend on matrix $\mathbf{A}$ in the decomposition $\Sigma = \mathbf{A}\mathbf{A}^{\mathrm{T}}$.

Irrelevance of the LMC model

Proposition. In case of perfect correlation or anticorrelation of functions, LMC GASP emulator marginals (if specified conditionally) coincide exactly with independently constructed emulators of each output (with MLE estimates of parameters).

Theorem. Marginal emulators coincide for independent and separable models (with MLE estimates of Σ), if correlation function $c(\cdot, \cdot)$ with its parameters is given.

Irrelevance of the LMC model (cont.)

Theorem. Let $\mathbf{A}$ be symmetric and let the correlation functions $c_j(\cdot,\cdot)$ be fully specified for each $j = 1, \ldots, p$ *(Paulo et al., 2012))*. Then the sum of predictive variances, at any new input $\mathbf{x}'$, does not depend on any features of the matrix $\mathbf{A}$ other than the marginal prior variances.

$$\sum \mathrm{diag}(\mathrm{V}y^M(\mathbf{x}') \mid \mathbf{y}) = K\,, \tag{17}$$

where $K = K(\sigma_1^2, \ldots, \sigma_p^2, \mathbf{x}')$.

Theorem. Cholesky decomposition of $\boldsymbol{\Sigma}$ *(Schmidt and Gelfand, 2003)*. If processes $W_1, \ldots, W_p$ are given, then lower triangular $\mathbf{A}$ keeps the first component marginal emulator the same as the independent model (with MLE estimates of $\mathbf{A}$). Upper triangular $\mathbf{A}$ — the last component.

Irrelevance of the LMC model

These theorems show analytically that LMC model is not advantageous over IND.

In particular, in case of perfect correlation of outputs, no benefit of LMC.

Generalization to other cases is complicated, because of the need for estimation of parameters of the LMC model.

Correlation between outputs behaves nonintuitively.

Correlation

Correlation between kth and lth outputs Y_k and Y_l is

$$\rho_{kl} = \frac{\sum_{i=1}^{p} A_{ki} A_{li}}{\sqrt{(\sum_{i=1}^{p} A_{ki}^2)(\sum_{i=1}^{p} A_{li}^2)}} .$$

where A_{ij} is the ith row, jth column element of matrix $\mathbf{A}$.

Proposition. Zero-correlation between outputs $\rho_{kl} = 0$ does not imply independence of the outputs.

Exceptions: Zero-correlation implies independence only in case of a separable model or in case of bivariate output together with symmetric $\mathbf{A}$.

Simulation study

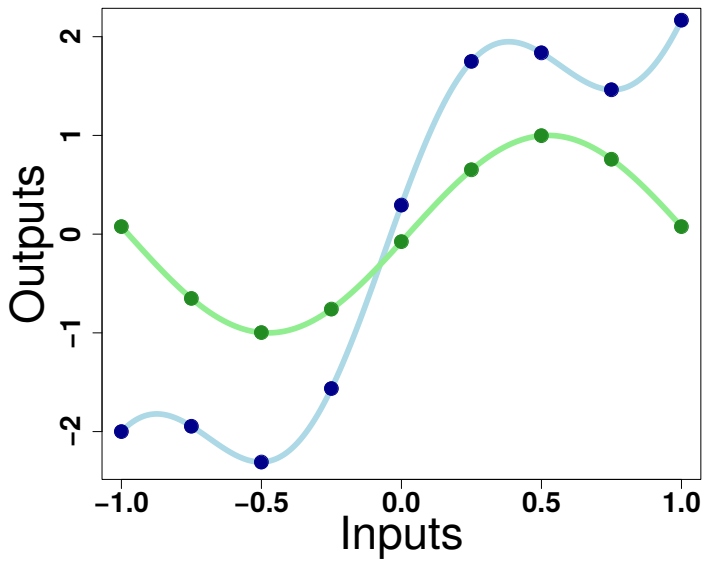
Two test functions are two outputs of a computer model.

$$f(x) = 3x + \cos(5(x + \kappa)), \tag{18}$$
$$g(x) = \sin(\pi(x + \kappa)), x \in [-1, 1]. \tag{19}$$

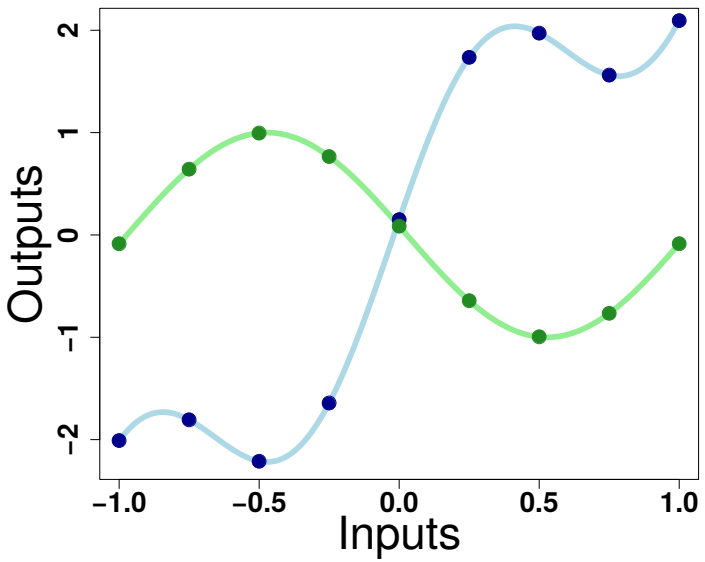
Each value of κ corresponds to one simulation study.

211 equidistant points of $\kappa \in [-\frac{3\pi}{2}, \frac{\pi}{2}]$.

Example of two outputs

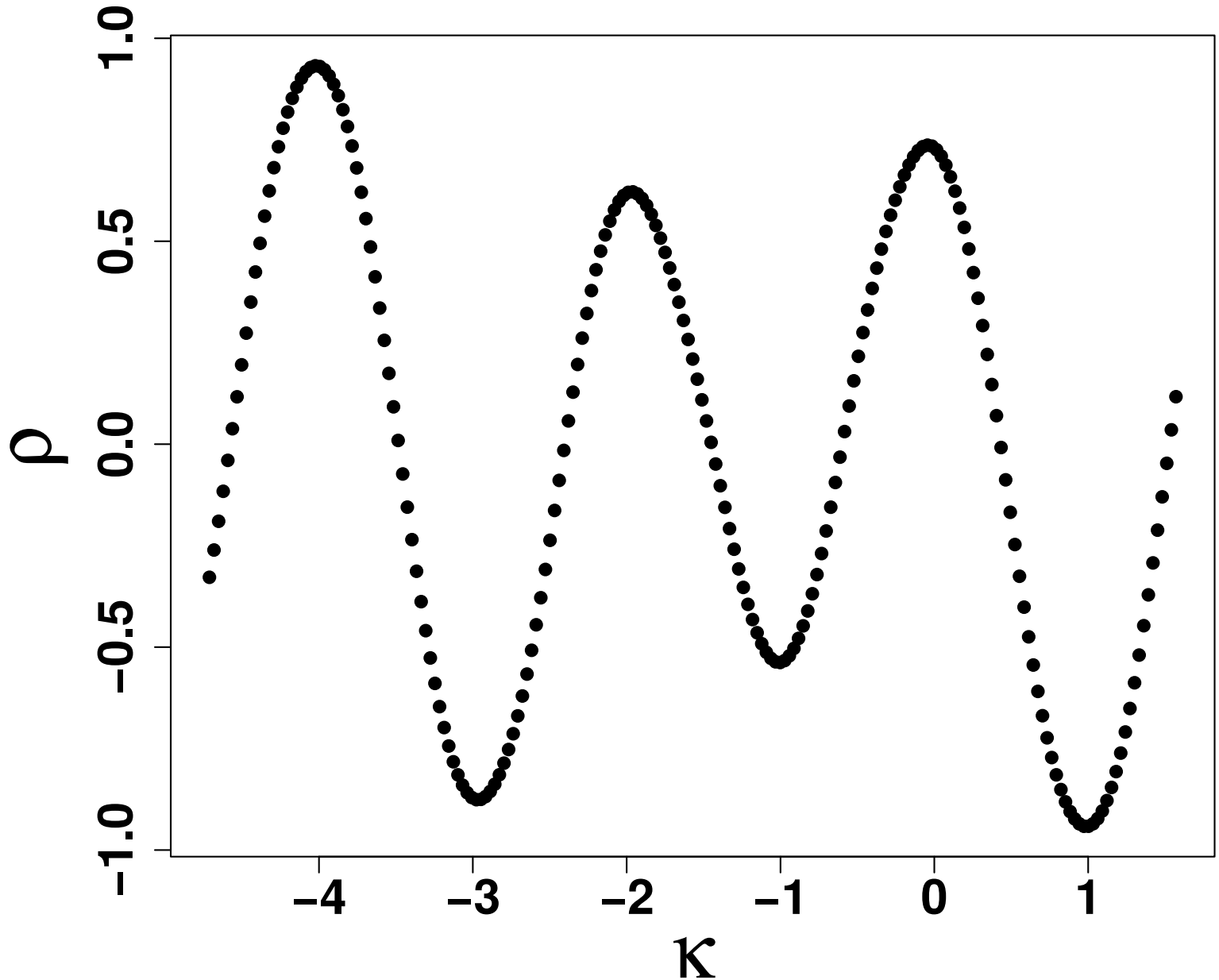


$$\kappa_{24} = \frac{23\pi}{105} - \frac{3\pi}{2}$$



$$\kappa_{191} = \frac{190\pi}{105} - \frac{3\pi}{2}$$

Correlation between functions



Correlation between functions

Definition. Correlation ρ, between two smooth functions $f_1(x)$ and $f_2(x)$ on some input space $\mathcal{X}$, is defined as

$$\rho = \frac{\int_{\mathcal{X}} [(f_1(\mathbf{x}) - \int_{\mathcal{X}} f_1(\mathbf{z})\,\mathrm{d}\mathbf{z})(f_2(\mathbf{x}) - \int_{\mathcal{X}} f_2(\mathbf{z})\,\mathrm{d}\mathbf{z})]\,\mathrm{d}\mathbf{x}}{\sqrt{\int_{\mathcal{X}} (f_1(\mathbf{x}) - \int_{\mathcal{X}} f_1(\mathbf{z})\,\mathrm{d}\mathbf{z})^2\,\mathrm{d}\mathbf{x} \int_{\mathcal{X}} (f_2(\mathbf{x}) - \int_{\mathcal{X}} f_2(\mathbf{z})\,\mathrm{d}\mathbf{z})^2\,\mathrm{d}\mathbf{x}}}\,.$$

LMC estimated correlation

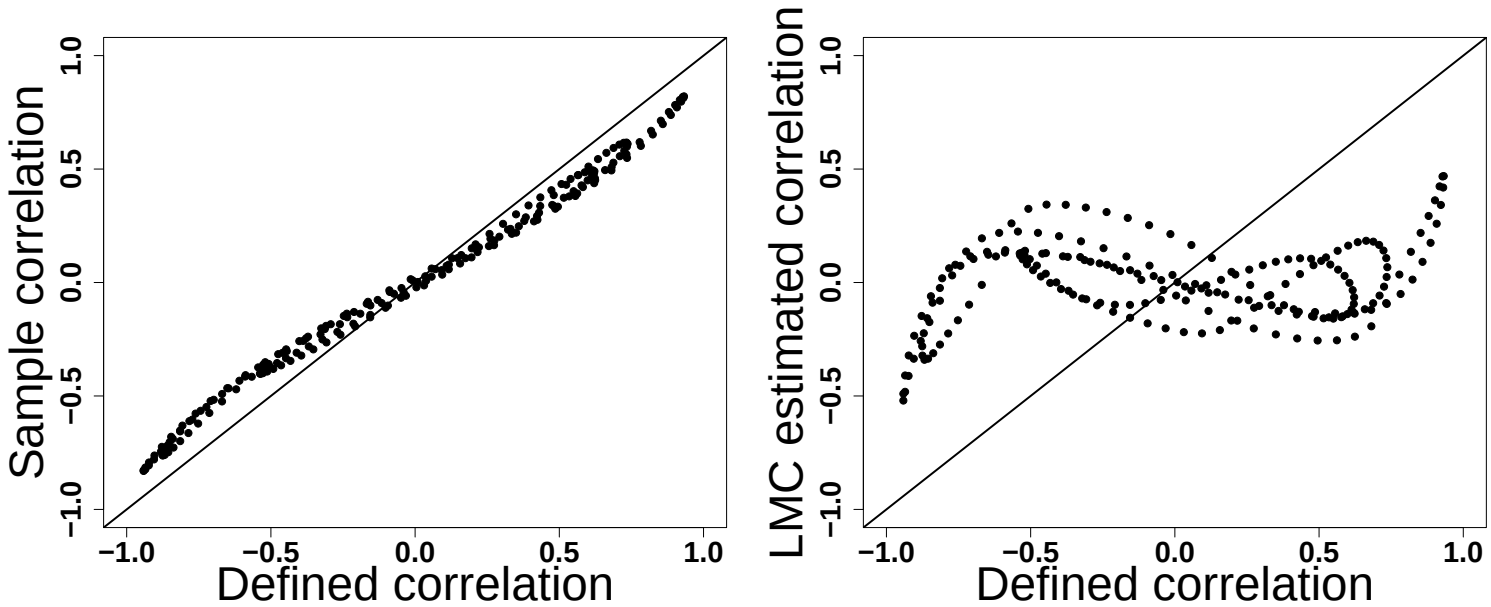


Figure: Left panel: true correlation between functions f and g vs. observed sample correlation for all simulation studies. Right panel: true correlation vs. the estimate from the LMC model for all simulation studies. The solid line on both panels is $y = x$.

Conclusion

Multivariate modeling of a computer model multivariate output does not lead to better emulation results than individual modeling of each output.

Application

Bent, volcano ash plume model, has four-dimensional input and five-dimensional output.

Puff, volcano ash transport and dispersal model, takes Bent output as input and produces scalar output.

Coupling of Bent and Puff through linking their GASP emulators. *(Kyzyurova et al., 2018)*.

Linked emulators of Puff with

independent emulators of each of Bent output/
multivariate emulator of Bent

coincide almost exactly.

Thank you

Questions

Previous research on LMC model

Irrelevance of multivariate modeling has been observed before.

1. LMC (including separable) model on case studies in a working paper of *(Fricker et al., 2013)*.
2. 7 LMC simulation examples and a case study with LMC model in *(Kleijnen and Mehdad, 2014)*.
3. Separable model gave the same predictive mean as independent modeling *(Gu and Berger, 2016)*.
4. Similar accuracy of independent and multivariate emulators reported in *(Parussini et al., 2017)*.

In this work theoretical properties of the LMC model and its emulator have been investigated.

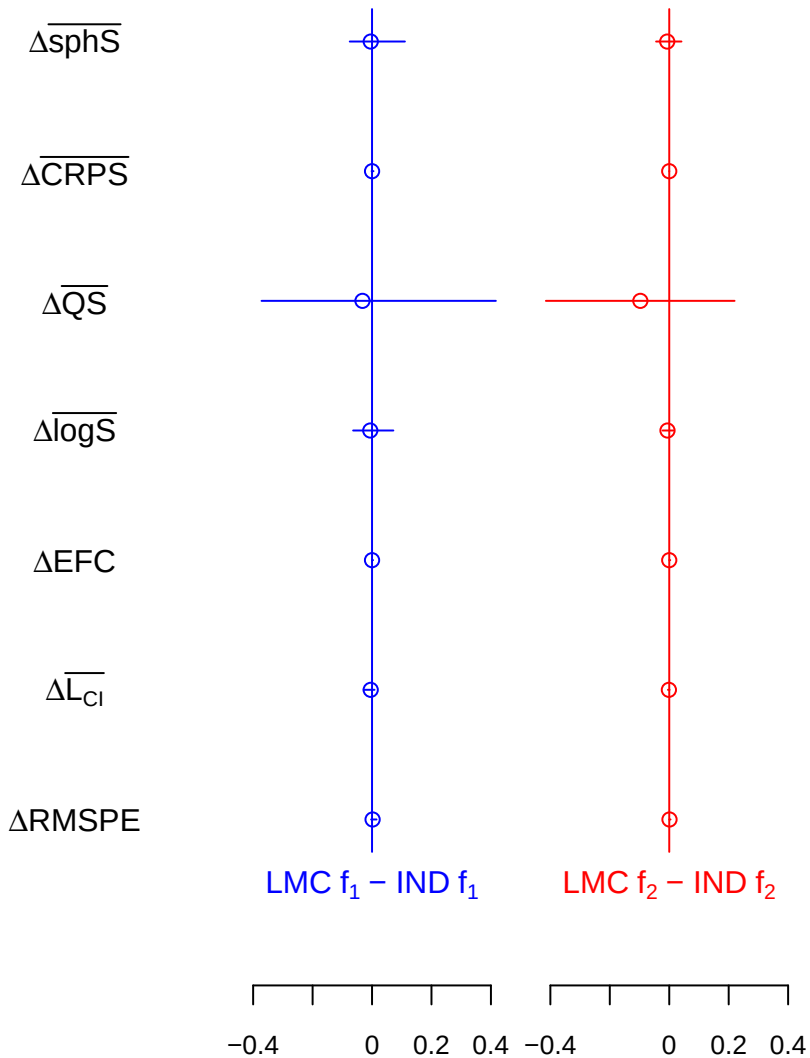
Simulation study. Simulation results

Table: Absolute differences in predictive checks and scores.

	LMC f - IND f	LMC g - IND g
$\Delta\mathrm{RMSPE}$	0.0019 (-0.0031,0.0152)	8e-04 (-1e-04,0.0044)
$\Delta\overline{\mathrm{L_{CI}}}$	-0.0045 (-0.0283,0.0084)	-0.0012 (-0.0044,0.0012)
$\Delta\mathrm{EFC}$	0 (0,0)	0 (0,0)
$\Delta\overline{\mathrm{logS}}$	-0.0059 (-0.0636,0.0717)	-0.0059 (-0.0225,0.0145)
$\Delta\overline{\mathrm{QS}}$	-0.0322 (-0.3718,0.4166)	-0.0972 (-0.415,0.2198)
$\Delta\overline{\mathrm{CRPS}}$	1e-04 (-0.0015,0.0049)	0 (-1e-04,4e-04)
$\Delta\overline{\mathrm{sphS}}$	-0.0043 (-0.075,0.1106)	-0.0068 (-0.0443,0.0411)

Each cell provides the average value of a score or a predictive check (with minimum and maximum values in parentheses) across all the simulation studies with various values of κ.

Pictorial representation



Estimation of $\mathbf{A}$ and Σ

Matrix $\mathbf{A}$ is not estimable from the data only.

Theorem MLE $\hat{\mathbf{A}}$ satisfies the following system, if correlation matrices of processes $W_1, \ldots, W_p$ are given,

$$\frac{1}{2n} \sum_{ij} \hat{\mathbf{A}}^{-\mathrm{T}} \tilde{\mathbf{D}}_{ij} \hat{\mathbf{A}}^{-1} (\mathbf{y}^{(j)} \mathbf{y}^{(i)\mathrm{T}} + \mathbf{y}^{(i)} \mathbf{y}^{(j)\mathrm{T}}) = \mathbf{I}_{p \times p}. \qquad (20)$$

Corollary There are 2^p MLEs.

$\mathbf{A}$ has p^2 number of parameters. This is computationally challenging for the estimation of parameters of the LMC model.

One proposal on estimation of Σ

Cholesky decomposition of Σ with lower triangular $\mathbf{A}$ *(Schmidt and Gelfand, 2003)*.

The likelihood depends on the order of data $\mathbf{y}$ associated with processes. *(Fricker et al., 2013)*

$$\begin{aligned} Y_1 &= \eta_1(\cdot) + A_{11} W_1, \\ Y_2 &= \eta_2(\cdot) + A_{21} W_1 + A_{22} W_2, \\ &\vdots \\ Y_p &= \eta_p(\cdot) + A_{p1} W_1 + A_{p2} W_2 + \ldots + A_{pp} W_p. \end{aligned} \tag{21}$$

Cholesky decomposition of Σ

Proposition If processes $W_1, \ldots, W_p$ are given, then depending on the choice of lower or upper triangular matrix A, one obtains analytically

- different likelihoods,
- different estimates of Σ and ρ,
- different predictive distributions at a new input to a computer model.

Lower triangular **A** keeps the first component marginal emulator as the same as the independent model. Upper triangular **A** keeps the last component the same as the independent model.

Another proposal on estimation of Σ

Symmetric $\mathbf{A}$ *(Fricker et al., 2013)*.

Proposition LMC likelihood does not depend on the order of data $\mathbf{y}$, if processes $W_1, \ldots, W_p$ are not ordered or fixed. If processes $W_1, \ldots, W_p$ are either ordered or fixed, then the resulting likelihood depends on the order of data $\mathbf{y}$.

Computationally challenging to estimate $p(p+1)/2$ parameters of $\mathbf{A}$ simultaneously with p sets of parameters in the correlation functions of processes $W_1, \ldots, W_p$. Issues are reported with $p = 3$ and $p = 6$ *(Fricker et al., 2013)*.

No compelling argument to restrict $\mathbf{A}$ to be symmetric or any other form.

Maria J Bayarri, James O Berger, Rui Paulo, Jerry Sacks, John A Cafeo, James Cavendish, Chin-Hsu Lin, and Jian Tu. A framework for validation of computer models. *Technometrics*, 49(2):138–154, 2007.

Marcus Bursik, Matthew Jones, Simon Carn, Ken Dean, Abani Patra, Michael Pavolonis, E Bruce Pitman, Tarunraj Singh, Puneet Singla, Peter Webley, et al. Estimation and propagation of volcanic source parameter uncertainty in an ash transport and dispersal model: application to the eyjafjallajokull plume of 14–16 april 2010. *Bulletin of volcanology*, 74(10):2321–2338, 2012.

Thomas E Fricker, Jeremy E Oakley, and Nathan M Urban. Multivariate gaussian process emulators with nonseparable covariance structures. *Technometrics*, 55(1):47–56, 2013.

Robert B Gramacy and Herbert KH Lee. Cases for the nugget in modeling computer experiments. *Statistics and Computing*, 22(3):713–722, 2012.

Mengyang Gu and James O Berger. Parallel partial gaussian process emulation for computer models with massive output. *The Annals of Applied Statistics*, 10(3):1317–1347, 2016.

Mengyang Gu, Jesus Palomo, and James Berger. *RobustGaSP: Robust Gaussian Stochastic Process Emulation*, 2016. URL `https://CRAN.R-project.org/package=RobustGaSP`. R package version 0.5.

Jack PC Kleijnen and Ehsan Mehdad. Multivariate versus univariate kriging metamodels for multi-response simulation models. *European Journal of Operational Research*, 236(2):573–582, 2014.

Ksenia N Kyzyurova, James O Berger, and Robert L Wolpert. Coupling computer models through linking their statistical emulators. *submitted*, 2018.

L Parussini, D Venturi, P Perdikaris, and GE Karniadakis. Multi-fidelity gaussian process regression for prediction of random fields. *Journal of Computational Physics*, 336:36–50, 2017.

Rui Paulo, Gonzalo García-Donato, and Jesús Palomo. Calibration of computer models with multivariate output. *Computational Statistics & Data Analysis*, 56(12):3959–3974, 2012.

Jerome Sacks, William J Welch, Toby J Mitchell, and Henry P Wynn. Design and analysis of computer experiments. *Statistical science*, pages 409–423, 1989.

Alexandra M Schmidt and Alan E Gelfand. A bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research: Atmospheres*, 108(D24), 2003.