On linear model of coregionalization

Ksenia N. Kyzyurova
kseniak.ucoz.net, ksenia.kyzyurova@gmail.com

December 14, 2019

Abstract

Linear model of coregionalization (LMC), while historically proposed to model several variables in spatial statistics, attempts to account for correlation among computer model outputs in the construction of Gaussian process emulators of the outputs with the goal of achieving a more accurate multivariate emulator. In this work properties of LMC are revealed. Theoretical evidence is found that multivariate emulator does not lead to "better" (that is, more accurate, precise or less uncertain) emulation results compared to independent modeling of each component of the output. Simulation studies support the conclusion.

In spatial statistics, therefore, multivariate modeling of multivariable data on a grid is equivalent to independent modeling of each variable.

1 Introduction

This work builds on the example provided in Chapter 2 in the monograph Kyzyurova (2019*a*) (initially presented in Kyzyurova (2017)) which has shown that marginal emulators of two computer model outputs which are perfectly correlated coincide exactly with those obtained in the independent case. The example suggests that for models of multivariate output the conclusion of no benefit holds as well. In this work the closed-form solution to the LMC emulator of a multivariate computer model output is given. The properties of the emulator are analyzed in the main part of the article. Evidence is found to support the conclusion of no benefit of multivariate modeling compared to independent modeling.

2 Linear model of coregionalization

Formal construction of an emulator for a continuous real-valued function $f(\cdot)$ starts by assuming a Gaussian stochastic process (GASP) prior on the function , i.e.

$$f^M(\cdot) \sim \mathcal{GASP}(\mu(\cdot), \sigma^2 c(\cdot,\cdot))\,,$$

where $\mu(\cdot)$ is the mean function, σ^2 is the unknown variance and $c(\cdot,\cdot)$ is the correlation function of the process. Typically $c(\cdot,\cdot)$ is assumed to follow some functional form, depending on the distance between inputs to the functions and also on parameters $\boldsymbol{\delta}$, the (possibly multivariate input) range and smoothness of the correlation function.

The construction of the LMC starts by assuming p independent Gaussian processes $W_i \sim \mathcal{GASP}(0, c_i(\cdot,\cdot))$, $i = 1,\ldots,p$, with unit variances. Let $\mathbf{A}$ be a $p \times p$ matrix. The p outputs of the computer model, $\mathbf{Y} = (Y_1,\ldots,Y_p)$ arranged in a vector, are then modeled as

$$\mathbf{Y} = \boldsymbol{\eta}(\cdot) + \mathbf{AW}\,, \tag{1}$$

where $\mathbf{W} = (W_1, \ldots, W_p)$ is a vector of p zero-mean, unit-variance processes and $\boldsymbol{\eta}(\cdot) = (\eta_1(\cdot), \ldots, \eta_p(\cdot))$ is the vector of the p mean functions. In this model, $\boldsymbol{\Sigma} = \mathbf{A}\mathbf{A}^{\mathrm{T}}$ is interpreted as the output covariance matrix (as opposed to the covariance matrices arising from varying the inputs).

Suppose the computer model has been evaluated at n, possibly multivariate, input points $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$ obtaining the corresponding output points $\mathbf{y}^{(1:n)} = (\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(n)})$, where $\mathbf{y}^{(j)} = (y_1^{(j)}, \ldots, y_p^{(j)})$ is the arranged vector of the p outputs at input $\mathbf{x}_j$. According to the LMC model, the joint distribution of the observed outputs is then

$$\begin{pmatrix} \mathbf{y}^{(1)} \\ \mathbf{y}^{(2)} \\ \vdots \\ \mathbf{y}^{(n)} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \\ \vdots \\ \boldsymbol{\mu}^{(n)} \end{pmatrix}, \tilde{\boldsymbol{\Sigma}} = \begin{pmatrix} \mathbf{A}\mathbf{A}^{\mathrm{T}} & \mathbf{A}D_{12}\mathbf{A}^{\mathrm{T}} & \cdots & \mathbf{A}D_{1n}\mathbf{A}^{\mathrm{T}} \\ \mathbf{A}D_{12}\mathbf{A}^{\mathrm{T}} & \mathbf{A}\mathbf{A}^{\mathrm{T}} & \cdots & \mathbf{A}D_{2n}\mathbf{A}^{\mathrm{T}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}D_{1n}\mathbf{A}^{\mathrm{T}} & \mathbf{A}D_{2n}\mathbf{A}^{\mathrm{T}} & \cdots & \mathbf{A}\mathbf{A}^{\mathrm{T}} \end{pmatrix} \right),$$

where the diagonal matrices $D_{ij} = \mathrm{diag}(c_1(\mathbf{x}_i, \mathbf{x}_j), c_2(\mathbf{x}_i, \mathbf{x}_j), \cdots, c_p(\mathbf{x}_i, \mathbf{x}_j))$. Due to the symmetry of correlation functions, $D_{ij} = D_{ji}$. The mean vectors are $\boldsymbol{\mu}^{(j)} = (\eta_1(\mathbf{x}_j), \ldots, \eta_p(\mathbf{x}_j))$, with $\eta_i(\cdot)$ being the mean function corresponding to the ith output variable, $i = 1, \ldots, p$.

The likelihood of the LMC model can be written as

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

where $\boldsymbol{\mu}^{(1:n)} = (\boldsymbol{\mu}^{(1)}, \ldots, \boldsymbol{\mu}^{(n)})$, $\otimes$ denotes the Kronecker product, $\mathbf{T}_j = \mathbf{A}_{\cdot j}\mathbf{A}_{\cdot j}^{\mathrm{T}}$ with $\mathbf{A}_{\cdot j}$ being the jth column of matrix $\mathbf{A}$, and $\mathbf{C}_j$ is the jth correlation matrix whose (k, l)th element is given by the correlation function $c_j(\mathbf{x}_k, \mathbf{x}_l)$ evaluated at inputs $\mathbf{x}_k$ and $\mathbf{x}_l$.

2.1 Crucial computational lemmas

The following two crucial computational lemmas are useful for the derivation of properties of the LMC model, including the explicit formulae for the LMC GASP emulator.

Lemma.

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

Proof. This can be seen using, for instance, results on block matrices (Powell 2011). ∎

Lemma.

$$\tilde{\boldsymbol{\Sigma}}^{-1} = \sum_{j=1}^{p} \mathbf{C}_j^{-1} \otimes \mathbf{S}_j .$$

Equivalently,

$$\tilde{\boldsymbol{\Sigma}}^{-1} = \begin{pmatrix} \mathbf{A}^{-\mathrm{T}}\tilde{D}_{11}\mathbf{A}^{-1} & \mathbf{A}^{-\mathrm{T}}\tilde{D}_{12}\mathbf{A}^{-1} & \cdots & \mathbf{A}^{-\mathrm{T}}\tilde{D}_{1n}\mathbf{A}^{-1} \\ \mathbf{A}^{-\mathrm{T}}\tilde{D}_{12}\mathbf{A}^{-1} & \mathbf{A}^{-\mathrm{T}}\tilde{D}_{22}\mathbf{A}^{-1} & \cdots & \mathbf{A}^{-\mathrm{T}}\tilde{D}_{2n}\mathbf{A}^{-1} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}^{-\mathrm{T}}\tilde{D}_{1n}\mathbf{A}^{-1} & \mathbf{A}^{-\mathrm{T}}\tilde{D}_{2n}\mathbf{A}^{-1} & \cdots & \mathbf{A}^{-\mathrm{T}}\tilde{D}_{nn}\mathbf{A}^{-1} \end{pmatrix},$$

where $\tilde{D}_{ij} = \mathrm{diag}({\mathbf{C}_1^{-1}}_{ij}, {\mathbf{C}_2^{-1}}_{ij}, \cdots, {\mathbf{C}_p^{-1}}_{ij})$ is the diagonal matrix of (i, j)th elements of inverses of each of the correlation matrices $\mathbf{C}_k$, $k = 1, \ldots, p$, and $\mathbf{S}_j = (\mathbf{A}^{-\mathrm{T}}{}_{\cdot j})\,(\mathbf{A}^{-\mathrm{T}}{}_{\cdot j})^{\mathrm{T}}$.

Proof. Relies on the next lemma and standard results for Kronecker products. ∎

Lemma. Defining $\boldsymbol{S}_j = (\mathbf{A}^{-\mathrm{T}}{}_{\cdot j})\,(\mathbf{A}^{-\mathrm{T}}{}_{\cdot j})^{\mathrm{T}}$, where $(\mathbf{A}^{-\mathrm{T}}{}_{\cdot j})$ is the jth column of $\mathbf{A}^{-\mathrm{T}}$, the following statements are true.

1. $\sum_{j=1}^{p} \mathbf{T}_j = \boldsymbol{\Sigma}$,

2. $\mathbf{T}_j = \mathbf{T}_j^{\mathrm{T}}$,

3. $\mathbf{T}_j \mathbf{S}_i = \mathbf{0}\ \forall\ i \neq j$,

4. $\sum_{j=1}^{p} \mathbf{T}_j \mathbf{S}_j = \mathbf{I}_{p \times p}$,

5. $\mathbf{T}_j \mathbf{S}_j \mathbf{T}_j = \mathbf{T}_j$,

where $\mathbf{I}_{p\times p}$ is the $p \times p$-dimensional identity matrix.

2.2 LMC GASP emulator

Using the explicit expression for the inverse of the covariance matrix $\boldsymbol{\Sigma}$, given above, analytical expressions for the joint and conditional distributions of the LMC model are provided. Expressions for the LMC GASP emulator conclude the section.

Lemma. Partition $\mathbf{y}^{(1:n)}$ into two vectors $\mathbf{y}^{(1:n)} = \left(\mathbf{y}^{(1:m)}, \mathbf{y}^{((m+1):n)}\right)$, where $\mathbf{y}^{(1:m)} = (\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(m)})$ and $\mathbf{y}^{((m+1):n)} = (\mathbf{y}^{(m+1)}, \ldots, \mathbf{y}^{(n)})$. Then the joint LMC distribution is

$$\begin{pmatrix} \mathbf{y}^{(1:m)} \\ \mathbf{y}^{((m+1):n)} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu}^{(1:m)} \\ \boldsymbol{\mu}^{((m+1):n)} \end{pmatrix}, \begin{pmatrix} \sum_{j=1}^{p} \mathbf{R}_{j_{11}} \otimes \mathbf{T}_j & \sum_{j=1}^{p} \mathbf{R}_{j_{12}} \otimes \mathbf{T}_j \\ \sum_{j=1}^{p} \mathbf{R}_{j_{12}}^{\mathrm{T}} \otimes \mathbf{T}_j & \sum_{j=1}^{p} \mathbf{R}_{j_{22}} \otimes \mathbf{T}_j \end{pmatrix} \right) ,$$

where $\mathbf{R}_{j_{12}}$ is the cross-correlation between outputs at inputs $\mathbf{x}_{1:m}$ and $\mathbf{x}_{(m+1):n}$, $\mathbf{R}_{j_{11}}$ is the correlation matrix of inputs $\mathbf{x}_{1:m}$, and $\mathbf{R}_{j_{22}}$ is the correlation matrix between outputs at inputs $\mathbf{x}_{(m+1):n}$. The (k, l)th element in each of these matrices is the correlation function $c_j(\mathbf{x}_k, \mathbf{x}_l)$ evaluated at $\mathbf{x}_k$ and $\mathbf{x}_l$.

Proof. Relies on the definition of the LMC model, i.e. the LMC likelihood. ■

Corollary. The conditional distribution of $\mathbf{y}^{((m+1):n)} \mid \mathbf{y}^{(1:m)} \sim \mathcal{N}(\bar{\boldsymbol{\mu}}, \bar{\mathbf{R}})$, with

$$\bar{\boldsymbol{\mu}} = \boldsymbol{\mu}^{((m+1):n)} + \left(\sum_{j=1}^{p} \mathbf{R}_{j_{12}}^{\mathrm{T}} \left(\mathbf{R}_{j_{11}}\right)^{-1} \otimes \mathbf{T}_j \mathbf{S}_j \right) \left(\mathbf{y}^{(1:m)} - \boldsymbol{\mu}^{(1:m)} \right) ,$$

$$\bar{\mathbf{R}} = \sum_{j=1}^{p} \left(\mathbf{R}_{j_{22}} - \mathbf{R}_{j_{12}}^{\mathrm{T}} \left(\mathbf{R}_{j_{11}}\right)^{-1} \mathbf{R}_{j_{12}} \right) \otimes \mathbf{T}_j .$$

Proof. Relies on the lemma about the inverse of the LMC covariance matrix, its proof and standard results for Kronecker products. ■

Theorem. The LMC Gaussian stochastic process emulator, conditional on the computer model evaluations $\mathbf{y}^{(1:n)}$ and at any new point $\mathbf{x}'$, is $\mathbf{y}'^{M} = (y'^{M}_1, \ldots, y'^{M}_p) = y^M(\mathbf{x}') \sim \mathcal{N}(\boldsymbol{\mu}', \mathbf{R}')$, where

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

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

with $\mathbf{R}_{j_{x,x'}}$ being the cross-correlation between a new input $\mathbf{x}'$ and the vector of inputs $\mathbf{x}_{1:n}$ and $\mathbf{R}_{j_{x,x}}$ being the correlation matrix of inputs $\mathbf{x}_{1:n}$.

Proof. Relies on the previous lemma and its corollary. ■

These formulae do not suggest that the LMC emulator is a better emulator than independent emulators of each output of a computer model. Nothing in these formulae would indicate that a constructed LMC emulator is a more accurate emulator. To quantify this observation, in the next section special cases of the LMC are considered. The irrelevance of the modeling with the LMC emulator analytically in certain cases is demonstrated.

3 Irrelevance of the LMC joint modeling

3.1 Special cases of the LMC model

LMC model may be simplified to several commonly used models, namely, an independent model and a separable model.

3.1.1 Independent model

Suppose that in the LMC model (1), $\mathbf{A}$ is chosen to be diagonal, with diagonal elements $a_{jj} \neq 0$. Then the jth output of the model is independent from any other output,

$$Y_j \sim \mathcal{GASP}(\eta_j(\cdot), a_{jj}^2 c_j(\cdot,\cdot)) ,$$

and $\mathbf{\Sigma}$ is the diagonal matrix with diagonal elements $\Sigma_{jj} = a_{jj}^2$.

Lemma. If the correlation functions $c_j(\cdot,\cdot)$ are specified, the MLE for $\mathbf{\Sigma}$ is diagonal with diagonal elements

$$\hat{\Sigma}_{jj} = \frac{\mathbf{y_j}^{\mathrm{T}} \mathbf{C}_j^{-1} \mathbf{y_j}}{n} ,$$

where $\mathbf{y_j} = (y_j^{(1)}, \ldots, y_j^{(n)})$. This lemma is used below to compare predictive distribution arising from an independent model (called IND) to the one arising from another special case of the LMC, separable LMC model.

3.1.2 Separable LMC model

A separable LMC model appears if all correlation functions in the model are assumed to be the same, that is $c_j(\cdot,\cdot) = c_i(\cdot,\cdot) = c(\cdot,\cdot)$ for each $i, j = 1, \ldots, p$. Then the joint distribution of the p output variables, obtained from evaluation of the computer model at n input points $\mathbf{x}$, is

$$\mathbf{y}^{(1:n)} \sim \mathcal{N}(\boldsymbol{\mu}^{(1:n)}, \tilde{\mathbf{\Sigma}} = \mathbf{R} \otimes \mathbf{\Sigma}) ,$$

where $\mathbf{R}$ is the common correlation matrix arising from the correlation function $c(\cdot,\cdot)$ and inputs $\mathbf{x}$. $\mathbf{\Sigma}$ represents the covariances between the p outputs at any input. Likelihood of the separable model depends only on $\mathbf{\Sigma}$ and not on the matrix $\mathbf{A}$.

The joint distribution of $\mathbf{y}^{(1:n)} = (\mathbf{y}^{(1:m)}, \mathbf{y}^{((m+1):n)})$ of the separable model is

$$\begin{pmatrix} \mathbf{y}^{(1:m)} \\ \mathbf{y}^{((m+1):n)} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \boldsymbol{\mu}^{(1:m)} \\ \boldsymbol{\mu}^{((m+1):n)} \end{pmatrix}, \begin{pmatrix} \mathbf{R}_{11} \otimes \mathbf{\Sigma} & \mathbf{R}_{12} \otimes \mathbf{\Sigma} \\ \mathbf{R}_{12}^{\mathrm{T}} \otimes \mathbf{\Sigma} & \mathbf{R}_{22} \otimes \mathbf{\Sigma} \end{pmatrix} \right) ,$$

where $\mathbf{R}_{12}$ is the cross-covariance between $\mathbf{y}^{(1:m)}$ and $\mathbf{y}^{((m+1):n)}$, $\mathbf{R}_{11}$ is the covariance of $\mathbf{y}^{(1:m)}$, and $\mathbf{R}_{22}$ is the covariance of $\mathbf{y}^{((m+1):n)}$.

Using corollary from the section on LMC GASP emulator, the conditional distribution of $\mathbf{y}^{((m+1):n)} \mid \mathbf{y}^{(1:m)} \sim \mathcal{N}(\bar{\boldsymbol{\mu}}, \bar{\mathbf{R}})$, with $\bar{\boldsymbol{\mu}}$ and $\bar{\mathbf{R}}$ given in the following lemma.

Lemma.

$$\begin{aligned} \bar{\boldsymbol{\mu}} &= \boldsymbol{\mu}^{((m+1):n)} + \left(\mathbf{R}_{12}^{\mathrm{T}} \left(\mathbf{R}_{11}\right)^{-1} \otimes \mathbf{I}_{p\times p}\right) \left(\mathbf{y}^{(1:m)} - \boldsymbol{\mu}^{(1:m)}\right) , \\ \bar{\mathbf{R}} &= \left(\mathbf{R}_{22} - \mathbf{R}_{12}^{\mathrm{T}} \left(\mathbf{R}_{11}\right)^{-1} \mathbf{R}_{12}\right) \otimes \mathbf{\Sigma} . \end{aligned}$$

The following properties of the predictive separable LMC model stand out. Namely, the predictive mean $\mathrm{E}\mathbf{y}^{((m+1):n)} \mid \mathbf{y}^{(1:m)}$ of the separable model does not depend on the matrix $\mathbf{\Sigma}$. While the predictive variance $\mathrm{V}\mathbf{y}^{((m+1):n)} \mid \mathbf{y}^{(1:m)}$ does depend on the covariance matrix $\mathbf{\Sigma}$.

Corollary. The marginal means of the separable LMC model coincide exactly with those given by the IND emulators for all p outputs of the LMC model if correlation function on the inputs is specified (fully parametrized) and is the same in both models.

Next theorem shows that the marginal variances in the separable LMC model also coincide exactly with those given by the IND models if the MLE estimate of $\boldsymbol{\Sigma}$ is used.

Theorem. If the correlation function $c(\cdot,\cdot)$ is specified, then the separable model marginal predictive variances $\mathrm{diag}(\mathrm{V}\mathbf{y}^{((m+1):n)} \mid \mathbf{y}^{(1:m)})$ with MLE estimate of $\boldsymbol{\Sigma}$ plugged in equal to the MLEs of corresponding variances which could be obtained from independent modeling of p outputs, if the same correlation function is used for independent modeling of each output.

Proof. Since

$$\frac{d \log \det \tilde{\boldsymbol{\Sigma}}^{-1/2}}{d\boldsymbol{\Sigma}^{-1}} = \frac{n}{2}\boldsymbol{\Sigma}^{\mathrm{T}} = \frac{n}{2}\boldsymbol{\Sigma}$$

and

$$\mathrm{tr}(\mathbf{y}^{(1:n)^{\mathrm{T}}}\mathbf{C}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\mathbf{y}^{(1:n)}) = \sum_{ij=1}^{n} \mathbf{C}_{ij}^{-1}\mathbf{y}_j\mathbf{y}_i^{\mathrm{T}}\,,$$

in the separable case of a common correlation function, the MLE for $\boldsymbol{\Sigma}$ is the matrix whose (k,l)th element is

$$\hat{\Sigma}_{kl} = \frac{\mathbf{y}_k^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{y}_l}{n}\,.$$

If the correlation function is fully specified, $\mathrm{diag}(\hat{\boldsymbol{\Sigma}})$ coincides exactly with that arising from the independent model. ■

Theorem. Let $\mathbf{A}$ be symmetric and let correlation functions $c_j(\cdot,\cdot)$ be specified for each $j = 1,\ldots,p$. Let marginal variances of the p processes be the same, i.e. we fix $\sum_{j=1}^{p} A_{ij}^2 = \sigma_i^2$ for each $i = 1,\ldots,p$. Then the sum of marginal predictive variances of the LMC GASP emulator at any new input $\mathbf{x}'$ stays constant and does not depend on matrix $\mathbf{A}$ other than through $\{\sigma_i^2\}_{i=1,\ldots,p}$, i.e. $\sum \mathrm{diag}(\mathrm{V}y^M(\mathbf{x}') \mid \mathbf{y}^{(1:n)}) = \mathit{const}$.

Proof. Marginal predictive variance for the ith component is

$$\sum_{j=1}^{p}(1 - \boldsymbol{R}_{j12}^{\mathrm{T}}\boldsymbol{R}_{j22}^{-1}\boldsymbol{R}_{j12})\boldsymbol{A}_{ij}^2\,.$$

It can be shown that the sum of these variances equals

$$\sum \mathrm{diag}(\mathrm{V}y^M(x') \mid \mathbf{y}^{(1:n)}) = \sum_{j=1}^{p}\sigma_i^2(1 - \mathbf{R}_{j12}\mathbf{R}_{j22}^{-1}\mathbf{R}_{j12})\,. \ ■$$

This means that the amount of variance one of the p components decreases by is exactly the amount of variance by which all other components increase.

Another result concerns the Cholesky decomposition of $\boldsymbol{\Sigma}$. It has been proposed in the literature to use some special forms of matrix $\mathbf{A}$ to make the estimation of this matrix an easier problem. Several results in this article demonstrate that this is not so.

Theorem. Let the correlation functions $c_j(\cdot,\cdot)$ be specified for each $j = 1,\ldots,p$.

1. If $\mathbf{A}$ is a lower triangular matrix of the Cholesky decomposition of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}$ is estimated by maximum likelihood, then the mean of the first output $\mathrm{E}y^M(\mathbf{x}')_1$ and its variance $\mathrm{V}y^M(\mathbf{x}')_1$ stay the same as in the independent scenario.

2. If $\mathbf{A}$ is an upper triangular matrix of the Cholesky decomposition of $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}$ is estimated by maximum likelihood, then the pth output's mean $\mathrm{E}y^M(\mathbf{x}')_p$ and its variance $\mathrm{V}y^M(\mathbf{x}')_p$ stay the same as in the independent scenario.

Proof. The theorem in the Appendix on the maximum likelihood estimate (MLE) of the matrix $\mathbf{A}$ of the LMC model lets to conclude that in case of lower triangular matrix $\mathbf{A}$, the MLE of the $a_{11}^2 = \mathbf{y}_1^{\mathrm{T}}\mathbf{C}_1^{-1}\mathbf{y}_1/n$, which coincides exactly with the MLE estimate of the corresponding variance of the independent emulator of the first component, and resulting in exactly the same emulator.

The upper triangular case of matrix $\mathbf{A}$ is analogous. ■

To add, even if someone told us that improvement is possible by considering one output first, and then the second, conditionally on the first one (as with lower triangular Cholesky decomposition of $\mathbf{A}$), usually one would not know which output to take first. The above theorems show analytically that LMC model is not advantageous over independent modeling.

In other cases LMC model leads to a different predictive distribution than the independent modeling. The generalization is complicated, because of the need for estimation of parameters of the LMC model. However, no evidence is found that the predictive distributions in other cases from theoretical perspective should be closer (more narrow or precise) to true outputs.

4 Estimation of correlation in the LMC emulator

Another important complication which arise with the use of the LMC model is that correlation between outputs behaves nonintuitively. This misbehavior is demonstrated with comprehensive simulation examples. In the following section we present one such example. Three more examples are provided in Kyzyurova (2017). These examples are not included here because they all convey the same message. Taken together, all the examples provide quite compelling numerical evidence that use of the LMC model for emulation of multivariate outputs of a computer model is not needed.

4.1 Correlation between outputs

We provide definitions of correlation between multiple outputs of a computer model.

Definition. The correlation ρ, between two variables Y_1 and Y_2, is defined as

$$\rho = \frac{\mathrm{E}[(Y_1 - \mathrm{E}Y_1)(Y_2 - \mathrm{E}Y_2)]}{\sqrt{\mathrm{E}(Y_1 - \mathrm{E}Y_1)^2\mathrm{E}(Y_2 - \mathrm{E}Y_2)^2}}\,.$$

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

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

Perfect correlation, $\rho = 1$, or perfect anti-correlation, $\rho = -1$, occurs if and only if $f_1(x)$ and $f_2(x)$ (or Y_1 and Y_2), as defined in the above definitions, are linearly dependent, i.e., $f_2(\mathbf{x}) = \xi + \lambda f_1(\mathbf{x})$ (or $Y_2 = \xi + \lambda Y_1$).

Below the correlation between outputs of the LMC model is shown to behave in a nonintuitive and undesirable way. In particular, zero-correlation between outputs of a computer model, for general LMC model, does not imply independence between the outputs, as one might expect.

Proposition. The correlation between the kth and lth outputs, Y_k and Y_l, of the p-variate output of a computer model, at each input point of the LMC model, is

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

Proof.

$$\rho = \frac{\mathrm{E}(Y_k(x)Y_l(x))}{\sqrt{\mathrm{E}(Y_k^2(x))\mathrm{E}(Y_l^2(x))}} =$$

$$= \frac{\mathrm{E}(\sum_{i=1}^p A_{ki}W_i(x))(\sum_{i=1}^p A_{li}W_i(x)))}{\sqrt{(\mathrm{E}(\sum_{i=1}^p A_{ki}W_i(x))^2)(\mathrm{E}(\sum_{i=1}^p A_{li}W_i(x))^2)}} = \frac{\sum_{i=1}^p A_{ki}A_{li}}{\sqrt{(\sum_{i=1}^p A_{ki}^2)(\sum_{i=1}^p A_{li}^2)}}\,. \blacksquare$$

Corollary. If $\rho_{kl} = \pm 1$, then $\det \mathbf{A} = 0$.

If the kth and lth rows of the matrix $\mathbf{A}$ are linearly dependent, i.e. $A_{ki}/A_{li} = \eta\ \forall i$, then $\rho_{kl} = \pm 1$.

For bivariate output, $\rho = \pm 1$ if and only if $\det \mathbf{A} = 0$.

Corollary. If $A_{ki} = A_{lj} = 0\ \forall k \neq i, l \neq j$, then $\rho_{kl} = 0$. The converse, for general LMC model, is not true.

Intuitively, we would like zero-correlation to imply independence of the two processes for each output, but this is not the case, as shown in the above corollary. As exceptions, there are two special cases of the LMC model for which zero-correlation between outputs implies independence of the outputs.

Corollary. In the case of bivariate output, if $\rho_{12} = 0$ and $\mathbf{A}$ is symmetric, then the two processes are independent.

Corollary. In the case of a separable model, if $\rho_{kl} = 0$, then kth and lth outputs are independent.

Since $f_1(x)$ and $f_2(x)$ may only be computed at a modest number of inputs, we need to be concerned with sample correlations.

Definition. Let m values of each vector output $(\mathbf{y}_1, \mathbf{y}_2)$ from two functions f_1 and f_2 at some inputs $\mathbf{x}$, then the *sample correlation* is defined as

$$\rho_{smpl} = \frac{\sum_{i=1}^m (y_{1i} - \overline{\mathbf{y}_1})(y_{2i} - \overline{\mathbf{y}_2})}{\sqrt{\sum_{i=1}^m (y_{1i} - \overline{\mathbf{y}_1})^2 \sum_{i=1}^m (y_{2i} - \overline{\mathbf{y}_2})^2}}\,.$$

In simulation studies, the underlying true functions are known, and the actual correlation may be obtained via Monte-Carlo integration; indeed, as the number of randomly sampled points grows, $\rho_{smpl} \to \rho_{f_1 f_2}$.

Lemma. $\rho_{smpl} \to \rho_{f_1 f_2}$ as the number of randomly sampled points $n \to \infty$.

4.2 Simulation study

The simulation study numerically compares the predictions of the LMC emulator with those of the independent emulator (using maximum likelihood estimates of all the parameters in both models), concluding that there is little difference in predictive performance. This example shows that correlation between outputs is not well estimated (Figure 3).

Two functions $f_1(x) = 3x + \cos 5(x + \kappa)$ and $f_2(x) = \sin \pi(x + \kappa)$ are chosen. κ varies from $-\frac{3\pi}{2}$ to $\frac{\pi}{2}$ in both functions. Example of two functions f_1 and f_2 observed together are shown in Figure 1.

211 simulations $\kappa_1, \ldots, \kappa_{211}$ have been performed. Correlation between functions varies with varying κ (Figure 2). $n = 9$ equidistant training points $(x_1, x_2, \ldots, x_9)$ have been chosen in the domain $[-1, 1]$ (with $x_1 = -1$ and $x_9 = 1$); the functions are evaluated at these points resulting in function values $\mathbf{y}_1 = (f_1(x_1), \ldots, f_1(x_9))$ and $\mathbf{y}_2 = (f_2(x_1), \ldots, f_2(x_9))$. Nine triples (x_i, y_{1i}, y_{2i}), $i = 1, \ldots, 9$ of inputs-outputs are used to train the emulators. The circled points in Figure 1 are these input-output pairs for the two functions arising from two different choices of κ.

For the LMC and independent modeling, we use zero-mean Gaussian processes, with power-exponential correlation functions

$$c(x_k, x_l) = \exp\Big\{ - \Big(\frac{|x_k - x_l|}{\delta} \Big)^{\alpha} \Big\}\,,$$

having smoothness parameters $\alpha = 1.9$, for numerical stability. Often the smoothness parameter is fixed to avoid confounding with the range parameter and variance. The range parameter $\{\delta\}$ is considered in its transformed version $\tilde{\delta} = -\alpha \log \delta$, and estimated with maximum likelihood (along with other parameters in

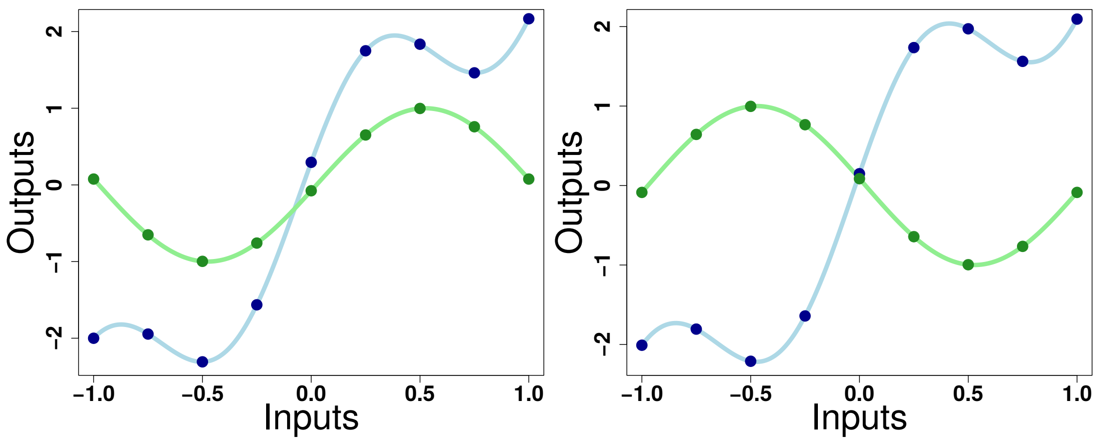

Figure 1: Examples of two output functions (blue and green curves) observed together at the same input points (shown with circles on the plot). Left panel: functions $f_1(x) = 3x + \cos 5(x + \kappa)$ and $f_2(x) = \sin \pi(x + \kappa)$ with $\kappa_{24} = \frac{23\pi}{105} - \frac{3\pi}{2}$. Right panel: the same functions $f_1(x)$ and $f_2(x)$ with $\kappa_{191} = \frac{190\pi}{105} - \frac{3\pi}{2}$. The corresponding observed sample correlations are about 0.82 and -0.83 respectively, with the true correlations being 0.93 and -0.94. These are maximum and minimum sample correlations among all simulations with various values of κ.

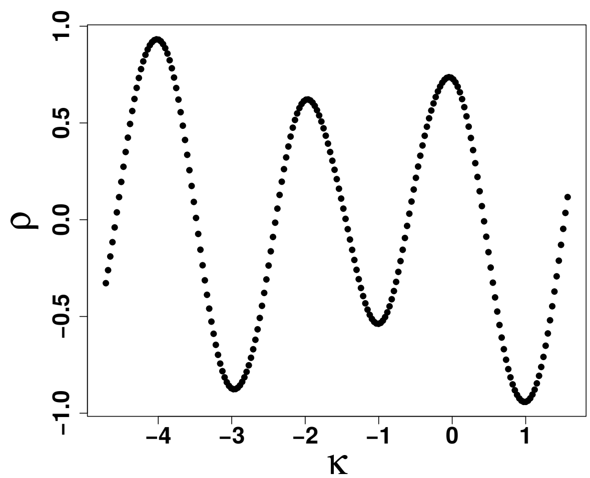

Figure 2: Correlation $\rho_{f_1 f_2}$ between two functions varies with the constant κ, ranging from about -0.94 to about 0.93.

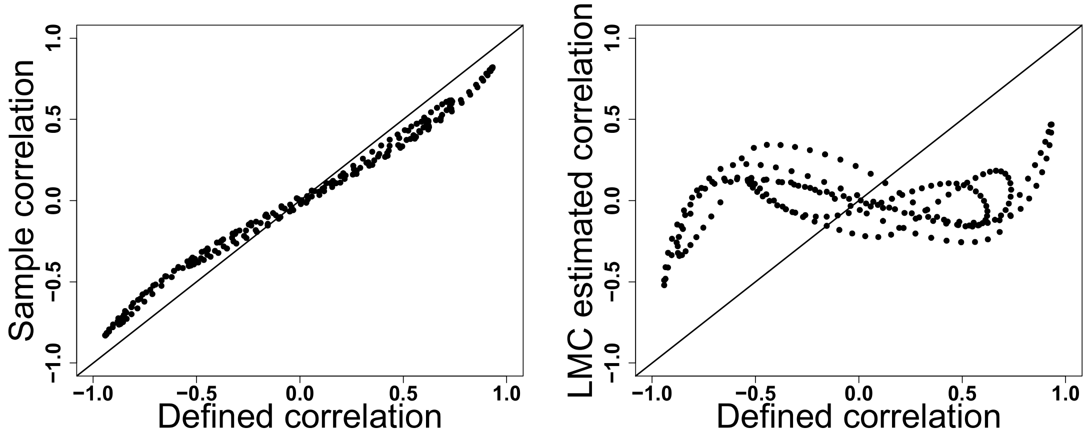

Figure 3: Left panel: true correlation between functions f_1 and f_2 vs. observed sample correlation. Right panel: true correlation vs. the estimate from the LMC model. The solid line on both panels is $y = x$.

the model). We intentionally avoid utilizing Bayesian framework with other possible priors on parameters in order to demonstrate the irrelevance phenomenon in principle for all groups of statisticians no matter what their favorite priors are.

We construct predictive distributions with LMC and the independent model emulators at $m = 192$ test points in $[-1, 1]$ with an increment of 0.01, excluding training points. We evaluate the predictions by several measures, such as root mean square predictive error (RMSPE), empirical frequency coverage, length of credible intervals and proper scoring rules (spherical, continuous ranked probability, quadratic and logarithmic (Kyzyurova 2019*b*)). A detailed explanation of comparison of LMC and independent model emulators is given in Kyzyurova (2017). In short, the predictive performance of two models is very similar to each other.

In this study we show that LMC model does not capture the correlation between two functions. The illustration is given in Figure 3. The number of input points $(x_1, \ldots, x_9)$ at which the output is obtained is enough for accurate estimation of the sample correlation between outputs (as it is close to the true correlation between functions), but not enough for the LMC model to capture the correlation between functions. In practice, with computationally expensive computer models, we do not know if we have enough observations for the proposed model to capture the correlation between functions.

5 Conclusion

Multivariate modeling of a computer model multivariate output has been shown not to lead to better emulation results than individual modeling of each output. No reduction of uncertainty (variance of the emulator) or systematic changes in other characteristics which would be desirable (to make an emulator more precise or accurate) were found. Interestingly, performance of the LMC model for spatial data on a grid is exactly the same as have been discussed in this work, that is the use of the LMC is not advantageous over independent modeling of a multivariate output variable for the predictive purpose.

6 Appendix

Other results have appeared which are tangential to the main message of this work. One of such results concerns estimation of matrix $\mathbf{A}$ in the LMC model (virtually impossible task in practice). In particular, in the supplementary materials we show that $\mathbf{A}$ is *not estimable from the data only*, in a sense that one can not find MLE estimate of $\mathbf{A}$ independently from other parameters in the model.

6.1 Estimation of covariance matrix

In this work we would like to avoid specification of the priors on the parameters to eliminate their possible influence on the estimates of the parameters and further behaviour of predictive distribution which constitutes the emulator.

The following theorem shows that $\mathbf{A}$ is *not estimable from the data only*, in a sense that one can not find MLE estimate of $\mathbf{A}$ independently from other parameters in the LMC model, typically and in our case these would be parameters in correlation functions. This theorem finds the MLE estimate of matrix $\mathbf{A}$ conditional on given processes Ws (i.e. fixed correlation functions with the parameters of the functions).

Theorem. If the LMC model has $\hat{\mathbf{A}}$, an MLE of $\mathbf{A}$ which is in no special a priori form, while correlation functions $c_j(\cdot,\cdot)$ of processes W_js are fixed $\forall\ j=1,\dots,p$, then $\hat{\mathbf{A}}$ must satisfy the following system of equations

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

Proof. Log-likelihood of the LMC model is

$$\log\det\tilde{\mathbf{\Sigma}}^{-1/2}-\frac{1}{2}\mathbf{y}^{(1:n)^{\mathrm{T}}}\tilde{\mathbf{\Sigma}}^{-1}\mathbf{y}^{(1:n)}\,.$$

Using computational lemma about the determinant of the covariance matrix of the LMC, the derivative of the log-likelihood over A^{-1} is as follows

$$n\mathbf{A}^{\mathrm{T}}-\frac{1}{2}\sum_{ij}\tilde{D}_{ij}\mathbf{A}^{-1}(\mathbf{y}^{(j)}\mathbf{y}^{(i)^{\mathrm{T}}}+\mathbf{y}^{(i)}\mathbf{y}^{(j)^{\mathrm{T}}}).$$

Setting the derivative of the log-likelihood to zero completes the proof. ■

Corollary.

1. If $\hat{\mathbf{A}}$ is the MLE of the LMC model parameter $\mathbf{A}$, then a matrix which is obtained from $\hat{\mathbf{A}}$ by multiplying any columns of $\hat{\mathbf{A}}$ by -1 is also a solution to the system and is also an MLE estimate of $\mathbf{A}$.
2. There are 2^p MLE estimates of $\mathbf{A}$.

Large number of parameters in the model results in flat likelihood surface, making estimation of parameters in the model a challenging problem.

Corollary. If $\rho=\pm1$, i.e. $\det\mathbf{A}=0$, then the likelihood of the LMC model does not exist, so MLEs cannot be defined.

Corollary. If the observed vectors are proportional to each other, i.e. $\mathbf{y}_2=\lambda\mathbf{y}_1, \lambda\neq0$, then the likelihood of the LMC model does not exist, so MLEs cannot be defined.

6.1.1 Previous estimates of A and Σ

Most previous papers working with the LMC model have focused on identifying the matrix $\mathbf{\Sigma}$, assuming that the covariance matrix of outputs is the key quantity. This makes sense, say, from the viewpoint of elicitation, since the output covariance is all that one could reasonably assess.

Interestingly, however, the LMC model depends on $\mathbf{A}$ itself explicitly, and not just on $\mathbf{\Sigma}=\mathbf{A}\mathbf{A}^{\mathrm{T}}$. The decomposition $\mathbf{\Sigma}=\mathbf{A}\mathbf{A}^{\mathrm{T}}$ is not unique, so that for the same covariance matrix $\mathbf{\Sigma}$, depending on the chosen decomposition, different values of the likelihood may be obtained. Therefore, depending on the chosen decomposition, different LMC GASP emulators may be obtained.

The representation introduced in the main part suggests that there should be no any restriction on entries of $\mathbf{A}$ (except that $a_{i1},a_{i2},\dots,a_{ip}$ are not zeros at the same time for each ith row ($i=1,\dots,p$) of matrix $\mathbf{A}$).

Inference with the specification of lower triangular matrix $\mathbf{A}$ of Cholesky decomposition depends on the order of data vectors $\mathbf{y}^{(1:n)}$ in which one writes the LMC likelihood. With the lower triangular matrix $\mathbf{A}$ one necessarily models the first component of the vector $\mathbf{Y}_1$ as one Gaussian process, the second one — as the sum of two Gaussian processes and so on. Clearly the procedure then depends on the order of data vectors $\mathbf{y}^{(1:n)}$ and various estimates of $\mathbf{A}$ and $\mathbf{\Sigma}$ (and including predictive LMC GASP emulator) may be obtained depending just on the order of data.

The formal propositions on comparison of the inference from the LMC model made with Cholesky decomposition in completely arbitrary setting of parameters of the LMC model (matrix $\mathbf{A}$ and parameters in correlation functions of the processes $W_i, i = 1, \ldots, p$) is not tractable. For specific examples, refer to Kyzyurova (2017).

Some literature suggests instead of Cholesky decomposition to use spectral decomposition of $\mathbf{\Sigma} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{\mathrm{T}}$ to avoid the dependence on the order, thus providing the recommendation to look at only symmetric $\mathbf{A} = \mathbf{P}\mathbf{\Lambda}^{1/2}\mathbf{P}^{\mathrm{T}}$ (which also aids the simplification of computations). Although it is reported that even with examples of $p = 3$ and $p = 6$ outputs, identifying parameters of the LMC model is challenging, highlighting that the numerical procedure used for obtaining the parameter estimates had no guarantee of achieving the global maximum of the chosen LMC likelihood. Although we discuss properties of the LMC model with symmetric matrix $\mathbf{A}$, there is no any motivation to restricting $\mathbf{A}$ to a certain form other than for "computational benefits", which have been demonstrated to be more of spurious benefits rather than the true benefits. The idea of the LMC model is to model each vector of data from a corresponding function out of p available outputs as a linear combination of p underlying unknown processes, thus the processes $W_1, W_2, \ldots, W_p$ should not be associated with any specific function.

Any prior information assumed on processes $W_j, j = 1, \ldots, p$ and matrix $\mathbf{A}$ makes LMC likelihood of data $\mathbf{y}^{(1:n)}$ to depend on the order of elements of $\mathbf{y}^{(1:n)}$ in the likelihood.

References

Kyzyurova, K. N. (2017), On Uncertainty Quantification for Systems of Computer Models, PhD thesis, Duke University.

Kyzyurova, K. N. (2019*a*), *Analysis of scientific computer models. Methodology in numerical simulator data analysis.*

Kyzyurova, K. N. (2019*b*), *On scoring rules.*

Powell, P. D. (2011), 'Calculating determinants of block matrices', *arXiv preprint arXiv:1112.4379* .