Proceedings of the 2017 Winter Simulation Conference
W. K. V. Chan, A. D'Ambrogio, G. Zacharewicz, N. Mustafee, G. Wainer, and E. Page, eds.

A JOINT GAUSSIAN PROCESS METAMODEL TO IMPROVE QUANTILE PREDICTIONS

Songhao Wang
Szu Hui Ng

Department of Industrial Systems Engineering and Management
National University of Singapore
10 Kent Ridge Crescent
Singapore 119260, SINGAPORE

ABSTRACT

Developing metamodels for quantiles can be inaccurate when the input estimates of the quantiles used to fit the model are noisy. In this paper, a multiple response model is developed to jointly model the quantile with a correlated and less-noisy expectation to improve the fit and predictions from the quantile metamodel. We first extend the standard stochastic Gaussian process model to the multi-response case and then use a simple *m*-design-point example to analytically study the benefits of the joint model over the single model. Several other numerical experiments are also conducted, and the results show that the joint model can provide better performance and thus improve quantile predictions.

1 INTRODUCTION

Although the mean of a simulation response distribution is typically used as an output measure of a system, the quantiles of the distribution have also become an increasingly important performance measure in many fields, including engineering safety, finance and healthcare. When the computer simulations are expensive, it is, however, very costly to conduct sufficient simulation replications (especially for heavy tailed distributions) to estimate the quantiles for every considered design or situation. To overcome this problem, metamodels, which in general provide an approximation to the input-output relationship of computer models, have been applied to provide global predictions for quantiles (Koenker 2005; Chen 2009; Dabo-Niang and Thiam 2010; Chen and Kim 2016).

The Gaussian process (GP) model, also termed as kriging model, has become a popular choice of metamodels. Although often used in deterministic computer experiments (Pham and Wagner 1999; Gupta et al. 2006; Joseph 2006), kriging has been more recently extended to approximate stochastic simulation models. These include the stochastic kriging (SK) model by Ankenman, Nelson, and Staum (2010) and the modified nugget effect kriging (MNEK) model by Yin, Ng, and Ng (2011). Most of these works, however, focused on developing GP models for mean performance measure as it is an easy measure to estimate from the simulation outputs with nice properties and decision approaches based on them are comparable to their deterministic counterpart.

To facilitate prediction of risk measures, Chen and Kim (2016) extended the SK model to approximate quantiles from stochastic simulation models. They studied various point and variance estimators of quantiles for inputs to the extended SK model, including batching (Seila 1982), sectioning (Asmussen and Glynn 2007), sectioning-batching (Nakayama 2014) and jackknifing (Nance and Sargent 2002; Gordy and Juneja 2010; Kleijnen 2008) methods for quantile estimates. They concluded that this extended SK approach for quantiles is promising and can provide competitive predictions when compared to the quantile regression approach.

As noted in Yin, Ng, and Ng (2009) and Hernandez and Grover (2013), the noise levels in the stochastic simulation can greatly affect the parameter estimates of the stochastic Gaussian process models and thus influence the performance of these models. When applying these models for quantile prediction, these issues can be more severe as the variance and biases of the quantile input estimates to the model can be large when the number of simulation replications is limited. One way to reduce the biases and variances of these quantile estimators is to increase the number of simulation replications. However, the improvements are limited when the simulation is expensive. In this paper, we propose an alternative approach to model and enhance the prediction of quantiles by developing a joint metamodel for the quantile and expectation. The main idea is to leverage and learn from more stable estimators of the expectation to enhance the prediction of the quantile. These two variables are likely to be correlated as they come from the same underlying distribution. For example, in the M/M/1 queueing system, the expected number of customers at a steady state will be highly correlated with the α-quantile, where the true distribution is geometric with parameter $1-\lambda/\mu$ (λ is the arrival rate, μ is the service rate). In portfolio analysis, larger risks are likely to bring larger expected returns with higher variations, i.e. larger mean values of return tend be associated with larger upper-α-quantile. In addition, the expectation estimator is unbiased with a noise level decreasing with the rate $\frac{1}{n(n-1)}$ while the quantile estimator is biased with a noise level decreasing with the rate $\frac{1}{n}$. Hence, with this joint model, we investigate further whether the quantile prediction can be improved with the information learned from this correlated, less-noisy and unbiased expectation prediction.

The multiple response Gaussian process model is a popular way to jointly model a number of output responses and provide an estimation of the correlations among them. It has also been applied to jointly model multiple output measures of deterministic computer models. Following the work by Kennedy and O'Hagan (2001), a lot of researchers have used the multiple response GP to model the dependencies between outputs (Conti and OHagan 2010; McFarland et al. 2008; Rougier 2008; Álvarez and Lawrence 2011; Kleijnen and Mehdad 2014). Fricker, Oakley, and Urban (2013) provide a review of the different types of multiple response GP models including the separable model (Conti and OHagan 2010), the convolution method model (Ver Hoef and Barry 1998; Higdon 2002) and the linear model of coregionalization (LMC) (Journel and Huijbregts 1978; Goulard and Voltz 1992). From their study, it was found that LMC model performed better than the others as it is more able to capture the complexities of the responses. In this paper, we also adopt this model for better flexibility to model quantile and expectation.

This paper is organized as follows. In section 2, we extend the traditional single response stochastic GP model into the multiple response case to build a joint model for quantile and expectation. In section 3, we use a *m*-design-point problem to give some insights of this model. In section 4, some numerical examples are provided to show the power of this model. In section 5, we summarize the work and present some future research topics.

2 MULTIRESPONSE STOCHASTIC GAUSSIAN PROCESS

In this section, we first review the basics of the stochastic GP and LMC models and then extend them to build a joint model for quantile and expectation.

2.1 Stochastic Gaussian Process Model Basics

The standard stochastic GP model assumes that the response of stochastic simulation can be represented in the following form:

$$Y(x) = Z(x) + \varepsilon(x) = \boldsymbol{f}(x)^T\boldsymbol{\beta} + \eta(x) + \varepsilon(x),$$

where $Z(x)$ represents the noise-free response which can be further decomposed into the mean function, $\boldsymbol{f}(x)^T\boldsymbol{\beta}$, and a second-order stationary GP, $\eta(x)$, (Cressie 2015). $\boldsymbol{f}(x)$ is a $p\times 1$ vector of known functions and $\boldsymbol{\beta}$ is a $p\times 1$ vector of model parameters. In this work, we consider the Gaussian correlation function to capture the spatial correlation of the GP, $\eta(x)$: $\text{cov}(\eta(x_1),\eta(x_2)) = \sigma_z^2\text{corr}(x_1,x_2) = \sigma_z^2\exp\sum_{j=1}^{d}\theta_j(x_{1,j}-x_{2,j})^2$, where $\sigma_z^2 = \text{var}(\eta(x))$, $x_{i,j}$ is the jth coordinate of x_i and $\boldsymbol{\theta}$ is the sensitivity parameter. ε is the random

noise with mean zero and variance $\sigma_\varepsilon^2(x)$. This noise is assumed to be independent of $\eta(x)$. Here we also assume that no common random numbers are applied and hence the noise at different input levels are also independent, i.e., $\text{cov}(\varepsilon(x_i),\varepsilon(x_j))=0$, for $i\neq j$.

Typically, the stochastic GP model is used to predict the mean function at an unknown input point x_0, $Z(x_0)$, given a series of observations at design point $x_1,...,x_m$. Suppose there are n_i independent simulation replications at point x_i, then the predictor $\widehat{Z}(x_0)$ and the associated mean square error (MSE) could be expressed as follows based on the sample mean $\overline{\boldsymbol{Y}}=(\overline{Y}(x_1),...,\overline{Y}(x_m))^T$:

$$\widehat{Z}(x_0)=\boldsymbol{f}(x)^T\widehat{\boldsymbol{\beta}}+\boldsymbol{c}(x_0)^T\boldsymbol{R}^{-1}(\overline{\boldsymbol{Y}}-\boldsymbol{F}\widehat{\boldsymbol{\beta}}), \tag{1}$$

$$\text{MSE}(\widehat{Z}(x_0))=\sigma_z^2-\boldsymbol{c}(x_0)^T\boldsymbol{R}^{-1}\boldsymbol{c}(x_0)-(\boldsymbol{f}(x_0)-\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{c}(x_0))^T(\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{F})^{-1}(\boldsymbol{f}(x_0)-\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{c}(x_0)), \tag{2}$$

where $\boldsymbol{c}(x_0)$ is the covariance vector of x_0 with existing design points: $\boldsymbol{c}(x_0)=\sigma_z^2(\text{corr}(x_0,x_1),...,\text{corr}(x_0,x_m))^T$, $\boldsymbol{F}=(\boldsymbol{f}(x_1)^T,...,\boldsymbol{f}(x_m)^T)^T$. $\widehat{\boldsymbol{\beta}}=(\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}^T\boldsymbol{R}^{-1}\overline{\boldsymbol{Y}}$ is the generalized least squares estimators for $\boldsymbol{\beta}$, and $\boldsymbol{R}$ is the covariance matrix of the design points:

$$\boldsymbol{R}=\boldsymbol{R}_z+\boldsymbol{R}_\varepsilon=\sigma_z^2\begin{bmatrix}1 & \text{corr}(x_1,x_2) & \cdots & \text{corr}(x_1,x_m)\\ \text{corr}(x_2,x_1) & 1 & \cdots & \text{corr}(x_2,x_m)\\ \vdots & \vdots & \ddots & \vdots\\ \text{corr}(x_m,x_1) & \text{corr}(x_m,x_2) & \cdots & 1\end{bmatrix}+\sigma_z^2\begin{bmatrix}s_1^2 & 0 & \cdots & 0\\ 0 & s_2^2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & s_m^2\end{bmatrix}, \tag{3}$$

where s_j^2 represents the ratio of the noise variance over the spatial uncertainty, i.e., $s_j^2=\text{var}(\overline{\varepsilon}(x_j))/\sigma_z^2=\sigma_\varepsilon^2(x_j)/(n_j\sigma_z^2)$. When the remaining parameters are unknown, $\sigma_\varepsilon^2(x_i)$ can be estimated by the sample variance of the simulation runs at point x_i. σ_z^2 and $\boldsymbol{\theta}$ can be estimated by maximizing the likelihood (details can be found in Ankenman, Nelson, and Staum (2010); Yin, Ng, and Ng (2011)).

As mentioned above, the standard stochastic GP model is developed to model and predict the mean of a stochastic simulation and requires the point estimates of the mean and their respective variance at the m design points. To apply it for quantile prediction, Chen and Kim (2016) proposed to replace the point estimates of the mean (which is the $\overline{\boldsymbol{Y}}$ in (1)) and the corresponding variance ($\boldsymbol{R}_\varepsilon$ in (3)) with the quantile estimates and its noise variance. They then compared different types of point and variance estimators to measure the predictive performance of VaR and CVaR. As their study found that the sectioning method provides better predictive properties and results, in this paper, we adopt this same method to estimate the quantiles and their corresponding variances. Below, we briefly introduce the sectioning methods applied to estimate these values.

Denote the n_i simulation runs at x_i as $H_1(x_i),...,H_{n_i}(x_i)$. The point estimator of the α-quantile is just the sample quantile: $\widehat{q}_\alpha(x_i)=H_{(\lceil\alpha n_i\rceil)}(x_i)$, where $H_{(j)}$ is the jth order statistic. The resulting predictor will share the similar formula as (1) except that the $\overline{\boldsymbol{Y}}$ is replaced with $\widehat{\boldsymbol{Q}}_\alpha$, where $\widehat{\boldsymbol{Q}}_\alpha=(\widehat{q}_\alpha(x_1),...,\widehat{q}_\alpha(x_m))$. The corresponding noise variance matrix $\boldsymbol{R}_\varepsilon$ is also modified with $s_i^2=\text{var}(\widehat{q}_\alpha(x_i))/\sigma_z^2$. However, as the sample quantile estimators are biased (Chen and Kim 2014), the MSE of quantile predictor will be larger. Specifically, the bias of quantile predictor at x_0 can be derived to:

$$\text{bias}(\widehat{Z}(x_0))=E[\widehat{Z}(x_0)-Z(x_0)]=\{\boldsymbol{f}(x)^T(\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}^T\boldsymbol{R}^{-1}+\boldsymbol{c}(x_0)^T\boldsymbol{R}^{-1}(\boldsymbol{I}-\boldsymbol{F}(\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}^T\boldsymbol{R}^{-1})\}^T\boldsymbol{\xi},$$

where $\boldsymbol{\xi}$ represents the bias of $\widehat{\boldsymbol{Q}}_\alpha$, i.e., $\xi_i=E[\widehat{q}_\alpha(x_i)-q_\alpha(x_i)]$. (The details of this derivation can be found in Chen and Kim (2014)). Hence, the overall MSE of quantile predictor consists of (2) plus the square of the bias term given above. In the sectioning method applied, the variance $\text{var}(\widehat{q}_\alpha(x_i))$ is estimated in the following manner. First, the simulation runs $H_1(x_i),...,H_{n_i}(x_i)$ are divided into n_b non-overlapping groups

with n_c replications in each group. Then, the sample α-quantile within each group, $\widehat{q}_{\alpha,j}(x_i), j = 1,...,n_b$, is obtained. The variance estimator is then computed as:

$$\widehat{\text{var}}(\widehat{q}_\alpha(x_i)) = \frac{1}{n_b(n_b-1)} \sum_{j=1}^{n_b} (\widehat{q}_{\alpha,j}(x_i) - \widehat{q}_\alpha(x_i))^2. \tag{4}$$

2.2 Linear Model of Coregionalization Basics

In this subsection, we briefly review the LMC model for the deterministic case before extending it to the stochastic case in the next subsection. In this model, each output can be represented by "a linear combination of 'building block' processes" (Fricker, Oakley, and Urban 2013). These 'building block' processes are a series of independent Gaussian processes. An LMC with k blocks is given as:

$$\boldsymbol{Z}(x) = (\boldsymbol{I}_k \otimes \boldsymbol{f}(x)^T)\boldsymbol{\beta} + \boldsymbol{A}\boldsymbol{\eta}, \tag{5}$$

where $\boldsymbol{Z}$ is an $k \times 1$ vector of deterministic responses, $\boldsymbol{I}_k$ is a k-dimensional identity matrix, $\otimes$ represents Kronecker product, and $\boldsymbol{\beta}$ is a $kp \times 1$ vector of model parameters. $(\boldsymbol{I}_k \otimes \boldsymbol{f}^T(x))\boldsymbol{\beta}$ indicates that the mean part of these k outputs share the same set of known functions $\boldsymbol{f}$, but different mean parameters. $\boldsymbol{\eta}$ is a $k \times 1$ vector of independent zero-mean and single variance GPs and $\boldsymbol{A}$ is a $k \times k$ full-rank matrix. The covariance matrix of $\boldsymbol{Z}(x_i)$ and $\boldsymbol{Z}(x_j)$ is:

$$\text{cov}(\boldsymbol{Z}(x_i), \boldsymbol{Z}(x_j)) = \boldsymbol{A}\text{diag}\{\text{cov}(\eta_l(x_i), \eta_l(x_j))\}_{l=1}^k \boldsymbol{A}^T.$$

It is easy to see that the between-output covariance is $\boldsymbol{R}_0 = \text{cov}(\boldsymbol{Z}(x_i), \boldsymbol{Z}(x_i)) = \boldsymbol{A}\boldsymbol{A}^T$. According to Kleijnen and Mehdad (2014), the cross-correlation of the ith and jth output can be defined as:

$$r_{ij} = \frac{\sum_{l=1}^k \boldsymbol{A}_{il}\boldsymbol{A}_{jl}}{\sqrt{\sum_{l=1}^k \boldsymbol{A}_{il}\boldsymbol{A}_{il}}\sqrt{\sum_{l=1}^k \boldsymbol{A}_{jl}\boldsymbol{A}_{jl}}},$$

where larger value of r_{ij} indicates a larger correlation between ith and jth output.

2.3 A Joint Model for Quantile-Expectation Using LMC

In this subsection, we extend the stochastic GP model to a multivariate case based on the LMC model, so that the dependency of quantile and expectation can be captured. Denote $Y_1(x)$ and $Y_2(x)$ as the α-quantile and expectation of the stochastic simulation at x, while $Z_1(x)$ and $Z_2(x)$ represent their noise-free versions. The multivariate GP model for Y_1, Y_2 could be expressed as follows:

$$\boldsymbol{Y}(x) = \boldsymbol{Z}(x) + \boldsymbol{\varepsilon}(x) = (\boldsymbol{I}_2 \otimes \boldsymbol{f}(x)^T)\boldsymbol{\beta} + \boldsymbol{\delta}(x) + \boldsymbol{\varepsilon}(x),$$

where $\boldsymbol{Y}(x) = (Y_1(x), Y_2(x))^T$, $\boldsymbol{Z}(x) = (Z_1(x), Z_2(x))^T$. $\boldsymbol{I}_2$ is a 2-dimensional identity matrix. $\boldsymbol{\beta}$ is a $2p \times 1$ vector of model parameters. $\boldsymbol{\varepsilon}(x) = (\varepsilon_1(x), \varepsilon_2(x))^T$ represent the noises of the performance measures following multivariate normal distribution, which is a new term added to the deterministic LMC (5) to capture the stochastic behavior of the responses in this work. Since the sample mean and sample quantile might be estimated from the same simulation runs, these two noises $\varepsilon_1(x_i)$ and $\varepsilon_2(x_i)$ are dependent and should not be modeled independently. $\boldsymbol{\delta}(x)$ is assumed to be a realization of a 2-variate GP with zero mean. Following the LMC approach, $\delta_i(x)$ could be modeled as a linear combination of independent single-variate GPs:

$$\delta_1(x) = A_{11}\eta_1(x) + A_{12}\eta_2(x),$$
$$\delta_2(x) = A_{21}\eta_1(x) + A_{22}\eta_2(x),$$

where η_1, η_2 are two independent GP with zero mean and unit variance: $\text{cov}(\eta_i(x_1), \eta_i(x_2)) = \text{corr}_i(x_1, x_2), i = 1, 2$. The covariance functions of this model are:

$$\text{cov}_{11}(x_i, x_j) = \text{cov}(Z_1(x_i), Z_1(x_j)) = A_{11}^2 \text{corr}_1(x_1, x_2) + A_{12}^2 \text{corr}_2(x_1, x_2),$$

$$\text{cov}_{22}(x_i, x_j) = \text{cov}(Z_2(x_i), Z_2(x_j)) = A_{21}^2 \text{corr}_1(x_1, x_2) + A_{22}^2 \text{corr}_2(x_1, x_2),$$

$$\text{cov}_{12}(x_i, x_j) = \text{cov}(Z_1(x_i), Z_2(x_j)) = A_{11}A_{12} \text{corr}_1(x_1, x_2) + A_{12}A_{22} \text{corr}_2(x_1, x_2),$$

$$\sigma_{z1}^2 = A_{11}^2 + A_{12}^2, \quad \sigma_{z2}^2 = A_{21}^2 + A_{22}^2, \quad \sigma_{z12}^2 = A_{11}A_{21} + A_{12}A_{22}.$$

The cross-correlation between these two responses is:

$$r := \text{corr}(Z_1(x), Z_2(x)) = \frac{A_{11}A_{21} + A_{12}A_{22}}{\sqrt{(A_{11}^2 + A_{12}^2)(A_{21}^2 + A_{22}^2)}}.$$

The predictor of $\boldsymbol{Z}(x_0)$ is:

$$\widehat{\boldsymbol{Z}}(x_0) = (\boldsymbol{I}_2 \otimes \boldsymbol{f}(x)^T)\widehat{\boldsymbol{\beta}} + \boldsymbol{C}(x_0)^T \boldsymbol{R}^{-1}(\mathscr{Y} - \boldsymbol{F}\widehat{\boldsymbol{\beta}}), \tag{6}$$

where $\mathscr{Y} = (\widehat{\boldsymbol{Q}}_\alpha^T, \overline{\boldsymbol{Y}}^T)^T$ is the vector of point estimators of quantile and mean, $\boldsymbol{F} = \boldsymbol{I}_2 \otimes (\boldsymbol{f}(x_1)^T, ..., \boldsymbol{f}(x_m)^T)^T$, $\boldsymbol{C}(x_0)$ is the covariance matrix between $\boldsymbol{Z}(x_0)$ and observations $\mathscr{Y}$. $\widehat{\boldsymbol{\beta}} = (\boldsymbol{F}^T\boldsymbol{R}^{-1}\boldsymbol{F})^{-1}\boldsymbol{F}^T\boldsymbol{R}^{-1}\mathscr{Y}$. $\boldsymbol{R} = \boldsymbol{R}_z + \boldsymbol{R}_\varepsilon$ is the covariance matrix for $\mathscr{Y}$ with

$$\boldsymbol{R}_z = \begin{bmatrix}
\sigma_{z1}^2 & \text{cov}_{11}(x_1,x_2) & \cdots & \text{cov}_{11}(x_1,x_m) & \sigma_{z12}^2 & \text{cov}_{12}(x_1,x_2) & \cdots & \text{cov}_{12}(x_1,x_m) \\
\text{cov}_{11}(x_2,x_1) & \sigma_{z1}^2 & \cdots & \text{cov}_{11}(x_2,x_m) & \text{cov}_{12}(x_2,x_1) & \sigma_{z12}^2 & \cdots & \text{cov}_{12}(x_2,x_m) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\text{cov}_{11}(x_m,x_1) & \text{cov}_{11}(x_m,x_2) & \cdots & \sigma_{z1}^2 & \text{cov}_{12}(x_m,x_1) & \text{cov}_{12}(x_m,x_2) & \cdots & \sigma_{z12}^2 \\
\sigma_{z12}^2 & \text{cov}_{21}(x_1,x_2) & \cdots & \text{cov}_{21}(x_1,x_m) & \sigma_{z2}^2 & \text{cov}_{22}(x_1,x_2) & \cdots & \text{cov}_{22}(x_1,x_m) \\
\text{cov}_{21}(x_2,x_1) & \sigma_{z12}^2 & \cdots & \text{cov}_{21}(x_2,x_m) & \text{cov}_{22}(x_2,x_1) & \sigma_{z2}^2 & \cdots & \text{cov}_{22}(x_2,x_m) \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
\text{cov}_{21}(x_m,x_1) & \text{cov}_{21}(x_m,x_2) & \cdots & \sigma_{z12}^2 & \text{cov}_{22}(x_m,x_1) & \text{cov}_{22}(x_m,x_2) & \cdots & \sigma_{z2}^2
\end{bmatrix}$$

$$\boldsymbol{R}_\varepsilon = \begin{bmatrix}
\sigma_{z1}^2 \text{diag}\{s_{j,1}^2\}_{j=1}^m & \sigma_{z1}\sigma_{z2}\text{diag}\{\rho_j s_{j,1} s_{j,2}\}_{j=1}^m \\
\sigma_{z1}\sigma_{z2}\text{diag}\{\rho_j s_{j,1} s_{j,2}\}_{j=1}^m & \sigma_{z2}^2 \text{diag}\{s_{j,2}^2\}_{j=1}^m
\end{bmatrix},$$

where $s_{j,1}^2 = \text{var}(\widehat{q}_\alpha(x_j))/\sigma_{z1}^2$, $s_{j,2}^2 = \text{var}(\overline{Y}(x_j))/\sigma_{z2}^2$, which represent the ratio of the noise variance over the spatial uncertainty for output 1 (quantile) and output 2 (expectation), respectively. $\rho_j = \text{corr}(\widehat{q}_\alpha(x_j), \overline{Y}(x_j))$, which represents the correlation between the noises for output 1 and 2 at point x_j. When the estimators for quantile and expectation are drawn from the same simulation runs, this correlation will be non-zero. We propose the following method to estimate this correlation given simulation replications at x_i. Following the same procedures as sectioning method, these replications will first be divided into n_b groups. Then the correlation can be estimated based on the sample quantile and mean, $\widehat{q}_{\alpha,j}(x_i)$ and $\overline{Y}_j(x_i)$, $j = 1, ..., n_b$, within each group:

$$\widehat{\rho}_i = \frac{\sum_{j=1}^{n_b}(\widehat{q}_{\alpha,j}(x_i) - \widehat{q}_\alpha(x_i))(\overline{Y}_j(x_i) - \overline{Y}(x_i))}{\sqrt{\sum_{j=1}^{n_b}(\widehat{q}_{\alpha,j}(x_i) - \widehat{q}_\alpha(x_i))^2}\sqrt{\sum_{j=1}^{n_b}(\overline{Y}_j(x_i) - \overline{Y}(x_i))^2}}. \tag{7}$$

As quantile prediction ($\widehat{Z}_1(x_0)$) is our main interest, we derive below the predictor and associated variance and bias for $Z_1(x_0)$:

$$\widehat{Z}_1(x_0) = (\boldsymbol{e}_1 \otimes \boldsymbol{f}(x)^T)\widehat{\boldsymbol{\beta}} + \boldsymbol{c}(x_0)^T \boldsymbol{R}^{-1}(\mathscr{Y} - \boldsymbol{F}\widehat{\boldsymbol{\beta}}),$$

$$\text{var}(\widehat{Z}_1(x_0)) = E[\widehat{Z}_1(x_0)^2] - E[\widehat{Z}_1(x_0)]^2$$

$$= \sigma_{z1}^2 - \boldsymbol{c}(x_0)^T \boldsymbol{R}^{-1} \boldsymbol{c}(x_0) - (\boldsymbol{e}_1 \otimes \boldsymbol{f}(x_0) - \boldsymbol{F}^T \boldsymbol{R}^{-1} \boldsymbol{c}(x_0))^T (\boldsymbol{F}^T \boldsymbol{R}^{-1} \boldsymbol{F})^{-1} (\boldsymbol{e}_1 \otimes \boldsymbol{f}(x_0) - \boldsymbol{F}^T \boldsymbol{R}^{-1} \boldsymbol{c}(x_0)),$$

$$\text{bias}(\widehat{Z}_1(x_0)) = E[\widehat{Z}_1(x_0) - Z_1(x_0)]$$

$$= \{\boldsymbol{e}_1 \otimes \boldsymbol{f}(x)^T (\boldsymbol{F}^T \boldsymbol{R}^{-1} \boldsymbol{F})^{-1} \boldsymbol{F}^T \boldsymbol{R}^{-1} + \boldsymbol{c}(x_0)^T \boldsymbol{R}^{-1} (\boldsymbol{I} - \boldsymbol{F}(\boldsymbol{F}^T \boldsymbol{R}^{-1} \boldsymbol{F})^{-1} \boldsymbol{F}^T \boldsymbol{R}^{-1})\}^T \boldsymbol{\xi},$$

where $\boldsymbol{e}_1 = (1,0)$, $\boldsymbol{c}(x_0) = \{\text{cov}_{11}(x_0,x_1),...,\text{cov}_{11}(x_0,x_m),\text{cov}_{12}(x_0,x_1),...\text{cov}_{12}(x_0,x_m)\}^T$. $\boldsymbol{\xi} = (\boldsymbol{\xi}_1{}^T, \boldsymbol{0}_{m\times 1}^T)^T$, where $\boldsymbol{\xi}_1$ represents a vector of bias of the quantile estimators.

In this model, the parameters $\boldsymbol{A}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ are estimated by maximizing the likelihood (Álvarez and Lawrence 2011). Fricker, Oakley, and Urban (2013) proposed an alternative approach to estimate $\boldsymbol{A}$ by choosing $\boldsymbol{A}$ as an eigendecomposition of the between-output covariance matrix $\boldsymbol{R}_0 = \boldsymbol{A}\boldsymbol{A}^T$. This approach can reduce the number of parameters to be estimated when number of responses is large.

To summarize, the steps to estimate the inputs and parameters of the joint model are as follows:

1. Estimate the variance of quantile estimator by sectioning method (4) and the variance of expectation estimator by $1/n_i$ times the sample variance of the simulation runs at x_i. The correlation ρ_i of these estimators would be estimated by (7).
2. Estimate $\boldsymbol{A}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ by maximizing the likelihood function(Álvarez and Lawrence 2011).
3. The predictor (6) can be obtained by plugging in all the estimated parameters.

3 A m-DESIGN-POINT PROBLEM

In this section, we use a similar example from Mitchell, Morris, and Ylvisaker (1994) and Chen, Ankenman, and Nelson (2013) to gain some insights on the benefits of the joint model over a single response model of the quantile. We make the following assumptions to simplify the analysis:

1. The two 'building block' GPs, η_1, η_2, share the same parameters: $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 = \boldsymbol{\theta}$, $\text{cov}(\eta_i(x_1), \eta_i(x_2)) = \text{corr}(x_1,x_2) = \exp\sum_{j=1}^d \theta_j (x_{1,j} - x_{2,j})^2, i = 1,2$.
2. In this example, the m design points are distant enough so that the spatial correlation between any two points are negligible: $\text{cov}_{11}(x_i,x_j) = \text{cov}_{22}(x_i,x_j) = \text{cov}_{12}(x_i,x_j) = 0$, for $i \neq j$.
3. The unknown point x_0 is equally correlated to every design point: $\text{cov}_{11}(x_0,x_i) = \sigma_{z1}^2 r_0, \text{cov}_{12}(x_0,x_i) = \sigma_{z12}^2 r_0$, where r_0 represents the spatial correlation between the unknown point and design points in η_1, η_2.
4. All model parameters are known. The noise variances are equal for all design points, and the correlation between the noises are assumed to be zero: $s_{i,1}^2 = s_1^2,\ s_{i,2}^2 = s_2^2,\ \rho_i = 0,\ i = 1,...,m$.
5. The bias vector $\boldsymbol{\xi} = (\boldsymbol{\xi}_1{}^T, \boldsymbol{0}_{m\times 1}^T)^T$, and $\boldsymbol{\xi}_1 > \boldsymbol{0}$, where $\boldsymbol{\xi}_1$ is a vector of bias of quantile estimators.

The first assumption is to simplify the joint model (Kleijnen and Mehdad 2014) and make it comparable with the stochastic GP model. The second and third assumptions would be plausible if all design points are in the extreme regions of the input space while the unknown point is in the center (Ankenman, Nelson, and Staum 2010). Although the fourth assumption indicates that $\rho_i = 0$, which can be accomplished if the quantile and expectation are estimated from separate sets of simulation replications, and all parameters are given, we would relax this restriction in our numerical examples in the next section, i.e., all parameters are estimated and $\rho_i \neq 0, i = 1,...,m$. Assumption 5 indicates that the biases for the quantile estimators are all positive for $x_i, i = 1,...m$. In fact, the empirical quantile estimators would be positively (or negatively) biased depending on the convexity of the distribution function (Kim and Hardy 2007). In other words, it is reasonable to assume these biases share the same sign if the underlying distributions are of the same family or share the same convexity at all input points. Therefore, we assume positive bias since only the absolute value of bias is of interest to calculate the MSE. Moreover, assumption 5 also ensures that the expectation estimators are unbiased. This simplifying example will enable us to analytically study the benefits of leveraging on information from a correlated model with lower variance and bias to learn and improve on the quantile model and prediction.

Based on these assumptions, the covariance matrix $\boldsymbol{R}$ and $\boldsymbol{c}(x_0)$ can be simplified:

$$\boldsymbol{R} = \begin{bmatrix} \boldsymbol{R}_1 & \boldsymbol{R}_2 \\ \boldsymbol{R}_3 & \boldsymbol{R}_4 \end{bmatrix} = \begin{bmatrix} \boldsymbol{R}_{z1} & \boldsymbol{R}_{z2} \\ \boldsymbol{R}_{z3} & \boldsymbol{R}_{z4} \end{bmatrix} + \begin{bmatrix} \boldsymbol{R}_{\varepsilon 1} & \boldsymbol{R}_{\varepsilon 2} \\ \boldsymbol{R}_{\varepsilon 3} & \boldsymbol{R}_{\varepsilon 4} \end{bmatrix}, \quad \boldsymbol{c}(x_0) = \{\boldsymbol{c}_1(x_0)^T, \boldsymbol{c}_2(x_0)^T\}^T,$$

where

$$\boldsymbol{R}_{z1} = \sigma_{z1}^2 \boldsymbol{I}_m, \quad \boldsymbol{R}_{z2} = \boldsymbol{R}_{z3} = \sigma_{z12}^2 \boldsymbol{I}_m, \quad \boldsymbol{R}_{z4} = \sigma_{z2}^2 \boldsymbol{I}_m,$$
$$\boldsymbol{R}_{\varepsilon 1} = \sigma_{z1}^2 s_1^2 \boldsymbol{I}_m, \quad \boldsymbol{R}_{\varepsilon 2} = \boldsymbol{R}_{\varepsilon 3} = \boldsymbol{0}_{m\times m}, \quad \boldsymbol{R}_{\varepsilon 4} = \sigma_{z2}^2 s_2^2 \boldsymbol{I}_m,$$
$$\boldsymbol{c}_1(x_0) = \sigma_{z1}^2 r_0 \boldsymbol{e}_{m\times 1}, \quad \boldsymbol{c}_2(x_0) = \sigma_{z12}^2 r_0 \boldsymbol{e}_{m\times 1},$$

where $\boldsymbol{e}_{m\times 1}$ is a $m \times 1$ vector whose components are all 1.

Denote $\text{var}_{\text{single}}, \text{bias}_{\text{single}}, \text{MSE}_{\text{single}}$ and $\text{var}_{\text{joint}}, \text{bias}_{\text{joint}}, \text{MSE}_{\text{joint}}$ as the variance, bias and MSE for quantile predictor at x_0 in the stochastic GP and the joint model, respectively. The following properties and results can be obtained for these two models (detailed derivations are provided in the appendix A):

$$\text{bias}_{\text{single}} = \boldsymbol{c}_1(x_0)^T \boldsymbol{R}_1{}^{-1} \boldsymbol{\xi}_1 = \frac{r_0}{1+s_1^2} \sum_{i=1}^{m} \xi_1(x_m),$$

$$\text{bias}_{\text{joint}} = \text{bias}_{\text{single}} - (\boldsymbol{c}_2(x_0)^T - \boldsymbol{c}_1(x_0)^T \boldsymbol{R}_1{}^{-1} \boldsymbol{R}_2) \boldsymbol{P} \boldsymbol{R}_3 \boldsymbol{R}_1{}^{-1} \boldsymbol{\xi}_1 = \frac{r_0}{1+s_1^2} \left(1 - \frac{r^2 s_1^2}{(1+s_1^2)(1+s_2^2) - r^2}\right) \sum_{i=1}^{m} \xi_1(x_m),$$

$$\text{var}_{\text{single}} = \sigma_{z1}^2 - \boldsymbol{c}_1(x_0)^T \boldsymbol{R}_1{}^{-1} \boldsymbol{c}_1(x_0) = \sigma_{z1}^2 \left(1 - \frac{m r_0^2}{1+s_1^2}\right),$$

$$\text{var}_{\text{joint}} = \text{var}_{\text{single}} - \boldsymbol{\Gamma}^T \boldsymbol{P} \boldsymbol{\Gamma} = \sigma_{z1}^2 \left(1 - \frac{m r_0^2}{1+s_1^2} - \frac{m}{1+s_1^2} \frac{r^2 r_0^2 s_1^4}{(1+s_1^2)(1+s_2^2) - r^2}\right),$$

where $\boldsymbol{\Gamma} = \boldsymbol{R}_3 \boldsymbol{R}_1^{-1} \boldsymbol{c}_1(x_0) - \boldsymbol{c}_2(x_0)$. $\boldsymbol{P} = \left(\boldsymbol{R}_4 - \boldsymbol{R}_3 \boldsymbol{R}_1^{-1} \boldsymbol{R}_2\right)^{-1}$, which is a positive definite matrix. Therefore, $\boldsymbol{\Gamma}^T \boldsymbol{P} \boldsymbol{\Gamma} > 0$ and $\text{var}_{\text{joint}} < \text{var}_{\text{single}}$. It is also easy to show that $0 < \text{bias}_{\text{joint}} < \text{bias}_{\text{single}}$ (see appendix A). Hence, it follows that $\text{MSE}_{\text{single}} > \text{MSE}_{\text{joint}}$. This example clearly shows the benefits of the joint model for the quantile predictor as it is better able to leverage on additional lower variance and unbiased information of the mean to provide a better predictor (in terms of lower MSE, lower variance and lower bias).

Next, we further investigate the influence of r and s_2^2 on the ratio $\frac{\text{var}_{\text{joint}}}{\text{var}_{\text{single}}}$ and $\frac{\text{bias}_{\text{joint}}}{\text{bias}_{\text{single}}}$. From the above, we get:

$$\frac{\text{var}_{\text{joint}}}{\text{var}_{\text{single}}} = 1 - \frac{m}{1+s_1^2 - m r_0^2} \frac{r^2 r_0^2 s_1^4}{(1+s_1^2)(1+s_2^2) - r^2},$$

$$\frac{\text{bias}_{\text{joint}}}{\text{bias}_{\text{single}}} = 1 - \frac{r^2 s_1^2}{(1+s_1^2)(1+s_2^2) - r^2}.$$

Here, we note that smaller values of this two ratios indicate a larger benefit from the joint model. The results above show the following: First, both ratios will decrease as r increases, i.e., the larger cross-correlation the two variables have, the larger benefit of the joint model. This is intuitive as the more highly correlated the responses are, the more information can be 'drawn' from the other response to improve the quantile predictor. Second, both ratios will decrease as s_2^2 decreases, i.e., the less noisy the expectation estimator is, the larger benefit we get. This is also intuitive as the noisier response (quantile) can benefit from the more accurate data of other response (expectation). To better illustrate the influence of s_1^2, s_2^2, r, we assess the change of the two ratios with these parameters numerically. In this numerical test, we set $m = 10$ and r_0 =95% of the maximum value it could take (to make the covariance matrix of $Z_1(x_0), Z_1(x_1), ..., Z_1(x_m)$ positive definite, $r_0^2 < 1/m$). The results are shown in Figure 1. In the first test (top two figures in Figure

1), s_2^2 is set to be 0.1. It is easy to see that both ratios will decrease as s_1^2 increases for fixed r. In the second test (bottom two figures in Figure 1), the cross-correlation r is set to be 0.7. Here again, it is very clear that both ratios will decrease as s_1^2 increases for fixed s_2^2. These results indicate that under similar conditions, a quantile function with larger noise will receive larger benefits from the joint model. In addition, these numerical results can also confirm our aforementioned conclusion about the influence of r and s_2^2.

It is noteworthy that the joint model may not provide satisfying performance when the cross-correlation r is too small. This is intuitive as this implies that the correlation between $Z_1(x)$ and $Z_2(x)$ is very small. For example, in the top figure in Figure 1, the two ratios are equal to 1 when $r = 0$, indicating no benefit from joint model. Moreover, when $s_1^2 = 0$, i.e., the problem becomes deterministic, the benefit of joint model will also vanish. This result agrees with the conclusions of Kleijnen and Mehdad (2014) and further illustrates that the joint model can be useful when the response of interest (quantile) is noisier, in that we could leverage and learn from a less noisy response (expectation).

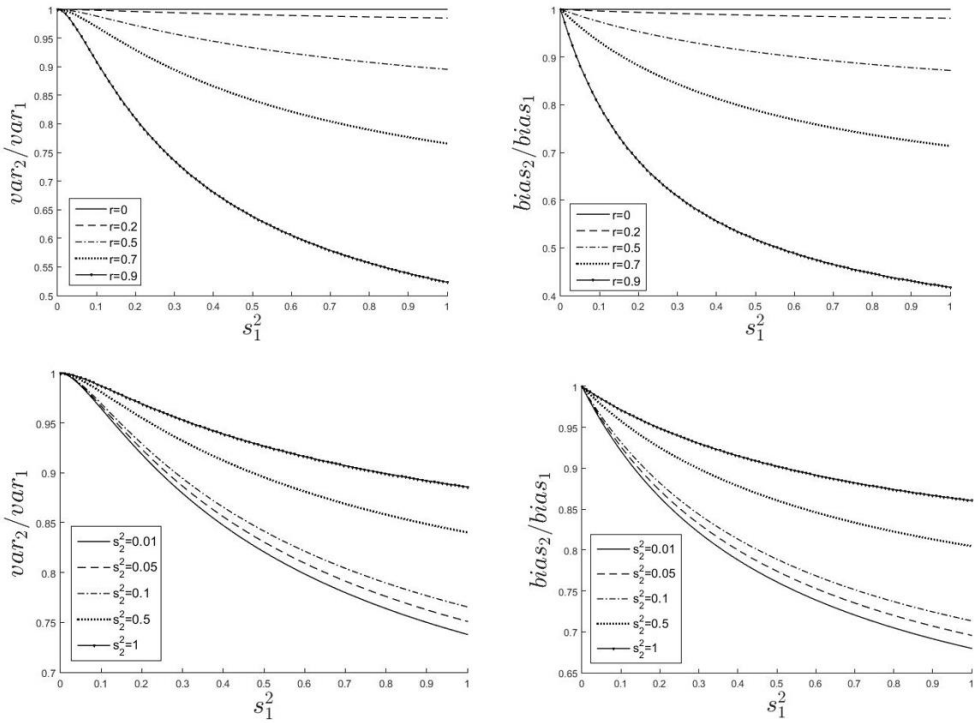

Figure 1: Change of the two ratios with s_1^2 (Top figures show the results with s_2^2=0.1, bottom figures show the results with $r = 0.7$).

4 NUMERICAL EXPERIMENT

In the simplifying example above, we assume that all parameters are known and $\rho = 0$. Here we relax this constraint and compare the joint model with the single model numerically where all the parameters of the models have to be estimated. The input space for this example is $\mathscr{X} = [0,1]$. At each input, the simulation output follows normal distribution with the following mean and variance:

$$\text{mean: } L(x) = (1+x^2)\sin(2x(5x+5)) - x,$$

$$\text{variance: } \sigma^2(x) = (6-5(x-0.75)^2)^2.$$

The 0.9-quantile and 0.95-quantile are to be determined in this example. The true mean and quantile functions are shown in Figure 2. Here, we also consider two different number of simulation runs, 500 and

5000. Within each experiment, there are 15 evenly distributed design points with identical runs at each point. Another 50 different points are selected to be the test points, and the root mean square error (RMSE) of these test points is used to compare the two models:

$$\text{RMSE} = \sqrt{\frac{1}{50}\sum_{i=1}^{50}\left(\widehat{Z}_1(x_i) - q_\alpha(x_i)\right)^2},$$

where $q_\alpha(x_i)$ represent the true α-quantile at x_i. Each experiment has been conducted by $M = 100$ micro-replications. The averaged results for each experiment is summarized in Table 1, where $\text{RMSE}_{\text{single}}$ and $\text{RMSE}_{\text{joint}}$ stands for the RMSE for stochastic GP and joint models, respectively.

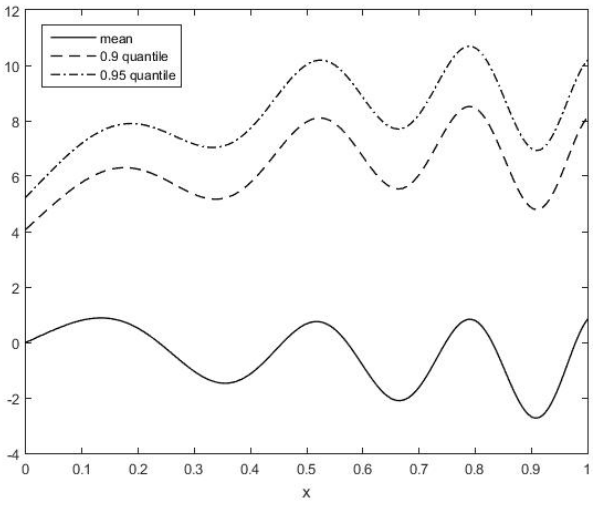

Figure 2: The true mean and quantile functions.

Table 1: The averaged results for each experiment with M=100 replications.

Test scenario	α	simulation runs	$\overline{\text{RMSE}}_{\text{single}}$	$\overline{\text{RMSE}}_{\text{joint}}$	$\overline{s_1^2}$	$\overline{s_2^2}$	$\bar{r}$
A	0.9	500	0.3712	0.3254^*	0.0938	0.0569	0.686
B	0.9	5000	0.1317	0.09882^*	0.00826	0.00576	0.704
C	0.95	500	0.4432	0.3622^*	0.156	0.0590	0.688
D	0.95	5000	0.1594	0.1149^*	0.0128	0.00584	0.703

From Table 1, we observe that the RMSE for both models are much smaller in test scenario B than that in A, which indicates that the predictions improve when the simulation replications increase. Similar results are observed from test scenarios C and D. In addition, the quantile prediction is enhanced in the joint model by the expectation information, providing a statistically smaller RMSE than the single model for all four test scenarios ($\overline{\text{RMSE}}^*_{\text{joint}}$ represents a statistically smaller $\overline{\text{RMSE}}_{\text{joint}}$ at 0.05 significance level). Besides, if we compare test A and C (or test B and D), the ratio $\overline{\text{RMSE}}_{\text{joint}}/\overline{\text{RMSE}}_{\text{single}}$ in A is 0.877 while it is 0.817 in test C with similar $\bar{r}$ and $\overline{s_2^2}$ in these two experiments. The only difference between test A and C is that test C has a much larger value of $\overline{s_1^2}$ (as the probability density of the 0.95 quantile is likely to be smaller and result in larger variance of the estimator, since this variance is approximately $\alpha(1-\alpha)/(nf(q_\alpha)^2)$ (Bahadur 1966), where α is the quantile level, n is the number of simulation runs and $f(q_\alpha)$ is the probability density of the true quantile). This comparison agrees with our conclusion in the m-design-point problem that the benefit of joint model will be larger with noisier quantile estimation.

Based on this numerical experiment, we can easily see the advantage of the joint model in predicting quantiles, especially when the number of simulation replications is small and limited, resulting in noisier quantile input point estimates. From this numerical example and the m-design-point problem, we see that when the correlation between quantile and expectation is high (which is true in many situations), the joint model will provide much better performance compared with a single model.

5 CONCLUSION

In this paper, we investigate quantile prediction with a joint quantile-expectation model. The idea is to leverage on the less noisy and unbiased expectation information to improve on the estimation and prediction of the quantile model. To develop this joint model, we first extend the standard stochastic GP model to the multi-response case for the quantile and expectation. To estimate the inputs to the model, we choose the sectioning method to obtain the point and variance estimators for the quantile and propose a similar method to estimate the correlation between the sample quantile and sample mean. Next, we analytically analyze a simple m-design-point problem to study the benefit of the joint stochastic GP model over a single response stochastic GP model and also numerically study a more complicated model where the parameters are unknown. The results from these two examples indicate that the larger the variances of the quantile input estimates are with respect to the variances of the mean, and the larger the correlations between responses, the larger the benefit of the joint model over the single response quantile model. This suggests that the joint model is promising and can be very useful if the expectation response (which is cheaper and easier to estimate) is highly correlated and less noisy than the quantile estimates.

In the analytical examples studied in this paper, we assumed that the correlation between the noises ρ equals to 0. In practice, however, the noises from the responses are likely to be correlated if they are estimated from the same simulation runs, hence further research can be done to study the effect of ρ. In addition, further investigations should also be conducted to study the impact of the noises on the parameter estimates, and the trade-offs between the additional parameters to estimate in the joint model and the precision of the estimates.

A Derivations For the m-Design-Point Problem

$$\boldsymbol{R}^{-1} = \begin{bmatrix} \boldsymbol{R}_1^{-1}(\boldsymbol{I}_m + \boldsymbol{R}_2\boldsymbol{P}\boldsymbol{R}_3\boldsymbol{R}_1^{-1}) & -\boldsymbol{R}_1^{-1}\boldsymbol{R}_2\boldsymbol{P} \\ -\boldsymbol{P}\boldsymbol{R}_3\boldsymbol{R}_1^{-1} & \boldsymbol{P} \end{bmatrix}.$$

Therefore

$$\text{bias}_2 = \boldsymbol{c}(x_0)^T\boldsymbol{R}^{-1}\boldsymbol{\xi} = (\boldsymbol{c}_1(x_0)^T, \boldsymbol{c}_2(x_0)^T) \begin{bmatrix} \boldsymbol{R}_1^{-1}(\boldsymbol{I}_m + \boldsymbol{R}_2\boldsymbol{P}\boldsymbol{R}_3\boldsymbol{R}_1^{-1}) & -\boldsymbol{R}_1^{-1}\boldsymbol{R}_2\boldsymbol{P} \\ -\boldsymbol{P}\boldsymbol{R}_3\boldsymbol{R}_1^{-1} & \boldsymbol{P} \end{bmatrix} \begin{pmatrix} \boldsymbol{\xi}_1 \\ \boldsymbol{0} \end{pmatrix}$$

$$= \boldsymbol{c}_1(x_0)^T\boldsymbol{R}_1^{-1}(\boldsymbol{I}_m + \boldsymbol{R}_2\boldsymbol{P}\boldsymbol{R}_3\boldsymbol{R}_1^{-1})\boldsymbol{\xi}_1 - \boldsymbol{c}_2(x_0)^T - \boldsymbol{P}\boldsymbol{R}_3\boldsymbol{R}_1^{-1}\boldsymbol{\xi}_1 = \text{bias}_1 - (\boldsymbol{c}_2(x_0)^T - \boldsymbol{c}_1(x_0)^T\boldsymbol{R}_1{}^{-1}\boldsymbol{R}_2)\boldsymbol{P}\boldsymbol{R}_3\boldsymbol{R}_1{}^{-1}\boldsymbol{\xi}_1$$

$$= \frac{r_0}{1+s_1^2}\left(1 - \frac{r^2 s_1^2}{(1+s_1^2)(1+s_2^2) - r^2}\right)\sum_{i=1}^{m}\xi_1(x_m).$$

Since $r^2 \leq 1$, then,

$$(1+s_1^2)(1+s_2^2) - r^2 - r^2 s_1^2 = (1+s_2^2 - r^2)(1+s_1^2) > 0 \Rightarrow 1 - \frac{r^2 s_1^2}{(1+s_1^2)(1+s_2^2) - r^2} > 0 \Rightarrow \text{bias}_2 > 0.$$

Also,

$$\text{bias}_1 - \text{bias}_2 = \frac{r_0}{1+s_1^2}\frac{r^2 s_1^2}{(1+s_1^2)(1+s_2^2) - r^2} > 0,$$

it follows that $0 < \text{bias}_2 < \text{bias}_1$.

$$\text{var}_2 = \sigma_{z1}^2 - \boldsymbol{c}(x_0)^T \boldsymbol{R}^{-1} \boldsymbol{c}(x_0) = \sigma_{z1}^2 - (\boldsymbol{c}_1(x_0)^T, \boldsymbol{c}_2(x_0)^T) \begin{bmatrix} \boldsymbol{R}_1^{-1}(\boldsymbol{I}_m + \boldsymbol{R}_2 \boldsymbol{P} \boldsymbol{R}_3 \boldsymbol{R}_1^{-1}) & -\boldsymbol{R}_1^{-1} \boldsymbol{R}_2 \boldsymbol{P} \\ -\boldsymbol{P} \boldsymbol{R}_3 \boldsymbol{R}_1^{-1} & \boldsymbol{P} \end{bmatrix} \begin{pmatrix} \boldsymbol{c}_1(x_0) \\ \boldsymbol{c}_2(x_0) \end{pmatrix}$$

$$= \sigma_{z1}^2 - \boldsymbol{c}_1(x_0)^T {\boldsymbol{R}_1}^{-1} \boldsymbol{c}_1(x_0) - \boldsymbol{\Gamma}^T \boldsymbol{P} \boldsymbol{\Gamma} = \text{var}_1 - \boldsymbol{\Gamma}^T \boldsymbol{P} \boldsymbol{\Gamma},$$

where,

$$\boldsymbol{P}^{-1} = \boldsymbol{R}_4 - \boldsymbol{R}_3 \boldsymbol{R}_1^{-1} \boldsymbol{R}_2 = \sigma_{z2}^2 (1 + s_2^2) \boldsymbol{I}_m - \frac{\sigma_{z12}^4}{\sigma_{z1}^2} \frac{1}{1 + s_1^2} \boldsymbol{I}_m = \sigma_{z2}^2 (1 + s_2^2 - \frac{r^2}{1 + s_1^2}) \boldsymbol{I}_m,$$

since $1 + s_2^2 > 1 > \frac{r^2}{1+s_1^2}$, it follows that $\boldsymbol{P}^{-1}$ is a diagonal matrix whose diagonal entries are all positive. Therefore $\boldsymbol{P}^{-1}$ is positive definite, and so as $\boldsymbol{P}$.

REFERENCES

Álvarez, M. A., and N. D. Lawrence. 2011. "Computationally efficient convolved multiple output gaussian processes". *Journal of Machine Learning Research* 12 (May): 1459–1500.

Ankenman, B., B. L. Nelson, and J. Staum. 2010. "Stochastic kriging for simulation metamodeling". *Operations research* 58 (2): 371–382.

Asmussen, S., and P. W. Glynn. 2007. *Stochastic simulation: algorithms and analysis*, Volume 57. Springer Science & Business Media.

Bahadur, R. R. 1966. "A note on quantiles in large samples". *The Annals of Mathematical Statistics* 37 (3): 577–580.

Chen, E. J. 2009. "Metamodels for estimating quantiles of systems with one controllable parameter". *Simulation* 85 (5): 307–317.

Chen, X., B. E. Ankenman, and B. L. Nelson. 2013. "Enhancing stochastic kriging metamodels with gradient estimators". *Operations Research* 61 (2): 512–528.

Chen, X., and K.-K. Kim. 2014. "Stochastic kriging with biased sample estimates". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 24 (2): 8.

Chen, X., and K.-K. Kim. 2016. "Efficient VaR and CVaR measurement via stochastic kriging". *INFORMS Journal on Computing* 28 (4): 629–644.

Conti, S., and A. OHagan. 2010. "Bayesian emulation of complex multi-output and dynamic computer models". *Journal of statistical planning and inference* 140 (3): 640–651.

Cressie, N. 2015. *Statistics for spatial data*. John Wiley & Sons.

Dabo-Niang, S., and B. Thiam. 2010. "Robust quantile estimation and prediction for spatial processes". *Statistics & probability letters* 80 (17): 1447–1458.

Fricker, T. E., J. E. Oakley, and N. M. Urban. 2013. "Multivariate Gaussian process emulators with nonseparable covariance structures". *Technometrics* 55 (1): 47–56.

Gordy, M. B., and S. Juneja. 2010. "Nested simulation in portfolio risk measurement". *Management Science* 56 (10): 1833–1848.

Goulard, M., and M. Voltz. 1992. "Linear coregionalization model: tools for estimation and choice of cross-variogram matrix". *Mathematical Geology* 24 (3): 269–286.

Gupta, A., Y. Ding, L. Xu, and T. Reinikainen. 2006. "Optimal parameter selection for electronic packaging using sequential computer simulations". *Journal of Manufacturing Science and Engineering* 128 (3): 705–715.

Hernandez, A. F., and M. A. Grover. 2013. "Error estimation properties of Gaussian process models in stochastic simulations". *European Journal of Operational Research* 228 (1): 131–140.

Higdon, D. 2002. "Space and space-time modeling using process convolutions". In *Quantitative methods for current environmental issues*, 37–56. Springer.

Joseph, V. R. 2006. "Limit kriging". *Technometrics* 48 (4): 458–466.

Journel, A. G., and C. J. Huijbregts. 1978. *Mining geostatistics*. Academic press.
Kennedy, M. C., and A. O'Hagan. 2001. "Bayesian calibration of computer models". *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63 (3): 425–464.
Kim, J. H. T., and M. R. Hardy. 2007. "Quantifying and correcting the bias in estimated risk measures". *Astin Bulletin* 37 (02): 365–386.
Kleijnen, J. P. 2008. *Design and analysis of simulation experiments*, Volume 20. Springer.
Kleijnen, J. P., and E. Mehdad. 2014. "Multivariate versus univariate Kriging metamodels for multi-response simulation models". *European Journal of Operational Research* 236 (2): 573–582.
Koenker, R. 2005. *Quantile regression*. Number 38. Cambridge university press.
McFarland, J., S. Mahadevan, V. Romero, and L. Swiler. 2008. "Calibration and uncertainty analysis for computer simulations with multivariate output". *AIAA journal* 46 (5): 1253–1265.
Mitchell, T., M. Morris, and D. Ylvisaker. 1994. "Asymptotically optimum experimental designs for prediction of deterministic functions given derivative information". *Journal of statistical planning and inference* 41 (3): 377–389.
Nakayama, M. K. 2014. "Confidence intervals for quantiles using sectioning when applying variance-reduction techniques". *ACM Transactions on Modeling and Computer Simulation (TOMACS)* 24 (4): 19.
Nance, R. E., and R. G. Sargent. 2002. "Perspectives on the evolution of simulation". *Operations Research* 50 (1): 161–172.
Pham, T., and M. Wagner. 1999. "Filtering noisy images using kriging". In *Signal Processing and Its Applications, 1999. ISSPA'99. Proceedings of the Fifth International Symposium on*, Volume 1, 427–430. IEEE.
Rougier, J. 2008. "Efficient emulators for multivariate deterministic functions". *Journal of Computational and Graphical Statistics* 17 (4): 827–843.
Seila, A. F. 1982. "A batching approach to quantile estimation in regenerative simulations". *Management Science* 28 (5): 573–581.
Ver Hoef, J. M., and R. P. Barry. 1998. "Constructing and fitting models for cokriging and multivariable spatial prediction". *Journal of Statistical Planning and Inference* 69 (2): 275–294.
Yin, J., S. H. Ng, and K. M. Ng. 2009. "A study on the effects of parameter estimation on Kriging model's prediction error in stochastic simulations". In *Proceedings of the 2009 Winter Simulation Conference*, edited by M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, 674–685. Austin, TX, USA: Institute of Electrical and Electronics Engineers, Inc.
Yin, J., S. H. Ng, and K. M. Ng. 2011. "Kriging metamodel with modified nugget-effect: The heteroscedastic variance case". *Computers & Industrial Engineering* 61 (3): 760–777.

AUTHOR BIOGRAPHIES

SONGHAO WANG is a Ph.D. student in the Department of Industrial Systems Engineering and Management at National University of Singapore. He received his B.S. in Modern Mechanics from University of Science and Technology of China in 2015. His research interests include experiment design and metamodeling. His email address is wangsonghao@u.nus.edu.

SZU HUI NG is an Associate Professor in the Department of Industrial Systems Engineering and Management at the National University of Singapore. She holds B.S., M.S., and Ph.D. degrees in Industrial and Operations Engineering from the University of Michigan. Her research interests include computer simulation modeling and analysis, design of experiments, and quality and reliability engineering. She is a member of IEEE and INFORMS and a senior member of IIE. Her email address is isensh@nus.edu.sg and her web page is https://www.isem.nus.edu.sg/staff/ngsh/index.html .