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 HETEROSCEDASTIC T-PROCESS SIMULATION METAMODELING APPROACH AND ITS APPLICATION IN INVENTORY CONTROL AND OPTIMIZATION

Guangrui Xie
Xi Chen

Grado Department of Industrial and Systems Engineering
Virginia Tech
Blacksburg, VA 24061, USA

ABSTRACT

In this paper we develop a heteroscedastic t-process metamodeling approach (TP) for approximating the mean response surface implied by a stochastic simulation and performing metamodel-based optimization. We provide details on how to construct a TP metamodel, make inference and perform prediction based on TP. We show that TP can retain the attractive properties of approaches that rely on Gaussian processes, but it also enjoys enhanced flexibility, at no additional computational cost. We further provide a closed-form expression for the TP-based expected improvement to perform metamodel-based optimization. We compare the predictive performance of TP and stochastic kriging (SK) via an M/M/1 inspired example, and demonstrate the performance of TP and SK-based algorithms for optimizing a simple periodic review (s,S) inventory system. The preliminary numerical evaluations indicate that TP can serve as a promising simulation analytics tool for solving problems encountered in production planning and supply chain management.

1 INTRODUCTION

Many complex stochastic systems can be accurately modeled and further analyzed by simulation experiments. In some situations, however, simulation models can be very expensive to run; and if there are many scenarios to evaluate, computational cost quickly becomes prohibitive. Due to the lack of efficiency of such simulation experiments, a metamodel is often built from the outputs of the simulation runs at some selected design points to "map" the performance response surface as a function of the controllable decision variables, or uncontrollable environmental variables. This metamodel can be used as an accurate, drop-in replacement for the simulation model as if the simulation can be run "on demand" to support decision making in a timely manner. Successful applications of simulation metamodeling have been recorded in many areas, such as inventory control, resource planning in the emergency departments, optimization of energy systems and financial risk management.

Simulation metamodeling has been used to solve problems encountered in production planning of manufacturing systems and supply chain management. For instance, Yang (2010) studied the use of neural network metamodeling for estimation of cycle time-throughput profiles in manufacturing systems; Song, Li, and Garcia-Diaz (2008) performed polynomial regression metamodel-based optimization of a multi-echelon supply chain via simulation; Mortazavi and Arshadi Khamseh (2014) proposed a simulation metamodel-based framework for multi-objective robust optimization of an inventory system; and Dellino, Kleijnen, and Meloni (2012) proposed a Taguchian robust optimization approach based on kriging metamodels and applied it to solve an inventory control problem. Our work will contribute to this line of research by providing a new robust metamodeling tool.

Commonly used metamodeling methods include polynomial regression, splines, radial basis functions, neural networks, and Gaussian process (GP) models, to name a few (Barton and Meckesheimer 2006). Among them, GP models have been recognized as the most widely adopted analytics tool with impressive

empirical performance in real-world applications. One primary reason for GP's popularity is that it unites sophisticated and consistent theoretical investigations with computational tractability (Rasmussen and Williams 2006).

Recent literature has highlighted the benefits of using alternative models instead of GP models, e.g., to achieve high robustness against outliers in both input and output spaces and heavy-tailed observation noise. One of such early attempts is to use a Student-t observation model in GP regression (GPR) by Neal (1997). Assuming that the observation noise terms have a joint multivariate Student-t distribution while using a GP to model the underlying mean function, this line of work focuses on making the GP models more robust with a Student-t likelihood (see, e.g., Jylänki, Vanhatalo, and Vehtari 2011). The main challenge, however, is that the resulting inference is no longer analytically tractable and approximative inference methods have to be used. Another stream of work assumes that not only observation noise/error terms are multivariate t-distributed but the underlying mean function is generated by a t-process as well. Such a modeling perspective is considered "perhaps not as exciting as one might have hoped" by Rasmussen and Williams (2006) since the t distribution is not closed under addition, and nice properties of GPR models such as analytically tractable inference are lost.

With a deeper understanding of properties of t distribution, t process and elliptical processes at large (Arellano-Valle and Bolfarine 1995; Girón and del Castillo 2010; Kotz and Nadarajah 2004; Roth 2013), research in the field has taken significant strides over the past decade—analytical and computational tractability has been shown to hold for some t-process models that are constructed hierarchically (Zhang and Yeung 2010; Shah, Wilson, and Ghahramani 2014; Wang, Shi, and Lee 2015). Inspired by the extended t-process regression models developed by Wang, Shi, and Lee (2015), in this paper we propose the heteroscedastic t-process metamodeling approach (TP) for stochastic simulation metamodeling. We show that TP is closed under addition and it retains many desirable properties as GP models do. In particular, the inference procedure for TP metamodels is relatively simple and a slight modification of the existing algorithms for a GP model (e.g., stochastic kriging) would suffice. Furthermore, we derive closed-form expressions that specify the predictive distributions and the expected improvement based on TP.

The remainder of this paper is organized as follows. Section 2 provides a detailed review of the extended t-process regression models. Section 3 establishes the heteroscedastic t-process (TP) metamodeling approach, by providing details on performing prediction, parameter estimation and metamodel-based optimization using TP. Section 4 presents numerical examples that demonstrate the performance of TP relative to stochastic kriging (Ankenman, Nelson, and Staum 2010) with respect to mean function estimation and metamodel-based optimization. Section 5 concludes this paper.

2 A REVIEW ON EXTENDED T-PROCESS REGRESSION MODELS

Consider the following hierarchical process to model Y, a real-valued function such that $\mathsf{Y}: \mathscr{X} \subset \mathbb{R}^d \to \mathbb{R}$.

$$\mathsf{Y}|r \sim GP(\mu, r\mathscr{R}_M), \qquad r \sim IG(\nu, \omega),$$

where $GP(\mu, r\mathscr{R}_M)$ represents a Gaussian process with mean function μ and covariance function $r\mathscr{R}_M$, with $\mathscr{R}_M$ denoting a kernel function which is equivalent to a covariance function used by, e.g., Santner, Williams, and Notz (2003). For instance, $\mathscr{R}_M(\mathbf{x}_i, \mathbf{x}_\ell) = \tau^2 \exp\left(\sum_{j=1}^d \theta_j \left(x_{ij} - x_{\ell j}\right)^2\right)$, with the θ_j's and τ^2 being the hyperparameters. $IG(\nu, \omega)$ denotes an inverse gamma distribution parameterized by two positive parameters ν and ω, and the corresponding density function is

$$g(r) = \frac{\omega^\nu}{\Gamma(\nu)} r^{-\nu-1} \exp\left(-\frac{\omega}{r}\right), \ r > 0$$

and $\Gamma(\cdot)$ is the gamma function. It turns out that Y follows an extended t-process (Wang, Shi, and Lee 2015), that is, $\mathsf{Y} \sim ETP(\nu, \omega, \mu, \mathscr{R}_M)$. This implies that for any collection of a finite number of points in $\mathscr{X}$,

say, $\mathbf{X}_k = (\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k)^\top$, we have that $\mathsf{Y}(\mathbf{X}_k) := (\mathsf{Y}(\mathbf{x}_1), \mathsf{Y}(\mathbf{x}_2), \ldots, \mathsf{Y}(\mathbf{x}_k))^\top$ has an extended multivariate t-distribution (EMTD), namely,

$$\mathsf{Y}(\mathbf{X}_k) \sim \mathrm{EMTD}_k(\nu, \omega, \mu(\mathbf{X}_k), \Sigma_{\mathsf{M}}),$$

with the following density function

$$p(z) = |2\pi\omega\Sigma_{\mathsf{M}}|^{-\frac{1}{2}} \frac{\Gamma(\frac{k}{2}+\nu)}{\Gamma(\nu)} \left(1 + \frac{(z-\mu(\mathbf{X}_k))^\top \Sigma_{\mathsf{M}}^{-1}(z-\mu(\mathbf{X}_k))}{2\omega}\right)^{-(\frac{k}{2}+\nu)}, \tag{1}$$

with $\mu(\mathbf{X}_k) = (\mu(\mathbf{x}_1), \ldots, \mu(\mathbf{x}_k))^\top$, $\Sigma_{\mathsf{M}} = (\mathscr{R}_M(\mathbf{x}_i, \mathbf{x}_j))_{k\times k}$. We have that $\mathrm{E}(\mathsf{Y}(\mathbf{X}_k)) = \mu(\mathbf{X}_k)$ when $\nu > 1/2$ and $\mathrm{Cov}(\mathsf{Y}(\mathbf{X}_k)) = \omega\Sigma_{\mathsf{M}}/(\nu-1)$ when $\nu > 1$. Following Lee and Nelder (2006) and Wang, Shi, and Lee (2015), from Section 3 and on we set $\omega = \nu - 1$ so that the covariance matrix $\mathrm{Cov}(\mathsf{Y}(\mathbf{X}_k)) = \Sigma_{\mathsf{M}}$ and does not depend on ν. In this case, the mean and variance-covariance matrix of an ETMD take the same form as those of a multivariate normal distribution.

The following properties of EMTD are crucial for the development of the heteroscedastic t-process metamodeling approach detailed in Section 3.

Proposition 1 (Wang, Shi, and Lee 2015)**.** *Let $\mathbf{Z} \in \mathscr{X}$ be a $k \times 1$ random vector such that $\mathbf{Z} \sim \mathrm{EMTD}_k(\nu, \omega, \mu_z, \Sigma_z)$.*

(a) For any matrix $\mathbf{A} \in \mathbb{R}^{\ell\times k}$ with rank $\ell \le p$, $\mathbf{AZ} \sim \mathrm{EMTD}_\ell(\nu, \omega, \mathbf{A}\mu_z, \mathbf{A}\Sigma_z\mathbf{A}^\top)$.

(b) Write $\mathbf{Z} = \left(\mathbf{Z}_1^\top, \mathbf{Z}_2^\top\right)^\top$ where the dimensions of $\mathbf{Z}_1$ and $\mathbf{Z}_2$ are respectively $k_1 \times 1$ and $k_2 \times 1$ such that $k_1 + k_2 = k$. Also write $\mu_z = \left(\mu_{z_1}^\top, \mu_{z_2}^\top\right)^\top$ and

$$\Sigma_z = \begin{pmatrix} \Sigma_{z_{11}} & \Sigma_{z_{12}} \\ \Sigma_{z_{12}}^\top & \Sigma_{z_{22}} \end{pmatrix}.$$

Then it follows that

$$\mathbf{Z}_1 \sim \mathrm{EMTD}_{k_1}(\nu, \omega, \mu_z, \Sigma_{z_{11}}),\ \mathbf{Z}_2|\mathbf{Z}_1 = \mathbf{z}_1 \sim \mathrm{EMTD}_{k_2}(\nu^*, \omega^*, \mu_{z_2}^*, \Sigma_{z_{22}}^*),$$

where $\nu^* = k_1/2 + \nu$, $\omega^* = k_1/2 + \omega$, $\mu_{z_2}^* = \Sigma_{z_{12}}^\top \Sigma_{z_{11}}^{-1}(\mathbf{z}_1 - \mu_{z_1}) + \mu_{z_2}$, and

$$\Sigma_{z_{22}}^* = \left(2\omega + (\mathbf{z}_1 - \mu_{z_1})^\top \Sigma_{z_{11}}^{-1}(\mathbf{z}_1 - \mu_{z_1})\right) \widetilde{\Sigma}_{z_{22}}/(2\omega + k_1)$$

with $\widetilde{\Sigma}_{z_{22}} = \Sigma_{z_{22}} - \Sigma_{z_{12}}^\top \Sigma_{z_{11}}^{-1} \Sigma_{z_{12}}$.

3 A HETEROSCEDASTIC T-PROCESS SIMULATION METAMODELING APPROACH

3.1 Basic Setup for the Heteroscedastic T-Process Simulation Metamodeling Approach

Consider the following model for a random simulation output obtained at a design point $\mathbf{x} \in \mathscr{X} \subset \mathbb{R}^d$,

$$\mathscr{Y}(\mathbf{x}) = \mathsf{Y}(\mathbf{x}) + \varepsilon(\mathbf{x}) = \mathbf{f}(\mathbf{x})^\top \beta + \mathsf{M}(\mathbf{x}) + \varepsilon(\mathbf{x}), \tag{2}$$

where $\mathsf{Y}(\mathbf{x})$ represents the true mean simulation response at $\mathbf{x}$, $\mathbf{f}(\cdot)$ denotes a $p \times 1$ vector of known regression functions, and β denotes a $p \times 1$ vector of unknown parameters. Let us assume that $\mathsf{M}(\cdot)$ and $\varepsilon(\cdot)$ have a joint extended t-process,

$$\begin{pmatrix} \mathsf{M} \\ \varepsilon \end{pmatrix} \sim ETP\left(\nu, \nu-1, \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} r\mathscr{R}_M & 0 \\ 0 & r\mathscr{R}_\varepsilon \end{pmatrix}\right), \tag{3}$$

and $\varepsilon(\mathbf{x})$ denotes the simulation noise incurred at $\mathbf{x}$ and $\mathscr{R}_\varepsilon$ represents the noise kernel function satisfying $\mathscr{R}_\varepsilon(\mathbf{x},\mathbf{x}') = 0$ if $\mathbf{x} \neq \mathbf{x}'$, and $\mathscr{R}_\varepsilon(\mathbf{x},\mathbf{x}) = \mathsf{V}(\mathbf{x})$. As described in Subsection 2, such a joint ETP can be constructed hierarchically as

$$\begin{pmatrix} \mathsf{M} \\ \varepsilon \end{pmatrix} \Big| r \sim GP\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, r \begin{pmatrix} \mathscr{R}_M & 0 \\ 0 & \mathscr{R}_\varepsilon \end{pmatrix} \right) \qquad \text{and} \qquad r \sim IG(\nu, \nu-1).$$

By properties of extended t-process, we have $\mathscr{Y} = \mathsf{Y} + \varepsilon \sim ETP(\nu, \nu-1, \mathbf{f}(\cdot)^\top\beta, \mathscr{R}_M + \mathscr{R}_\varepsilon)$. Therefore, similar to GP, additivity and many other properties can be shown to hold for ETP.

Let us consider a simulation experimental design that consists of $\{(\mathbf{x}_i, n_i), i = 1,2,\ldots,k\}$. That is, running n_i simulation replications at $\mathbf{x}_i \in \mathscr{X} \subseteq \mathbb{R}^d$ that results in independent and identically distributed (i.i.d.) simulation outputs $\{\mathscr{Y}_j(\mathbf{x}_i)\}_{j=1}^{n_i}$, $i = 1,2,\ldots,k$. Suppose model (2) and its related discussion can be used to describe the random simulation output generated on the jth simulation replication at design point $\mathbf{x}_i$, that is,

$$\mathscr{Y}_j(\mathbf{x}_i) = \mathsf{Y}(\mathbf{x}_i) + \varepsilon_j(\mathbf{x}_i), \;\; j = 1,2,\ldots,n_i; \;\; i = 1,2,\ldots,k. \tag{4}$$

Let the $k \times 1$ vector of sample average simulation responses be $\bar{\mathscr{Y}}(\mathbf{X}_k) = \left(\bar{\mathscr{Y}}(\mathbf{x}_1), \bar{\mathscr{Y}}(\mathbf{x}_2), \ldots, \bar{\mathscr{Y}}(\mathbf{x}_k)\right)^\top$ with $\bar{\mathscr{Y}}(\mathbf{x}_i) = \sum_{j=1}^{n_i} \mathscr{Y}_j(\mathbf{x}_i)/n_i$, and use $\bar{\varepsilon}$ as a shorthand for the $k \times 1$ vector of average simulation noise, $\bar{\varepsilon}(\mathbf{X}_k) = (\bar{\varepsilon}(\mathbf{x}_1), \bar{\varepsilon}(\mathbf{x}_2), \ldots, \bar{\varepsilon}(\mathbf{x}_k))^\top$, where $\bar{\varepsilon}(\mathbf{x}_i) = \sum_{j=1}^{n_i} \varepsilon_j(\mathbf{x}_i)/n_i$, for $i = 1,2,\ldots,k$; $\mathsf{Y}(\mathbf{X}_k)$ is defined in a similar fashion as $\bar{\mathscr{Y}}(\mathbf{X}_k)$ and $\bar{\varepsilon}(\mathbf{X}_k)$. It then follows from Proposition 1 given in Subsection 2 that

$$\begin{aligned} \mathsf{Y}(\mathbf{X}_k) &\sim \text{EMTD}_k\left(\nu, \nu-1, \mathbf{F}\beta, \Sigma_{\mathsf{M}}(\mathbf{X}_k)\right), \\ \bar{\mathscr{Y}}(\mathbf{X}_k)|\mathsf{Y}(\mathbf{X}_k) &\sim \text{EMTD}_k\left(\nu, \nu-1, \mathsf{Y}(\mathbf{X}_k), \Sigma_\varepsilon(\mathbf{X}_k)\right), \\ \bar{\mathscr{Y}}(\mathbf{X}_k) &\sim \text{EMTD}_k(\nu, \nu-1, \mathbf{F}\beta, \Sigma(\mathbf{X}_k)), \end{aligned} \tag{5}$$

where the $k \times p$ model matrix $\mathbf{F}$ of full rank is given by $\mathbf{F} = \left(\mathbf{f}(\mathbf{x}_1)^\top, \mathbf{f}(\mathbf{x}_2)^\top, \ldots, \mathbf{f}(\mathbf{x}_k)^\top\right)^\top$, and the $k \times k$ matrix $\Sigma(\mathbf{X}_k)$ is defined as $\Sigma(\mathbf{X}_k) = \Sigma_{\mathsf{M}}(\mathbf{X}_k) + \Sigma_\varepsilon(\mathbf{X}_k)$, with Σ_{M} being the $k \times k$ covariance matrix whose (i,h)th entry $\Sigma_{\mathsf{M}}(\mathbf{x}_i, \mathbf{x}_h)$ is given by $\text{Cov}(\mathsf{M}(\mathbf{x}_i), \mathsf{M}(\mathbf{x}_h))$ for $i,h = 1,2,\ldots,k$; and similar to stochastic kriging (Ankenman, Nelson, and Staum 2010), the intrinsic variance-covariance matrix is given by $\Sigma_\varepsilon(\mathbf{X}_k) = \text{diag}\{\mathsf{V}(\mathbf{x}_1)/n_1, \mathsf{V}(\mathbf{x}_2)/n_2, \ldots, \mathsf{V}(\mathbf{x}_k)/n_k\}$. The following result sheds some light on an important relationship between ETP and GP, the proof of which can be given following that of Lemma 2 in Shah, Wilson, and Ghahramani (2014).

Lemma 1. $\text{EMTD}_k(\nu, \nu-1, \mu(\mathbf{X}_k), \Sigma(\mathbf{X}_k))$ *converges to* $\mathscr{N}(\mu(\mathbf{X}_k), \Sigma(\mathbf{X}_k))$ *as* $\nu \to \infty$.

3.2 Prediction via the Heteroscedastic T-Process Metamodeling Approach

For a given prediction point $\mathbf{x}_0 \in \mathscr{X}$, the following holds

$$\begin{pmatrix} \bar{\mathscr{Y}}(\mathbf{X}_k) \\ \mathsf{Y}(\mathbf{x}_0) \end{pmatrix} \sim \text{EMTD}_{k+1}\left(\nu, \nu-1, \begin{pmatrix} \mathbf{F}\beta \\ \mathbf{f}(\mathbf{x}_0)^\top\beta \end{pmatrix}, \begin{pmatrix} \Sigma(\mathbf{X}_k) & \Sigma_{\mathsf{M}}(\mathbf{X}_k, \mathbf{x}_0) \\ \Sigma_{\mathsf{M}}(\mathbf{X}_k, \mathbf{x}_0)^\top & \Sigma_{\mathsf{M}}(\mathbf{x}_0, \mathbf{x}_0) \end{pmatrix}\right), \tag{6}$$

where $\Sigma_{\mathsf{M}}(\mathbf{X}_k, \mathbf{x}_0)$ denotes the $k \times 1$ vector of spatial covariances between $\mathsf{Y}(\mathbf{x}_0)$ and $\mathsf{Y}(\mathbf{x}_i)$ for $i = 1,2,\ldots,k$. Let the data set obtained from running a simulation experiment following the design $\{(\mathbf{x}_i, n_i), i = 1,2,\ldots,k\}$ be $\mathscr{D}_k = \{\mathbf{x}_i, \{\mathscr{Y}_j(\mathbf{x}_i)\}_{j=1}^{n_i}, i = 1,2,\ldots,k\}$. Assuming that all hyperparameters are known, it can be shown via arguments in light of Zellner (1976), Wang, Shi, and Lee (2015) and Section 3.3.1 of Santner, Williams, and Notz (2003) that the predictive distribution at $\mathbf{x}_0$ can be given as $\mathsf{Y}(\mathbf{x}_0)|\mathscr{D}_k \sim \text{EMTD}_1\left(\frac{k}{2} + \nu, \frac{k}{2} + \nu - 1, \mu_*(\mathbf{x}_0), \sigma_*^2(\mathbf{x}_0)\right)$, where

$$\mu_*(\mathbf{x}_0) = \Sigma_{\mathsf{M}}(\mathbf{X}_k, \mathbf{x}_0)^\top \Sigma(\mathbf{X}_k)^{-1}\left(\bar{\mathscr{Y}}(\mathbf{X}_k) - \mathbf{F}\widehat{\beta}\right) + \mathbf{f}(\mathbf{x}_0)^\top\widehat{\beta}, \tag{7}$$

$$\sigma_*^2(\mathbf{x}_0) = s(\mathbf{X}_k)\left(\Sigma_{\mathsf{M}}(\mathbf{x}_0,\mathbf{x}_0) - \Sigma_{\mathsf{M}}(\mathbf{X}_k,\mathbf{x}_0)^\top \Sigma(\mathbf{X}_k)^{-1}\Sigma_{\mathsf{M}}(\mathbf{X}_k,\mathbf{x}_0) + \eta(\mathbf{x}_0)^\top \left(\mathbf{F}^\top \Sigma(\mathbf{X}_k)^{-1}\mathbf{F}\right)^{-1}\eta(\mathbf{x}_0)\right), \quad (8)$$

with $\eta(\mathbf{x}_0) = \mathbf{f}(\mathbf{x}_0) - \Sigma_{\mathsf{M}}(\mathbf{X}_k,\mathbf{x}_0)^\top \Sigma(\mathbf{X}_k)^{-1}\mathbf{F}$, and $\widehat{\beta} = \left(\mathbf{F}^\top \Sigma(\mathbf{X}_k)^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^\top \Sigma(\mathbf{X}_k)^{-1}\bar{\mathscr{Y}}(\mathbf{X}_k)$ is the generalized least squares estimator of β; furthermore,

$$s(\mathbf{X}_k) = \frac{2(\nu-1) + \left(\bar{\mathscr{Y}}(\mathbf{X}_k) - \mathbf{F}\widehat{\beta}\right)^\top \left(\mathbf{A}\Sigma(\mathbf{X}_k)\mathbf{A}^\top\right)^{-1}\left(\bar{\mathscr{Y}}(\mathbf{X}_k) - \mathbf{F}\widehat{\beta}\right)}{2(\nu-1)+k},$$

with $\mathbf{A} = \mathbf{I}_k - \mathbf{F}\left(\mathbf{F}^\top \Sigma(\mathbf{X}_k)^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^\top \Sigma(\mathbf{X}_k)^{-1}$ and $\mathbf{I}_k$ being the $k \times k$ identity matrix.

In particular, the predictive distribution at the k design points, $\mathsf{Y}(\mathbf{X}_k)|\mathscr{D}_k$, is given by $\mathrm{EMTD}_k(\frac{k}{2}+\nu, \frac{k}{2}+\nu-1, \mu_*(\mathbf{X}_k), \sigma_*^2(\mathbf{X}_k))$, where

$$\mu_*(\mathbf{X}_k) = \Sigma_{\mathsf{M}}(\mathbf{X}_k,\mathbf{X}_k)^\top \Sigma(\mathbf{X}_k)^{-1}\left(\bar{\mathscr{Y}}(\mathbf{X}_k) - \mathbf{F}\widehat{\beta}\right) + \mathbf{F}\widehat{\beta}, \quad (9)$$

$$\sigma_*^2(\mathbf{X}_k) = s(\mathbf{X}_k)\left(\Sigma_{\mathsf{M}}(\mathbf{X}_k)^\top \Sigma(\mathbf{X}_k)^{-1}\Sigma_\varepsilon + \eta(\mathbf{X}_k)^\top \left(\mathbf{F}^\top \Sigma(\mathbf{X}_k)^{-1}\mathbf{F}\right)^{-1}\eta(\mathbf{X}_k)\right). \quad (10)$$

We make the following comments on the predictive distribution for $\mathsf{Y}(\mathbf{x}_0)|\mathscr{D}_k$. First, one feature that distinguishes TP-based metamodeling prediction from the GP-based is that the predictive variance given in (8) for $\widehat{\mathsf{Y}}(\mathbf{x}_0)$ increases with the misfit at the design points via $s(\mathbf{X}_k)$, whereas the GP-based predictive variance does not depend on the fit at the design points. Second, the predictive mean given in (7) shares the common form with the GP-based predictive mean, which is unbiased for predicting $\mathsf{Y}(\mathbf{x}_0)$; yet the GP and TP-based predictive variances differ and the latter one depends on the parameter ν. Third, we see from (9) that if there were no simulation errors at the design points, TP-based approach would interpolate at each of the design point as kriging does and the predictive covariance matrix would be a matrix of zeros.

3.3 Parameter Estimation for the Heteroscedastic T-Process Metamodel

The log-likelihood function for $\bar{\mathscr{Y}}(\mathbf{X}_k)$ can be given based on (1) and (5) as follows,

$$\begin{aligned}\mathscr{L}(\bar{\mathscr{Y}}(\mathbf{X}_k)|\nu,\Psi) = {} & -\frac{k}{2}\log(2\pi(\nu-1)) - \frac{1}{2}\log|\Sigma(\mathbf{X}_k)| + \log(\Gamma(\frac{k}{2}+\nu)) - \log(\Gamma(\nu)) \\ & - \left(\nu+\frac{k}{2}\right)\log\left(1 + \frac{(\bar{\mathscr{Y}}(\mathbf{X}_k) - \mathbf{F}\beta)^\top \Sigma(\mathbf{X}_k)^{-1}(\bar{\mathscr{Y}}(\mathbf{X}_k) - \mathbf{F}\beta)}{2(\nu-1)}\right), \end{aligned} \quad (11)$$

where Ψ entails all the hyperparameters in the spatial covariance kernel $\mathscr{R}_M$, and given Ψ the MLE of β is the generalized least squares estimate $\widehat{\beta} = \widehat{\beta}(\Psi)$. Upon substituting the sample variance-covariance matrix $\widehat{\Sigma}_\varepsilon$ into $\Sigma(\mathbf{X}_k)$ as done by Ankenman, Nelson, and Staum (2010) and Chen and Kim (2014) for stochastic kriging, the parameter estimates can be obtained via maximum likelihood estimation by using the derivative of (11) with respective to each parameter in Ψ. Specifically, the derivative with respect to a given parameter Ψ_i can be obtained analytically as follows

$$\frac{\partial}{\partial\Psi_i}\mathscr{L}(\bar{\mathscr{Y}}(\mathbf{X}_k)|\nu,\Phi,\mu) = \frac{1}{2}\mathrm{Tr}\left(\left(\frac{(k+2\nu)\alpha_k\alpha_k^\top}{2(\nu-1)+\gamma_k} - \Sigma(\mathbf{X}_k)^{-1}\right)\frac{\partial\Sigma(\mathbf{X}_k)}{\partial\Psi_i}\right), \quad (12)$$

where $\alpha_k = \Sigma(\mathbf{X}_k)^{-1}(\bar{\mathscr{Y}}(\mathbf{X}_k) - \mathbf{F}\beta)$ and $\gamma_k = (\bar{\mathscr{Y}}(\mathbf{X}_k) - \mathbf{F}\beta)^\top \Sigma(\mathbf{X}_k)^{-1}(\bar{\mathscr{Y}}(\mathbf{X}_k) - \mathbf{F}\beta)\ \frac{\partial\Sigma(\mathbf{X}_k)}{\partial\Psi_i}$. Notice that ν is not estimable as argued in Wang, Shi, and Lee (2015) and Lee and Nelder (2006). Hence, we fix the value of ν while performing the maximum likelihood estimation of the hyperparameters.

3.4 Optimization via the Heteroscedastic T-Process Metamodeling Approach

Efficient global optimization (EGO) is a well-known GP-based sequential algorithm proposed for unconstrained optimization in the context of deterministic computer experiments by Jones, Schonlau, and Welch (1998). In each iteration, EGO selects a new point to evaluate the computer model via the expected improvement criterion (EI). The new point selected provides the maximum expected improvement in objective function value when compared to the best point that has been found so far. Over the past decades, various GP-based sequential optimization algorithms have been proposed in the stochastic simulation setting, see, e.g., Huang et al. (2006), Quan et al. (2013), Picheny et al. (2013) and Liu, Pedrielli, and Ng (2014). Most of these algorithms are variants of the classic EI criterion, which can be written (for minimization problems) as

$$\begin{aligned} \mathrm{EI}(\mathbf{x};\mathscr{D}) &= \mathrm{E}\left(\max(\mathrm{Y}_{\min}-\mathrm{Y}(\mathbf{x}),0)|\mathscr{D}\right) \\ &= (\mathrm{Y}_{\min}-\mu(\mathbf{x}))\,\Phi\left(\frac{\mathrm{Y}_{\min}-\mu(\mathbf{x})}{\sqrt{\sigma^2(\mathbf{x})}}\right)+\sqrt{\sigma^2(\mathbf{x})}\,\phi\left(\frac{\mathrm{Y}_{\min}-\mu(\mathbf{x})}{\sqrt{\sigma^2(\mathbf{x})}}\right), \end{aligned} \tag{13}$$

where $\mathscr{D}$ denotes the set of experiment results collected up to the current iteration, and $\mathrm{Y}_{\min}=\min_{\mathbf{X}}\mathrm{Y}(\mathbf{x}_i)$ represents the optimal objective function value obtained over $\mathbf{X}$, consisting of points $\mathbf{x}_i$'s evaluated up to the current iteration. The predictive mean and variance at an arbitrary point $\mathbf{x}$ given by a GP approach adopted are respectively denoted by $\mu(\mathbf{x})$ and $\sigma^2(\mathbf{x})$; and $\Phi(\cdot)$ and $\phi(\cdot)$ denote the cumulative distribution function (CDF) and probability density function (PDF) of the standard normal distribution, respectively.

Recently, it has been reported by Shah, Wilson, and Ghahramani (2014) that a Student-t process-based sequential optimization approach using its corresponding EI criterion can be a superior choice to its GP counterpart for deterministic computer experiments. Inspired by Shah, Wilson, and Ghahramani (2014), we can derive the expression of TP-based EI at a given point $\mathbf{x}$ as follows

$$\begin{aligned} \mathrm{EI}(\mathbf{x};\mathscr{D}) &= \mathrm{E}\left(\max(\widehat{\mathrm{Y}}_{\min}-\mathrm{Y}(\mathbf{x}),0)|\mathscr{D}\right) \\ &= \int_{-\infty}^{\widehat{\mathrm{Y}}_{\min}}\frac{\widehat{\mathrm{Y}}_{\min}-u}{\sqrt{\sigma_*^2(\mathbf{x})}}\lambda_{(\nu+\frac{k}{2},0,1)}\left(\frac{u-\mu_*(\mathbf{x})}{\sqrt{\sigma_*^2(\mathbf{x})}}\right)du \\ &= \gamma(\mathbf{x})\int_{-\infty}^{\gamma(\mathbf{x})}\sqrt{\sigma_*^2(\mathbf{x})}\lambda_{(\nu+\frac{k}{2},0,1)}(t)dt-\sqrt{\sigma_*^2(\mathbf{x})}\int_{-\infty}^{\gamma(\mathbf{x})}t\lambda_{(\nu+\frac{k}{2},0,1)}(t)dt \\ &= \gamma(\mathbf{x})\sqrt{\sigma_*^2(\mathbf{x})}\Lambda_{(\nu+\frac{k}{2},0,1)}(\gamma(\mathbf{x}))+\sqrt{\sigma_*^2(\mathbf{x})}\left(1+\frac{\gamma(\mathbf{x})^2-1}{2\nu+k-1}\right)\lambda_{(\nu+\frac{k}{2},0,1)}(\gamma(\mathbf{x})), \end{aligned} \tag{14}$$

where in the setting of stochastic simulation $\widehat{\mathrm{Y}}_{\min}$ denotes the predicted minimum over $\mathbf{X}$ (see details on how to obtain $\widehat{\mathrm{Y}}_{\min}$ in Subsection 4.2.2), $\gamma(\mathbf{x})=\left(\widehat{\mathrm{Y}}_{\min}-\mu_*(\mathbf{x})\right)/\sqrt{\sigma_*^2(\mathbf{x})}$, with $\mu_*(\mathbf{x})$ and $\sigma_*^2(\mathbf{x})$ being the predictive mean and variance at $\mathbf{x}$ given by (7) and (8), respectively; $\lambda_{(\nu,\mu,\sigma^2)}$ and $\Lambda_{(\nu,\mu,\sigma^2)}$ denote the PDF and CDF of $\mathrm{EMTD}_1(\nu,\nu-1,\mu,\sigma^2)$, respectively. A TP-based sequential algorithm can be developed based on (14) for solving a stochastic simulation optimization problem.

Notice that EMTD (as specified by the density form given in (1)) with different combinations of ν and ω used in fact corresponds to various multivariate t distributions studied in the literature (Wang, Shi, and Lee 2015). In particular, $\mathrm{EMTD}_1(\frac{\nu}{2},\frac{\nu}{2}-1,\mu,\sigma^2)$ corresponds to the form of $MVT_1(\nu,\mu,\sigma^2)$ proposed by Shah, Wilson, and Ghahramani (2014) and $\mathrm{EMTD}_1(\frac{\nu}{2},\frac{\nu}{2},\mu,\sigma^2)$ is equivalent to the form of multivariate t distribution $St_1(\nu,\mu,\sigma^2)$ used by Roth (2013), Rasmussen and Williams (2006) and MATLAB. Therefore, (14) derived above can be easily utilized for optimization if other t-process based approaches are adopted.

4 NUMERICAL EXAMPLES

Through two numerical examples, in this section we compare the TP metamodeling approach with one popular simulation metamodeling approach, i.e., stochastic kriging (SK), with respect to mean response surface prediction and metamodel-based simulation optimization.

4.1 An M/M/1 Inspired Example

Let us consider estimating the steady-state mean number in an M/M/1 queue with service rate 1 and arrival rate x varying in $\mathscr{X} = [0.3, 0.9]$. It is known that the steady-state mean number of customers in an M/M/1 queue with service rate 1 and arrival rate x is $x/(1-x)$, for $x \in (0,1)$. If a simulation replication is initiated at time 0 in steady state given the arrival rate x, then the average number of customers in system during the first T time units is an unbiased estimator for $\mathsf{Y}(x)$, and its variance is approximately $2x(1+x)/(T(1-x)^4)$ for large T (Whitt 1989).

We consider the following M/M/1 inspired example. The simulated response on each replication at a point x is generated as $\mathscr{Y}_j(x) = \mathsf{Y}(x) + \varepsilon_j(x)$, with $\varepsilon_j(x)$'s being i.i.d. $\mathscr{N}(0, \mathsf{V}(x))$ random variables, where $\mathsf{Y}(x) = x/(1-x)$ and $\mathsf{V}(x) = 2x(1+x)/(T(1-x)^4)$ with $x \in [0.3, 0.9]$ and $T = 1000$. Notice that the simulation variance increases explosively with x which makes the prediction task considerably challenging.

The design-point set consists of k design points in $[0.3, 0.9]$ and we consider varying $k \in \{7, 25\}$. A total simulation budget N is allocated to the k design points and we want to evaluate the predictive performance of TP and SK by varying $N \in \{500, 2000\}$. A simulation experiment is performed with a total simulation budget of N replications to expend at k distinct design points, with n_i simulation replications assigned at design point x_i, for $i = 1, 2, \ldots, k$. Two budget allocation schemes, i.e., the equal and unequal allocations, are considered. Specifically, the unequal budget allocation scheme sets $n_i = \left\lceil \frac{\sqrt{\mathsf{V}(x_i)}}{\sum_{i=1}^{k} \sqrt{\mathsf{V}(x_i)}} N \right\rceil$, while under the equal allocation scheme $n_i = \lceil N/k \rceil$ is used; and $\lceil a \rceil$ denotes the least integer not less than a. The entire experiment is repeated for 100 independent macro-replications, and the corresponding performance measure, the empirical root mean squared errors (ERMSE), is calculated as follows,

$$\text{ERMSE}_\ell = \sqrt{\frac{1}{K} \sum_{i=1}^{K} \left(\widehat{\mathsf{Y}}_\ell(x_i) - \mathsf{Y}(x_i) \right)^2}, \quad \ell = 1, 2, \ldots, 100, \tag{15}$$

where $\widehat{\mathsf{Y}}_\ell(\cdot)$ represents the prediction given by TP on the ℓth macro-replication; K represents the number of prediction points and $K = 193$.

Table 1: Summary of ERMSEs obtained for the M/M/1 inspired example.

(k, N)		$(7, 500)$	$(25, 500)$	$(7, 2000)$	$(25, 2000)$
Equal allocation	SK	0.27 (0.01)	0.35 (0.01)	0.154 (0.005)	0.180 (0.007)
	TP	0.22 (0.01)	0.24 (0.01)	0.167 (0.006)	0.133 (0.006)
	ν	25	1.25	25	1.75
Unequal allocation	SK	0.188 (0.005)	0.146 (0.004)	0.168 (0.004)	0.079 (0.002)
	TP	0.196 (0.006)	0.122 (0.004)	0.172 (0.004)	0.065 (0.002)
	ν	25	21	25	11.5

Summary of results. The mean and standard error of 100 ERMSEs obtained by SK and TP under different (k, N) settings are shown in Table 1 (with standard error given in parentheses). For implementing the proposed TP approach under each (k, N) setting, we have varied the values of ν from 1.25 to 25 to investigate the impact of ν on TP's predictive performance. It is observed that the mean ERMSE achieved by TP

gets closer to that of SK as ν becomes larger, which is consistent with Lemma 1. The value of ν shown in Table 1 under each (k,N) setting is the one used by TP to obtain the minimum mean ERMSE across the 100 macro-replications. From these ν values, we observe that the minimum mean EMSRE is typically obtained with $\nu = 25$ for a sparse design (i.e., $k = 7$) whereas for a dense design (i.e., $k = 25$) the minimum ERMSEs are achieved with smaller ν values. Regarding the predictive performance of TP as compared to SK, we observe that TP outperforms SK when a dense design is used, regardless of the budget allocation scheme adopted. On the other hand, SK tends to outperform TP when the design is sparse (i.e., $k = 7$) and especially when the unequal allocation scheme is used.

4.2 A Periodic Review Inventory System

In this subsection we demonstrate the performance of TP and SK-based algorithms for optimizing a simple periodic review (s,S) inventory system.

4.2.1 The Model: Dynamics and Simulation

The system is assumed to have i.i.d. continuous demands, zero lead times, full backlogging, and linear ordering, holding and shortage costs. The scenario considered here is similar to that discussed in Fu and Healy (1997), from which much of this example is constructed. Let X_i be the inventory position (inventory level plus outstanding orders in period i) and W_i be the inventory level (on-hand minus on backorder). The assumption of zero lead times gives $X_i = W_i$. The (s,S) inventory system works as follows. If X_i is below s units, an order of amount $(S - X_i)$ will be made and a fixed ordering cost K and a purchase cost $c(S - X_i)$ will be incurred. The inventory holding cost and shortage cost are also taken into account. The demand in period i, D_i, has distribution function $F_D(\cdot)$, which is absolutely continuous with density function $f_D(\cdot)$; denote the mean demand by $\mathrm{E}[D]$. The one-period cost is the sum of ordering, holding and backorder costs as follows

$$J_i = \mathbf{1}\{X_i < s\}(K + c(S - X_i)) + h\max\{0, W_i\} + p\max\{0, -W_i\},$$

and the quantity of interest is the long-run average cost per period

$$J = \lim_{n\to\infty} J_n, \text{ and } J_n = \frac{1}{n}\sum_{i=1}^{n} J_i.$$

Define $\delta = S - s$ and let $\lambda = 1/\mathrm{E}[D]$. Consider the two-dimensional problem of estimating the unknown response $J(\delta, s)$ at a given point (δ, s). If the demands are i.i.d. exponentially distributed with mean $\mathrm{E}[D]$, then the following analytic expressions for the long-run average cost and the optimal solutions can be obtained

$$J(\delta, s) = c\mathrm{E}[D] + \frac{K + h(s - \mathrm{E}[D] + \lambda\delta(s + \delta/2)) + (h + p)\mathrm{E}[D]e^{-\lambda s}}{1 + \lambda\delta}, \tag{16}$$

$$\delta_{\mathsf{opt}} = \sqrt{2K/\lambda h}, \quad s_{\mathsf{opt}} = -\mathrm{E}[D]\ln\left(\frac{h + \sqrt{2Kh\lambda}}{h + p}\right),$$

where $\lambda = 1/\mathrm{E}[D]$.

4.2.2 Experimental Design, Optimization Algorithms and Results

Experimental design. The design space is set as $\Omega_\delta \times \Omega_s = [100, 300] \times [200, 500]$. The initial design is a 'maximin' Latin Hypercube sample of 20 design points. The list of parameters used are shown in Table 2. TP and SK-based metamodeling approaches are examined under both low ($N = 1000$) and high ($N = 5000$) budget schemes. Under a given budget scheme with N simulation replications to expend, the optimization

Table 2: Parameters for the (s,S) inventory system.

E[D]	c	p	K	h
200	1	10	100	1

process starts with running n_0 simulation replications at each of the 20 design points in the initial design. In each simulation replication, the run length of $T = 1000$ is used to estimate the response $J(\delta,s)$ at a given point (δ,s). Pretending that we have little information about the true response surface, a constant trend model $\mathbf{f}^\top(\delta,s)\beta = \beta_0$ is used; we fit initial TP and SK metamodels using maximum likelihood estimation. In each subsequent iteration, Δn replications are allocated to each point selected from the candidate solution set $\mathbf{X}^{\mathsf{tot}}$, which consists of 2500 regularly spaced points in $\Omega_\delta \times \Omega_s$; and the TP and SK metamodels are updated correspondingly. We choose to set $n_0 = \Delta n \in \{25,50,100,125\}$ under the high budget scheme, and $n_0 = \Delta n \in \{5,10,20,25\}$ under the low budget scheme. This results in 4 distinct experiment configurations under either budget scheme.

Optimization algorithms. In this paper we focus on examining the performance of TP and SK-based EI and augmented expected improvement (AEI, Huang et al. 2006) algorithms. We next elaborate on the two selection criteria used by EI and AEI algorithms. Notice that below $\mathbf{x}$ is used to denote (δ,s) for notational convenience.

Selection criterion. In each iteration, the EI (respectively, AEI) algorithm selects the point in $\mathbf{X}^{\mathsf{tot}}$ that maximizes EI (resp., AEI) specified below to perform simulation runs in the next iteration. Take SK for example. At $\mathbf{x} \in \mathbf{X}^{\mathsf{tot}}$, we have

- $\mathrm{EI}(\mathbf{x}) = \left(\widehat{\mathrm{Y}}_{\min} - \mu(\mathbf{x})\right)\cdot\Phi\left(\frac{(\widehat{\mathrm{Y}}_{\min}-\mu(\mathbf{x})}{\sqrt{\sigma^2(\mathbf{x})}}\right) + \sqrt{\sigma^2(\mathbf{x})}\phi\left(\frac{\widehat{\mathrm{Y}}_{\min}-\mu(\mathbf{x})}{\sqrt{\sigma^2(\mathbf{x})}}\right)$, where $\Phi(\cdot)$ and $\phi(\cdot)$ denote the CDF and PDF of the standard normal distribution, respectively; the predictive mean and variance at $\mathbf{x}$ given by SK are respectively denoted by $\mu(\mathbf{x})$ and $\sigma^2(\mathbf{x})$; and $\widehat{\mathrm{Y}}_{\min} = \mu(\mathbf{x}^*)$ with $\mathbf{x}^*$ denoting the current effective best point found by the minimum-mean-response identification criterion specified below.
- $\mathrm{AEI}(\mathbf{x}) = \mathrm{EI}(\mathbf{x})\cdot\left(1 - \frac{\sqrt{\mathsf{V}(\mathbf{x})}}{\sqrt{\sigma^2(\mathbf{x})+\mathsf{V}(\mathbf{x})}}\right)$, where $\mathsf{V}(\mathbf{x})$ is the sampling variance at $\mathbf{x}$.

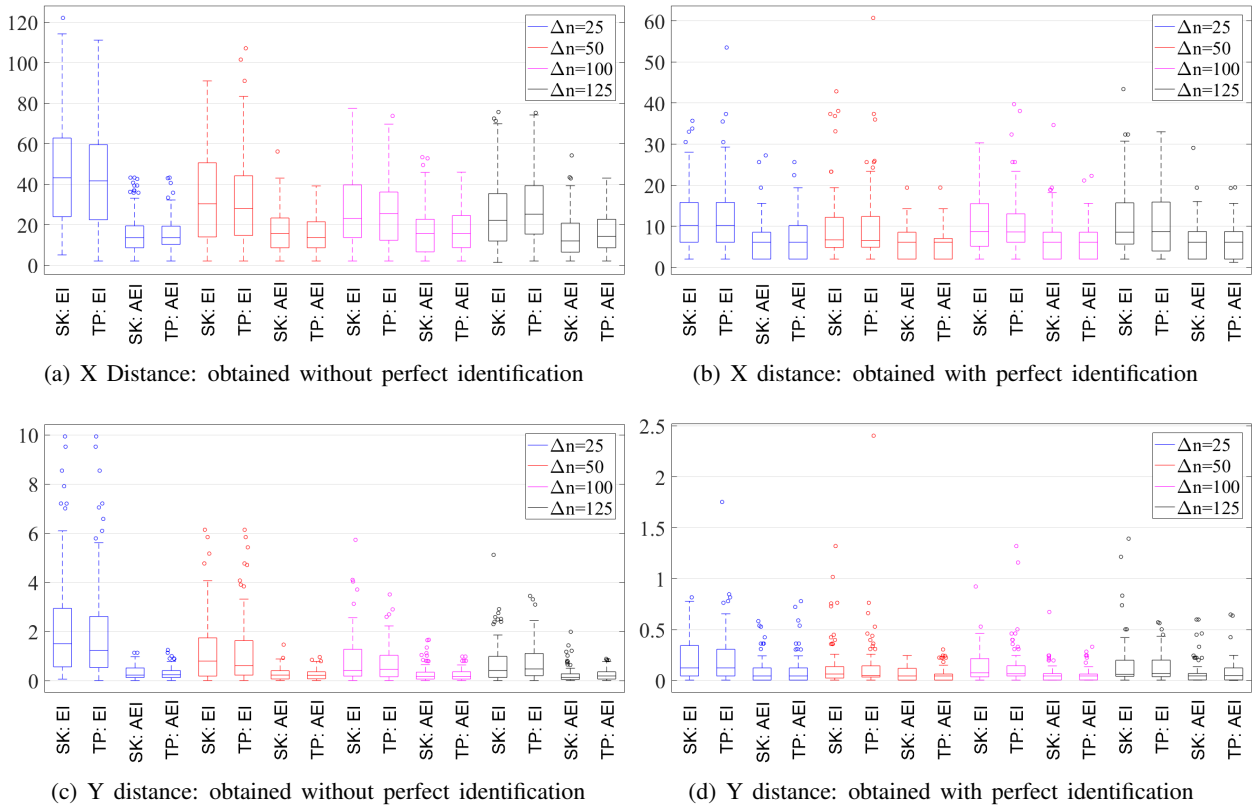
Identification criterion. By the end of each iteration, the current effective best point is chosen as $\mathbf{x}^* = \mathrm{argmin}_{\mathbf{x}\in\mathbf{X}} q(\mathbf{x})$, with $\mathbf{X}$ representing the set of points that have been evaluated up to the current iteration; and $q(\cdot)$ is specified according to the following two criteria. Specifically,

- the minimum-mean-response criterion uses $q(\mathbf{x}) = \bar{\mathscr{Y}}(\mathbf{x})$ for $\mathbf{x} \in \mathbf{X}$; and
- the minimum-predictive-quantile criterion adopts $q(\mathbf{x}) = \mu(\mathbf{x}) + \Phi^{-1}(\alpha)\sqrt{\sigma^2(\mathbf{x})}$ with $\alpha \in [0.5,1]$. In our implementation $\Phi^{-1}(\alpha) = 1$ is used following Huang et al. (2006)'s suggestion.

For TP-based EI and AEI algorithms, the selection and identification criteria can be specified in a similar fashion by using the expression of EI given in (14) and relevant quantities defined for ETMD. We note that to implement the TP-based (respectively, SK-based) AEI algorithm, a TP (resp., kriging) metamodel with nugget effect is built for estimating the variance function $\mathsf{V}(\cdot)$.

We apply the SK and TP-based EI and AEI algorithms, with each algorithm implemented under each experiment configuration for 100 independent macro-replications. To implement the TP-based algorithms, we set ν to 1.25 so that TP can behave as differently from SK as possible. Two measures are used to evaluate the performance of each algorithm, namely, the X distance and Y distance. Let $\mathbf{x}_{\mathsf{opt}} = (\delta_{\mathsf{opt}}, s_{\mathsf{opt}})$ denote the true optimal solution. The X distance refers to the Euclidean distance between $\mathbf{x}^{**}$ returned by a given algorithm according to an identification criterion adopted and $\mathbf{x}_{\mathsf{opt}}$, whereas Y distance is given by the discrepancy between the true objective function values at $\mathbf{x}^{**}$ and $\mathbf{x}_{\mathsf{opt}}$, $|J(\mathbf{x}^{**}) - J(\mathbf{x}_{\mathsf{opt}})|$.

Inspired by Jalali, Nieuwenhuyse, and Picheny (2017), we also examine the identification ability of each algorithm. Specifically, by the end of sequential iterations under each experiment configuration, instead of applying either identification criterion given above, the true best point $\mathbf{x}$ among the set of evaluated points by each algorithm is returned as the final solution, assuming that the true objective function given in (16) is known. The X and Y distances achieved in this case are referred to as being obtained with perfect identification, in contrast to those obtained without using the knowledge of the true objective function.



(a) X Distance: obtained without perfect identification

(b) X distance: obtained with perfect identification

(c) Y distance: obtained without perfect identification

(d) Y distance: obtained with perfect identification

Figure 1: Summary of performance of TP and SK-based algorithms under the high budget scheme.

Summary of results. The X and Y distances achieved by each algorithm with various Δn under the high budget scheme are shown in Figure 1. For the low budget scheme, we comment without showing details to economize on space. The following observations can be made. First, significant identification gaps exist for all algorithms under all experiment configurations examined. More specifically, under the high budget scheme shown in Figure 1, a comparison of (a) with (b) and (c) with (d) clearly shows that the X and Y distances achieved by a given algorithm with perfect identification are much smaller. Similar comments apply to results obtained under the low budget scheme. Furthermore, we note without showing details that these algorithms have selected promising points yet they fail to return the best point evaluated as the final solution. Second, we observe that under all experiment configurations, AEI algorithms outperform their EI counterparts, regardless which metamodeling approach is used and whether perfect identification is in effect. This observation confirms those reported earlier in the literature (e.g., Picheny and Ginsbourger 2014); and it is due to the fact that AEI algorithms tend to select more distinct points than their EI counterparts.

Under the high budget scheme, we see from Figure 1 that TP-based algorithms outperform the SK-based ones when Δn is relatively low (i.e., when $\Delta n = 25$ and 50) without perfect identification, yet the reverse holds true when Δn is relatively high (i.e., $\Delta n = 100$ and 125). Under the low budget scheme, we observe that when Δn is relatively low (i.e., $\Delta n = 5, 10$ and 20), the SK-based EI algorithm performs slightly

better than its TP counterpart without perfect identification, and the SK-based AEI algorithm performs slightly worse than its TP counterpart. However, when $\Delta n = 25$, TP-based algorithms are comparable to the SK-based ones, if not much better.

In summary, we find that when Δn is moderately high and a sufficiently large number of iterations is permitted under a given total budget, TP-based algorithms can outperform their SK counterparts. In other cases, however, superiority of TP-based algorithms over their SK-based counterparts has not been observed. Furthermore, we reach a similar conclusion as that mentioned by Jalali, Nieuwenhuyse, and Picheny (2017) regarding the performance of all kriging-based algorithms for simulation optimization: All SK and TP-based algorithms examined suffer from the identification problem.

5 CONCLUSION

In this paper we have developed a heteroscedastic t-process metamodeling approach (TP) for approximating the mean response surface implied by a stochastic simulation and performing metamodel-based optimization. The preliminary numerical evaluations indicate that TP can serve as a promising simulation analytics tool for solving problems encountered in production planning and supply chain management. In particular, TP can outperform stochastic kriging in terms of prediction and optimization when a dense design is used with a sufficiently large mount of simulation effort being allocated to each point.

ACKNOWLEDGMENT

This work was partially supported by the Virginia Tech ICTAS Junior Faculty Award.

REFERENCES

Ankenman, B. E., B. L. Nelson, and J. Staum. 2010. "Stochastic Kriging for Simulation Metamodeling". *Operations Research* 58:371–382.

Arellano-Valle, R. B., and H. Bolfarine. 1995. "On Some Characterizations of the t-Distribution". *Statistics & Probability Letters* 25:79–85.

Barton, R. R., and M. Meckesheimer. 2006. "Chapter 18 Metamodel-Based Simulation Optimization". In *Simulation*, edited by S. G. Henderson and B. L. Nelson, Handbooks in Operations Research and Management Science Vol. 13, 535–574. Oxford, UK: Elsevier.

Chen, X., and K. Kim. 2014. "Stochastic Kriging with Biased Sample Estimates". *ACM Transactions on Modeling and Computer Simulation* 24:8/1–8/23.

Dellino, G., J. P. C. Kleijnen, and C. Meloni. 2012. "Robust Optimization in Simulation: Taguchi and Krige Combined". *INFORMS Journal on Computing* 24:471–484.

Fu, M. C., and K. J. Healy. 1997. "Techniques for Simulation Optimization : An Experimental Study on an (s,S) Inventory System". *IIE Transactions* 29:191–199.

Girón, F. J., and C. del Castillo. 2010. "The Multivariate Behrens Fisher Distribution". *Journal of Multivariate Analysis* 101:2091–2102.

Huang, D., T. T. Allen, W. I. Notz, and N. Zeng. 2006. "Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Metamodels". *Journal of Global Optimization* 34:441–466.

Jalali, H., I. V. Nieuwenhuyse, and V. Picheny. 2017. "Comparison of Kriging-Based Algorithms for Simulation Optimization with Heterogeneous Noise". *European Journal of Operational Research* 261:279–301.

Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient Global Optimization Expensive Black-Box Functions". *Journal of Global Optimization* 21:455–492.

Jylänki, P., J. Vanhatalo, and A. Vehtari. 2011. "Robust Gaussian Process Regression with a Student-t Likelihood". *Journal of Machine Learning Research* 12:3227–3257.

Kotz, S., and S. Nadarajah. 2004. *Multivariate t Distributions and Their Applications*. Cambridge: Cambridge University Press.

Lee, Y., and J. A. Nelder. 2006. "Double Hierarchical Generalized Linear Models (With Discussion)". *Journal of the Royal Statistical Society: C (Applied Statistics)* 55:139–185.

Liu, C., G. Pedrielli, and S. H. Ng. 2014. "eTSSO: Adaptive Search Method for Stochastic Global Optimization under Finite Budget". Working paper, Department of Industrial & Systems Engineering, National University of Singapore.

Mortazavi, A., and A. Arshadi Khamseh. 2014. "Robust Inventory System Optimization Based on Simulation and Multiple Criteria Decision Making". *Modelling and Simulation in Engineering* 2014:1–17.

Neal, R. M. 1997. "Monte Carlo Implementation of Gaussian Process Models for Bayesian Regression and Classification". Technical report, University of Toronto, Department of Statistics and Department of Computer Science.

Picheny, V., and D. Ginsbourger. 2014. "Noisy Kriging-Based Optimization Methods: A Unified Implementation within the Diceoptim Package". *Computational Statistics and Data Analysis* 71:1035–1053.

Picheny, V., D. Ginsbourger, Y. Richet, and G. Caplin. 2013. "Quantile-Based Optimization of Noisy Computer Experiments with Tunable Precision". *Technometrics* 55:2–13.

Quan, N., J. Yin, S. H. Ng, and L. Lee. 2013. "Simulation Optimization via Kriging: A Sequential Search Using Expected Improvement with Computing Budget Constraints". *IIE Transactions* 45:763–780.

Rasmussen, C., and C. K. I. Williams. 2006. *Gaussian Processes for Machine Learning*. New York: MIT Press.

Roth, M. 2013. "On the Multivariate t Distribution". Technical report, Linköpings universitet, Department of Electrical Engineering.

Santner, T. J., B. J. Williams, and W. I. Notz. 2003. *The Design and Analysis of Computer Experiments*. New York: Springer.

Shah, A., A. G. Wilson, and Z. Ghahramani. 2014. "Student-t Processes as Alternatives to Gaussian Processes". In *Proceedings of the 17th International Conference on Artificial Intelligence and Statistics*, 877–885. Reykjavik, Iceland.

Song, L., X. Li, and A. Garcia-Diaz. 2008. "Multi-Echelon Supply Chain Simulation Using Metamodel". In *Proceedings of the 2008 Winter Simulation Conference*, edited by S. J. Mason, R. R. Hill, L. Mönch, T. J. O. Rose, and J. W. Fowler, 2691–2699. Piscataway, NJ: Institute of Electrical and Eletronics Engineers.

Wang, Z., J. Shi, and Y. Lee. 2015. "Extended t-Process Regression Models". *ArXiv e-prints*.

Whitt, W. 1989. "Planning Queueing Simulations". *Management Science* 35:1341–1366.

Yang, F. 2010. "Neural Network Metamodeling for Cycle Time-Throughput Profiles in Manufacturing". *European Journal of Operational Research* 205:172–185.

Zellner, A. 1976. "Bayesian and Non-Bayesian Analysis of the Regression Model with Multivariate Student-t Error Terms". *Journal of the American Statistical Association* 71:400–405.

Zhang, Y., and D. Y. Yeung. 2010. "Multi-Task Learning Using Generalized t-Process". In *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics*, 964–971. Sardinia, Italy.

AUTHOR BIOGRAPHIES

GUANGRUI XIE is a PhD student in the Grado Department of Industrial and Systems Engineering at Virginia Tech. He received his B.E. degree in Industrial Engineering at Tianjin University in China. His research interests lie in stochastic optimization and simulation. His email address is guanx92@vt.edu.

XI CHEN is an Assistant Professor in the Grado Department of Industrial and Systems Engineering at Virginia Tech. Her research interests include stochastic modeling and simulation, applied probability and statistics, computer experiment design and analysis, and simulation optimization. Her email address is xchen6@vt.edu and her web page is https://sites.google.com/vt.edu/xi-chen-ise/home.