Proceedings of the 2018 Winter Simulation Conference
M. Rabe, A. A. Juan, N. Mustafee, A. Skoogh, S. Jain, and B. Johansson, eds.

GENERALIZED METHOD OF MOMENTS APPROACH TO HYPERPARAMETER ESTIMATION FOR GAUSSIAN MARKOV RANDOM FIELDS

Eunhye Song
Yi Dong

Department of Industrial and Manufacturing Engineering
Penn State University
310 Leonhard Building
University Park, PA 16802, USA

ABSTRACT

When a Gaussian Markov random field (GMRF) is used as a metamodel of an unknown response surface for a discrete optimization via simulation (DOvS) problem, the hyperparameters of the GMRF are estimated based on a few initial design points in a large feasible solution space. Although the maximum likelihood estimators (MLEs) are most commonly adopted to estimate these hyperparameters, its computation time increases polynomially in the size of the feasible solution space. We introduce new generalized method of moments (GMM) estimators of the hyperparameters of GMRFs and their initial sampling schemes, and show they are consistent under some conditions. Unlike MLEs, the computation time for these GMM estimators does not depend on the size of the feasible solution space. We show empirically that the GMM estimators have smaller biases and standard errors than MLE for a wide range of initial simulation budget while requiring orders of magnitude smaller computation time.

1 INTRODUCTION

Gaussian process (GP) is a popular choice for a metamodel to represent an unknown response surface for both deterministic computer experiment and stochastic simulation when the goal is to find the optimal parameters that minimizes (or maximizes) the response. After Jones et al. (1998) first introduced the idea of using a GP metamodel for global optimization of a deterministic computer model, GP has been applied to solve optimization via simulation (OvS) problems with continuous (Scott et al. 2011) and discrete feasible solution spaces (Quan et al. 2013, Xie et al. 2016). These approaches can be categorized as GP-based adaptive random search (ARS) algorithms, which iteratively update the GP model conditional on the cumulative simulation results and decide which solution to simulate next by drawing inference on the response surface from the GP model. A good GP model with strong inferential power allows the algorithm to identify good solutions quickly by simulating only a small fraction of the feasible solution space.

Salemi et al. (2018) propose an ARS method for DOvS that uses a GMRF as a metamodel. A GMRF is a special case of GP defined on undirected graph $\mathscr{G} = \{\mathscr{V}, \mathscr{E}\}$, where $\mathscr{V}$ is the set of feasible solutions and $\mathscr{E}$ is the set of edges connecting the solutions. The Markov property of GMRF makes the response at each solution independent from the rest of the graph conditional on the solution's immediate neighbors (Rue and Held 2005). Salemi et al. (2018) empirically show a GMRF provides better inference on the remaining optimality gap between the current best solution and the global optimum than a continuous GP when applied to a DOvS problem.

Similar to other GP-based ARS, Salemi et al. (2018) parameterize the initial GMRF model and estimate its hyperparameters by running simulations at the initial design points (solutions). Hyperparameter estimation for GMRF has been extensively studied in the context of image processing, where GMRFs

are used for modeling textures in a two-dimensional image by representing each pixel as a node on the graph Manjunath and Chellappa 1991, Dryden et al. 2002. In this setting, there is no need for "simulating a solution" as observations at all solutions are given and MLE is commonly adopted for hyperparmeter estimation. On the other hand, a DOvS problem for which ARS is considered as a solution method tends to have a large, high-dimensional feasible solution space and therefore simulating all solutions is not an option. Also, the size of the feasible solution space can make computation for MLE cumbersome as the computational complexity of the likelihood function evaluation increases polynomially in the size of $\mathscr{V}$ (Rue 2001). *The focus of this paper is to develop an estimation method for hyperparameters of GMRFs with good small-sample performance that is computationally efficient for a large-scale DOvS problem.*

We propose two generalized method of moments (GMM) estimators of the hyperparameters of GMRFs whose computational complexity does not depend on the size of $\mathscr{V}$. GMM is a popular parameter estimation method in econometrics due to its computatinoal efficiency and robustness to the choice of the parametric model (Mátyás 1999). The moment functions for GMM are derived from the conditional moments of the GMRF. The first estimator, *full-GMM*, requires sampling all neighbors of the initial design points and the neighbors of those neighbors, while *slim-GMM* requires a reduced number of solutions to be sampled.

In the next section, we introduce the concept of GMRF and discuss its MLE formulation. In Section 3.1, we provide a brief overview of GMM and introduce the full- and slim-GMM estimators in Sections 3.2–3.3 followed by discussions on their consistency (Section 4) and small-sample performance (Section 5). We empirically compare the performance of full-GMM and slim-GMM with MLE in Section 6.

2 PARAMETERIZATION OF GAUSSIAN MARKOV RANDOM FIELDS

Suppose we simulate each solution $\mathbf{x}$ of a DOvS problem to obtain simulation output $Y(\mathbf{x}) = y(\mathbf{x}) + \varepsilon(\mathbf{x})$, where $y(\mathbf{x}) \triangleq \mathrm{E}[Y(\mathbf{x})]$ and simulation error $\varepsilon(\mathbf{x})$ has mean 0 and finite variance $\sigma^2(\mathbf{x})$. Given the set of feasible solutions, $\mathscr{X}$, we can formulate the DOvS problem as $\min_{\mathbf{x}\in\mathscr{X}} y(\mathbf{x})$. We assume $\mathscr{X}$ is an integer-ordered set in $\mathbb{R}^d$ containing n solutions.

Salemi et al. (2018) model the unknown response surface $\mathbf{y} = \{y(\mathbf{x}_1), y(\mathbf{x}_2), \ldots, y(\mathbf{x}_n)\}$ as a realization of the GRMF defined on graph $\mathscr{G} = (\mathscr{X}, \mathscr{E})$

$$\mathbb{Y} \triangleq \{\mathbb{Y}(\mathbf{x}_1), \mathbb{Y}(\mathbf{x}_2), \ldots, \mathbb{Y}(\mathbf{x}_n)\}^\top \sim \mathrm{N}(\boldsymbol{\mu}, \mathbf{Q}^{-1}), \tag{1}$$

where $\boldsymbol{\mu}$ is the mean vector and $\mathbf{Q}$ is the precision matrix, the inverse of the variance-covariance matrix. The structure of $\mathbf{Q}$ is determined by the set of edges, $\mathscr{E}$, which represents the connectivity among the solutions in $\mathscr{X}$ on the graph. Defining $\mathscr{E}$ is a modeling decision as it is not inherent to the DOvS problem itself. Salemi et al. (2018) suggest to define the set of neighboring solutions of $\mathbf{x} \in \mathscr{X}$ to be $N(\mathbf{x}) \triangleq \{\mathbf{x}' \in \mathscr{X} : ||\mathbf{x} - \mathbf{x}'||_2 = 1\}$, i.e., a solution can have at most $2d$ neighbors. We generalize this notation to define a set of neighbors of $\mathscr{A} \subset \mathscr{X}$: $N(\mathscr{A}) \triangleq \{\mathbf{x}' \in \mathscr{X} \setminus \mathscr{A} | \exists \mathbf{x} \in \mathscr{A} \text{ s.t. } ||\mathbf{x} - \mathbf{x}'||_2 = 1\}$.

Given a neighborhood structure, the Markov property of GMRF can be written as $\mathbb{Y}(\mathbf{x}) \perp \mathbb{Y}(\mathscr{V}\setminus\{\mathbf{x}, N(\mathbf{x})\}) | \mathbb{Y}(N(\mathbf{x}))$ for $\mathbf{x} \in \mathscr{X}$, which implies $Q_{ii} = \mathrm{Prec}(\mathbb{Y}(\mathbf{x}_i) | \mathbb{Y}(N(\mathbf{x}_i))) = \mathrm{V}^{-1}(\mathbb{Y}(\mathbf{x}_i) | \mathbb{Y}(N(\mathbf{x}_i))), \mathrm{Corr}(\mathbb{Y}(\mathbf{x}_i), \mathbb{Y}(\mathbf{x}_j) | \mathbb{Y}(N(\{\mathbf{x}_i, \mathbf{x}_j\})) = -Q_{ij}/\sqrt{Q_{ii}Q_{jj}}$, and $Q_{ij} \neq 0$ if and only if $\mathbf{x}_i \in N(\mathbf{x}_j)$, where Q_{ij} represents the (i,j)th element of $\mathbf{Q}$. Therefore, $\mathbf{Q}$ becomes a very sparse matrix even if n is large as each row of $\mathbf{Q}$ has at most $2d+1$ nonzero elements.

Given $N(\mathbf{x})$, Salemi et al. (2018) parameterize $\mathbf{Q}$ by introducing $\boldsymbol{\theta} = \{\theta_0, \theta_1, ..., \theta_d\}$ and function

$$p(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\theta}) = \begin{cases} \theta_0, & \text{if } \mathbf{x}_i = \mathbf{x}_j, \\ -\theta_0\theta_\ell, & \text{if } |\mathbf{x}_i - \mathbf{x}_j| = \mathbf{e}_\ell, \\ 0, & \text{otherwise}, \end{cases} \tag{2}$$

where $\mathbf{e}_\ell$ is the ℓth standard basis vector of $\mathbb{R}^d$. Setting $Q_{ij} = p(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\theta})$, we get $\mathrm{V}(\mathbb{Y}(\mathbf{x}_i) | \mathbb{Y}(N(\mathbf{x}_i))) = \theta_0^{-1}$ and $\mathrm{Corr}(\mathbb{Y}(\mathbf{x}_i), \mathbb{Y}(\mathbf{x}_j) | \mathbb{Y}_{\mathscr{V}\setminus\{\mathbf{x}_i, \mathbf{x}_j\}}) = \theta_\ell$, if $|\mathbf{x}_i - \mathbf{x}_j| = \mathbf{e}_\ell$. Thus, θ_0 is positive and $0 \leq \theta_\ell \leq 1$, to ensure $\mathbb{Y}(\mathbf{x})$ are positively correlated, which is a common modeling assumption for GP (Santner et al. 2003).

Additionally, $\theta_1, \theta_2, \ldots, \theta_d$ are constrained such that $\mathbf{Q}$ is positive definite. They also parameterize the mean vector of GMRF to be $\boldsymbol{\mu} = \beta \mathbf{1}_n$, where $\beta \in \mathbb{R}$ and $\mathbf{1}_n$ is the n-dimensional vector of ones, i.e., no mean trend of $\mathbb{Y}$ is assumed. In this paper, we adopt the same parameterization for $\mathbf{Q}$ and $\boldsymbol{\mu}$.

To represent the stochastic simulation output of a DOvS problem, we define GMRF $\mathbb{Y}^{\varepsilon}$ such that $\mathbb{Y}^{\varepsilon} - \mathbb{Y} \sim \mathrm{N}(\mathbf{0}, \mathbf{Q}_{\varepsilon}) \perp \mathbb{Y}$, where $\mathbf{Q}_{\varepsilon}$ is the precision matrix of the stochastic errors at the solutions in $\mathscr{X}$. We treat $\bar{Y}(\mathbf{x}) = \sum_{m=1}^{r(\mathbf{x})} Y_m(\mathbf{x})/r(\mathbf{x})$ as a realization of $\mathbb{Y}^{\varepsilon}(\mathbf{x})$, where $r(\mathbf{x})$ is the number of replications run at solution $\mathbf{x}$. Therefore, the diagonal element of $\mathbf{Q}_{\varepsilon}$ corresponding to $\mathbf{x}$ is $\sigma^2(\mathbf{x})/r(\mathbf{x})$. We simulate all solutions independently, i.e., $\mathbf{Q}_{\varepsilon}$ is a diagonal matrix.

In an ARS algorithm, (1) is updated by conditioning on the observations from the simulated solutions at each iteration. The values of $\boldsymbol{\theta}$ and β can be estimated via MLE in the initialization phase of the algorithm by running simulations at the initial design points, which we review below.

Suppose $\mathscr{X}$ is partitioned into two sets $\mathscr{X}_1$ and $\mathscr{X}_2$, where the latter includes the initial design points. Similarly, we can partition $\mathbb{Y} = (\mathbb{Y}_1^{\top}, \mathbb{Y}_2^{\top})^{\top}$ and $\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_{11} & \mathbf{Q}_{12}; \mathbf{Q}_{12}^{\top} & \mathbf{Q}_{22} \end{bmatrix}$. We also define $\mathbb{Y}_2^{\varepsilon}$ as a subset of $\mathbb{Y}^{\varepsilon}$ corresponding to $\mathscr{X}_2$ and $\bar{\mathbf{Y}}_2$ as the realization of $\mathbb{Y}_2^{\varepsilon}$. A plug-in estimator of $\mathbf{Q}_{\varepsilon}$ is $\widehat{\mathbf{Q}}_{\varepsilon} = \mathrm{diag}\left(r(\mathbf{x}_1)/S^2(\mathbf{x}_1), r(\mathbf{x}_2)/S^2(\mathbf{x}_2), ..., r(\mathbf{x}_k)/S^2(\mathbf{x}_k)\right)$, where $S^2(\mathbf{x}) = \sum_{j=1}^{r(\mathbf{x})} \left(Y_j(\mathbf{x}) - \bar{Y}(\mathbf{x})\right)^2/(r(\mathbf{x})-1)$. Then, the log-likelihood function of (1) given $\bar{\mathbf{Y}}_2$ is

$$\mathscr{L}(\boldsymbol{\theta}|\bar{\mathbf{Y}}_2) \triangleq -\frac{1}{2}\log|\Sigma_{22} + \widehat{\mathbf{Q}}_{\varepsilon}^{-1}| - \frac{1}{2}(\bar{\mathbf{Y}}_2 - \widehat{\beta}\mathbf{1}_{|\mathscr{X}_2|})^{\top}(\Sigma_{22} + \widehat{\mathbf{Q}}_{\varepsilon}^{-1})^{-1}(\bar{\mathbf{Y}}_2 - \widehat{\beta}\mathbf{1}_{|\mathscr{X}_2|}), \tag{3}$$

where $\widehat{\beta} = \left(\mathbf{1}_{|\mathscr{X}_2|}^{\top}(\Sigma_{22} + \widehat{\mathbf{Q}}_{\varepsilon}^{-1})^{-1}\mathbf{1}_{|\mathscr{X}_2|}\right)^{-1}\mathbf{1}_{|\mathscr{X}_2|}^{\top}(\Sigma_{22} + \widehat{\mathbf{Q}}_{\varepsilon}^{-1})^{-1}\bar{\mathbf{Y}}_2$ and $\Sigma_{22} = (\mathbf{Q}_{22} - \mathbf{Q}_{12}^{\top}\mathbf{Q}_{11}^{-1}\mathbf{Q}_{12})^{-1}$. Note that Σ_{22} is the variance-covariance matrix of $\mathbb{Y}_2$ conditional on $\mathbb{Y}_2^{\varepsilon} = \bar{\mathbf{Y}}_2$. Since $\mathbf{Q}_{\varepsilon}$ is a diagonal matrix, $\mathbf{Q}_{\varepsilon}^{-1}$ is cheap to compute. We can compute $\mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}$ efficiently by first factorizing $\mathbf{Q}_{11}$ and use its factors to solve $\mathbf{Q}_{11}\mathbf{z} = \mathbf{Q}_{12}$ for $\mathbf{z} = \mathbf{Q}_{11}^{-1}\mathbf{Q}_{12}$. Although $\mathbf{Q}_{11}$ is a sparse matrix, these computations are expensive when n is large since $|\mathscr{X}_2| \ll |\mathscr{X}_1| \approx n$ for typical ARS. On the other hand, $\Sigma_{22} + \widehat{\mathbf{Q}}_{\varepsilon}^{-1}$ is a dense, but small matrix, therefore, computing $\log|\Sigma_{22} + \widehat{\mathbf{Q}}_{\varepsilon}^{-1}|$ and $(\Sigma_{22} + \widehat{\mathbf{Q}}_{\varepsilon}^{-1})^{-1}(\bar{\mathbf{Y}}_2 - \widehat{\beta}\mathbf{1}_{|\mathscr{X}_2|})$ in (3) is relatively cheap.

3 GENERALIZED METHOD OF MOMENTS ESTIMATOR FOR GMRF

In this chapter, we first provide a brief review of GMM method and propose new GMM estimators of the hyperparameters of the GMRF. Two versions of GMM estimators are presented: full-GMM and slim-GMM estimators. The latter requires a fewer number of solutions to be simulated. To represent the dependence of $\mathbf{Q}$ on $\boldsymbol{\theta}$, we use $\mathbf{Q} \equiv \mathbf{Q}(\boldsymbol{\theta})$ in the following sections.

3.1 Generalized Method of Moments

GMM estimates the parameters of a statistical model by matching the moments formulated by unknown parameters with the sample moments (Mátyás 1999). Let $\boldsymbol{\lambda}$ and ω_i represent the vector of parameters of the model and the ith i.i.d. observation for $i = 1, 2, \ldots, k$, respectively. We define $g : (\boldsymbol{\lambda}, \omega_i) \to \mathbb{R}^p$ as a moment function, which is chosen to satisfy the following moment condition

$$h(\tilde{\boldsymbol{\lambda}}) = \mathrm{E}[g(\omega_i, \tilde{\boldsymbol{\lambda}})] = \mathbf{0}_p, \tag{4}$$

where $\tilde{\boldsymbol{\lambda}}$ is the true parameter vector and $\mathbf{0}_p$ is a p-dimensional vector of zeroes. The expectation in (4) can be estimated by a sample average $\bar{g}$ given observations $\omega_1, \omega_2, \cdots, \omega_k : \bar{g}(\boldsymbol{\lambda}) = \sum_{i=1}^{k} g(\omega_i, \boldsymbol{\lambda})/k$. The GMM estimator $\widehat{\boldsymbol{\lambda}}$ of $\tilde{\boldsymbol{\lambda}}$ is defined as

$$\widehat{\boldsymbol{\lambda}} = \arg\min_{\boldsymbol{\lambda} \in \Lambda} \ \bar{g}(\boldsymbol{\lambda})^{\top} W \bar{g}(\boldsymbol{\lambda}), \tag{5}$$

where W is a $p \times p$ positive definite matrix and $\Lambda \in \mathbb{R}^r$ is a feasible set of $\boldsymbol{\lambda}$. Newey and McFadden (1994) provide a set of regularity conditions for $\widehat{\boldsymbol{\lambda}}$ to be a consistent estimator as below.

Theorem 1 (Newey and McFadden 1994) Suppose ω_i, $i = 1,2,\ldots,$ are i.i.d. and (i) $\mathrm{E}[g(\omega_i,\boldsymbol{\lambda})] = 0$, if and only if, $\boldsymbol{\lambda} = \tilde{\boldsymbol{\lambda}}$; (ii) $\tilde{\boldsymbol{\lambda}} \in \Lambda$ and Λ is compact; (iii) $g(\omega_i,\boldsymbol{\lambda})$ is continuous at each $\boldsymbol{\lambda} \in \Lambda$ with probability 1; iv) $\mathrm{E}[\sup_{\boldsymbol{\lambda}\in\Lambda} ||g(\omega_i,\boldsymbol{\lambda})||] < \infty$. Then $\boldsymbol{\lambda} \xrightarrow{p} \tilde{\boldsymbol{\lambda}}$.

The condition that ω_i's are i.i.d. can be dropped if ω_i is observed from an ergodic process. Condition (i), also referred to as an identification condition, is difficult to establish for general $\tilde{\boldsymbol{\lambda}}$ when g is a nonlinear function of $\boldsymbol{\lambda}$. Rothenberg (1971) show that if $g(\omega_i,\boldsymbol{\lambda})$ is continuously differentiable and $\mathrm{E}[\nabla_{\boldsymbol{\lambda}} g(\omega_i,\boldsymbol{\lambda})] = \nabla_{\boldsymbol{\lambda}} \mathrm{E}[g(\omega_i,\boldsymbol{\lambda})]$, then a sufficient condition for the local identification of $\tilde{\boldsymbol{\lambda}}$ is that $W\mathrm{E}[\nabla_{\boldsymbol{\lambda}} g(\omega_i,\boldsymbol{\lambda})]$ has a full column rank in a neighborhood of $\tilde{\boldsymbol{\lambda}}$. Due to the difficulties of showing global identification, (i) is often simply assumed to hold (Newey and McFadden 1994).

If the number of moment conditions, p, is smaller than the number of parameters, r, then typically there are multiple $\boldsymbol{\lambda}$ that satisfy $\mathrm{E}[g(\omega_i,\boldsymbol{\lambda})] = \mathbf{0}_p$. The full-GMM and slim-GMM estimators presented in the following sections have the same number of moment conditions as the number of parameters, $d+2$.

The asymptotic variance of $\widehat{\boldsymbol{\lambda}}$ is minimized when we choose $W \propto \Omega^{-1}$, where $\Omega = \mathrm{E}[g(\omega_i,\tilde{\boldsymbol{\lambda}})g(\omega_i,\tilde{\boldsymbol{\lambda}})^T]$ (Newey and McFadden, 1994). Although Ω is a function on unknown $\tilde{\boldsymbol{\lambda}}$, it can be estimated by 1) solving (5) with $\widehat{W} = I$ to obtain $\widehat{\boldsymbol{\lambda}}$; and 2) plugging $\widehat{\boldsymbol{\lambda}}$ in to compute $\widehat{\Omega} = \sum_{j=1}^{k} g(\omega_j,\widehat{\boldsymbol{\lambda}})g(\omega_j,\widehat{\boldsymbol{\lambda}})^\top/k$. The estimator obtained by resolving (5) with $\widehat{\Omega}$ is referred to as a two-step GMM estimator. If this procedure is repeated multiple times, then the estimator is referred to as an iterative GMM estimator. Since $\widehat{\Omega}$ is a consistent estimator of Ω, the two-step or iterative GMM estimator has the minimum asymptotic variance, however, in practice, estimator of Ω tends to have large variance when k is small. For our numerical example in Section 6, we tested both two-step and iterative estimators, however, their empirical mean squared errors tend to be larger than the GMM estimator with $W = I$ for all test cases.

3.2 Full-GMM Estimator

In this section, we derive a moment function $g : (\boldsymbol{\theta},\beta) \to \mathbb{R}^{d+2}$ to estimate the true hyperparameters $\tilde{\boldsymbol{\theta}}$ and $\tilde{\beta}$ of the GMRF given the neighborhood structure defined in Section 2. Let $\Theta \in \mathbb{R}^{d+2}$ denote the feasible region of $(\boldsymbol{\theta},\beta)$; as mentioned in Section 2, $\theta_0 > 0$ and $0 \le \theta_\ell \le 1$. To make Θ compact, the former condition can be modified to $\theta_0 \ge \delta_0$, where δ_0 is an arbitrarily small positive constant. For the experiments in Section 6, we used $\delta_0 = 10^{-6}$. A sufficient condition for $\mathbf{Q}(\boldsymbol{\theta})$ to be positive definite is diagonal dominance (Geršgorin 1931), i.e., $\sum_{j=1,j\neq i}^{d} |Q_{ij}| < |Q_{ii}|$, which is equivalent to $\sum_{j=1}^{d} \theta_j < 0.5$ under the parameterization in (2). Similar to θ_0, we may choose arbitrarily small $\delta > 0$ and use $\sum_{j=1}^{d} \theta_j \le 0.5 - \delta$ instead. Although β is unconstrained, we can put practical bounds on β such that $\beta_L \le \beta \le \beta_U$.

We define $g \triangleq (g_0, g_1, \ldots, g_d, g_{d+1})^\top$, where each g_i is a real-valued function of $(\boldsymbol{\theta},\beta)$. Exploiting the Markov property of GMRF, the proposed moment function for the full-GMM estimation involves simulating a central design point and its neighbors as well as the neighbors of the neighbors. Thus, the ith sample for the full-GMM is defined as

$$\omega_i = \mathbb{Y}^{\varepsilon}(\mathbf{x}_i) \cup \mathbb{Y}^{\varepsilon}(N(\mathbf{x}_i)) \cup \left(\cup_{\ell=1}^{d} \mathbb{Y}^{\varepsilon}(N(\mathbf{x}_{i\ell}))\right), \tag{6}$$

where $\mathbf{x}_{i\ell}$ is a ℓth-directional neighbor of $\mathbf{x}_i$. All selected solutions are simulated r times, i.e., $r(\mathbf{x}) = r$. For ease of exposition, we first present g for the case when $\mathbb{Y}(\mathbf{x})$ can be observed without simulation error.

Suppose $\mathbb{Y} \sim \mathrm{N}(\tilde{\beta}\mathbf{1}_n, \mathbf{Q}^{-1}(\tilde{\boldsymbol{\theta}}))$. Then, as mentioned in Section 2

$$\mathrm{V}(\mathbb{Y}(\mathbf{x}_i)|\mathbb{Y}(N(\mathbf{x}_i))) = \tilde{\theta}_0^{-1}, \tag{7}$$

$$\mathrm{Corr}(\mathbb{Y}(\mathbf{x}_i), \mathbb{Y}(\mathbf{x}_{i\ell})|\mathbb{Y}(N(\{\mathbf{x}_i, \mathbf{x}_{i\ell}\}))) = \tilde{\theta}_\ell. \tag{8}$$

Applying the law of total variance, (7) can be rewritten as $\tilde{\theta}_0^{-1} = \mathrm{V}[\mathbb{Y}(\mathbf{x}_i)] - \mathrm{V}[\mathrm{E}[\mathbb{Y}(\mathbf{x}_i)|\mathbb{Y}(N(\mathbf{x}_i))]] = \mathrm{E}[\mathbb{Y}(\mathbf{x}_i)^2] - \mathrm{E}[\mathrm{E}[\mathbb{Y}(\mathbf{x}_i)|\mathbb{Y}(N(\mathbf{x}_i))]^2]$. From the conditional distribution of $\mathbb{Y}(\mathbf{x}_i)|\mathbb{Y}(N(\mathbf{x}_i))$, Rue and Held

(2005) show that

$$\mathrm{E}[\mathbb{Y}(\mathbf{x}_i)|\mathbb{Y}(N(\mathbf{x}_i))] = \tilde{\beta} + \sum_{\mathbf{x}_s \in N(\mathbf{x}_i)} \tilde{\theta}(i,s)\mathbb{Y}(\mathbf{x}_s), \tag{9}$$

where $\tilde{\theta}(i,s) = \tilde{\theta}_\ell$, if $|\mathbf{x}_i - \mathbf{x}_s| = \mathbf{e}_\ell$. Therefore, the following satisfies $\mathrm{E}[g_0(\omega_i, \tilde{\boldsymbol{\theta}}, \tilde{\beta})] = 0$:

$$g_0(\omega_i, \boldsymbol{\theta}, \beta) = \theta_0^{-1/2} - \theta_0^{1/2}\mathbb{Y}(\mathbf{x}_i)^2 + \theta_0^{1/2}\left(\beta + \sum_{\mathbf{x}_s \in N(\mathbf{x}_i)} \theta(i,s)\mathbb{Y}(\mathbf{x}_s)\right)^2. \tag{10}$$

Note that $\theta_0^{-1/2} g_0(\omega_i, \boldsymbol{\theta}, \beta)$ also satisfy the moment constraint, however, it tends to show poorer finite-sample performance than (10) when k is small (see Section 5. Estimating (10) from the sample requires observing all neighbors of $\mathbf{x}_i$. We modify (8) to define g_ℓ for $\ell = 1, 2, \ldots, d$. From the conditional distribution of $\mathbb{Y}(\{\mathbf{x}_i, \mathbf{x}_{i\ell}\})|\mathbb{Y}(N(\{\mathbf{x}_i, \mathbf{x}_{i\ell}\}))$ we can derive

$$\mathrm{V}[\mathbb{Y}(\mathbf{x}_i)|\mathbb{Y}(N(\{\mathbf{x}_i, \mathbf{x}_{i\ell}\}))] = \mathrm{V}[\mathbb{Y}(\mathbf{x}_{i\ell})|\mathbb{Y}(N(\{\mathbf{x}_i, \mathbf{x}_{i\ell}\}))] = \tilde{\theta}_0^{-1}(1 - \tilde{\theta}_\ell^2)^{-1}, \tag{11}$$

$$\mathrm{E}[\mathbb{Y}(\mathbf{x}_i)|\mathbb{Y}(N(\{\mathbf{x}_i, \mathbf{x}_{i\ell}\}))] = \tilde{\beta} + \frac{1}{1 - \tilde{\theta}_\ell^2}\left(\sum_{\mathbf{x}_s \in N(\mathbf{x}_i)\setminus \mathbf{x}_{i\ell}} \tilde{\theta}(i,s)(\mathbb{Y}(\mathbf{x}_s) - \tilde{\beta}) + \sum_{\mathbf{x}_s \in N(\mathbf{x}_{i\ell})\setminus \mathbf{x}_i} \tilde{\theta}_\ell \tilde{\theta}(i,s)(\mathbb{Y}(\mathbf{x}_s) - \tilde{\beta})\right). \tag{12}$$

Using (11), (8) can be rewritten as $\mathrm{Cov}(\mathbb{Y}(\mathbf{x}_i), \mathbb{Y}(\mathbf{x}_{i\ell})|\mathbb{Y}(N(\{\mathbf{x}_i, \mathbf{x}_{i\ell}\}))) = \tilde{\theta}_\ell/(\tilde{\theta}_0(1 - \tilde{\theta}_\ell^2))$. From the law of total covariance,

$$\begin{aligned}
&\mathrm{Cov}(\mathbb{Y}(\mathbf{x}_i), \mathbb{Y}(\mathbf{x}_{i\ell})|\mathbb{Y}(N(\{\mathbf{x}_i, \mathbf{x}_{i\ell}\}))) \\
&= \mathrm{Cov}(\mathbb{Y}(\mathbf{x}_i), \mathbb{Y}(\mathbf{x}_{i\ell})) - \mathrm{Cov}(\mathrm{E}[\mathbb{Y}(\mathbf{x}_i)|\mathbb{Y}(N(\{\mathbf{x}_i, \mathbf{x}_{i\ell}\}))], \mathrm{E}[\mathbb{Y}(\mathbf{x}_{i\ell})|\mathbb{Y}(N(\{\mathbf{x}_i, \mathbf{x}_{i\ell}\}))]) \\
&= \mathrm{E}[\mathbb{Y}(\mathbf{x}_i)\mathbb{Y}(\mathbf{x}_{i\ell})] - \mathrm{E}\left[\mathrm{E}[\mathbb{Y}(\mathbf{x}_i)|\mathbb{Y}(N(\{\mathbf{x}_i, \mathbf{x}_{i\ell}\}))]\mathrm{E}[\mathbb{Y}(\mathbf{x}_{i\ell})|\mathbb{Y}(N(\{\mathbf{x}_i, \mathbf{x}_{i\ell}\}))]\right].
\end{aligned} \tag{13}$$

Notice that the expression for the product of conditional expectations in (13) can be derived from (12). Thus, we define $g_\ell(\omega_i, \boldsymbol{\theta}, \beta)$ as

$$\begin{aligned}
g_\ell(\omega_i, \boldsymbol{\theta}, \beta) = &-\theta_\ell/(\theta_0^{1/2}(1 - \theta_\ell^2)) + \theta_0^{1/2}\mathbb{Y}(\mathbf{x}_i)\mathbb{Y}(\mathbf{x}_{i\ell}) \\
&- \theta_0^{1/2}\left(\beta + \frac{1}{1 - \theta_\ell^2}\left(\sum_{\mathbf{x}_s \in N(\mathbf{x}_i)\setminus \mathbf{x}_{i\ell}} \theta(i,s)(\mathbb{Y}(\mathbf{x}_s) - \beta) + \sum_{\mathbf{x}_s \in N(\mathbf{x}_{i\ell})\setminus \mathbf{x}_i} \theta_\ell \theta(i,s)(\mathbb{Y}(\mathbf{x}_s) - \beta)\right)\right) \\
&\times \left(\beta + \frac{1}{1 - \theta_\ell^2}\left(\sum_{\mathbf{x}_s \in N(\mathbf{x}_{i\ell})\setminus \mathbf{x}_i} \theta(i,s)(\mathbb{Y}(\mathbf{x}_s) - \beta) + \sum_{\mathbf{x}_s \in N(\mathbf{x}_i)\setminus \mathbf{x}_{i\ell}} \theta_\ell \theta(i,s)(\mathbb{Y}(\mathbf{x}_s) - \beta)\right)\right),
\end{aligned} \tag{14}$$

which satisfies $\mathrm{E}[g_\ell(\omega_i, \tilde{\boldsymbol{\theta}}, \tilde{\beta}))] = 0$ for $\ell = 1, 2, \ldots, d$. Notice that $\theta_0^{-1/2} g_\ell(\omega_i, \tilde{\boldsymbol{\theta}}, \tilde{\beta}))$ also satisfies the moment condition, however, we choose (14) for the same reason as for (10) (see Section 5). Unlike (10), (14) involves sampling the neighbors of $\mathbf{x}_{i\ell}$. Given our neighborhood structure, there are two candidates for $\mathbf{x}_{i\ell}$ for each ℓ. We select one of the two candidates with probability $1/2$. Additionally, we have a moment condition with respect to $\tilde{\beta}$: $\mathrm{E}[\mathrm{E}[\mathbb{Y}(\mathbf{x}_i)|\mathbb{Y}(N(\mathbf{x}_i))]] = \tilde{\beta}$. Combined with (9), this condition implies that $\mathrm{E}\left[\tilde{\beta} + \sum_{\mathbf{x}_s \in N(\mathbf{x}_i)} \tilde{\theta}(i,s)\mathbb{Y}(\mathbf{x}_s)\right] - \tilde{\beta} = 0$. Therefore, the following satisfies $\mathrm{E}[g_{d+1}(\omega_i, \tilde{\boldsymbol{\theta}}, \tilde{\beta}))] = 0$:

$$g_{d+1}(\omega_i, \boldsymbol{\theta}, \beta) = \sum_{\mathbf{x}_s \in N(\mathbf{x}_i)} \theta(i,s)\mathbb{Y}(\mathbf{x}_s). \tag{15}$$

In the context of DOvS, we observe $\bar{Y}(\mathbf{x}_i)$, a realization of $\mathbb{Y}^{\varepsilon}(\mathbf{x}_i)$, instead of $\mathbb{Y}(\mathbf{x}_i)$. We define the corresponding sample moment functions, $g_0^{\varepsilon}, g_1^{\varepsilon}, \ldots, g_{d+1}^{\varepsilon}$, as follows.

$$g_0^{\varepsilon}(\omega_i, \boldsymbol{\theta}, \beta) = \frac{1}{{\theta_0}^{1/2}} - \theta_0^{1/2}\left(\bar{Y}^2(\mathbf{x}_i) - \left(\beta + \sum_{\mathbf{x}_s \in N(\mathbf{x}_i)} \theta(i,s)(\bar{Y}(\mathbf{x}_s) - \beta)\right)^2 - \frac{S^2(\mathbf{x}_i)}{r} + \sum_{\mathbf{x}_s \in N(\mathbf{x}_i)} \theta^2(i,s)\frac{S^2(\mathbf{x}_s)}{r}\right), \tag{16}$$

$$\begin{aligned} g_{\ell}^{\varepsilon}(\omega_i, \boldsymbol{\theta}, \beta) = &-\frac{\theta_{\ell}}{\theta_0^{1/2}(1-\theta_{\ell}^2)} + \theta_0^{1/2}\bar{Y}(\mathbf{x}_i)\bar{Y}(\mathbf{x}_{i\ell}) \\ &- \theta_0^{1/2}\left(\beta + \frac{1}{1-\theta_{\ell}^2}\left(\sum_{\mathbf{x}_s \in N(\mathbf{x}_i)\setminus \mathbf{x}_{i\ell}} \theta(i,s)(\bar{Y}(\mathbf{x}_s) - \beta) + \sum_{\mathbf{x}_s \in N(\mathbf{x}_{i\ell})\setminus \mathbf{x}_i} \theta_{\ell}\theta(i,s)(\bar{Y}(\mathbf{x}_s) - \beta)\right)\right) \\ &\times \left(\beta + \frac{1}{1-\theta_{\ell}^2}\left(\sum_{\mathbf{x}_s \in N(\mathbf{x}_{i\ell})\setminus \mathbf{x}_i} \theta(i,s)(\bar{Y}(\mathbf{x}_s) - \beta) + \sum_{\mathbf{x}_s \in N(\mathbf{x}_i)\setminus \mathbf{x}_{i\ell}} \theta_{\ell}\theta(i,s)(\bar{Y}(\mathbf{x}_s) - \beta)\right)\right) \\ &+ \frac{\theta_0^{1/2}\theta_{\ell}}{(1-\theta_{\ell}^2)^2}\left(\sum_{\mathbf{x}_s \in N(\mathbf{x}_i)\setminus \mathbf{x}_{i\ell}} \theta^2(i,s)\frac{S^2(\mathbf{x}_s)}{r} + \sum_{\mathbf{x}_s \in N(\mathbf{x}_{i\ell})\setminus \mathbf{x}_i} \theta^2(i,s)\frac{S^2(\mathbf{x}_s)}{r}\right), \end{aligned} \tag{17}$$

$$g_{d+1}^{\varepsilon}(\omega_i, \boldsymbol{\theta}, \beta) = \sum_{\mathbf{x}_s \in N(\mathbf{x}_i)} \theta(i,s)(\bar{Y}(\mathbf{x}_s) - \beta). \tag{18}$$

The following lemma shows that $g^{\varepsilon} \triangleq (g_0^{\varepsilon}, g_1^{\varepsilon}, \ldots, g_{d+1}^{\varepsilon})$ provides the desired moment condition.

Lemma 1. *Given true parameters $\tilde{\boldsymbol{\theta}}$ and $\tilde{\beta}$, $\mathrm{E}[g^{\varepsilon}(\omega_i, \tilde{\boldsymbol{\theta}}, \tilde{\beta})] = \mathbf{0}_p$.*

Proof. Recall that $\bar{Y}(\mathbf{x})$ is a realization of $\mathbb{Y}^{\varepsilon}(\mathbf{x})$. From the definition in Section 2, $\mathbb{Y}^{\varepsilon}(\mathbf{x}) = \mathbb{Y}(\mathbf{x}) + \bar{\varepsilon}(\mathbf{x})$. Clearly, $\mathrm{E}[g_{d+1}^{\varepsilon}(\omega_i, \tilde{\boldsymbol{\theta}}, \tilde{\beta})] = 0$ as $\mathrm{E}[\bar{Y}(\mathbf{x})] = \beta$. Since $\mathbb{Y}(\mathbf{x}) \perp \varepsilon(\mathbf{x})$ and all solutions are simulated independently $\mathrm{E}[\bar{Y}^2(\mathbf{x}_i)] = \mathrm{E}[(\mathbb{Y}(\mathbf{x}_i) + \bar{\varepsilon}(\mathbf{x}_i))^2] = \mathrm{E}[\mathbb{Y}^2(\mathbf{x}_i)] + \sigma^2(\mathbf{x}_i)/r$, and $\mathrm{E}[\bar{Y}(\mathbf{x}_i)\bar{Y}(\mathbf{x}_{i\ell})] = \mathrm{E}[(\mathbb{Y}(\mathbf{x}_i) + \bar{\varepsilon}(\mathbf{x}_i))(\mathbb{Y}(\mathbf{x}_{i\ell}) + \bar{\varepsilon}(\mathbf{x}_{i\ell}))] = \mathrm{E}[\mathbb{Y}(\mathbf{x}_i)\mathbb{Y}(\mathbf{x}_{i\ell})]$. Therefore, after some algebra we can show that

$$\mathrm{E}\left[\frac{1}{\tilde{\theta}_0} - \bar{Y}^2(\mathbf{x}_i) + \left(\tilde{\beta} + \sum_{\mathbf{x}_s \in N(\mathbf{x}_i)} \tilde{\theta}(i,s)(\bar{Y}(\mathbf{x}_s) - \tilde{\beta})\right)^2\right] = -\frac{\sigma^2(\mathbf{x}_i)}{r} + \sum_{\mathbf{x}_s \in N(\mathbf{x}_i)} \tilde{\theta}^2(i,s)\frac{\sigma^2(\mathbf{x}_s)}{r}. \tag{19}$$

Because $\mathrm{E}[S^2(\mathbf{x})] = \sigma^2(\mathbf{x})$, it follows from (19) that $\mathrm{E}[g_0^{\varepsilon}(\omega_i, \tilde{\boldsymbol{\theta}}, \tilde{\beta})] = 0$. Similarly,

$$\begin{aligned} &\mathrm{E}\Bigg[\bar{Y}(\mathbf{x}_i)\bar{Y}(\mathbf{x}_{i\ell}) - \left(\tilde{\beta} + \frac{1}{1-\tilde{\theta}_{\ell}^2}\left(\sum_{\mathbf{x}_s \in N(\mathbf{x}_i)\setminus \mathbf{x}_{i\ell}} \tilde{\theta}(i,s)(\bar{Y}(\mathbf{x}_s) - \tilde{\beta}) + \sum_{\mathbf{x}_s \in N(\mathbf{x}_{i\ell})\setminus \mathbf{x}_i} \tilde{\theta}_{\ell}\tilde{\theta}(i,s)(\bar{Y}(\mathbf{x}_s) - \tilde{\beta})\right)\right) \\ &\qquad \times \left(\tilde{\beta} + \frac{1}{1-\tilde{\theta}_{\ell}^2}\left(\sum_{\mathbf{x}_s \in N(\mathbf{x}_{i\ell})\setminus \mathbf{x}_i} \tilde{\theta}(i,s)(\bar{Y}(\mathbf{x}_s) - \tilde{\beta}) + \sum_{\mathbf{x}_s \in N(\mathbf{x}_i)\setminus \mathbf{x}_{i\ell}} \tilde{\theta}_{\ell}\tilde{\theta}(i,s)(\bar{Y}(\mathbf{x}_s) - \tilde{\beta})\right)\right)\Bigg] \\ &= \frac{\tilde{\theta}_{\ell}}{\tilde{\theta}_0(1-\tilde{\theta}_{\ell}^2)} - \frac{\tilde{\theta}_{\ell}}{(1-\tilde{\theta}_{\ell}^2)^2}\left(\sum_{\mathbf{x}_s \in N(\mathbf{x}_i)\setminus \mathbf{x}_{i\ell}} \tilde{\theta}^2(i,s)\frac{\sigma^2(\mathbf{x}_s)}{r} + \sum_{\mathbf{x}_s \in N(\mathbf{x}_{i\ell})\setminus \mathbf{x}_i} \tilde{\theta}^2(i,s)\frac{\sigma^2(\mathbf{x}_s)}{r}\right). \end{aligned}$$

Hence, combined with that $\mathrm{E}[S^2(\mathbf{x}_s)] = \sigma^2(\mathbf{x}_s), \mathrm{E}[g_{\ell}^{\varepsilon}(\omega_i, \tilde{\boldsymbol{\theta}}, \tilde{\beta})] = 0$ for $\ell = 1, 2, \ldots, d$. □

Given sample average of the moment function $\bar{g}^{\varepsilon}(\boldsymbol{\theta}, \beta) = \sum_{i=1}^{k} g^{\varepsilon}(\omega_i, \boldsymbol{\theta}, \beta)/k$, the full-GMM estimators are $(\widehat{\boldsymbol{\theta}}, \widehat{\beta}) \triangleq \arg\min_{\boldsymbol{\theta}, \beta \in \Theta} ||\bar{g}^{\varepsilon}(\boldsymbol{\theta}, \beta)||^2$, where Θ is the feasible space for $(\widehat{\boldsymbol{\theta}}, \widehat{\beta})$ defined in Section 3.2. Note that this formulation is the same as choosing an identity matrix for W in (5).

Figure 1a illustrates an example of ω_i in two-dimensional $\mathscr{X}$ given the central design point, $\mathbf{x}_i$. The simulated solutions are in solid circles. All neighbors of $\mathbf{x}_i$, $\mathbf{x}_{i1}$ and $\mathbf{x}_{i2}$ are simulated, where $\mathbf{x}_{i1}$ and $\mathbf{x}_{i2}$ are

(a) Full-GMM (b) Slim-GMM

Figure 1: Sampling schemes for full-GMM and slim-GMM in a two-dimensional solution space. Simulated solutions are represented by solid circles.

the uniformly selected neighbors of $\mathbf{x}_i$ along the first and the second coordinate directions, respectively. Notice that one solution belongs to both $N(\mathbf{x}_{i1})$ and $N(\mathbf{x}_{i2})$. In a d-dimensional case, $\mathbf{x}_i$ has $2d$ neighbors among which we select d neighbors, $\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{id}$, to construct $g_1, g_2, \ldots, g_d$. For each $\mathbf{x}_{i\ell}$, there are $(2d-1)$ neighbors that are not included in $N(\mathbf{x}_i)$. Also, each pair $(\mathbf{x}_{i\ell}, \mathbf{x}_{i\ell'}), \ell \neq \ell'$, shares one neighbor. Thus, the total number of feasible solutions we simulate for full-GMM estimation given k central design points is $\left(1+2d+d(2d-1)-\binom{d}{2}\right)k = \left(1+\frac{3}{2}d+\frac{3}{2}d^2\right)k$. This implies that the total simulation effort for full-GMM estimation increases in $\Omega(d^2kr)$, which is not desirable for a high-dimensional DOvS problem. In the next section, we introduce slim-GMM, which requires $\Omega(dkr)$ simulation budget instead of $\Omega(d^2kr)$.

3.3 Slim-GMM Estimator

The expressions for g_0 and g_{d+1} in (16) and (18) do not involve the neighbors of $\mathbf{x}_{i\ell}$, whereas g_ℓ in (17) has summations over the solutions in $N(\mathbf{x}_{i\ell})$. As shown in Section 3.2, the neighbors of $\mathbf{x}_{i\ell}$ are simulated to estimate the conditional expectation in (12). To reduce the number of solutions to simulate, we modify $g_\ell(\omega_i, \boldsymbol{\theta}, \beta)$ by introducing the following $(2d-1)$-dimensional multinomial random vector

$$\mathbf{Z}_{i\ell} \sim \text{multinomial}\left(1, (2d-1)^{-1}, (2d-1)^{-1}, ..., (2d-1)^{-1}\right). \tag{20}$$

By definition, $\mathbf{Z}_{i\ell}$ has only one nonzero element equals 1. We pair each element of $\mathbf{Z}_{i\ell}$ with each of $2d-1$ solutions in $N(\mathbf{x}_{i\ell}) \setminus \mathbf{x}_i$ and simulate the solution only if the corresponding element of $\mathbf{Z}_{i\ell}$ is 1. Suppose $\mathbf{Z}^1_{i\ell}$ and $\mathbf{Z}^2_{i\ell}$ are i.i.d. random vectors of distribution (20) drawn independently from ω_i, then we define the modified ℓth moment function for slim-GMM as

$$\begin{aligned}
&\ddot{g}^{\varepsilon}_\ell(\omega_i, \boldsymbol{\theta}, \beta) \\
&= -\frac{\theta_\ell}{\theta_0^{1/2}(1-\theta_\ell^2)} + \theta_0^{1/2}\bar{Y}(\mathbf{x}_i)\bar{Y}(\mathbf{x}_{i\ell}) + \frac{\theta_0^{1/2}\theta_\ell}{(1-\theta_\ell^2)^2}\left(\sum_{\mathbf{x}_s \in N(\mathbf{x}_i)\setminus \mathbf{x}_{i\ell}} \theta^2(i,s)\frac{S^2(\mathbf{x}_s)}{r}\right) \\
&- \theta_0^{1/2}\left(\beta + \frac{1}{1-\theta_\ell^2}\left(\sum_{\mathbf{x}_s \in N(\mathbf{x}_i)\setminus \mathbf{x}_{i\ell}} \theta(i,s)(\bar{Y}(\mathbf{x}_s)-\beta) + (2d-1)\sum_{\mathbf{x}_s \in N(\mathbf{x}_{i\ell})\setminus \mathbf{x}_i} Z^1_{i\ell}(s)\theta_\ell\theta(i,s)(\bar{Y}^{(1)}(\mathbf{x}_s)-\beta)\right)\right) \\
&\times \left(\beta + \frac{1}{1-\theta_\ell^2}\left((2d-1)\sum_{\mathbf{x}_s \in N(\mathbf{x}_{i\ell})\setminus \mathbf{x}_i} Z^2_{i\ell}(s)\theta(i,s)(\bar{Y}^{(2)}(\mathbf{x}_s)-\beta) + \sum_{\mathbf{x}_s \in N(\mathbf{x}_i)\setminus \mathbf{x}_{i\ell}} \theta_\ell\theta(i,s)(\bar{Y}(\mathbf{x}_s)-\beta)\right)\right),
\end{aligned}$$

where $Z^1_{i\ell}(s)$ (or $Z^2_{i\ell}(s)$) denotes the element of $\mathbf{Z}^1_{i\ell}$ (or $\mathbf{Z}^2_{i\ell}$) corresponding to $\mathbf{x}_s$ and $\bar{Y}^{(1)}(\mathbf{x}_s) \perp \bar{Y}^{(2)}(\mathbf{x}_s)$. When $\mathbf{Z}^1_{i\ell}$ and $\mathbf{Z}^2_{i\ell}$ pick different solutions to simulate, then the latter condition is satisfied. When $\mathbf{Z}^1_{i\ell}$ and $\mathbf{Z}^2_{i\ell}$ select the same $\mathbf{x}_s \in N(\mathbf{x}_{i\ell}) \setminus \mathbf{x}_i$, $\bar{Y}^{(1)}(\mathbf{x}_s)$ is from the first r replications, and $\bar{Y}^{(2)}(\mathbf{x}_s)$ is from another r replications. Thus, ω_i for the slim-GMM includes observations in (6) as well as $\{\mathbf{Z}^1_{i\ell}, \mathbf{Z}^2_{i\ell}, \ell = 1,2,\ldots,d\}$. The following lemma shows that $\ddot{g}_\ell(\omega_i, \boldsymbol{\theta}, \beta)$ satisfies the moment condition.

Lemma 2. *If* $\mathbb{Y} \sim \mathrm{N}(\tilde{\beta}\mathbf{1}_n, \mathbf{Q}^{-1}(\tilde{\boldsymbol{\theta}}))$, $\mathrm{E}[\ddot{g}^\varepsilon_\ell(\omega_i, \tilde{\boldsymbol{\theta}}, \tilde{\beta})] = 0$.

Proof. Because $\mathbf{Z}^1_{i\ell}, \mathbf{Z}^2_{i\ell} \perp \omega_i$ and $\mathbf{Z}^1_{i\ell} \perp \mathbf{Z}^2_{i\ell}$,

$$
\begin{aligned}
&\mathrm{E}\Bigg[\left(\tilde{\beta} + \frac{1}{1-\tilde{\theta}^2_\ell}\left(\sum_{\mathbf{x}_s \in N(\mathbf{x}_i)\setminus\mathbf{x}_{i\ell}} \tilde{\theta}(i,s)(\bar{Y}(\mathbf{x}_s) - \tilde{\beta}) + (2d-1)\sum_{\mathbf{x}_s \in N(\mathbf{x}_{i\ell})\setminus\mathbf{x}_i} Z^1_{i\ell}(s)\tilde{\theta}_\ell\tilde{\theta}(i,s)(\bar{Y}^{(1)}(\mathbf{x}_s) - \tilde{\beta})\right)\right) \\
&\times\left(\tilde{\beta} + \frac{1}{1-\tilde{\theta}^2_\ell}\left((2d-1)\sum_{\mathbf{x}_s \in N(\mathbf{x}_{i\ell})\setminus\mathbf{x}_i} Z^2_{i\ell}(s)\tilde{\theta}(i,s)(\bar{Y}^{(2)}(\mathbf{x}_s) - \tilde{\beta}) + \sum_{\mathbf{x}_s \in N(\mathbf{x}_i)\setminus\mathbf{x}_{i\ell}} \tilde{\theta}_\ell\tilde{\theta}(i,s)(\bar{Y}(\mathbf{x}_s) - \tilde{\beta})\right)\right)\Bigg| \omega_i\Bigg] \\
&= \left(\beta + \frac{1}{1-\tilde{\theta}^2_\ell}\left(\sum_{\mathbf{x}_s \in N(\mathbf{x}_i)\setminus\mathbf{x}_{i\ell}} \tilde{\theta}(i,s)(\bar{Y}(\mathbf{x}_s) - \tilde{\beta}) + \sum_{\mathbf{x}_s \in N(\mathbf{x}_{i\ell})\setminus\mathbf{x}_i} \tilde{\theta}_\ell\tilde{\theta}(i,s)(\bar{Y}^{(1)}(\mathbf{x}_s) - \tilde{\beta})\right)\right) \\
&\quad\times\left(\tilde{\beta} + \frac{1}{1-\tilde{\theta}^2_\ell}\left(\sum_{\mathbf{x}_s \in N(\mathbf{x}_{i\ell})\setminus\mathbf{x}_i} \tilde{\theta}(i,s)(\bar{Y}^{(2)}(\mathbf{x}_s) - \tilde{\beta}) + \sum_{\mathbf{x}_s \in N(\mathbf{x}_i)\setminus\mathbf{x}_{i\ell}} \tilde{\theta}_\ell\tilde{\theta}(i,s)(\bar{Y}(\mathbf{x}_s) - \tilde{\beta})\right)\right),
\end{aligned}
$$

where the expectation is with respect to $\mathbf{Z}^1_{i\ell}$ and $\mathbf{Z}^2_{i\ell}$. Therefore,

$$
\begin{aligned}
&\mathrm{E}\Bigg[\bar{Y}(\mathbf{x}_i)\bar{Y}(\mathbf{x}_{i\ell}) - \left(\tilde{\beta} + \frac{1}{1-\tilde{\theta}^2_\ell}\left(\sum_{\mathbf{x}_s \in N(\mathbf{x}_i)\setminus\mathbf{x}_{i\ell}} \tilde{\theta}(i,s)(\bar{Y}(\mathbf{x}_s) - \tilde{\beta}) + \sum_{\mathbf{x}_s \in N(\mathbf{x}_{i\ell})\setminus\mathbf{x}_i} \tilde{\theta}_\ell\tilde{\theta}(i,s)(\bar{Y}^{(1)}(\mathbf{x}_s) - \tilde{\beta})\right)\right) \\
&\quad\times\left(\tilde{\beta} + \frac{1}{1-\tilde{\theta}^2_\ell}\left(\sum_{\mathbf{x}_s \in N(\mathbf{x}_{i\ell})\setminus\mathbf{x}_i} \tilde{\theta}(i,s)(\bar{Y}^{(2)}(\mathbf{x}_s) - \tilde{\beta}) + \sum_{\mathbf{x}_s \in N(\mathbf{x}_i)\setminus\mathbf{x}_{i\ell}} \tilde{\theta}_\ell\tilde{\theta}(i,s)(\bar{Y}(\mathbf{x}_s) - \tilde{\beta})\right)\right)\Bigg] \\
&= \frac{\tilde{\theta}_\ell}{\tilde{\theta}_0(1-\tilde{\theta}^2_\ell)} - \frac{\tilde{\theta}_\ell}{(1-\tilde{\theta}^2_\ell)^2}\left(\sum_{\mathbf{x}_s \in N(\mathbf{x}_i)\setminus\mathbf{x}_{i\ell}} \tilde{\theta}^2(i,s)\frac{\sigma^2(\mathbf{x}_s)}{r}\right). \qquad (21)
\end{aligned}
$$

Notice that only the simulation error variances of solutions in $N(\mathbf{x}_i) \setminus \mathbf{x}_{i\ell}$ show up in (21), because $\bar{Y}^{(1)}(\mathbf{x}_s) \perp \bar{Y}^{(2)}(\mathbf{x}_s)$, i.e., the mean of the product of their simulation errors equals 0. Hence, $\mathrm{E}[\ddot{g}^\varepsilon_\ell(\omega_i, \tilde{\boldsymbol{\theta}}, \tilde{\beta})] = 0$. □

The slim-GMM estimators, $\ddot{\boldsymbol{\theta}}$ and $\ddot{\beta}$, are obtained by using $\ddot{g}^\varepsilon = (g^\varepsilon_0, \ddot{g}^\varepsilon_1, \ldots, \ddot{g}^\varepsilon_d, g^\varepsilon_{d+1})^\top$ as a moment function instead of g^ε. For slim-GMM, at most two neighbors of $\mathbf{x}_{i\ell}$ are selected for simulation regardless of the dimension of the solution space. Figure 1b shows an example of the sampling scheme for slim-GMM in the same two-dimensional solution space in Figure 1a. One solution in $N(\mathbf{x}_{i1})$ is represented by a dashed circle indicating it was not simulated, i.e., corresponding elements of $\mathbf{Z}^1_{i1}$ and $\mathbf{Z}^2_{i1}$ for this solution were 0. On the other hand, only one solution in $N(\mathbf{x}_{i2})$ was selected by both $\mathbf{Z}^1_{i2}$ and $\mathbf{Z}^2_{i2}$, which happens to be one of the solutions selected for x_{i1}; this solution is simulated $3r$ times as described earlier.

In a d-dimensional case, we simulate $(1+2d+2d)k = (1+4d)k$ solutions given k central design points allowing some solutions to be chosen multiple times; we sample all $2d$ neighbors of the central design point and only two neighbors of each of $\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{id}$. Therefore, the required total simulation effort for slim-GMM is $\Omega(dkr)$, which is significantly smaller than that of the full-GMM for large d.

4 CONSISTENCY OF FULL-GMM AND SLIM-GMM

Since the observations from the GMRF are correlated, we need to assume the GMRF to be stationary to show consistency of the full-GMM and slim-GMM estimators as discussed in Section 3.1). As the

moment functions for both estimators are highly nonlinear, it is difficult to establish a general identification condition as mentioned in Section 3.1. Another challenge is that $\mathrm{E}[g^{\varepsilon}(\omega_i,\boldsymbol{\theta},\beta)]$ depends on variances and covariances of the sampled solutions, which are the elements of $\mathbf{Q}^{-1}(\tilde{\boldsymbol{\theta}})$. Although $\mathbf{Q}(\boldsymbol{\theta})$ has a nice structure in (2), once inverted, the elements of $\mathbf{Q}^{-1}(\boldsymbol{\theta})$ are nonlinear functions of $\boldsymbol{\theta}$ that depends on n as well as the locations of the sampled solutions within $\mathscr{X}$. The following theorem shows that the full-GMM and slim-GMM estimators are consistent under the stationarity and identification assumptions.

Theorem 2 Suppose $\mathbb{Y} \sim \mathrm{N}(\tilde{\beta}\mathbf{1}_n, \mathbf{Q}^{-1}(\tilde{\boldsymbol{\theta}}))$ is stationary. If $\mathrm{E}[g^{\varepsilon}(\omega_i,\boldsymbol{\theta},\beta)] = \mathbf{0}_{d+2}$ has a unique solution at $(\boldsymbol{\theta},\beta) = (\tilde{\boldsymbol{\theta}},\tilde{\beta})$, then (i) $(\widehat{\boldsymbol{\theta}},\widehat{\beta}) \xrightarrow{p} (\tilde{\boldsymbol{\theta}},\tilde{\beta})$; (ii) $(\ddot{\boldsymbol{\theta}},\ddot{\beta}) \xrightarrow{p} (\tilde{\boldsymbol{\theta}},\tilde{\beta})$.

Proof. We first show consistency of $(\widehat{\boldsymbol{\theta}},\widehat{\beta})$ by verifying the conditions in Theorem 1. Following the discussion in Section 2, we can choose Θ to be a compact space by adjusting the lower bounds on $\boldsymbol{\theta}$ and upper bounds on θ_0 and β. From definition (6), ω_i is a vector of simulation outputs. Given $\boldsymbol{\theta}$ and β, and for each $\ell = 0,1,\ldots,d+1$, $g^{\varepsilon}_{\ell}(\omega_i,\boldsymbol{\theta},\beta)$ is a continuous functions of $\boldsymbol{\theta}$ and β, and second-order polynomial in the elements of ω_i: namely, $g^{\varepsilon}_{\ell}(\omega_i,\boldsymbol{\theta},\beta) = \sum_{s,s'} c_{i,s,s'}(\boldsymbol{\theta},\beta)\omega_{is}\omega_{is'}$, where ω_{is} is the sth element of ω_i and $c_{i,s,s'}(\boldsymbol{\theta},\beta)$'s are the parameters that only depend on $\boldsymbol{\theta}$ and β. Since $g^{\varepsilon}_{\ell}(\omega_i,\boldsymbol{\theta},\beta)$ is a continuous function of $\boldsymbol{\theta}$ and β, $c_{i,s,s'}(\boldsymbol{\theta},\beta)$'s are also continuous functions of $\boldsymbol{\theta}$ and β, and therefore bounded for compact Θ. Because $\omega_{is}\omega_{is'} \le (\omega_{is}^2 + \omega_{is'}^2)/2$, $g^{\varepsilon}_{\ell}(\omega_i,\boldsymbol{\theta},\beta) \le \sum_{s,s'} \max\{0, c_{i,s,s'}(\boldsymbol{\theta},\beta)\}\left((\omega_{is})^2 + (\omega_{is'})^2\right)$. Since there are a finite number of (s,s') pairs, we can find $C_i < \infty$ such that $C_i = \sup_{(\boldsymbol{\theta},\beta)\in\Theta} \max_{(s,s')}\{\max\{0, c_{i,s,s'}(\boldsymbol{\theta},\beta)\}\}$, thus for any $\boldsymbol{\theta}$ and β $g^{\varepsilon}_{\ell}(\omega_i,\boldsymbol{\theta},\beta) \le \sum_{s\neq s'} C_i\omega_{is}\omega_{is'}$. Hence, combined with that the elements of ω_i have finite second moments as they are normally distributed, we can show $\mathrm{E}[\sup_{\boldsymbol{\lambda}\in\Lambda} ||g(\omega_i,\boldsymbol{\lambda})||] < \infty$. Condition (ii)–(iv) can be verified similarly for $\ddot{g}^{\varepsilon}$. □

Rue and Held (2005) provide two common examples of stationary GMRFs: (i) $\mathscr{X}$ is wrapped onto a torus; (ii) $\mathscr{X}$ is an infinite lattice. In both cases, $\mathscr{X}$ has no boundaries, hence the distribution of $\mathbb{Y}$ has no boundary effects granting stationarity. Toroidal $\mathscr{X}$ is a common assumption in image texture analysis as it simplifies the computation of the likelihood function (Dryden, Ippoliti, and Romagnoli 2002). However, for a DOvS problem, wrapping $\mathscr{X}$ onto a torus is not plausible in general as the objective function values of the solutions at different boundaries of $\mathscr{X}$ may differ significantly. Assuming $\mathscr{X}$ is an infinite lattice may not be plausible, either. However, when $\mathscr{X}$ is large, the GMM estimators may still approximate the true parameters pretty well as can be confirmed by experiment results in Section 6.

5 SMALL-SAMPLE PERFORMANCE

There are several possible choices for g_0, g_ℓ, and g_{d+1} other than (10), (14) and (15) as discussed in Section 3.2; $\theta_0^{-1/2} g_0(\omega_i,\boldsymbol{\theta},\beta)$ and $\theta_0^{-1/2} g_\ell(\omega_i,\boldsymbol{\theta},\beta)$ also satisfy the moment conditions in (4). However, these functions show poorer finite-sample performances than $g_0(\omega_i,\boldsymbol{\theta},\beta)$ and $g_\ell(\omega_i,\boldsymbol{\theta},\beta)$ for small k; there is a positive probability that $\widehat{\theta}_0$ diverges to infinity (or pushed to the boundary of Θ) due to sampling error for any finite k.

To see this, suppose we use $\theta_0^{-1/2} g_0(\omega_i,\boldsymbol{\theta},\beta) = \theta_0^{-1} - \mathbb{Y}(\mathbf{x}_i)^2 + \left(\beta + \sum_{\mathbf{x}_s\in N(\mathbf{x}_i)} \theta(i,s)\mathbb{Y}(\mathbf{x}_s)\right)^2$ as a moment function instead of $g_0(\omega_i,\boldsymbol{\theta},\beta)$ and $\tilde{\theta}_\ell, 1 \le \ell \le d$, and $\tilde{\beta}$ are known. We define event $\Xi \triangleq \{\omega_i | \mathbb{Y}(\mathbf{x}_i)^2 - (\tilde{\beta} + \sum_{\mathbf{x}_s\in N(\mathbf{x}_i)} \tilde{\theta}(i,s)\mathbb{Y}(\mathbf{x}_s))^2 < 0\}$. When Ξ occurs, θ_0 is pushed to infinity to make $|\theta_0^{-1/2} g_0(\omega_i,\boldsymbol{\theta},\beta)|$ small as θ_0 is constrained to be positive. Note that $\mathbb{Y}(\mathbf{x}_i) | \mathbb{Y}(N(\mathbf{x}_i)) \overset{D}{=} V + W$, where $V = \beta + \sum_{\mathbf{x}_s\in N(\mathbf{x}_i)} \tilde{\theta}(i,s)\mathbb{Y}(\mathbf{x}_s)$ and $W \sim \mathrm{N}(0, \tilde{\theta}_0^{-1})$. Therefore, Ξ occurs, if and only if, $(V+W)^2 - V^2 =$

Table 1: Estimated parameters for Case 1–4 from full-GMM and slim-GMM with $k = 5, 10$ and 50 and from MLE with $n_0 = 45, 90$ and 450. Starred (*) statistics are computed without the outliers, where the estimators are greater than $1000\theta_0$. All outliers are observed for the slim-GMM; Case 1 with $k = 10$, Case 2 and 4 with $k = 5$ each has one outlier; Case 4 with $k = 10$ has two outliers. We highlighted the results of all three estimators for Case 1 in bold when $k = 5$ and $n_0 = 45$.

			Case1			Case2			Case3			Case4		
			$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_0$	$\hat{\theta}_1$	$\hat{\theta}_2$
			0.001	0.1	0.1	1	0.1	0.1	0.001	0.4	0.05	1	0.4	0.05
full-GMM	$k=5$	Bias	**0.000563**	**0.016**	**0.012**	0.883	0.016	0.013	0.000083	−0.178	0.054	0.446	−0.174	0.048
		SE	**0.000060**	**0.004**	**0.004**	0.121	0.005	0.005	0.000050	0.005	0.004	0.232	0.005	0.004
	k=10	Bias	0.000669	0.016	0.019	0.659	0.016	0.017	0.000573	−0.136	0.058	0.799	−0.132	0.053
		SE	0.000138	0.004	0.004	0.096	0.004	0.004	0.000211	0.005	0.004	0.484	0.005	0.004
	k=50	Bias	0.000130	0.006	0.009	0.144	0.006	0.009	0.000150	−0.033	0.020	0.250	−0.031	0.019
		SE	0.000013	0.003	0.003	0.015	0.003	0.003	0.000034	0.002	0.002	0.112	0.002	0.002
slim-GMM	k=5	Bias	**0.000744**	**0.013**	**0.011**	1.376*	0.012	0.011	0.000276	−0.189	0.056	0.598*	−0.186	0.050
		SE	**0.000174**	**0.004**	**0.004**	0.557	0.004	0.004	0.000068	0.005	0.004	0.165	0.005	0.004
	k=10	Bias	0.000460*	0.016	0.015	1.018	0.017	0.016	0.000335	−0.145	0.059	0.237*	−0.146	0.057
		SE	0.000054	0.004	0.004	0.358	0.004	0.004	0.000149	0.004	0.004	0.056	0.005	0.004
	k=50	Bias	0.000125	0.006	0.009	0.136	0.009	0.010	0.000201	−0.044	0.028	0.090	−0.042	0.028
		SE	0.000011	0.003	0.003	0.012	0.003	0.003	0.000139	0.003	0.003	0.026	0.003	0.003
MLE	n_0=45	Bias	**0.002729**	**0.066**	**0.111**	1.869	0.061	0.055	0.001541	−0.218	0.163	1.280	−0.199	0.143
		SE	**0.000114**	**0.006**	**0.007**	0.088	0.006	0.006	0.000079	0.006	0.007	0.071	0.006	0.006
	n_0=90	Bias	0.000974	0.074	0.066	0.922	0.065	0.063	0.000427	−0.124	0.087	0.401	−0.125	0.081
		SE	0.000037	0.006	0.006	0.037	0.006	0.006	0.000025	0.006	0.005	0.024	0.006	0.005
	n_0=450	Bias	0.000043	0.001	0.001	0.043	0.001	0.002	0.000013	−0.006	0.001	0.014	−0.006	0.001
		SE	0.000003	0.003	0.002	0.003	0.003	0.002	0.000004	0.001	0.001	0.004	0.001	0.001

$(2V+W)W \leq 0$ and the probability of this event is

$$\begin{aligned}\Pr\{(2V+W)W \leq 0\} &= \Pr\{0 \leq W \leq -2V | V \leq 0\} \Pr\{V \leq 0\} + \Pr\{-2V \leq W \leq 0 | V \geq 0\} \Pr\{V \geq 0\} \\ &= \int_{-\infty}^{\infty} \left(\Phi(\sqrt{\tilde{\theta}_0}|2v|) - 0.5 \right) f_V(v) dv,\end{aligned}$$

where f_V is the probability density function of V and Φ is the cumulative distribution function of a standard normal random variable. Hence, for any $\tilde{\theta}_0 > 0$, event Ξ occurs with a positive probability and $\Pr\{\Xi\}$ is an increasing function of θ_0. In the context of DOvS, θ_0 tends to be small ($\ll 1$) representing large spatial uncertainty –small precision– about the response surface. Thus, the probability of having large $\widehat{\theta}_0$ is pretty small for a practical DOvS problem.

Note that Ξ is defined for a single ω_i. As k increases, the probability that the estimated θ_0 is pushed to infinity converges to 0. For our choice of $g_0(\omega_i, \boldsymbol{\theta}, \beta)$, even if Ξ occurs, θ_0 is not pushed to infinity because $\theta_0^{-1/2}$ and $\theta_0^{1/2} \left(\mathbb{Y}(\mathbf{x}_i)^2 - (\beta + \sum_{\mathbf{x}_s \in N(\mathbf{x}_i)} \theta(i,s) \mathbb{Y}(\mathbf{x}_s)) \right)^2$ balance each other. A similar observation can be made for $\theta_0^{-1/2} g_\ell(\omega_i, \boldsymbol{\theta}, \beta)$.

6 EMPIRICAL RESULTS

In this section, we compare the empirical performance of the full-GMM and slim-GMM estimators with the MLE by estimating the parameters of sampled observations from a GMRF with known parameters. For all test cases, $\mathscr{X}$ is a two-dimensional 50×50 lattice. Hence, the sampled GMRFs are non-stationary, which makes all three estimators inexact. When a solution is "simulated," a normally distributed noise with mean 0 and variance 0.5^2 was added to the sampled response surface to induce simulation noise.

Four settings of $\tilde{\boldsymbol{\theta}}$ were tested, $\tilde{\boldsymbol{\theta}} = (0.001, 0.1, 0.1)$, $(1, 0.1, 0.1)$, $(0.001, 0.4, 0.05)$, $(1, 0.4, 0.05)$, while $\tilde{\beta}$ is fixed to 0 for all cases. Note that the $\tilde{\boldsymbol{\theta}}$ values were selected to examine the performance of the

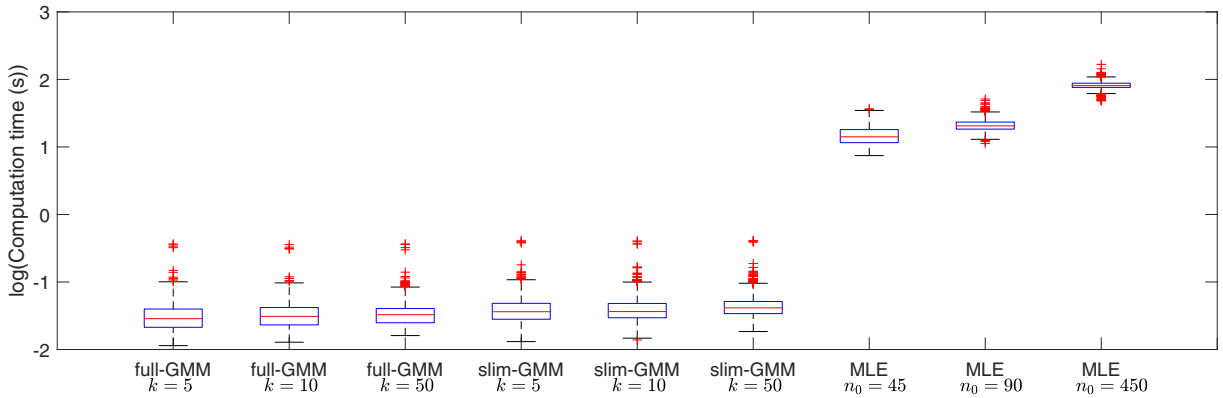

Figure 2: Computation (wall-clock) time of full-GMM, slim-GMM, and MLE for Case 3: $\tilde{\boldsymbol{\theta}} = (0.001, 0.4, 0.05)$ given $k = 5, 10, 50$ and $n_0 = 45, 90, 450$.

estimators under low vs. high precision ($\tilde{\theta}_0$) as well as low vs. high spatial correlations ($\tilde{\theta}_1$ and $\tilde{\theta}_2$). For full-GMM and slim-GMM, $k = 5, 10$, and 50 central design points were selected by Latin hypercube sampling. For MLE, we selected $n_0 = 45, 90$, and 450 initial design points by Latin hypercube sampling. Notice that these are the numbers of solutions (allowing duplicates) we simulate for slim-GMM when $k = 5, 10$, and 50, respectively. We used $r = 10$ for all experiments.

Table 1 shows the estimated bias and the standard error of each estimator computed from 1,000 replications for each test case. Some test cases that had outliers (i.e., estimated θ_0 is more than 1000 times larger than $\tilde{\theta}_0$) are marked with '$*$' and their biases and standard errors are computed without these outliers (See the caption of Table 1 for details). For $k = 5$ and 10 or equivalently $n_0 = 45$ and 90, both full-GMM and slim-GMM estimators have smaller bias and standard error in all test cases. When k is increased to 50 ($n_0 = 450$), MLEs have smaller biases and standard errors in all test cases. In the context of ARS, sampling 450 initial solutions out of 2500 feasible solutions for the purpose of hyperperameter estimation appear quite inefficient. Aside from a few outliers, full-GMM and slim-GMM do not have significant difference in performance in all test cases, which votes in favor of using the slim-GMM as it requires a smaller number of simulation runs given k. Note that one can always increase k or change from slim-GMM to full-GMM estimation by sampling additional solutions in these outlier cases. The value of $\tilde{\theta}_0$ does not affect the bias and standard error of the estimators of $\tilde{\theta}_1$ and $\tilde{\theta}_2$ for all three estimators. The relative bias of estimators of $\tilde{\theta}_1$ and $\tilde{\theta}_2$ are larger when they are unbalanced (Cases 3 and 4) for all three estimation procedures.

Figure 2 shows the computation time measured in wall-clock time for each estimation method for Case 3 on a machine equipped with Intel® Core™ i7-7700HQ processor and 2.80GHz RAM. All other cases showed similar patterns. The computation for GMM is about three orders of magnitude faster when $k = 5$ compared to that of MLE when $n_0 = 45$. When k and n_0 increase, the difference in computation time for GMM and MLE increases. The computational saving would be more dramatic for a larger solution space as GMM is not affected by the size of the solution space unlike MLE (see Section 2).

7 CONCLUSION

In this paper, we proposed the full-GMM and slim-GMM estimators of the hyperparameters of a GMRF and showed their consistency under mild assumptions. The computational complexity of these GMM estimators does not depend on the size of the feasible solution space whereas, computation required for MLE grows polynomially in the size of the feasible solution space. The empirical results show that for a reasonable range of sample size, the full-GMM and slim-GMM estimators have smaller empirical MSE than MLE.

REFERENCES

Dryden, I., L. Ippoliti, and L. Romagnoli. 2002. "Adjusted Maximum Likelihood and Pseudo-Likelihood Estimation for Noisy Gaussian Markov Random Fields". *Journal of Computational and Graphical Statistics* 11(2):370–388.

Geršgorin, S. 1931. "Über die Abgrenzung der Eigenwerte einer Matrix". *Bulletin de l'Académie des Sciences de l'URSS. Classe des sciences mathématiques et na* (6):749–754.

Jones, D. R., M. Schonlau, and W. J. Welch. 1998. "Efficient Global Optimization of Expensive Black-Box Functions". *Journal of Global Optimization* 13(4):455–492.

Manjunath, B. S., and R. Chellappa. 1991. "Unsupervised texture segmentation using Markov random field models". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13(5):478–482.

Mátyás, L. 1999. *Generalized method of moments estimation*, Volume 5. Cambridge University Press.

Newey, W. K., and D. McFadden. 1994. *Large sample estimation and hypothesis testing*, Volume 4 of *Handbook of Econometrics*, Chapter 36, 2111 – 2245. Elsevier.

Quan, N., J. Yin, S. H. Ng, and L. H. Lee. 2013. "Simulation Optimization via Kriging: A Sequential Search Using Expected Improvement with Computing Budget Constraints". *IIE Transactions* 45(7):763–780.

Rothenberg, T. J. 1971. "Identification in Parametric Models". *Econometrica* 39(3):577–91.

Rue, H. 2001. "Fast Sampling of Gaussian Markov Random Fields". *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* 63(2):325–338.

Rue, H., and L. Held. 2005. *Gaussian Markov Random Fields: Theory and Applications*. New York: Chapman and Hall/CRC.

Salemi, P., E. Song, B. L. Nelson, and J. Staum. 2018. "Gaussian Markov Random Fields for Discrete Optimization via Simulation: Framework and Algorithms". *Operations Research*. In print.

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

Scott, W., P. Frazier, and W. Powell. 2011. "The Correlated Knowledge Gradient for Simulation Optimization of Continuous Parameters using Gaussian Process Regression". *SIAM Journal on Optimization* 21(3):996–1026.

Xie, J., P. I. Frazier, and S. E. Chick. 2016. "Bayesian Optimization via Simulation with Pairwise Sampling and Correlated Prior Beliefs". *Operations Research* 64(2):542–559.

ACKNOWLEDGEMENT

The authors thank Xinmeng Wang for testing the full-GMM and slim-GMM estimation algorithms as a part of her M.S. degree paper at the Penn State University. Computations for this research were performed on the Penn State University's Institute for CyberScience Advanced CyberInfrastructure (ICS-ACI).

AUTHOR BIOGRAPHIES

EUNHYE SONG is Harold and Inge Marcus Early Career Assistant Professor in the Department of Industrial and Manufacturing Engineering at the Penn State University. Her research interests include simulation design of experiments, simulation uncertainty and risk quantification, optimization via simulation under input model risk and large-scale discrete optimization via simulation. Her email address is eus358@psu.edu.

YI DONG is a graduate student in the Department of Industrial and Manufacturing Engineering at the Penn State University. His research interests include statistical metamodeling and machine learning. His email address is enochdongyi@hotmail.com.