Supplementary Materials to “Pairwise Meta-Modeling of Multivariate Output
Computer Models Using Nonseparable Covariance Function”

Yongxiang Li, Qiang Zhou^{*}
Dept. of Systems Engineering and Engineering Management
City University of Hong Kong

^{*}Corresponding author email: q.zhou@cityu.edu.hk

Appendix A: Maximum Likelihood Estimation for the Pairwise Likelihood

Define the log likelihood for the pairwise likelihood $L_{i,j}$ as

$$l_{i,j} = \begin{cases} \ln L_{i,1}(\theta_i, \theta_1, \omega_{i,1}; \mathbf{y}_i, \mathbf{y}_1) & 1 = j < i \leq k \\ \ln L_{i,j}(\theta_i, \theta_j, \omega_{i,j} | \hat{\mathbf{v}}_{i,j}; \mathbf{y}_i, \mathbf{y}_j) & 2 \leq j < i \leq k \end{cases},$$

By equating the partial derivatives of the log likelihood with regard to σ_i, σ_j and $\boldsymbol{\beta}_i, \boldsymbol{\beta}_j$, to zero, their maximum likelihood estimates conditional on $\boldsymbol{\Phi}_i, \boldsymbol{\Phi}_j$ and $\omega_{i,j}$ are given by

$$\begin{bmatrix} \hat{\boldsymbol{\beta}}_i / \hat{\sigma}_i \\ \hat{\boldsymbol{\beta}}_j / \hat{\sigma}_j \end{bmatrix} = \left(\mathbf{E}^T (\mathbf{R}^{(i,j)})^{-1} \mathbf{E} \right)^{-1} \mathbf{E}^T (\mathbf{R}^{(i,j)})^{-1} \begin{bmatrix} \mathbf{y}_i / \hat{\sigma}_i \\ \mathbf{y}_j / \hat{\sigma}_j \end{bmatrix}, \quad (\text{A.1})$$

and

$$\begin{bmatrix} \hat{\sigma}_i \\ \hat{\sigma}_j \end{bmatrix} = \begin{bmatrix} n_i & 0 \\ 0 & n_j \end{bmatrix}^{-1} \boldsymbol{\Psi}^T (\mathbf{R}^{(i,j)})^{-1} \boldsymbol{\Psi} \begin{bmatrix} 1 / \hat{\sigma}_i \\ 1 / \hat{\sigma}_j \end{bmatrix}, \quad (\text{A.2})$$

where $\mathbf{E} = \mathbf{F}^{(i,j)} = \text{diag}(\mathbf{F}_i, \mathbf{F}_j)$, $\boldsymbol{\Psi} = \text{diag}(\mathbf{y}_i - \mathbf{F}_i \hat{\boldsymbol{\beta}}_i, \mathbf{y}_j - \mathbf{F}_j \hat{\boldsymbol{\beta}}_j)$ and

$$\mathbf{R}^{(i,j)} = \begin{bmatrix} \mathbf{K}_{i,i} & f_{i,j}(u_{i,j}) \mathbf{K}_{i,j} \\ f_{i,j}(u_{i,j}) \mathbf{K}_{j,i} & \mathbf{K}_{j,j} \end{bmatrix}.$$

By plugging (A.1) into (A.2), the equation for $\hat{\boldsymbol{\sigma}}$ can be rewritten as

$$\begin{bmatrix} \hat{\sigma}_i \\ \hat{\sigma}_j \end{bmatrix} = \begin{bmatrix} n_i & 0 \\ 0 & n_j \end{bmatrix}^{-1} \left(\mathbf{Y}^T (\mathbf{R}^{(i,j)})^{-1} \mathbf{Y} - \mathbf{Y}^T (\mathbf{R}^{(i,j)})^{-1} \mathbf{E} \left(\mathbf{E}^T (\mathbf{R}^{(i,j)})^{-1} \mathbf{E} \right)^{-1} \mathbf{E}^T (\mathbf{R}^{(i,j)})^{-1} \mathbf{Y} \right) \begin{bmatrix} 1/\hat{\sigma}_i \\ 1/\hat{\sigma}_j \end{bmatrix}, \quad (\text{A.3})$$

where $\mathbf{Y} = \text{diag}(\mathbf{y}_i, \mathbf{y}_j)$. The nonlinear equation (A.3) can be solved numerically by standard tools in R or Matlab. Then $\hat{\sigma}_i$ and $\hat{\sigma}_j$ are substituted back into (A.1) to obtain $\hat{\boldsymbol{\beta}}_i$ and $\hat{\boldsymbol{\beta}}_j$. One can get from (A.2) that, for the estimated $\hat{\sigma}_i$ and $\hat{\sigma}_j$, $(\mathbf{y}^{(i,j)} - \mathbf{F}^{(i,j)} \hat{\boldsymbol{\beta}}^{(i,j)})^T (\mathbf{C}^{(i,j)})^{-1} (\mathbf{y}^{(i,j)} - \mathbf{F}^{(i,j)} \hat{\boldsymbol{\beta}}^{(i,j)}) \equiv n_i + n_j$, which can be used to simplify the log-likelihood function. Therefore, $\boldsymbol{\Phi}_i$, $\boldsymbol{\Phi}_j$ and $\omega_{i,j}$ (given $\hat{v}_{i,j}$, if $1 < s$) can be finally estimated by maximizing

$$(\hat{\boldsymbol{\Phi}}_i, \hat{\boldsymbol{\Phi}}_j, \hat{\omega}_{i,j}) \propto \underset{(\boldsymbol{\Phi}_i, \boldsymbol{\Phi}_j, \omega_{i,j})}{\text{argmin}} \left\{ \ln |\mathbf{R}^{(i,j)}| + n_i \ln \hat{\sigma}_i^2 + n_j \ln \hat{\sigma}_j^2 \right\},$$

with simple constraints $\omega_{i,j} \in (0, \pi)$ and $\phi_{i,t}, \phi_{j,t} \geq 0$ in $\boldsymbol{\Phi}_i$ and $\boldsymbol{\Phi}_j$.

Appendix B: Discussion on Error Propagation

Proposition 1. *For the two pairwise methods introduced in Section 3.2.1, i.e., the FV method that produces estimate $\hat{\mathbf{T}}^{FV} = \{\hat{T}_{i,j}^{FV}\}$ and our proposed method that produces estimate $\hat{\mathbf{T}} = \{\hat{T}_{i,j}\}$, if $\hat{\mathbf{T}}^{FV}$ is positive definite, then $\hat{\mathbf{T}} = \hat{\mathbf{T}}^{FV}$ under some regularity conditions.*

The proof of Proposition 1 is provided in the Appendix C below. It is evident that the FV method does not accumulate error as all pairwise sub-models are built independently, without any necessary sequence. Hence, it is possible that by pulling all estimated elements together, $\hat{\mathbf{T}}^{FV}$ is not positive definite, leading to an invalid covariance function. Proposition 1 states that when $\hat{\mathbf{T}}^{FV}$ is indeed positive definite, the proposed method would produce the same parameter estimates. This implies that

the potential error propagation only exists in the cases when the FV method produces an invalid estimate. In fact, as can be seen from the proof, it is only when $\hat{\mathbf{T}}^{FV}$ is non-positive definite that some constraints imposed by the HD become active, forcing some $\hat{T}_{i,j}$'s to be different from their counterparts in $\hat{T}_{i,j}^{FV}$. The probability of $\hat{\mathbf{T}}^{FV}$ being positive-definite depends on the actual computer model response and the number of sample points used for model fitting. When the sample size is small, there may be a high probability that $\hat{\mathbf{T}}^{FV}$ is not positive-definite. It is also worthwhile to note that both methods produce the same estimates for the 1st column of $\mathbf{T}$ in all cases.

Proposition 2. *Assume a PDUDE $\mathbf{T}$ is parameterized to the space of Ω by the HD, i.e., $T_{i,j} = f_{i,j}(u_{i,j})$, then for any $T_{i,j}, 1 \leq j < i \leq k$ and any $\omega_{r,s}, 1 \leq s < r \leq k$*

$$\left| \frac{\partial T_{i,j}}{\partial \omega_{r,s}} \right| \leq \begin{cases} \sqrt{1-T_{i,j}^2} & s=1, r=i \text{ or } j \\ \sqrt{1-T_{i,j}^2} \prod_{t=1}^{s-1} \sin \omega_{r,t} & 1 < s \leq j \text{ \& } s < r, r=i \text{ or } j. \\ 0 & \text{else} \end{cases}$$

The proof is given in the Appendix D below. By the Taylor expansion, we get

$$\hat{T}_{i,j} - T_{i,j} = f_{i,j}(\hat{u}_{i,j}) - f_{i,j}(u_{i,j}) \approx \partial f_{i,j}(u_{i,j}) / \partial u_{i,j}^T (\hat{u}_{i,j} - u_{i,j}) = \partial T_{i,j} / \partial u_{i,j}^T (\hat{u}_{i,j} - u_{i,j}),$$

where $\partial T_{i,j} / \partial u_{i,j}^T$ is a row vector of partial derivatives and $u_{i,j}$ is a column vector

here. According to Proposition 2, if $|T_{i,j}|$ is close to 1, all its partial derivatives are

close to zero which limits the errors passing through parameters in the modeling

sequence. The MGP model is much more efficient than the univariate GP model when

the cross-correlations among different output levels are high, i.e., $|T_{i,j}|$ is close to 1.

Therefore, in applications where the MGP model is more likely to be used, we would

expect that the $|T_{i,j}|$'s are large, making the result from Proposition 2 applicable.

Appendix C: Proof of Proposition 1

As described in Section 3.1, $\{f_{i,j}(\cdot)\}: \mathbf{\Omega} \rightarrow \mathbf{T}$ is a one-to-one mapping. Thus, if the matrix $\hat{\mathbf{T}}^{FV}$ estimated from the FV method is positive definite, there exists a unique set of $\{\tilde{\omega}_{i,j}\}$, denoted by $\tilde{\mathbf{\Omega}}$, such that $\hat{T}_{i,j}^{FV} = f_{i,j}(\tilde{u}_{i,j})$. Clearly $\hat{T}_{i,1} = \hat{T}_{i,1}^{FV}$, $i = 2, \dots, k$, because $\hat{\omega}_{i,1} = \tilde{\omega}_{i,1}$ as no conditional estimation is involved when estimating the parameters in the first column. For any $j \in [2, k]$, assume all $\{\hat{u}_{i,j-1}\}$ have been obtained and $\hat{u}_{i,j-1} = \tilde{u}_{i,j-1}$, $j < i \leq k$, then clearly $\hat{v}_{i,j} = \tilde{v}_{i,j}$, thus it suffices to show that $\hat{u}_{i,j} = \tilde{u}_{i,j}$ for all i . Now we show that it is necessary that $\hat{\omega}_{i,j} = \tilde{\omega}_{i,j}$, otherwise $\hat{\omega}_{i,j}$ cannot be the maximum likelihood estimate, because the estimate is assumed unique. By fixing $v_{i,j}$, $f_{i,j}(v_{i,j}, \omega_{i,j})$ is a monotonically decreasing function with respect to $\omega_{i,j}$ in $(0, \pi)$. Therefore, $f_{i,j}(\tilde{v}_{i,j}, \pi) < \hat{T}_{i,j}^{FV} < f_{i,j}(\tilde{v}_{i,j}, 0)$ and $f_{i,j}(\hat{v}_{i,j}, \pi) < \hat{T}_{i,j} < f_{i,j}(\hat{v}_{i,j}, 0)$. Because $\hat{v}_{i,j} = \tilde{v}_{i,j}$, the search regions $(f_{i,j}(\tilde{v}_{i,j}, \pi), f_{i,j}(\tilde{v}_{i,j}, 0))$ and $(f_{i,j}(\hat{v}_{i,j}, \pi), f_{i,j}(\hat{v}_{i,j}, 0))$ coincide. Due to the monotonicity of the function $f_{i,j}(\cdot, \cdot)$ and the uniqueness assumption of the maximum likelihood estimator, it must be that $\hat{\omega}_{i,j} = \tilde{\omega}_{i,j}$. According to the sequential estimation procedure, it can then be seen that $\hat{u}_{i,j} = \tilde{u}_{i,j}$ for all i and j by applying mathematical induction. Therefore, $\hat{T}_{i,j} = f(\hat{u}_{i,j}) = f(\tilde{u}_{i,j}) = \hat{T}_{i,j}^{FV}$ for all i and j . This completes the proof.

Appendix D: Proof of Proposition 2

Before we give the proof for the Proposition 2, we first prove an inequality below.

Inequality 1: Let $\mathbf{u}$ and $\mathbf{v}$ be two orthogonal vectors in an n -dimensional vector space

$\mathbf{V}$, i.e., $\mathbf{u}^T \mathbf{v} = 0$ and denote $\mathbf{W}$ the subspace of $\mathbf{V}$ spanned by $\mathbf{u}$ and $\mathbf{v}$. Then for any vector $\mathbf{w}$ in $\mathbf{V}$, the following inequality holds

$$\frac{(\mathbf{w}^T \mathbf{u})^2}{\|\mathbf{u}\|^2} + \frac{(\mathbf{w}^T \mathbf{v})^2}{\|\mathbf{v}\|^2} \leq \|\mathbf{w}\|^2 \quad (\text{A.4})$$

where the equality holds iff $\mathbf{w} \in \mathbf{W}$.

Proof: The vector $\mathbf{w}$ in $\mathbf{V}$, can be expressed as $\mathbf{w} = \mathbf{w}_1 + \mathbf{w}_2$, where $\mathbf{w}_1 \in \mathbf{W}$, referred as the projection of $\mathbf{w}$ on $\mathbf{W}$ and $\mathbf{w}_2 \in \mathbf{W}^\perp$. It can be seen that

$$\mathbf{w}_1 = \frac{\mathbf{w}^T \mathbf{u}}{\|\mathbf{u}\|^2} \mathbf{u} + \frac{\mathbf{w}^T \mathbf{v}}{\|\mathbf{v}\|^2} \mathbf{v} \quad \text{and} \quad \|\mathbf{w}_1\| \leq \|\mathbf{w}\|. \quad \text{Thus}$$

$$\|\mathbf{w}_1\| = \sqrt{\left(\frac{\mathbf{w}^T \mathbf{u}}{\|\mathbf{u}\|}\right)^2 + \left(\frac{\mathbf{w}^T \mathbf{v}}{\|\mathbf{v}\|}\right)^2} \leq \|\mathbf{w}\|.$$

This completes the proof of inequality 1.

Denote $\mathbf{a}_i$ the i^{th} column vector of $\mathbf{A}^T$ and thus $\|\mathbf{a}_i\| = 1$. By definition,

$T_{i,j} = \mathbf{a}_i^T \mathbf{a}_j, j < i$, whose partial derivative with respect to $\omega_{r,s}$ is

$$\frac{\partial T_{i,j}}{\partial \omega_{r,s}} = \frac{\partial \mathbf{a}_i^T}{\partial \omega_{r,s}} \mathbf{a}_j + \mathbf{a}_i^T \frac{\partial \mathbf{a}_j}{\partial \omega_{r,s}}.$$

The partial derivatives $\partial \mathbf{a}_i^T / \partial \omega_{r,s}$ have following properties:

$$\frac{\partial \mathbf{a}_i^T}{\partial \omega_{r,s}} = \mathbf{0},$$

for $r \neq i$ and

$$\frac{\partial \mathbf{a}_i^T}{\partial \omega_{i,s}} = \begin{cases} \begin{bmatrix} -\sin \omega_{i,1} & \cos \omega_{i,1} \cos \omega_{i,2} & \cos \omega_{i,1} \cos \omega_{i,3} \sin \omega_{i,2} & \cdots & \cos \omega_{i,1} \prod_{t=2}^{i-1} \sin \omega_{i,t} \end{bmatrix} & s=1 \\ \sin \omega_{i,1} \begin{bmatrix} 0 & -\sin \omega_{i,2} & \cos \omega_{i,2} \cos \omega_{i,3} & \cdots & \cos \omega_{i,2} \prod_{t=3}^{i-1} \sin \omega_{i,t} \end{bmatrix} & s=2 \\ \sin \omega_{i,1} \sin \omega_{i,2} \begin{bmatrix} 0 & 0 & -\sin \omega_{i,3} & \cdots & \cos \omega_{i,3} \prod_{t=4}^{i-1} \sin \omega_{i,t} \end{bmatrix} & s=3 \\ \vdots & \vdots \\ \prod_{t=1}^{s-1} \sin \omega_{i,t} \begin{bmatrix} 0 & \cdots & 0 & -\sin \omega_{i,i-1} & \cos \omega_{i,i-1} \end{bmatrix} & s=i-1 \end{cases}.$$

It can be seen from these properties that

$$\left\| \frac{\partial \mathbf{a}_i^T}{\partial \omega_{r,s}} \right\| = \begin{cases} 1 & r = i, s = 1 \\ \prod_{t=1}^{s-1} \sin \omega_{i,t} & r = i, 1 < s < i \\ 0 & \text{else} \end{cases}$$

and that at least one of $\partial \mathbf{a}_i^T / \partial \omega_{r,s}$ and $\partial \mathbf{a}_j^T / \partial \omega_{r,s}$ is a zero vector for any $\omega_{r,s}$

since $i \neq j$. Note that $T_{i,i} = \mathbf{a}_i^T \mathbf{a}_i \equiv 1$ for $i = 1, 2, \dots, k$, thus

$$\frac{\partial T_{i,i}}{\partial \omega_{r,s}} = 2 \frac{\partial \mathbf{a}_i^T}{\partial \omega_{r,s}} \mathbf{a}_i \equiv 0.$$

This means that $\partial \mathbf{a}_i^T / \partial \omega_{r,s}$ and $\mathbf{a}_i$ are mutually orthogonal vectors. Since

$T_{i,j} = \mathbf{a}_i^T \mathbf{a}_j$ and $\|\mathbf{a}_i\| = \|\mathbf{a}_j\| = 1$, by taking $\mathbf{u} = \partial \mathbf{a}_i^T / \partial \omega_{r,s}$, $\mathbf{v} = \mathbf{a}_i$ and $\mathbf{w} = \mathbf{a}_j$, it can

be seen from (A.4) in inequality 1 that

$$\begin{cases} \left| \mathbf{a}_i^T \frac{\partial \mathbf{a}_j}{\partial \omega_{j,s}} \right| \leq \left\| \frac{\partial \mathbf{a}_j}{\partial \omega_{j,s}} \right\| \sqrt{1 - T_{i,j}^2} \\ \left| \mathbf{a}_j^T \frac{\partial \mathbf{a}_i}{\partial \omega_{i,s}} \right| \leq \left\| \frac{\partial \mathbf{a}_i}{\partial \omega_{i,s}} \right\| \sqrt{1 - T_{i,j}^2} \end{cases}.$$

As a result, for $1 \leq j < i \leq k$ and $1 \leq s < r \leq k$, one gets

$$\left| \frac{\partial T_{i,j}}{\partial \omega_{r,s}} \right| \leq \begin{cases} \sqrt{1 - T_{i,j}^2} & s = 1, r = i \text{ or } j \\ \sqrt{1 - T_{i,j}^2} \prod_{t=1}^{s-1} \sin \omega_{r,t} & 1 < s \leq j \text{ \& } s < r, r = i \text{ or } j. \\ 0 & \text{else} \end{cases}$$