Supplementary Materials for “A Clustered Gaussian Process Model for Computer Experiments”

S1. Proof of Proposition 1

For notational convention, denote \( \Sigma_j = \Phi_{\gamma_j}(X_{P_j \setminus \{i\}}, X_{P_j \setminus \{i\}}) \) and \( W_j = Y_{P_j \setminus \{i\}} - \mu_j(X_{P_j \setminus \{i\}}) \) for \( j = 1, \ldots, K \). Then, for any \( j \neq k \),

\[
f_j(Y_{P_j \setminus \{i\}}|X_{P_j \setminus \{i\}}; \theta_j) = \frac{1}{\sqrt{2\pi \det(\Sigma_j)}} \exp \left\{ -\frac{1}{2} W_j^T \Sigma_j^{-1} W_j \right\}, \tag{S1.1}
\]

by the fact that \( f_j \) is the probability density function of a multivariate normal distribution with parameters \( \theta_j = (\mu_j(\cdot), \sigma_j^2, \gamma_j) \). For \( j = k \), by partitioned matrix inverse and determinant formulas,

\[
f_k(Y_{P_k \cup \{i\}}|X_{P_k \cup \{i\}})
= \frac{1}{\sqrt{2\pi \det \left( \begin{bmatrix} 1 & r_{i,-i}^T \\ \Sigma_k & \sigma_k^2 \end{bmatrix} \right)}} \exp \left\{ -\frac{1}{2} \begin{bmatrix} W_k \\ y_i - \mu_k(x_i) \end{bmatrix}^T \begin{bmatrix} \Sigma_k & r_{i,-i}^T \\ r_{i,-i} & \sigma_k^2 \end{bmatrix}^{-1} \begin{bmatrix} W_k \\ y_i - \mu_k(x_i) \end{bmatrix} \right\}
= f_k(Y_{P_k \setminus \{i\}}|X_{P_k \setminus \{i\}}) \times \frac{1}{\sqrt{(\sigma_k^*)^2}} \exp \left\{ -\frac{1}{2} (y_i - \mu_k^*)^2 / (\sigma_k^*)^2 \right\}, \tag{S1.2}
\]

where \( r_{i,-i} = \Phi_{\gamma_k}(x_i, X_{P_k \setminus \{i\}}), \mu_k^* = \mu_k(x_i) + r_{i,-i} \Sigma_k^{-1} W_k \) and \( (\sigma_k^*)^2 = \sigma_k^2(1 - r_{i,-i} \Sigma_k^{-1} r_{i,-i}^T) \).

Therefore, combining (3.10), (S1.1) and (S1.2),

\[
f(z_i = k|X, Y, Z_{-i}) \propto f_k(Y_{P_k \cup \{i\}}|X_{P_k \cup \{i\}}; \theta_k) \prod_{j \neq k} f_j(Y_{P_j \setminus \{i\}}|X_{P_j \setminus \{i\}}; \theta_j) g_k(x_i; \varphi_k)
= \prod_{k=1}^K f_k(Y_{P_k \setminus \{i\}}|X_{P_k \setminus \{i\}}; \theta_k) \exp \left\{ -\frac{1}{2} (y_i - \mu_k^*)^2 / (\sigma_k^*)^2 \right\} g_k(x_i; \varphi_k)
\propto \phi((y_i - \mu_k^*) / (\sigma_k^*) g_k(x_i; \varphi_k).
S2. Efficient Update for the Stochastic E-step

In this section, partitioned matrix inverse formula is introduced to efficiently update the mean and variance of (3.12) when looping through observation \( i \) in the stochastic E-step. Suppose that the current assignment of observation \( i \) is \( z(x_i) = k \) but the new assignment of it is \( z(x_i) = s \) where \( s \neq k \), then the sets \( P_k \) and \( P_s \) will be updated, that is, \( P_k' \leftarrow P_k \setminus \{i\} \) and \( P_s' \leftarrow P_s \cup \{i\} \). The matrix inverses of \( \Phi_{\gamma_k}(X_{P_k'}, X_{P_k'}) \) and \( \Phi_{\gamma_s}(X_{P_s'}, X_{P_s'}) \) can be updated accordingly via partitioned matrix inverse formula as follows. Let \( U \in \mathbb{R}^{n_k \times 2} \), where \( n_k \) is the number of observations in the set \( P_k \), and \( U_{i,1} = 1 \) and \( U_{i,2} = \Phi_{\gamma_k}(X_{P_k}, x_i) \), otherwise \( U_{i,j} = 0 \). For notational simplicity, denote \( A = \Phi_{\gamma_k}(X_{P_k}, X_{P_k})^{-1} \). Then, by the Woodbury formula (Harville 1998), the matrix inverses of \( \Phi_{\gamma_k}(X_{P_k'}, X_{P_k'}) \) can be updated by

\[
\Phi_{\gamma_k}(X_{P_k'}, X_{P_k'})^{-1} = (A + AU(I_2 - U^T A)^{-1} U^T)_{-i,-i},
\]

where \( I_2 \) is a diagonal matrix.

Let \( V = \Phi_{\gamma_s}(X_{P_s}, x_i) \in \mathbb{R}^{n_s \times 1} \) and denote \( B = \Phi_{\gamma_s}(X_{P_s}, X_{P_s})^{-1} \). Then, by the partitioned matrix inverse formula (Harville 1998), the matrix inverses of \( \Phi_{\gamma_s}(X_{P_s'}, X_{P_s'}) \) can be updated by

\[
\begin{align*}
(\Phi_{\gamma_k}(X_{P_s'}, X_{P_s'})^{-1})_{i,i} &= 1/(1 - V^T B), \\
(\Phi_{\gamma_k}(X_{P_s'}, X_{P_s'})^{-1})_{i,-i} &= -V^T B/(1 - V^T B) = (\Phi_{\gamma_k}(X_{P_s'}, X_{P_s'})^{-1})^T_{-i,i}, \\
(\Phi_{\gamma_k}(X_{P_s'}, X_{P_s'})^{-1})_{-i,-i} &= B + B V V^T B/(1 - V^T B).
\end{align*}
\]

S3. Stochastic EM algorithm for clustered Gaussian process

Initialization:
Set \( K \) clusters with random memberships \( \{z(x_i)\}_{i=1}^n \)
Set \( P_k \leftarrow \{i : z(x_i) = k\} \) for each \( k \)
Set initial parameters \( \theta_k = \{\mu_k(\cdot), \sigma_k^2, \gamma_k\} \) and \( \varphi_k \) for \( k = 1, \ldots, K \)

Stochastic E-Step:
For \( i = 1 \) to \( i = n \),
For $k = 1$ to $K$ do parallel,
\[
\mu_k^* \leftarrow \mu_k(x_i) + \Phi_{\gamma_k}(x_i, X_{P_k \setminus \{i\}})\Phi_{\gamma_k}(X_{P_k \setminus \{i\}}, X_{P_k \setminus \{i\}})^{-1}(Y_{P_k \setminus \{i\}} - \mu_k(X_{P_k \setminus \{i\}}))
\]
\[
\sigma_k^2 \leftarrow \sigma_k^2(1 - \Phi_{\gamma_k}(x_i, X_{P_k \setminus \{i\}})\Phi_{\gamma_k}(X_{P_k \setminus \{i\}}, X_{P_k \setminus \{i\}})^{-1}\Phi_{\gamma_k}(X_{P_k \setminus \{i\}}, x_i))
\]
\[
p_{ik} \leftarrow \frac{\phi((y_i - \mu_k^*)/\sigma_k^*)}{\sum_{k=1}^K \phi((y_i - \mu_k^*)/\sigma_k^*)}\]

Draw $z$ from a random multinomial cluster assignment with probabilities $(p_{i1}, \ldots, p_{iK})$

Update $z(x_i) \leftarrow z$

Update $\mathcal{P}_k \leftarrow \{i : z(x_i) = k\}$ for each $k$

**M-Step:**

For $k = 1$ to $K$ do parallel,

Update $\theta_k \leftarrow \arg \max_{\theta_k} \log f_k(Y_{\mathcal{P}_k}, X_{\mathcal{P}_k}; \theta_k)\pi(\theta_k)$

Update $\{\varphi_k\}_{k=1}^K \leftarrow \arg \max_{\varphi} \sum_{k=1}^K \left(\sum_{i \in \mathcal{P}_k} \log g_k(x_i; \varphi_k) + \log \pi(\varphi_k)\right)$

**Iteration:** Iterate stochastic E-step and M-step until some stopping rule is met.

**Output** $\{z(x_i)\}_{i=1}^n, \{\theta_k, \varphi_k\}_{k=1}^K$

### S4. One-dimensional examples

Two more one-dimensional examples of Section 5.1 are presented here. Consider another example from [Xiong et al. (2007)](#), where the true function is
\[
f(x) = \sin(30(x - 0.9)^4) \cos(2(x - 0.9)) + (x - 0.9)/2
\]
and 17 unequally spaced points from $[0, 1]$ are chosen to evaluate. Similarly, the top panels of Figure [ST] show that the clustered GP (right) outperforms the stationary GP (left) in terms of prediction accuracy and uncertain quantitation. The two clusters are separated at location around $x = 0.40$. In particular, the predictor in the region $[0.42, 1.00]$ has better prediction accuracy with much smaller prediction uncertainty. The same argument applies to this example: the constant mean and variance assumptions are violated in this function so the stationary GP results in the erratic prediction in the region $[0.42, 1.00]$.

Lastly, consider the inhomogeneous smooth function in [Montagna and Tokdar (2016)](#),
\[
f(x) = \sin(x) + 2 \exp(-30x^2),
\]
and 15 unequally spaced points from $[-2, 2]$ are chosen to evaluate. The bottom panels of Figure S1 demonstrate a stationary GP (left), where the prediction mean curve has large oscillations with confidence intervals except the tall peak in the middle. This is due to the rippling effect of the discovery of a tall peak, and Montagna and Tokdar (2016) called the phenomenon a spline tension effect in the predictor form. The clustered GP (right) overcomes the issue by separating the input locations into three clusters and fits a stationary GP in each cluster. The result shows that the prediction mean curve quite matches the true curve with a narrower confidence band.

Figure S1: One-dimensional synthetic data from (top) Xiong et al. (2007) and (bottom) Montagna and Tokdar (2016). The left, middle and right panels illustrate the predictors by the stationary GP, the composite GP (Ba and Joseph, 2012), and the clustered GP, respectively. Black line is the true function, black circles are input locations, and blue dotted lines are the predictors, with the gray shaded region providing a pointwise 95% confidence band. Red, green, and blue dots in the right panels represent different clusters.
S5. Supporting Tables and Figures in Sections 5 and 6

The figures and tables that present the results in Sections 5 and 6 are provided in this section.

Table S1: Borehole function example with $n$ training samples $n_{\text{test}} = 10,000$ testing locations.

<table>
<thead>
<tr>
<th>Method</th>
<th>$n$</th>
<th>Fitting Time (sec.)</th>
<th>Prediction Time (sec.)</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>mlegp</td>
<td>1,000</td>
<td>5204</td>
<td>24</td>
<td>1.0902</td>
</tr>
<tr>
<td>laGP</td>
<td>1,000</td>
<td>-</td>
<td>153</td>
<td>1.1806</td>
</tr>
<tr>
<td></td>
<td>10,000</td>
<td>-</td>
<td>137</td>
<td>0.4149</td>
</tr>
<tr>
<td></td>
<td>100,000</td>
<td>-</td>
<td>144</td>
<td>0.1617</td>
</tr>
<tr>
<td>MRFA</td>
<td>1,000</td>
<td>116</td>
<td>17</td>
<td>0.4668</td>
</tr>
<tr>
<td></td>
<td>10,000</td>
<td>723</td>
<td>16</td>
<td>0.0844</td>
</tr>
<tr>
<td></td>
<td>100,000</td>
<td>6789</td>
<td>18</td>
<td>0.0827</td>
</tr>
<tr>
<td>clusterGP</td>
<td>1,000</td>
<td>255</td>
<td>9</td>
<td>0.1124</td>
</tr>
<tr>
<td></td>
<td>10,000</td>
<td>2950</td>
<td>55</td>
<td>0.0689</td>
</tr>
<tr>
<td></td>
<td>100,000</td>
<td>28434</td>
<td>535</td>
<td>0.0523</td>
</tr>
</tbody>
</table>

References


Figure S2: Solar irradiance simulation from the North American Mesoscale Forecast System (NAM). The black dots are the Remote Automatic Weather Station (RAWS) measurement sites in the contiguous United States from which the NAM simulations are extracted. The regional colors represent the solar irradiance in the subfield of a particular measurement site.

Figure S3: The LOOCV RMSEs with $K = 15, 25, 35$ and 45 during the 20 iteration of the stochastic EM algorithm (left), and the minimum LOOCV RMSEs of $K = 15, 20, 25, 30, 35, 40, 45, 50$ (right).
Figure S4: Visualization of the cluster assignments with $K = 35$.

Figure S5: Comparison of solar irradiance predictions. The true solar irradiance (top left), and the LOOCV predictions of a stationary GP (top right), a multi-resolution global/local GP (bottom left), and a clustered GP with $K = 35$ (bottom right) are presented, along with their corresponding LOOCV RMSEs in the figure titles.