
A Clustered Gaussian Process Model for Computer Experiments

Chih-Li Sung[1], Benjamin Haaland[2], Youngdeok Hwang[3], Siyuan Lu[4]

[1]*Michigan State University,* [2]*University of Utah*

[3]*City University of New York,* [4]*IBM Thomas J. Watson Research Center*

Supplementary Material

S1 Proof of Proposition 1

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

$$f_j(Y_{\mathcal{P}_j\setminus\{i\}}|X_{\mathcal{P}_j\setminus\{i\}}; \theta_j) = \frac{1}{\sqrt{2\pi \det(\Sigma_j)}} \exp\left\{-\frac{1}{2} W_j^T \Sigma_j^{-1} W_j\right\}, \qquad \text{(S1.1)}$$

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 formulae,

$$
\begin{aligned}
& f_k(Y_{\mathcal{P}_k\cup\{i\}}|X_{\mathcal{P}_k\cup\{i\}}) \\
=&\frac{1}{\sqrt{2\pi\det\left(\begin{bmatrix}\Sigma_k & r_{i,-i}^T\\ r_{i,-i} & \sigma_k^2\end{bmatrix}\right)}}\exp\left\{-\frac{1}{2}\begin{bmatrix}W_k\\ y_i-\mu_k(x_i)\end{bmatrix}^T\begin{bmatrix}\Sigma_k & r_{i,-i}^T\\ r_{i,-i} & \sigma_k^2\end{bmatrix}^{-1}\begin{bmatrix}W_k\\ y_i-\mu_k(x_i)\end{bmatrix}\right\} \\
=&f_k(Y_{\mathcal{P}_k\setminus\{i\}}|X_{\mathcal{P}_k\setminus\{i\}})\times\frac{1}{\sqrt{(\sigma_k^*)^2}}\exp\left\{-\frac{1}{2}(y_i-\mu_k^*)^2/(\sigma_k^*)^2\right\}, \qquad \text{(S1.2)}
\end{aligned}
$$

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 (3.3), (S1.1) and (S1.2),

$$
\begin{aligned}
f(z_i=k|X,Y,Z_{-i}) \propto& f_k(Y_{\mathcal{P}_k\cup\{i\}}|X_{\mathcal{P}_k\cup\{i\}};\theta_k)\prod_{j\neq k}f_j(Y_{\mathcal{P}_j\setminus\{i\}}|X_{\mathcal{P}_j\setminus\{i\}};\theta_j)g_k(x_i;\varphi_k) \\
=&\prod_{k=1}^K f_k(Y_{\mathcal{P}_k\setminus\{i\}}|X_{\mathcal{P}_k\setminus\{i\}};\theta_k)\exp\left\{-\frac{1}{2}(y_i-\mu_k^*)^2/(\sigma_k^*)^2\right\}g_k(x_i;\varphi_k) \\
\propto&\phi((y_i-\mu_k^*)/\sigma_k^*)g_k(x_i;\varphi_k).
\end{aligned}
$$

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 (3.5) 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 $\mathcal{P}_k$ and $\mathcal{P}_s$ will be updated, that is, $\mathcal{P}'_k \leftarrow \mathcal{P}_k \setminus \{i\}$ and $\mathcal{P}'_s \leftarrow \mathcal{P}_s \cup \{i\}$. The matrix inverses of $\Phi_{\gamma_k}(X_{\mathcal{P}'_k}, X_{\mathcal{P}'_k})$ and $\Phi_{\gamma_s}(X_{\mathcal{P}'_s}, X_{\mathcal{P}'_s})$ 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 $\mathcal{P}_k$, 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 $A = \Phi_{\gamma_k}(X_{\mathcal{P}_k}, X_{\mathcal{P}_k})^{-1}$. Then, by the Woodbury formula (Harville, 1998), the matrix inverses of $\Phi_{\gamma_k}(X_{\mathcal{P}'_k}, X_{\mathcal{P}'_k})$ can be updated by

$$\Phi_{\gamma_k}(X_{\mathcal{P}'_k}, X_{\mathcal{P}'_k})^{-1} = \left(A + AU(I_2 - U^T AU)^{-1} U^T A\right)_{-i,-i},$$

where $I_2 \in \mathbb{R}^2$ is a diagonal matrix.

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

$$\begin{aligned}
\left(\Phi_{\gamma_k}(X_{\mathcal{P}'_s}, X_{\mathcal{P}'_s})^{-1}\right)_{i,i} &= 1/(1 - V^T BV), \\
\left(\Phi_{\gamma_k}(X_{\mathcal{P}'_s}, X_{\mathcal{P}'_s})^{-1}\right)_{i,-i} &= -V^T B/(1 - V^T BV) = \left(\Phi_{\gamma_k}(X_{\mathcal{P}'_s}, X_{\mathcal{P}'_s})^{-1}\right)^T_{-i,i}, \\
\left(\Phi_{\gamma_k}(X_{\mathcal{P}'_s}, X_{\mathcal{P}'_s})^{-1}\right)_{-i,-i} &= B + BVV^T B/(1 - V^T BV).
\end{aligned}$$

S3 Stochastic EM algorithm for clustered Gaussian process

Initialization:

Set K clusters with random memberships $\{z(x_i)\}_{i=1}^n$

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

Set initial parameters $\theta_k = \{\mu_k(\cdot), \sigma_k^2, \gamma_k\}$ and φ_k for $k = 1, \ldots, K$

Stochastic E-Step:

For $i = 1$ to $i = n$,

For $k = 1$ to K do parallel,

$$\mu_k^* \leftarrow \mu_k(x_i) + \Phi_{\gamma_k}(x_i, X_{\mathcal{P}_k \setminus \{i\}}) \Phi_{\gamma_k}(X_{\mathcal{P}_k \setminus \{i\}}, X_{\mathcal{P}_k \setminus \{i\}})^{-1} \left(Y_{\mathcal{P}_k \setminus \{i\}} - \mu_k(X_{\mathcal{P}_k \setminus \{i\}})\right)$$

$$(\sigma_k^*)^2 \leftarrow \sigma_k^2 \left(1 - \Phi_{\gamma_k}(x_i, X_{\mathcal{P}_k \setminus \{i\}}) \Phi_{\gamma_k}(X_{\mathcal{P}_k \setminus \{i\}}, X_{\mathcal{P}_k \setminus \{i\}})^{-1} \Phi_{\gamma_k}(X_{\mathcal{P}_k \setminus \{i\}}, x_i)\right)$$

$$p_{ik} \leftarrow \frac{\phi((y_i - \mu_k^*)/\sigma_k^*) g_k(x_i; \varphi_k)}{\sum_{k=1}^K \phi((y_i - \mu_k^*)/\sigma_k^*) g_k(x_i; \varphi_k)}$$

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

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 Additional 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 S1 show that the clustered GP (right) outperforms the stationary GP (left) in terms of prediction accuracy and uncertain quantification. 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 demonstrates a stationary GP (left), where the prediction mean curve has large oscillations with confidence intervals

except the tall peak in the middle. The 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.

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.

Bibliography

Ba, S. and Joseph, V. R. (2012). Composite Gaussian process models for emulating expensive functions. *The Annals of Applied Statistics*, 6(4):1838–1860.

Harville, D. A. (1998). *Matrix Algebra from a Statistician's Perspective*. Springer, New York.

Montagna, S. and Tokdar, S. T. (2016). Computer emulation with nonsta-

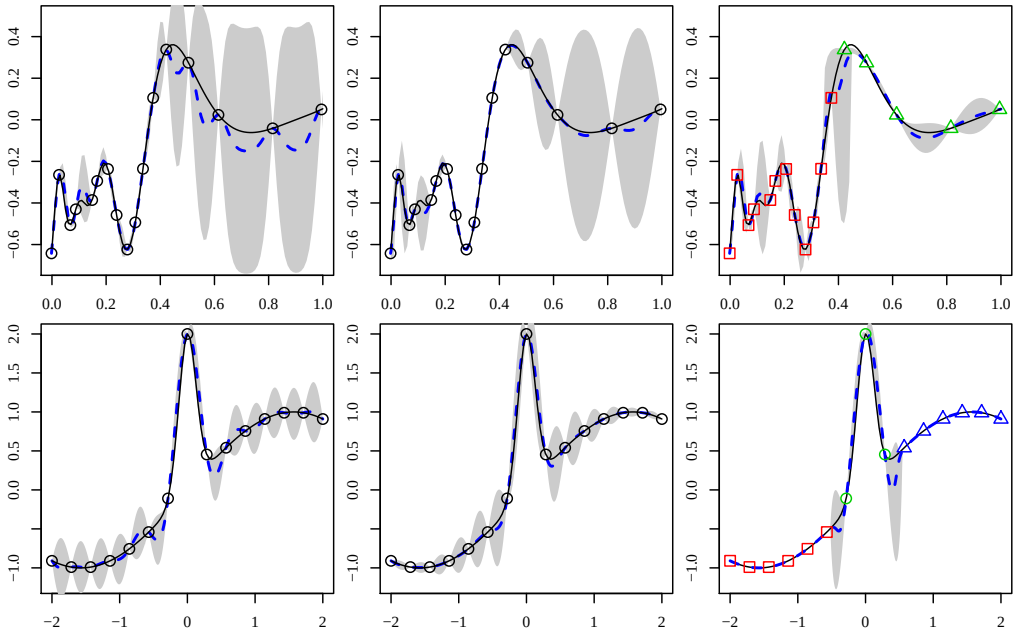

Figure S1: One-dimensional synthetic data from (top) Xiong et al. (2007) and (bottom) Montagna and Tokdar (2016). The left, middle and right panels illustrate the predictors by the stationary GP, composite GP (Ba and Joseph, 2012), and clustered GP, respectively. The solid line is the true function, circles are input locations, and dashed lines are the predictors, with the shaded region providing a pointwise 95% confidence band. The squares, dots, and triangles in the right panels represent different clusters.

tionary Gaussian processes. *SIAM/ASA Journal on Uncertainty Quantification*, 4(1):26–47.

Xiong, Y., Chen, W., Apley, D., and Ding, X. (2007). A non-stationarycovariance-based kriging method for metamodelling in engineering design. *International Journal for Numerical Methods in Engi-*

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

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

neering, 71(6):733–756.

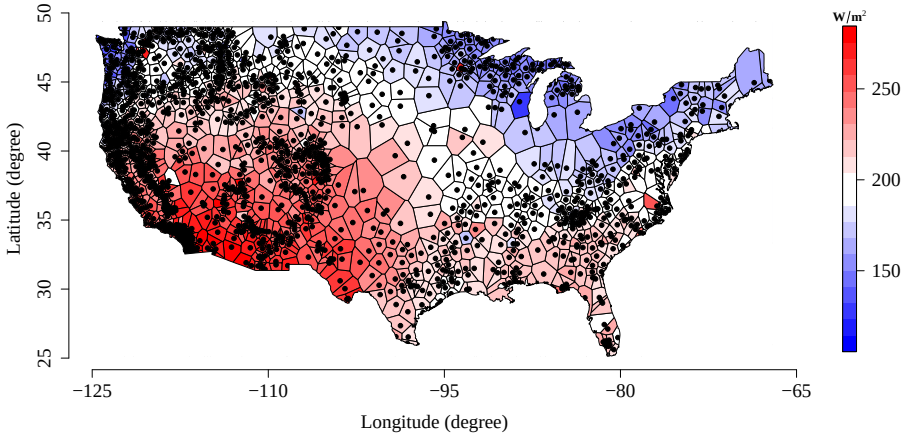

Figure S2: Solar irradiance simulation from the North American Mesoscale Forecast System (NAM). The black dots are the Remote Automatic Weather Station (RAWS) measurement sites in the contiguous United States from which the NAM simulations are extracted. The regional colors represent the solar irradiance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

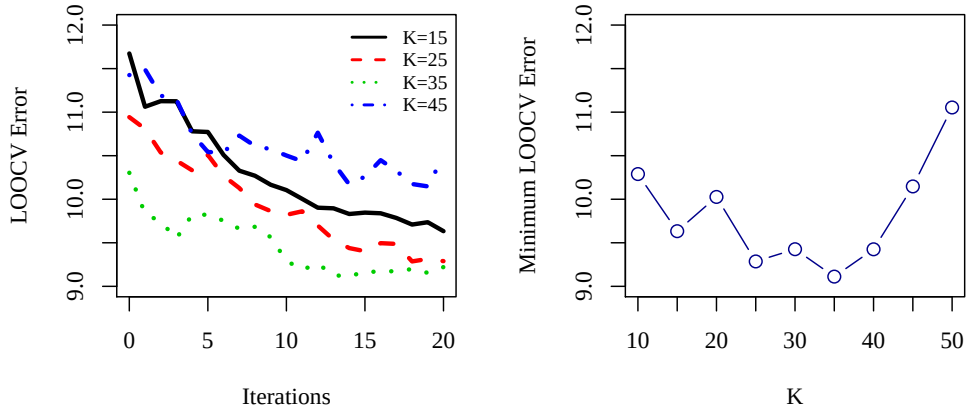

Figure S3: The LOOCV RMSEs with $K = 15, 25, 35$, and 45 during the 20 iteration of the stochastic EM algorithm (left), and the minimum LOOCV RMSEs of $K = 15, 20, 25, 30, 35, 40, 45, 50$ (right).

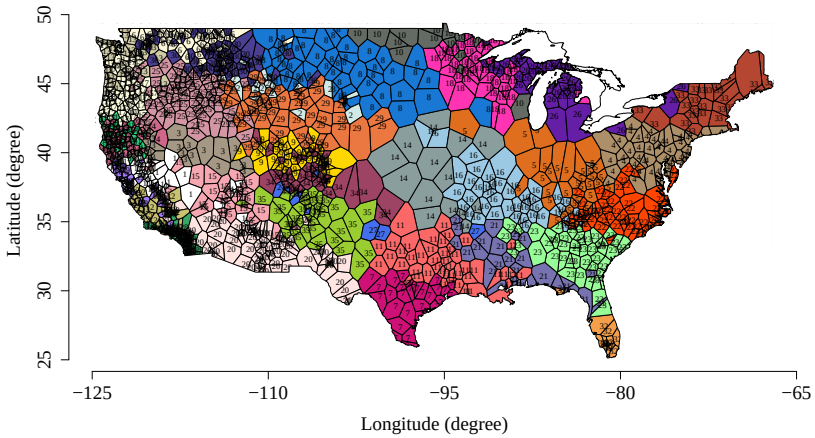

Figure S4: Visualization of the cluster assignments with $K = 35$.

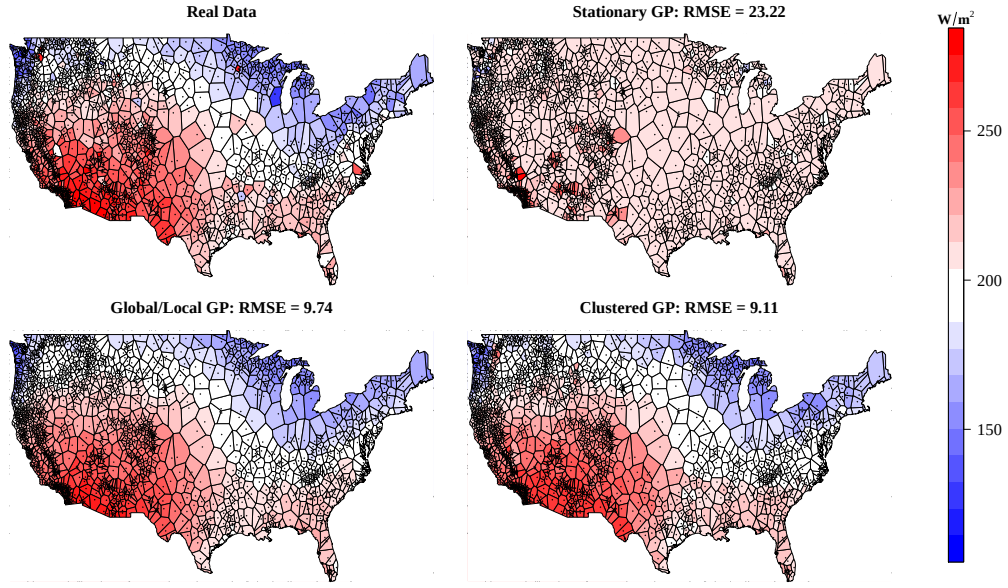

Figure S5: Comparison of solar irradiance predictions. The true solar irradiance (top left), and the LOOCV predictions of a stationary GP (top right), multi-resolution global/local GP (bottom left), and clustered GP with $K = 35$ (bottom right) are presented, along with their corresponding LOOCV RMSEs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