Gaussian Process Prediction using Design-Based Subsampling

Rutgers University

Supplementary Material

Appendix A: Assumptions

Let $\boldsymbol{y_i} = (y_s(\boldsymbol{x}_s), \boldsymbol{x}_s \in \mathcal{B}_n(\boldsymbol{i}))$ and $\boldsymbol{X_i} = (\boldsymbol{x}_s, \boldsymbol{x}_s \in \mathcal{B}_n(\boldsymbol{i}))^T$ denote the data in the ith block. Define $R_{\boldsymbol{i},\boldsymbol{j}}(\boldsymbol{\theta}) = \big[\psi(y(\boldsymbol{x}_s), y(\boldsymbol{x}_t); \boldsymbol{\theta}),\ \boldsymbol{x}_s \in \mathcal{B}_n(\boldsymbol{i}), \boldsymbol{x}_t \in \mathcal{B}_n(\boldsymbol{j})\big]$, $D_n(\boldsymbol{\theta}) = \mathrm{diag}(R_{\boldsymbol{i},\boldsymbol{i}}(\boldsymbol{\theta}))$ with $\boldsymbol{i} = (i_1, \ldots, i_d)$ in lexicographical order, and $E_n(\boldsymbol{\theta}) = R_n(\boldsymbol{\theta}) - D_n(\boldsymbol{\theta})$. We need the following assumptions for the proposed prediction procedures.

1. $m = o(n^{1/d})$ and $m \to \infty$.

2. $\lim_{n\to\infty} \sup_{\boldsymbol{\theta}} \lambda_{\max}(E_n(\boldsymbol{\theta})) = 0$, when the block space $b = l/m \to \infty$.

These two assumptions are also the necessary conditions for the consistency of the bootstrap estimators (Zhao et al. 2018). The second assumption aims to control the correlation between bootstrap blocks.

Appendix B: Proof of Theorem 1

By definition, the bootstrap predictive function is

$$\begin{aligned} h^*(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n) = &\ E^*\{h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}^*_N)\} \\ = &\ \int h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}^*_N) dP^*(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N \mid \boldsymbol{X}_n, \boldsymbol{y}_n). \end{aligned}$$

Take Taylor expansion of $h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}^*_N)$ at $\hat{\boldsymbol{\phi}}_n$ for each bootstrap, we have

$$\begin{aligned} h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}^*_N) \ = \ & h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n) \\ & +\nabla^T_{\boldsymbol{\phi}} h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)(\hat{\boldsymbol{\phi}}^*_N - \hat{\boldsymbol{\phi}}_n) \\ & +\frac{1}{2}(\hat{\boldsymbol{\phi}}^*_N - \hat{\boldsymbol{\phi}}_n)^T \nabla^2_{\boldsymbol{\phi}} h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)(\hat{\boldsymbol{\phi}}^*_N - \hat{\boldsymbol{\phi}}_n) \\ & +\frac{1}{6}\nabla\{(\hat{\boldsymbol{\phi}}^*_N - \hat{\boldsymbol{\phi}}_n)^T \nabla^2_{\boldsymbol{\phi}} h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)(\hat{\boldsymbol{\phi}}^*_N - \hat{\boldsymbol{\phi}}_n)\}(\hat{\boldsymbol{\phi}}^*_N - \hat{\boldsymbol{\phi}}_n) \\ & +O_{p^*}(\|\hat{\boldsymbol{\phi}}^*_N - \hat{\boldsymbol{\phi}}_n\|^4_2). \end{aligned}$$

So we only need to calculate the expectation of each term on the right-hand side of the equation above.

Again, we treat $\nabla \ell(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}^*_N)$ as a function of $\boldsymbol{\phi}$ and take the second order Taylor expansion at $\hat{\boldsymbol{\phi}}_n$. Recall that $\nabla \ell(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}^*_N) = 0$ and

$$\begin{aligned} 0 = \ & \nabla \ell(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n) + \nabla^2 \ell(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)\boldsymbol{\omega} \\ & +\tfrac{1}{2}\nabla\{\nabla^2 \ell(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)\boldsymbol{\omega}\}\boldsymbol{\omega} + O_{P^*}(\|\boldsymbol{\omega}\|^3_2) \end{aligned} \tag{0.01}$$

where $\boldsymbol{\omega} = \hat{\boldsymbol{\phi}}^*_N - \hat{\boldsymbol{\phi}}_n$. Multiplying (0.01) by $\nabla^T h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)$, we have

$$
\begin{aligned}
0 = \quad & \nabla \ell(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) \nabla^T h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) \\
& + \nabla^2 \ell(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) \boldsymbol{\omega} \nabla^T h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) \\
& + \tfrac{1}{2} \nabla \{ \nabla^2 \ell(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) \boldsymbol{\omega} \} \boldsymbol{\omega} \nabla^T h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) + O_{P^*}(\|\boldsymbol{\omega}\|_2^3).
\end{aligned} \tag{0.02}
$$

Using the fact that $\hat{\boldsymbol{\phi}}_N^* - \hat{\boldsymbol{\phi}}_n = I^{-1} \nabla \ell(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) + O_{P^*}(N^{-1/2})$, take expectations of each term in the equation above. For presentation simplicity, $\boldsymbol{X}_N^*, \boldsymbol{y}_N^*$ are omitted in the calculation below.

$$
\begin{aligned}
& E^* \{ \nabla^2 \ell(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) \boldsymbol{\omega} \nabla^T h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) \} \\
= \quad & E^* \{ \nabla^2 \ell(\hat{\boldsymbol{\phi}}_n) \} E^* \{ \boldsymbol{\omega} \nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n) \} + Cov^* \{ \nabla^2 \ell(\hat{\boldsymbol{\phi}}_n), \boldsymbol{\omega} \nabla^T \nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n) \} \\
= \quad & -I E^* \{ \boldsymbol{\omega} \nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n) \} + I^{-1} E^* \{ \nabla^2 \ell(\hat{\boldsymbol{\phi}}_n) \nabla \ell(\hat{\boldsymbol{\phi}}_n) \nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n) \} \\
& - I^{-1} E^* \{ \nabla^2 \ell(\hat{\boldsymbol{\phi}}_n) \} E^* \{ \nabla \ell(\hat{\boldsymbol{\phi}}_n) \nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n) \}
\end{aligned}
$$

Using the same technique, we have

$$
\begin{aligned}
& E^* \{ \nabla \{ \nabla^2 \ell(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) \boldsymbol{\omega} \} \boldsymbol{\omega} \nabla^T h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\boldsymbol{\phi}}_n) \} \\
= \quad & (K_{irs} L_{r,s}^j)_{i,j=1,\ldots,N} + O_{P^*}(N^{-2}).
\end{aligned}
$$

Plugging the equations back into (0.02), we have

$$\begin{aligned}0 = \quad & E^*\{\nabla\ell(\hat{\boldsymbol{\phi}}_n)\nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)\} - IE^*\{\boldsymbol{\omega}\nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)\} \\ & +I^{-1}E^*\{\nabla^2\ell(\hat{\boldsymbol{\phi}}_n)\nabla\ell(\hat{\boldsymbol{\phi}}_n)\nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)\} \\ & -I^{-1}E^*\{\nabla^2\ell(\hat{\boldsymbol{\phi}}_n)\}E^*\{\nabla\ell(\hat{\boldsymbol{\phi}}_n)\nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)\} \\ & +\tfrac{1}{2}(K_{irs}L^j_{r,s})_{i,j=1,\ldots,N} + O_{P^*}(N^{-2}).\end{aligned}$$

Thus,

$$\begin{aligned}E^*\{\boldsymbol{\omega}\nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)\} \quad = \quad & I^{-2}E^*\{\nabla^2\ell(\hat{\boldsymbol{\phi}}_n)\nabla\ell(\hat{\boldsymbol{\phi}}_n)\nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)\} \\ & +\frac{1}{2}I^{-1}(K_{irs}L^j_{r,s})_{i,j=1,\ldots,N} + O_{P^*}(N^{-2}).\end{aligned}$$

Taking trace on both side of the equation, we have

$$E^*\{\boldsymbol{\omega}\nabla^T h(\boldsymbol{x}_{n+1} \mid \hat{\boldsymbol{\phi}}_n)\} = I^{si}I^{jk}M_{s,j,ik} + \frac{1}{2}I^{ij}I^{jk}K_{irs}L^j_{r,s}(h) + O_{P^*}(N^{-2}).$$

Similarly,

$$\begin{aligned}& E^*(\hat{\boldsymbol{\phi}}^*_N - \hat{\boldsymbol{\phi}}_n)^T\nabla^2_\phi h(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\boldsymbol{\phi}}_n)(\hat{\boldsymbol{\phi}}^*_N - \hat{\boldsymbol{\phi}}_n) \\ = \quad & \{I^{rj}I^{si}J_{rs,ij}(h)\} + O_{P^*}(N^{-2}).\end{aligned}$$

The result follows by plugging the two equations into the Taylor expansion of $h^*(\cdot)$. □

Appendix C: Proof of Theorem 2

To investigate the asymptotic properties of the predictions from LHD-based block bootstrap, we decompose the likelihood function into blocks. For each block, denote $\boldsymbol{y_i} = (y_s(\boldsymbol{x}_s), \boldsymbol{x}_s \in \mathcal{B}_n(\boldsymbol{i}))$, $\boldsymbol{X_i} = (\boldsymbol{x}_s, \boldsymbol{x}_s \in \mathcal{B}_n(\boldsymbol{i}))^T$, $R_{\boldsymbol{i},\boldsymbol{j}}(\boldsymbol{\theta})=\left[\psi(y(\boldsymbol{x}_s), y(\boldsymbol{x}_t); \boldsymbol{\theta}), \boldsymbol{x}_s \in \mathcal{B}_n(\boldsymbol{i}), \boldsymbol{x}_t \in \mathcal{B}_n(\boldsymbol{j})\right]$ and $\boldsymbol{z_i} = R_{\boldsymbol{i},\boldsymbol{i}}^{-1/2}(\boldsymbol{\theta})(\boldsymbol{y_i} - \boldsymbol{X_i}\boldsymbol{\beta})$. Then, we can rewrite the normalised log-likelihood function $n^{-1}\ell(\boldsymbol{X}_n, \boldsymbol{y}_n, \boldsymbol{\phi})$ as

$$\begin{aligned} Q_n(\boldsymbol{X}_n, \boldsymbol{y}_n, \boldsymbol{\phi}) = & -(2n\sigma^2)^{-1}\textstyle\sum_{s=1}^n z_s^2 - (2n)^{-1}\sum_{s=1}^n \log(\lambda_s) \\ & -(2n)^{-1}\textstyle\sum_{s=1}^n \log(\sigma^2) + n^{-1}r_n(\boldsymbol{X}_n, \boldsymbol{y}_n, \boldsymbol{\phi}) \\ = & \; n^{-1}\textstyle\sum_{s=1}^n q_s(\omega, \boldsymbol{\phi}) + n^{-1}r_n(\omega, \boldsymbol{\phi}), \end{aligned}$$

where $\{\lambda_s, s = 1, \dots, n\} = \{\text{eigenvalues of } |R_{\boldsymbol{i},\boldsymbol{i}}(\boldsymbol{\theta})|, \boldsymbol{i} = (i_1, \dots, i_d)\}$ with $(i_1, \dots, i_d)$ in lexicographical order and eigenvalues from the largest to the smallest. Note that $r_n(\omega, \boldsymbol{\phi}) = \ell(\boldsymbol{X}_n, \boldsymbol{y}_n, \boldsymbol{\phi}) - \sum_{s=1}^n q_s(z_s, \boldsymbol{\phi})$ contains all terms involving the off block-diagonal terms. Define $D_n(\boldsymbol{\theta}) = \text{diag}(R_{\boldsymbol{i},\boldsymbol{i}}(\boldsymbol{\theta}))$ and $E_n(\boldsymbol{\theta}) = R_n(\boldsymbol{\theta}) - D_n(\boldsymbol{\theta})$. Assuming that $E_n(\boldsymbol{\theta}) = U_n(\boldsymbol{\theta})U_n^T(\boldsymbol{\theta})$, we have

$$\begin{aligned} r_n(\omega, \boldsymbol{\phi}) = & \frac{1}{2\sigma^2(1+g)}(\boldsymbol{y}_n - \boldsymbol{X}_n\boldsymbol{\beta})^T D_n^{-1}(\boldsymbol{\theta})E_n(\boldsymbol{\theta})D_n^{-1}(\boldsymbol{\theta})(\boldsymbol{y}_n - \boldsymbol{X}_n\boldsymbol{\beta}) \\ & + \frac{1}{2}\log|I_n + U_n^T(\boldsymbol{\theta})D_n^{-1}(\boldsymbol{\theta})U_n(\boldsymbol{\theta})|, \end{aligned}$$

where $g = \text{trace}(E_n(\boldsymbol{\theta})D_n^{-1}(\boldsymbol{\theta}))$.

The maximum likelihood estimator is obtained by $\hat{\boldsymbol{\phi}}_n = \arg\max_{\boldsymbol{\phi}} Q_n(\boldsymbol{X}_n, \boldsymbol{y}_n, \boldsymbol{\phi})$. Analogue to the decomposition for $Q_n(\boldsymbol{X}_n, \boldsymbol{y}_n, \boldsymbol{\phi})$, the log-likelihood function for LHD-based block bootstrap samples can be written as

$$Q_N^*(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \boldsymbol{\phi}) = N^{-1}\sum_{s=1}^N q_s^*(\cdot, \omega, \boldsymbol{\phi}) + N^{-1}r_N^*(\cdot, \omega, \boldsymbol{\phi}), \quad (0.03)$$

where $r_N^*(\cdot,\omega,\boldsymbol{\phi})$ contains all terms involving the off block-diagonal terms with bootstrapped samples. Specifically,

$$
\begin{aligned}
r_N^*(\cdot,\omega,\boldsymbol{\phi}) &= \frac{1}{2\sigma^2(1+g^*)}(\boldsymbol{y}_N^* - \boldsymbol{X}_N^*\boldsymbol{\beta})^T {D_N^*}^{-1}(\boldsymbol{\theta}) E_N^*(\boldsymbol{\theta}) {D_N^*}^{-1}(\boldsymbol{\theta})(\boldsymbol{y}_N^* - \boldsymbol{X}_N^*\boldsymbol{\beta}) \\
&\quad + \frac{1}{2}\log|I_N + {U_N^*}^T(\boldsymbol{\theta}){D_N^*}^{-1}(\boldsymbol{\theta})U_N^*(\boldsymbol{\theta})|,
\end{aligned}
$$

where $D_N^*(\boldsymbol{\theta}) = \mathrm{diag}(R_{\boldsymbol{i}_j^*,\boldsymbol{i}_j^*}(\boldsymbol{\theta}), j=1,\ldots,m)$ and $E_N^*(\boldsymbol{\theta}) = R_N^*(\boldsymbol{\theta}) - D_N^*(\boldsymbol{\theta})$ with $E_N^*(\boldsymbol{\theta}) = U_N^*(\boldsymbol{\theta}){U_N^*}^T(\boldsymbol{\theta})$; $g^* = \mathrm{trace}(E_N^*(\boldsymbol{\theta}){D_N^*}^{-1}(\boldsymbol{\theta}))$. The bootstrapped version of $\hat{\boldsymbol{\phi}}_n$ is $\hat{\boldsymbol{\phi}}_N^* = \arg\max_{\boldsymbol{\phi}} Q_N^*(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \boldsymbol{\phi})$, which is a consistent estimate of $\hat{\boldsymbol{\phi}}_n$ according to Zhao et al. (2018).

Similar to the decomposition of the bootstrapped likelihood (0.03), we rewrite the weighted average of the bootstrapped data. Recall $D_N^*(\hat{\boldsymbol{\theta}}_n) = \mathrm{diag}(R_{\boldsymbol{i}_j^*,\boldsymbol{i}_j^*}(\hat{\boldsymbol{\theta}}_n), j=1,\ldots,m)$ and $E_N^*(\hat{\boldsymbol{\theta}}_n) = R_N^*(\hat{\boldsymbol{\theta}}_n) - D_N^*(\hat{\boldsymbol{\theta}}_n)$ with $E_N^*(\hat{\boldsymbol{\theta}}_n) = U_N^*(\hat{\boldsymbol{\theta}}_n){U_N^*}^T(\hat{\boldsymbol{\theta}}_n)$; $\hat{g}^* = \mathrm{trace}(E_N^*(\hat{\boldsymbol{\theta}}_n){D_N^*}^{-1}(\hat{\boldsymbol{\theta}}_n))$. Then $\gamma_N^*(\hat{\boldsymbol{\theta}}_n)^T R^{*-1}(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{y}_N^* - \boldsymbol{X}_N^*\hat{\boldsymbol{\beta}}_n)$ can be written as

$$
\begin{aligned}
&\gamma_N^*(\hat{\boldsymbol{\theta}}_n)^T R^{*-1}(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{y}_N^* - \boldsymbol{X}_N^*\hat{\boldsymbol{\beta}}_n) \\
=& \sum_{j=1}^{m} \gamma_{\boldsymbol{i}_j^*}(\hat{\boldsymbol{\theta}}_n)^T R^{-1}_{\boldsymbol{i}_j^*,\boldsymbol{i}_j^*}(\hat{\boldsymbol{\theta}}_n)(y_{\boldsymbol{i}_j^*} - \boldsymbol{X}_{\boldsymbol{i}_j^*}\hat{\boldsymbol{\beta}}_n) + s_N^*(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\beta}}_n),
\end{aligned}
$$

where $\gamma_N^*(\hat{\boldsymbol{\theta}}_n)$ is the correlation between $\boldsymbol{x}_{n+1}$ and the bootstrapped data $\boldsymbol{X}_N^*$ calculated at $\hat{\boldsymbol{\theta}}_n$ and R^* is the correlation matrix of the bootstrapped data $\boldsymbol{X}_N^*$ calculated at $\hat{\boldsymbol{\theta}}_n$ as well; and $s_N^*(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\beta}}_n)$ contains all terms involving the off block-diagonal terms with bootstrapped samples. Specifically,

$$s_N^*(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\beta}}_n) = \frac{1}{(1+\hat{g}^*)}\gamma_N^*(\hat{\boldsymbol{\theta}}_n)^T {D_N^*}^{-1}(\hat{\boldsymbol{\theta}}_n) E_N^*(\hat{\boldsymbol{\theta}}_n) {D_N^*}^{-1}(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{y}_N^* - \boldsymbol{X}_N^*\hat{\boldsymbol{\beta}}_n).$$

According to Theorem 1, for both direct density prediction method and normal prediction method, the predictive distribution has mean

$$\begin{aligned} & E^*\{\boldsymbol{x}_{n+1}^T\hat{\boldsymbol{\beta}}_n + \gamma_N^*(\hat{\boldsymbol{\theta}}_n)^T R^{*-1}(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{y}_N^* - \boldsymbol{X}_N^*\hat{\boldsymbol{\beta}}_n)\} + o_p(1) \\ = \; & \boldsymbol{x}_{n+1}^T\hat{\boldsymbol{\beta}}_n + \frac{1}{m^{d-1}}\sum_{\boldsymbol{i}} \gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{y}_{\boldsymbol{i}} - \boldsymbol{X}_{\boldsymbol{i}}\hat{\boldsymbol{\beta}}_n) \\ & + E^*(s_N^*(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\beta}}_n)) + o_p(1). \end{aligned}$$

By the same treatment as the proof of $r_n(\cdot)$ and $r_N^*(\cdot)$ in Lemma 4 in Zhao et al. (2018), under condition A.3, we have $s_n(\cdot) = \frac{1}{(1+g)}\gamma_n(\hat{\boldsymbol{\theta}}_n)^T D_n^{-1}(\hat{\boldsymbol{\theta}}) E_n(\hat{\boldsymbol{\theta}}) D_n^{-1}\gamma_n(\hat{\boldsymbol{\theta}}_n) \to 0$ in P as well as $s_N^*(\cdot) \to 0$ prob-$P_{N,\omega}^*$ prob-P and $E^*(s_N^*(\hat{\boldsymbol{\theta}}_n, \hat{\boldsymbol{\beta}}_n)) \to 0$ in P. Decompose the predictive mean of plug-in predictor using the same technique, we show that

$$\begin{aligned} & \boldsymbol{E}\{\mu(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n, \phi_n) - \mu_1^*\} = \boldsymbol{E}\{\mu(\boldsymbol{x}_{n+1} \mid \boldsymbol{X}_n, \boldsymbol{y}_n, \phi_n) - \hat{\mu}_2^*\} \\ = \; & \boldsymbol{E}\frac{m^{d-1}-1}{m^{d-1}}\sum_{\boldsymbol{i}} \gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{y}_{\boldsymbol{i}} - \boldsymbol{X}_{\boldsymbol{i}}\hat{\boldsymbol{\beta}}_n) + o_p(1) \\ \to \; & 0. \end{aligned}$$

where $\sum_{\boldsymbol{i}}$ is the summation of all m^d blocks.

The predictive distribution of direct density prediction method, which follows normal mixture, has variance

$$
\begin{aligned}
\sigma_1^{2*} &= E^*\{\sigma^2(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^{*(u)}, \boldsymbol{y}_N^{*(u)}, \hat{\phi}_n) + [\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^{*(u)}, \boldsymbol{y}_N^{*(u)}, \hat{\phi}_n) \\
&\quad -\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2\} + o_p(1) \\
&= \frac{1}{(m!)^{d-1}} \sum_{\boldsymbol{\pi}_1,\ldots,\boldsymbol{\pi}_d} \{\sigma^2(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) + [\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) \\
&\quad -\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2\} + o_p(1) \\
&= \hat{\sigma}_n^2 \Big\{1 - \frac{1}{m^{d-1}} \sum_{\boldsymbol{i}} \gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1} \gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n) - E^*(t_N^*(\hat{\boldsymbol{\theta}}_n))\Big\} \\
&\quad + \frac{1}{(m!)^{d-1}} \sum_{\boldsymbol{\pi}_1,\ldots,\boldsymbol{\pi}_d} [\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) - \mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2 + o_p(1),
\end{aligned}
$$

where $t_N^*(\hat{\boldsymbol{\theta}}_n) = \frac{1}{(1+\hat{g}^*)} \gamma_N^*(\hat{\boldsymbol{\theta}}_n)^T D_N^{*\,-1}(\hat{\boldsymbol{\theta}}_n) E_N^*(\hat{\boldsymbol{\theta}}_n) D_N^{*\,-1}(\hat{\boldsymbol{\theta}}_n) \gamma_N^*(\hat{\boldsymbol{\theta}}_n)$ and $\sum_{\boldsymbol{\pi}}$ is the summation of independent permutation over $\{0, 1, \ldots, m-1\}$.

The predictive distribution of normal prediction method has variance

$$
\begin{aligned}
\sigma_2^{2*} &= E^*\{\sigma^2(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^{*(u)}, \boldsymbol{y}_N^{*(u)}, \hat{\phi}_n) + o_p(1) \\
&= \frac{1}{(m!)^{d-1}} \sum_{\boldsymbol{\pi}_1,\ldots,\boldsymbol{\pi}_d} \{\sigma^2(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) + o_p(1) \\
&= \hat{\sigma}_n^2 \Big\{1 - \frac{1}{m^{d-1}} \sum_{\boldsymbol{i}} \gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1} \gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n) - E^*(t_N^*(\hat{\boldsymbol{\theta}}_n))\Big\} + o_p(1)
\end{aligned}
$$

Under condition A.3, we have $t_N^*(\cdot) \to 0$ prob-$P_{N,\omega}^*$ prob-P. Then the result follows. Comparing the predictive variance under both methods, it is straightforward to show that

$$
\sigma_1^{2*} - \sigma_2^{2*} = E^*[\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^{*(u)}, \boldsymbol{y}_N^{*(u)}, \hat{\phi}_n) - \mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2 + o_p(1)
$$

i.e. $P(\sigma_1^{2*} \geq \sigma_2^{2*}) \to 1$ as $n \to \infty$ $\square$

Appendix D: Proof of Theorem 3

Using the same technique in proof of Theorem 2, it is easy to show that the variance of the plug-in predictive distribution can be written as

$$\hat{\sigma}_n^2\Big\{1-\sum_{\boldsymbol{i}}\gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}\gamma_{\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)-\frac{1}{(1+g)}\gamma_n(\hat{\boldsymbol{\theta}}_n)^T D_n^{-1}(\hat{\boldsymbol{\theta}}_n)E_n(\hat{\boldsymbol{\theta}}_n)D_n^{-1}\gamma_n(\hat{\boldsymbol{\theta}}_n)\Big\},$$

Under condition A.3, we have $t_n(\cdot)=\frac{1}{(1+g)}\gamma_n(\hat{\boldsymbol{\theta}}_n)^T D_n^{-1}(\hat{\boldsymbol{\theta}}_n)E_n(\hat{\boldsymbol{\theta}}_n)D_n^{-1}\gamma_n(\hat{\boldsymbol{\theta}}_n)$ $\to 0$ in P. Deducting the predictive variance σ_1^{2*} and σ_2^{2*} calculated in Theorem 2, the result follows immediately. □

Appendix D: Proof of Theorem 4

Under the regularity assumptions given in Appendix, we compare the predictive variance on both in-sample and out-of-sample case under direct density approach and normal approximation approach. For the direct density approach, denote the variance within the sampled data by $\sigma_1^{2*(I)}$ and the variance for out-of-sample by $\sigma_1^{2*(O)}$. Similarly, we have $\sigma_2^{2*(I)}$ and $\sigma_2^{2*(O)}$ for the normal approximation method. We predict y at a given value x_{n+1}.

In one single m-LHD subsamples $(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*)$,

when x_{n+1} is within $(\boldsymbol{X}_N^*, \boldsymbol{y}_N^*)$, by the interpolation property of Gaussian

Process Model, we have

$$\hat{\sigma}^2_{n+1} = 0$$

when x_{n+1} is out of $(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N)$, according to proof of Theorem 2 we have

$$\begin{aligned}\hat{\sigma}^2_{n+1} &= \sigma^2(\boldsymbol{x}_{n+1}|\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\phi}_n) + [\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\phi}_n) - \mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2 \\ &\quad + o_p(1).\end{aligned}$$

Under the regularity assumptions given in Appendix

$$\begin{aligned}\sigma_1^{2*(I)} &= (1 - \frac{1}{m^{d-1}})\hat{\sigma}^2_n\Big\{1 - \frac{1}{m^{d-1}}\sum_{\boldsymbol{i}} \gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R^{-1}_{\boldsymbol{i},\boldsymbol{i}}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n) - E^*(t^*_N(\hat{\boldsymbol{\theta}}_n))\Big\} \\ &\quad +(1 - \frac{1}{m^{d-1}})\frac{1}{(m!)^{d-1}}\sum_{\boldsymbol{\pi}_1,\ldots,\boldsymbol{\pi}_d}[\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\phi}_n) - \mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2 \\ &\quad + \frac{1}{m^{d-1}} * 0 + o_p(1) \\ &= (1 - \frac{1}{m^{d-1}})\Big\{\hat{\sigma}^2_n[1 - \frac{1}{m^{d-1}}\sum_{\boldsymbol{i}} \gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R^{-1}_{\boldsymbol{i},\boldsymbol{i}}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)] \\ &\quad + \frac{1}{(m!)^{d-1}}\sum_{\boldsymbol{\pi}_1,\ldots,\boldsymbol{\pi}_d}[\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\phi}_n) - \mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2\Big\} + o_p(1)\end{aligned}$$

Under the normal approximation approach, similarly, when x_{n+1} is within $(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N)$, by property of interpolation of Gaussian Process Model,

$$\hat{\sigma}^2_{n+1} = 0$$

when x_{n+1} is out of $(\boldsymbol{X}^*_N, \boldsymbol{y}^*_N)$, according to proof of Theorem 2 we have

$$\hat{\sigma}^2_{n+1} = \sigma^2(\boldsymbol{x}_{n+1}|\boldsymbol{X}^*_N, \boldsymbol{y}^*_N, \hat{\phi}_n) + o_p(1)$$

Under the regularity assumptions given in Appendix

$$\begin{aligned}\sigma_2^{2*(I)} &= (1-\frac{1}{m^{d-1}})\hat{\sigma}_n^2\Big\{1-\frac{1}{m^{d-1}}\sum_{\boldsymbol{i}}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n) - E^*(t_N^*(\hat{\boldsymbol{\theta}}_n))\Big\} \\ &\quad +\frac{1}{m^{d-1}} * 0 + o_p(1) \\ &= (1-\frac{1}{m^{d-1}})\hat{\sigma}_n^2[1-\frac{1}{m^{d-1}}\sum_{\boldsymbol{i}}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)] + o_p(1)\end{aligned}$$

According to Theorem 2,

$$\begin{aligned}\sigma_1^{2*(O)} &= \hat{\sigma}_n^2\Big\{1-\frac{1}{m^{d-1}}\sum_{\boldsymbol{i}}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)\Big\} \\ &\quad +\frac{1}{(m!)^{d-1}}\sum_{\boldsymbol{\pi}_1,\ldots,\boldsymbol{\pi}_d}[\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*,\boldsymbol{y}_N^*,\hat{\phi}_n) - \mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n,\boldsymbol{y}_n,\hat{\phi}_n)]^2 + o_p(1)\end{aligned}$$

and

$$\sigma_2^{2*(O)} = \hat{\sigma}_n^2\Big\{1-\frac{1}{m^{d-1}}\sum_{\boldsymbol{i}}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n) + o_p(1)$$

To compare the in-sample and out-of-sample predictive variance, simply take the difference under the corresponding approach and the result follows immediately, we have

$$\sigma_2^{2*(O)} - \sigma_2^{2*(I)} = \frac{\hat{\sigma}_n^2}{m^{d-1}}[1-\frac{1}{m^{d-1}}\sum_{\boldsymbol{i}}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)] + o_p(1)$$

i.e. $P(\sigma_2^{2*(O)} \geq \sigma_2^{2*(I)}) \to 1$ as $n \to \infty$

and

$$\begin{aligned}
\sigma_1^{2*(O)} - \sigma_1^{2*(I)} &= \frac{1}{m^{d-1}}\Big\{\hat{\sigma}_n^2[1 - \frac{1}{m^{d-1}}\sum_{\boldsymbol{i}} \gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)^T R_{\boldsymbol{i},\boldsymbol{i}}^{-1}\gamma_{n,\boldsymbol{i}}(\hat{\boldsymbol{\theta}}_n)] \\
&\quad + \frac{1}{(m!)^{d-1}} \sum_{\boldsymbol{\pi}_1,\ldots,\boldsymbol{\pi}_d} [\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) - \mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2\Big\} \\
&\quad + o_p(1) \\
&= \sigma_2^{2*(O)} - \sigma_2^{2*(I)} + o_p(1) \\
&\quad + (mm!)^{1-d}\sum_{\boldsymbol{\pi}}[\mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_N^*, \boldsymbol{y}_N^*, \hat{\phi}_n) - \mu(\boldsymbol{x}_{n+1}|\boldsymbol{X}_n, \boldsymbol{y}_n, \hat{\phi}_n)]^2 \\
&\geqslant 0
\end{aligned}$$

i.e. $P(\sigma_1^{2*(O)} \geq \sigma_1^{2*(I)}) \to 1$ as $n \to \infty$

□

Appendix E: Figures and Tables

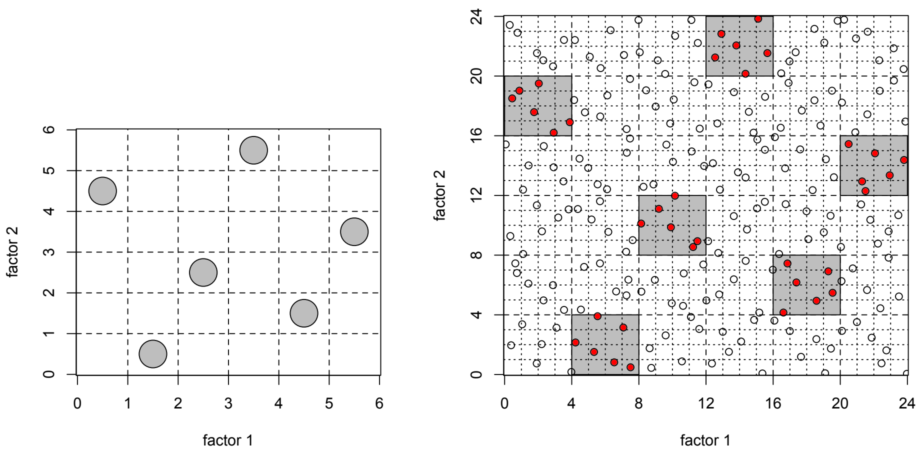

Figure 1: An example of LHD-based block bootstrap

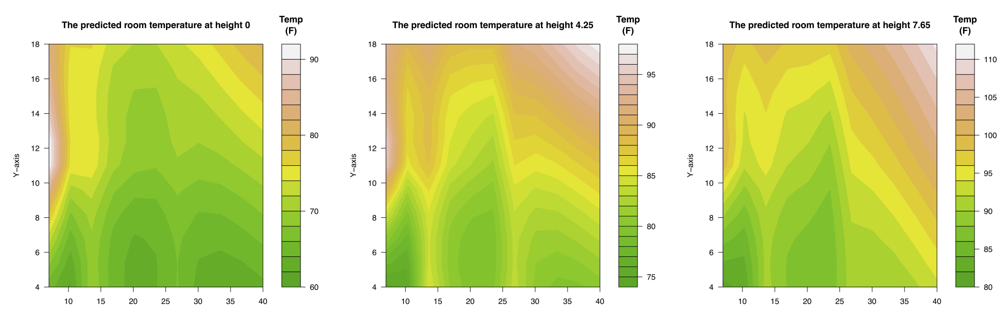

Figure 2: Bootstrap predictive heat map in a data center

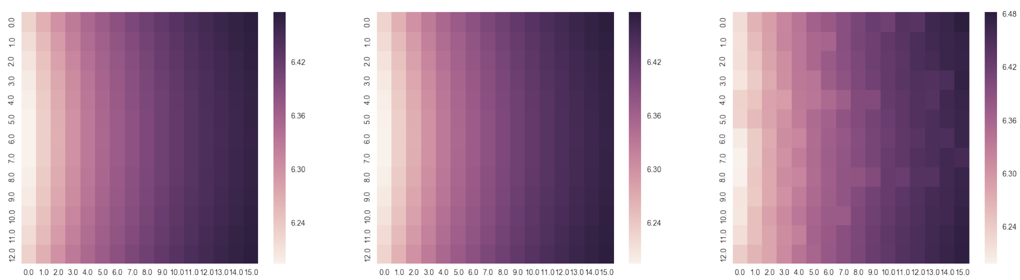

Figure 3: Thickness predictions of icesheet. Left: Truth. Middle: Prediction using conventional GP. Right: LHD-based method.

Table 1: Comparisons with regular MLE replications (standard deviation in parenthesis).

	AllData	LHD		
		m=3	m=4	m=5
		$n = 2000$		
θ_1	0.40(0.03)	0.43(0.09)	0.48(0.27)	0.90(2.68)
θ_2	0.40(0.03)	0.42(0.10)	0.45(0.24)	0.46(0.26)
θ_3	0.39(0.03)	0.45(0.13)	0.42(0.15)	0.50(0.41)
β_1	2.02(0.52)	2.04(0.68)	2.13(0.67)	2.06(0.72)
β_2	-2.04(0.57)	-1.98(0.70)	-2.03(0.64)	-2.00(0.73)
β_3	1.05(0.55)	1.03(0.69)	1.02(0.72)	1.04(0.68)
$MSPE$	0.10(0.14)	0.24(0.32)	0.33(0.46)	0.44(0.61)
$Time$	76.78(5.12)	10.96(4.03)	7.84(3.60)	4.83(1.67)
		$n = 4000$		
θ_1	0.40(0.02)	0.44(0.09)	0.43(0.13)	0.41(0.13)
θ_2	0.40(0.03)	0.44(0.09)	0.46(0.11)	0.41(0.14)
θ_3	0.40(0.02)	0.42(0.08)	0.44(0.12)	0.40(0.12)
β_1	2.07(0.53)	2.11(0.60)	2.10(0.68)	2.24(0.64)
β_2	-2.01(0.52)	-2.05(0.56)	-2.04(0.60)	-2.15(0.65)
β_3	1.04(0.49)	1.02(0.71)	1.02(0.60)	1.00(0.67)
$MSPE$	0.07(0.09)	0.16(0.22)	0.22(0.31)	0.27(0.38)
$Time$	605.83(35.93)	58.38(5.21)	20.53(4.41)	12.39(3.61)

Table 2: Comparisons of predictive variance (standard deviation in parenthesis).

	LHD (m=3)	Regular Bootstrap	Plugin
$n = 1000$			
Direct Density	0.53(0.19)	0.35(0.15)	0.15(0.04)
Normal	0.41(0.08)	0.22(0.05)	
Time	5.50(2.09)	405.30(119.61)	19.11(6.45)
$n = 2000$			
Direct Density	0.39(0.22)	0.33(0.16)	0.10(0.02)
Normal	0.31(0.08)	0.20(0.06)	
Time	10.96(4.03)	1917.15(543.02)	76.78(5.12)

Table 3: LHD Bootstrap analysis of thermal management data

	Variable	Levels	$\hat{\beta}$	$\hat{\theta}$
x_1	CRAC unit 1 flow rate (cfm)	(0,7000,8500,10000	-8.58(0.96)	0.85(0.17)
		11500,13000)		
x_2	CRAC unit 2 flow rate (cfm)	(0,7000,8500,10000	-11.12(1.26)	0.77(0.23)
		11500,13000)		
x_3	CRAC unit 3 flow rate (cfm)	(0,2500,4000,5500)	-6.83(0.80)	1.14(0.27)
x_4	CRAC unit 4 flow rate (cfm)	(0,2500,4000,5500)	-6.26(0.98)	1.70(0.71)
x_5	Room temperature setting (F)	(65,67,69,71,73, 75)	-0.82(0.66)	3.39(0.94)
x_6	Tile open area percentage (%)	$(15, 25, 35, 45$	0.15(3.63)	1.24(0.91)
		$(55, 65, 75)$		
x_7	Location in x-axis	8 unequally spaced	-5.09(2.72)	0.14(0.11)
x_8	Location in y-axis	4 unequally spaced	3.70(2.18)	0.62(0.25)
x_9	Height	18 equally spaced	33.43(3.90)	21.61(0.22)