A Appendix

A.1 Theoretical Results

In this section, we illustrate the list of required assumptions. The complete proof and the relevant lemmas for our main theorem will be also provided.

Assumption 1 MIXING CONDITION *We assume $Y_i = m(X_i) + E_i$ where the error process $\{E_i\}$ is a stationary, absolutely regular (β-mixing) process (Bradley 2005) with finite $(2+\delta)$th moment for some $\delta > 0$.*

Assumption 1 focuses on absolutely regular or β-mixing processes, which enable us to extend uniform law of large numbers from independent process to this dependent process under moderate restriction on the class of functions under consideration. No additional assumptions is required on the decay rate of the β-mixing coefficients.

Note that Assumption 1 implies that the error process $\{E_i\}$ also guarantees the α-mixing property that there exist a function φ such that $\varphi(t) \downarrow 0$ as $t \to \infty$, and a function $\psi : \mathcal{N}^2 \to R^+$ symmetric and increasing in each of its two arguments, such that

$$\begin{aligned}\alpha(\mathcal{B}(\mathcal{S}'), \mathcal{B}(\mathcal{S}'')) &:= \sup\{|P(AB) - P(A)P(B)|, A \in \mathcal{B}(\mathcal{S}'), B \in \mathcal{B}(\mathcal{S}'')\} \\ &\leq \psi(Card(\mathcal{S}'), Card(\mathcal{S}''))\varphi(d(\mathcal{S}', \mathcal{S}'')),\end{aligned}$$

for any $\mathcal{S}', \mathcal{S}'' \subset R^2$. The function φ moreover is such that

$$\lim_{m\to\infty} m^\gamma \sum_{j=m}^{\infty} j^2 \{\varphi(j)\}^{\kappa/(2+\kappa)} = 0$$

for some constant $\gamma > \max\{1, 2\kappa/(2+\kappa)\}$ and some $\kappa > 0$.

Assumption 2 REGULARITY OF THE WORKING PRECISION MATRIX *The working precision matrix $Q = \Gamma^{-1}$ admits a regular and sparse lower-triangular Cholesky factor $\Gamma^{-1/2}$ such that*

$$\Gamma^{-1/2} = \begin{pmatrix} L_{q\times q} & 0 & 0 & \cdots & \cdots \\ \rho^T_{1\times(q+1)} & & 0 & \cdots & \cdots \\ 0 & \rho^T_{1\times(q+1)} & & 0 & \cdots \\ \vdots & \ddots & & & \vdots \\ \cdots & 0 & 0 & \rho^T_{1\times(q+1)} & \end{pmatrix}$$

where $\rho = (\rho_q, \rho_{q-1}, \cdots, \rho_0)^T \in R^{q+1}$ for some fixed $q \in N$, and L is a fixed lower-triangular $q \times q$ matrix.

Assumption 2 requires the cholesky factor of the precision matrix to be sparse and regular. For spatial data, exponential covariance family on a one-dimensional grid satisfies this. Other covariances like the Matérn family (except the exponential covariance) do not generally satisfy this assumption. However, NNGP covariance matrices satisfy this and are now commonly used as an excellent approximation to the full GP covariance matrices (Datta et al. 2016a). We can always use an approximate working covariance matrix arising from NNGP to the true covariance of the process to satisfy this assumption.

Assumption 3 DIAGONAL DOMINANCE OF THE WORKING PRECISION MATRIX *Q is diagonally dominant satisfying $Q_{ii} - \sum_{j\neq i} |Q_{ij}| > \xi$, for all i, for some constant $\xi > 0$.*

Diagonal dominance (Assumption 3) implies $\lambda_{min}(Q)$ is bounded away from zero as $n \to \infty$ which is needed to ensure stability of the GLS estimate. Note that under Assumption 2, checking that the first $(q+1)$ rows of Q are diagonally dominant is enough to verify Assumption 3.

Assumption 4 TAIL BEHAVIOR OF THE ERROR DISTRIBUTION

(a) *There exist* $\{\xi_n\}_{n\geq 1}$ *such that*

$$\xi_n \to \infty, \frac{t_n(logn)\xi_n^4}{n} \to 0, \text{ and}$$
$$E\left[(\max_i \epsilon_i^2)1(\max_i\{\epsilon_i^2 > \xi_i^2\})\right] \to 0 \text{ as } n \to \infty$$

(b) *There exist constant* $C_\pi > 0$ *and* $n_0 \in N^*$ *such that with probability* $1-\pi$, *for* $\forall n > n_0$,

$$\max_i |\epsilon_i| \leq C_\pi \sqrt{\log n}$$

(c) *Let* $\mathcal{I}_n \subseteq \{1, 2, \cdots, n\}$ *with* $|\mathcal{I}_n| := a_n$ *and* $a_n \to \infty$ *as* $n \to \infty$. *Then* $\frac{1}{a_n}|\sum_{i\in\mathcal{I}_n} \epsilon_i| > \delta$ *with probability at most* $C\exp(-ca_n)$, *and* $\frac{1}{n}|\sum_i \epsilon_i^2| > \sigma_0^2$ *with probability at most* $C\exp(-cn)$ *for any* $\delta > 0$, *and some constants* $c, C, \sigma_0^2 > 0$.

For Gaussian errors, ξ_n needs to be $\mathcal{O}(logn)^2$ which makes the scaling condition in Assumption 4(a) as $\frac{t_n(logn)^9}{n} \to 0$. This is the same scaling used in Scornet et al. (2015) for Gaussian errors and using the entire sample. In general, the choice of ξ_n will be dependent on the error distribution. Assumption 4(a), 4(b), and 4(c) will all be satisfied by sub-Gaussian errors.

Assumption 5 ADDITIVE MODEL ON THE COORDINATES *The true mean function* $m(x, s)$ *is additive on the coordinates* s_d, *that is,* $m(x, s) = \sum_{d=1}^D m_d(x, s_d)$, *where each component* m_d *is continuous.*

Assumption 6 SAMPLING SITES *The observations are positioned at* $\{s_i, i = 1, 2, \cdots, N\} \subset R^2$, *for which are defined under domain-expanding infill asymptotics, where*

$$\delta_N := \max_{1\leq j\leq N} \delta_{j,N} \to 0,$$
$$\textit{with } \delta_{j,N} := \min\{\|s_i - s_j\| : 1 \leq i \leq n, i \neq j\}$$

that is, the distance between neighboring observations all tends to zero, as $N \to \infty$, *and*

$$\Delta_N := \min_{1\leq j\leq N} \Delta_{j,N} \to \infty,$$
$$\textit{with } \Delta_{j,N} := \max\{\|s_i - s_j\| : 1 \leq i \leq n, i \neq j\}$$

that is, the domain at each location is expanding to ∞, *as* $N \to \infty$, *where* $\|\cdot\|$ *denotes the Euclidean norm in* R^2. *We suppose* $\min_{1\leq j\leq N} \delta_{j,N}/\delta_N \geq c_1 > 0$, *and* $\max_{i\leq j\leq n} \Delta_{j,N}/\Delta_N \leq C_1 < \infty$, *for all N. Also, there exists a continuous sampling intensity function* f_S *defined on* R^2 *such that*

(a) *for any measurable set* $A \subset R^2$, $N^{-1}\sum_{i=1}^N I(s_i \in A) \to \int_A f_S(s)ds$ *as* $N \to \infty$

(b) $f_S(s)$ *is bounded and has second derivatives which are continuous on* R^2.

Assumption 7 KERNEL FUNCTIONS *The kernel function* $K(\cdot)$ *satisfies that* $\int K(u)du = 1$, $\int uK(u)du = 0$, *and* $\mu_{K,2} := \int u^2K(u)du < \infty$, $\nu_K := \int K^2(u)du < \infty$.

Assumption 8 BANDWIDTHS I *As* $N \to \infty$, *(a)* $h_N \to 0$; *(b)* $Nh_N \to \infty$; *and (c)* $\delta_N^{-2(1+2/\gamma)} h^{2(\gamma-2\kappa/(2+\kappa)\gamma} \to 0$.

Assumption 9 BANDWIDTHS II *Let* $c_N = \{\delta_n^2 h^{\kappa/(2+\kappa)}\}^{-1/\gamma}$, *which tends to* ∞ *as* $N \to \infty$. *(a)* $\limsup_{N\to\infty} Nh^2 > 0$, $Nh^5 = \mathcal{O}(1)$; *(b)* $\liminf_{m\to\infty} m^{4+3\gamma}\sum_{t=m}^\infty t^2[\varphi(t)]^{\kappa/(2+\kappa)} < \infty$; *(c)* $N\phi(1, N)\varphi(c_N) \to 0$, *as* $N \to \infty$

Assumption 10 *Covariate X has the uniform distribution over $[0,1]^p$.*

Assumption 10 is largely for notational convenience. More generally, one could assume that the density of X is positive and bounded. Likewise, the spatial domain D could be assumed to be compact and bounded region in R^2

Assumption 11 *Let $k_\theta(l) := |\{Z_i \in \mathcal{D}_n^*; X_i \in R_{l_{(x,\theta)}}\}|$ denote the number of units from its bootstrap sample $\mathcal{D}_n^*$ in its terminal node l containing x.*

(a) The proportion of observations from $\mathcal{D}_n^*$ in any given node, relative all observations from $\mathcal{D}_n^*$, is decreasing in n, that is, $\max_{l,\theta} k_\theta(l) = o(n)$

(b) The minimum number of observations from $\mathcal{D}_n^*$ in a node is increasing in n, that is, $1/\min_{l,\theta} k_\theta(l) = o(n)$

(c) The probability that variable $m \in \{1, \cdots, p\}$ is chosen for a given split point is bounded from below for every node by a positive constant.

(d) When a node is split, the proportion of observations belong to $\mathcal{D}_n^*$ in the original node that fall into each of the resulting sub-nodes is bounded from below by a positive constant.

The conditions in Assumption 11 are adapted from assumptions used to prove consistency of quantile regression forests [19]. Tree construction algorithms that satisfy these properties or variants of them have been referred to in recent random forest literature as "regular", "balanced", or "random-split" [25, 1, 8]

Assumption 12 *$F_E(e|X=x)$ is Lipschitz continuous with parameter L. That is, for all $x, x' \in [0,1]^p$,*

$$\sup_{e \in R} |F_E(\epsilon|x) - F_E(e|x')| \leq L\|x - x'\|_1$$

As Wager and Athey (2018) noted, all existing results on pointwise consistency of random forests have required an analogous smoothness condition in the distribution of interest, including Bias (2012), meinshausen (2006), and Wager and Athey (2018).

Assumption 13 *$F_E(e|X=x)$ is strictly monotone in e for all $x \in [0,1]^p$.*

We assume that the distribution of prediction errors is strictly monotone so that consistency of quantile estimates follows from consistency of distribution estimates.

Assumption 14 BEHAVIOR OF THE OUT-OF-BAG WEIGHT *Define $\mathcal{M}_i(\delta) := \{\hat{m}_\Gamma(X_i) - m(X_i)\}$ be the event for any given $\delta > 0$. We say that δ-stability of the i^{th} unit has been realized if and only if $\mathcal{M}_i(\delta)$ holds. For all $x \in [0,1]^p$, there exists $\delta_0 > 0$ such that for any $\delta_0 \in (0, \delta_0)$, $E[v_i(x)|\mathcal{M}_i(\delta)] = \mathcal{O}(n^{-1})$ and $E[v_i(x)| - \mathcal{M}_i(\delta)] = \mathcal{O}(n^{-1})$*

Assumption 14 characterizes the stability of the random forest and the underlying population distribution. It states that the expected out-of-bag weight of the i^{th} observation is of order $1/n$ whether δ-stability has been realized for the observation or not. The expected values are taken over all training units and all random parameters governing the sample-splitting and tree-growing mechanisms.

[**Proof of Theorem 2.1**] Fix $x \in [0,1]^p$ (Assumption 10). Denote $E = Y - m(X)$ as the true underlying error random variable (Assumption 1). Let E_i^* follow the distribution $F_E(\epsilon|X_i)$, and E_i follows $F_E(\epsilon|x)$, $i = 1, \cdots, n$. The goal is to prove that $\hat{F}_n(\epsilon|x)$ is a consistent estimator of $F_E(\epsilon|x)$ by showing its convergence in probability, i.e. $|\hat{F}_n(\epsilon|x) - F_E(\epsilon|x)| \xrightarrow{p} 0$ as $n \to \infty$, for any $\epsilon \in R$. We have

$$\begin{aligned}\hat{F}_n(\epsilon|x) &= \sum_{i=1}^n v_i(x) \int_{-\infty}^{\epsilon} K_h(E_i^* - e)de \\ &= \sum_{i=1}^n v_i(x) \int_{-\infty}^{\epsilon} K_h(E_i - e)de + \sum_{i=1}^n v_i(x) \int_{-\infty}^{\epsilon} \left(K_h(E_i^* - e) - K_h(E_i - e)\right) de\end{aligned}$$

$$\begin{aligned}\left|\hat{F}_n(\epsilon|x) - F_E(\epsilon|x)\right| \leq & \left|\sum_{i=1}^n v_i(x) \int_{-\infty}^{\epsilon} K_h(E_i - e)de - F_E(\epsilon|x)\right| \\ & + \left|\sum_{i=1}^n v_i(x) \int_{-\infty}^{\epsilon} \left(K_h(E_i^* - e) - K_h(E_i - e)\right) de\right|\end{aligned} \tag{3}$$

We wish to show that the right side of (3) converges to zero in probability. By following the terminologies in Meinshausen(2006), we will call the first term on the right side as a "Variance Term", and the second term as a "Shift Term", that is, $\left|\hat{F}_n(\epsilon|x) - F_E(\epsilon|x)\right| \leq$ (Variance Term)+(Shift Term). Next, we will verify that each term converges to zero in probability.

Bounding the Variance Term We will use the asymptotics for the marginal density function estimator for spatial data (Lu *et al.* 2014). In order to verify the variance term converges to zero in probability, it suffices to show that

$$\sum_{i=1}^n v_i(x) \int_{-\infty}^{\epsilon} K_{h_n}(E_i - e)de - F_E(\epsilon|x) \xrightarrow{d} \Phi\left(\frac{1}{2}\mu_{K,2}\ddot{f}_E(\epsilon|x)h_n^2, \frac{\nu_K f_E(\epsilon|x)}{nh_n}\right), \tag{4}$$

which implies that both the mean and the variance of the limiting distribution are close to zero for sufficiently large sample size.

Since the weights $\{v_i(x)\}$ are built on a randomly chosen subset of bootstrapped samples that are not containing Y_i, conditioning on $X = x$ yields sufficient independence $(v_i \perp\!\!\!\perp E_i)|X = x$. Then we can evaluate the expectation of the weighted average in (4) as follows:

$$\begin{aligned}E\left(\sum_{i=1}^n v_i(x)K_h(E_i - e)\right) &= E\left[E\left(\sum_{i=1}^n v_i(x)K_h(E_i - e)\middle| X = x\right)\right] \\ &= \sum_{i=1}^n E\left[E\left(v_i(x)\middle| X = x\right) E\left(K_h(E_i - e)\middle| X = x\right)\right]\end{aligned} \tag{5}$$

After separating the one conditional expectation into the out-of-bag weight part and the kernel part as shown in (5), we can apply the similar process illustrated in Zudi Lu *et al.* [17] and the assumption of kernel functions (Assumption 7) to the kernel part:

$$
\begin{aligned}
E\left(K_h(E_i-e)\middle|X=x\right) &= h^{-1}\int K((u-e)/h)f_E(u|x)du \\
&= \int K(u)f_E(e+hu|x)du \\
&= \int K(u)\left[f_E(e|x)+\dot{f}_E(e|x)(hu)+\frac{1}{2}\ddot{f}_E(e+\xi hu|x)(hu)^2\right]du \\
&= f_E(e|x)+\frac{1}{2}\ddot{f}_E(e|x)h^2\int u^2K(u)du
\end{aligned}
$$

This completes the equation (5) as follows:

$$
\begin{aligned}
E\left(\sum_{i=1}^{n}v_i(x)K_h(E_i-e)\right) &= E\left[E\left(\sum_{i=1}^{n}v_i(x)K_h(E_i-e)\middle|X=x\right)\right] \\
&= \sum_{i=1}^{n}E\left[E\left(v_i(x)\middle|X=x\right)E\left(K_h(E_i-e)\middle|X=x\right)\right] \\
&= \sum_{i=1}^{n}E\left[E\left(v_i(x)\middle|X=x\right)\left(f_E(e|x)+\frac{1}{2}\ddot{f}_E(e|x)h^2\int u^2K(u)du\right)\right] \\
&= \left(f_E(e|x)+\frac{1}{2}\ddot{f}_E(e|x)h^2\int u^2K(u)du\right)E\sum_{i=1}^{n}v_i(x) \\
&= f_E(e|x)+\frac{1}{2}\mu_{K,2}\ddot{f}_E(e|x)h^2
\end{aligned} \tag{6}
$$

Furthermore, since the variance of a summation is equal to the summation of the covariances,

$$
\begin{aligned}
Var\left(\sum_{i=1}^{n}v_i(x)K_h(E_i-e)\right) &= \sum_{i=1}^{n}Var\Big(v_i(x)K_h(E_i-e)\Big) \\
&\quad+\sum_{i\neq j}Cov\Big(v_i(x)K_h(E_i-e),v_j(x)K_h(E_j-e)\Big)
\end{aligned}
$$

Each summation of the right hand side converges to zero by Lemma A.1 and Lemma A.2.

Bounding the Shift Term We will first show that

$$
\sum_{i=1}^{n}v_i(x)K_{h_n}(E_i^*-e)-\sum_{i=1}^{n}v_i(x)f_E(e|X_i)\xrightarrow{p}0, \tag{7}
$$

as $n \to \infty$. Since $v_i(x)$ and $K_{h_n}(E_i^* - e)$ are not exactly but asymptotically independent conditioning on X_i (Lemma A.3),

$$
\begin{aligned}
&E\left(K_h(E_i^* - e)\middle| X_i\right) = E\left(K_h(\tilde{E}_i^* + o_p(1) - e)\middle| X_i\right)\\
&= h^{-1}\int K((u + o_p(1) - e)/h) f_{\tilde{E}}(u|x)du = \int K(u) f_{\tilde{E}}(e - o_p(1) + hu|x)du\\
&= \int K(u)\left[f_{\tilde{E}}(\epsilon|x) + \dot{f}_{\tilde{E}}(\epsilon|x)(hu - o_p(1)) + \frac{1}{2}\ddot{f}_{\tilde{E}}(e + \xi hu|x)(hu - o_p(1))^2\right]du\\
&= f_{\tilde{E}}(\epsilon|x) - \dot{f}_{\tilde{E}}(\epsilon|x)o_p(1) + \frac{1}{2}\ddot{f}_{\tilde{E}}(\epsilon|x)h^2(1 + o_p(1))\int u^2 K(u)du\\
&\to f_E(\epsilon|x) + \frac{1}{2}\mu_{K,2}\ddot{f}_E(\epsilon|x)h^2
\end{aligned}
\tag{8}
$$

Consequently, we find the left side of the equation (7) has expectation zero by the linearity of expectation such that

$$
\begin{aligned}
&E\left[\sum_{i=1}^n v_i(x)K_{h_n}(E_i^* - e) - \sum_{i=1}^n v_i(x) f_E(e|X_i)\right]\\
&= E\left[E\left(\sum_{i=1}^n v_i(x)K_h(\tilde{E}_i^* + o_p(1) - e)\middle| X_i\right) - \sum_{i=1}^n v_i(x) f_E(e|X_i)\right]\\
&= E\left[\sum_{i=1}^n E\left(v_i(x)\middle| X_i\right) E\left(K_h(\tilde{E}_i^* + o_p(1) - e)\middle| X_i\right) - \sum_{i=1}^n v_i(x) f_E(e|X_i)\right]\\
&\to E\left[\sum_{i=1}^n E\left(v_i(x)\middle| X_i\right)\left(f_E(\epsilon|X_i) + \frac{1}{2}\mu_{K,2}\ddot{f}_E(\epsilon|X_i)h^2\right) - \sum_{i=1}^n v_i(x) f_E(e|X_i)\right]\\
&= E\left[E\left(\sum_{i=1}^n v_i(x) f_E(\epsilon|X_i)\middle| X_i\right) - \sum_{i=1}^n v_i(x) f_E(e|X_i)\right] + \frac{1}{2}\mu_{K,2}\ddot{f}_E(\epsilon|X_i)h^2\left(E\sum_{i=1}^n v_i(x)\right)\\
&= \frac{1}{2}\mu_{K,2}\ddot{f}_E(\epsilon|X_i)h^2
\end{aligned}
\tag{9}
$$

Also, we can consider the left side of the equation (7) has decreasing variance by showing that

$$
\begin{aligned}
&Var\left[\sum_{i=1}^n v_i(x)K_{h_n}(E_i^* - e) - \sum_{i=1}^n v_i(x) f_E(e|X_i)\right]\\
&= \sum_{i=1}^n Var\Big[v_i(x)\Big(K_{h_n}(E_i^* - e) - f_E(e|X_i)\Big)\Big]\\
&+ \sum_{i\neq j} Cov\Big[v_i(x)\Big(K_{h_n}(E_i^* - e) - f_E(e|X_i)\Big), v_j(x)\Big(K_{h_n}(E_j^* - e) - f_E(e|X_j)\Big)\Big]
\end{aligned}
$$

with each summation of the right side converging to zero by Lemma A.4 and Lemma A.5. By the triangle inequality,

$$
\begin{aligned}
&\Big|\sum_{i=1}^{n} v_i(x)\left(K_h(E_i^*-e)-K_h(E_i-e)\right)-\sum_{i=1}^{n} v_i(x)\left(f_E(e|X_i)-f_E(\epsilon|x)\right)\Big| \\
&=\Big|\sum_{i=1}^{n} v_i(x)\left(K_h(E_i^*-e)-f_E(e|X_i)\right)-\sum_{i=1}^{n} v_i(x)\left(K_h(E_i-e)-f_E(\epsilon|x)\right)\Big| \\
&\leq \Big|\sum_{i=1}^{n} v_i(x)\left(K_h(E_i^*-e)-f_E(e|X_i)\right)\Big|+\Big|\sum_{i=1}^{n} v_i(x)\left(K_h(E_i-e)-f_E(\epsilon|x)\right)\Big|
\end{aligned}
$$

Recall that $\Big|\sum_{i=1}^{n} v_i(x)\left(K_h(E_i-e)-f_E(\epsilon|x)\right)\Big| = o_p(1)$ by equation (4), and $\Big|\sum_{i=1}^{n} v_i(x)\left(K_h(E_i^*-e)-f_E(e|X_i)\right)\Big| = o_p(1)$ by equation (7). Therefore, we can reduce the task of bounding the shift term by

$$
\sum_{i=1}^{n} v_i(x)\left(K_h(E_i^*-e)-K_h(E_i-e)\right) \xrightarrow{p} \sum_{i=1}^{n} v_i(x)\left(f_E(e|X_i)-f_E(\epsilon|x)\right),
$$

The Lipschitz continuity of the conditional prediction error distribution (Assumption 12) shows

$$
\Big|\sum_{i=1}^{n} v_i(x)\left(f_E(e|X_i)-f_E(\epsilon|x)\right)\Big| \leq \sum_{i=1}^{n} v_i(x)\|X_i-x\|_1 \tag{10}
$$

To complete the proof, we need to verify $\sum_{i=1}^{n} v_i(x)\|X_i-x\|_1 = o_p(1)$. This is followed by the argument in the Lemma 2 of Meinshausen [19]. In particular, we want to show that

$$
\lim_{B\to\infty}\frac{1}{B}\sum_{b=1}^{B}\sum_{i=1}^{n}\frac{\#\{Z_i\notin\mathcal{D}_n^*\}1(X_i\in R_{l(x,\theta_b)})}{\sum_{j=1}^{n}\#\{Z_j\notin\mathcal{D}_n^*\}1(X_j\in R_{l(x,\theta_b)})}\|X_i-x\|_1 \xrightarrow{p} 0
$$

Then it suffices to show that, for any given $b\in\{1,\cdots,B\}$,

$$
\sum_{i=1}^{n}\frac{\#\{Z_i\notin\mathcal{D}_n^*\}1(X_i\in R_{l(x,\theta_b)})}{\sum_{j=1}^{n}\#\{Z_j\notin\mathcal{D}_n^*\}1(X_j\in R_{l(x,\theta_b)})}\|X_i-x\|_1 \xrightarrow{p} 0
$$

Note that we can decompose the rectangular subspace $R_{l(x,\theta)}\subseteq[0,1]^p$ of leaf $l(x,\theta)$ of the tree into the intervals $I(x,m,\theta)\subseteq[0,1]$ for $m=1,\cdots,p$: $R_{l(x,\theta)}=\bigotimes_{m=1}^{p} I(x,m,\theta)$. Then the arguments in Lemma2 of Meinshausen [19] can assure that $\max_m|I(x,m,\theta)| = o_p(1)$, which completes our proof.

Lemma A.1 *Under Assumptions 10-14, as $n \to \infty$,*

$$\sum_{i=1}^{n} Var\Big(v_i(x) K_{h_n}(E_i - e)\Big) \to 0$$

[**Proof of Lemma A.1**]

$$\begin{aligned}
&\sum_{i=1}^{n} Var\Big(v_i(x) K_{h_n}(E_i - e)\Big) \\
&\quad = \sum_{i=1}^{n} \Big[Var\Big\{ E\Big(v_i(x) K_{h_n}(E_i - e) \Big| \Omega \setminus \{Y_i\}\Big)\Big\} + E\Big\{ Var\Big(v_i(x) K_{h_n}(E_i - e) \Big| \Omega \setminus \{Y_i\}\Big)\Big\}\Big] \\
&\quad = \sum_{i=1}^{n} \Big[Var\Big\{ v_i(x) E\Big(K_{h_n}(E_i - e) \Big| \Omega \setminus \{Y_i\}\Big)\Big\} + E\Big\{ v_i^2(x) Var\Big(K_{h_n}(E_i - e) \Big| \Omega \setminus \{Y_i\}\Big)\Big\}\Big] \\
&\quad = \sum_{i=1}^{n} \Big[(f_E(e|x) + \frac{1}{2}\mu_{K,2} \ddot{f}_E(e|x) h^2)^2 Var(v_i(x)) + \frac{\nu_K f_E(\epsilon|x)}{n h_n} E(v_i^2(x)) \Big] \\
&\quad = A \sum_{i=1}^{n} Var(v_i(x)) + B \sum_{i=1}^{n} E(v_i^2(x)),
\end{aligned} \tag{11}$$

where $A := (f_E(e|x) + \frac{1}{2}\mu_{K,2}\ddot{f}_E(e|x)h^2)^2$, $B := \frac{\nu_K f_E(\epsilon|x)}{n h_n}$. Let $C = \max\{A, B\}$, and M_n be the maximum possible weight given to any observation, which is decreasing in n (Assumption 11). Thus, equation 10 yields

$$\begin{aligned}
0 \le \sum_{i=1}^{n} Var\Big(v_i(x) K_{h_n}(E_i - e)\Big) &= A \sum_{i=1}^{n} Var(v_i(x)) + B \sum_{i=1}^{n} E(v_i^2(x)) \\
&\le (A + B) \sum_{i=1}^{n} E(v_i^2(x)) \\
&\le 2C \sum_{i=1}^{n} E(v_i^2(x)) \\
&\le 2C \sum_{i=1}^{n} M_n E(v_i(x)) \\
&= 2C M_n \xrightarrow{n \to \infty} 0
\end{aligned}$$

Lemma A.2 *Under Assumptions 10-14, as $n \to \infty$,*

$$\sum_{i \ne j} Cov\Big(v_i(x) K_h(E_i - e), v_j(x) K_h(E_j - e)\Big) \to 0$$

[**Proof of Lemma A.2**]

$$
\begin{aligned}
&\sum_{i\neq j} Cov\Big(v_i(x)K_h(E_i-e), v_j(x)K_h(E_j-e)\Big)\\
&\quad = \sum_{i\neq j}\Big[E\big\{v_i(x)K_h(E_i-e)v_j(x)K_h(E_j-e)\big\} - E\big\{v_i(x)K_h(E_i-e)\big\}E\big\{v_j(x)K_h(E_j-e)\big\}\Big]\\
&\quad = \sum_{i\neq j}E\big\{v_i(x)v_j(x)K_h(E_i-e)K_h(E_j-e)\big\} - E\big\{\sum_{i=1}^{n}v_i(x)K_h(E_i-e)\big\}E\big\{\sum_{j=1}^{n}v_j(x)K_h(E_j-e)\big\}\Big]\\
&\quad + \sum_{i=1}^{n}E\big\{v_i^2(x)K_h^2(E_i-e)\big\}\\
&\quad \to \sum_{i\neq j}E\big\{v_i(x)v_j(x)K_h(E_i-e)K_h(E_j-e)\big\} - \left(f_E(e|x)+\frac{1}{2}\mu_{K,2}\ddot{f}_E(e|x)h^2\right)^2,
\end{aligned}
$$

since, by equation (4),

$$
E\big\{\sum_{j=1}^{n}v_j(x)K_h(E_j-e)\big\} = E\big\{\sum_{j=1}^{n}v_j(x)K_h(E_j-e)\big\} = f_E(e|x)+\frac{1}{2}\mu_{K,2}\ddot{f}_E(e|x)h_n^2,
$$

and

$$
\begin{aligned}
\sum_{i=1}^{n}E\big\{v_i^2(x)K_h^2(E_i-e)\big\} &= \sum_{i=1}^{n}E\Big[E\big\{v_i^2(x)K_h^2(E_i-e)\Big|X=x\big\}\Big]\\
&= \sum_{i=1}^{n}E\Big[E\big\{v_i^2(x)\Big|X=x\big\}E\big\{K_h^2(E_i-e)\Big|X=x\big\}\Big]\\
&= \mu_{K,2}\sum_{i=1}^{n}E\big\{v_i^2(x)\big\}\to 0
\end{aligned}
$$

Let $A := f_E(e|x)+\frac{1}{2}\mu_{K,2}\ddot{f}_E(e|x)h^2$. Then it suffices to show

$$
\lim_{n\to\infty}\left|\sum_{i\neq j}E\big\{v_i(x)v_j(x)K_h(E_i-e)K_h(E_j-e)\big\} - A^2\right| = 0 \tag{12}
$$

We apply the second Taylor's expansion to the expectation of the kernel products condition on x.

$$
\begin{aligned}
&E\big\{K_h(E_i-e)K_h(E_j-e)\Big|X=x\big\}\\
&\quad = h^{-2}\int\int K((u-e)/h)K((v-e)/h)f_E(u|x)f_E(v|x)dudv\\
&\quad = \int\int K(u)K(v)f_E(e+uh|x)f_E(e+vh|x)dudv\\
&\quad = \int\int K(u)K(v)\Big[f_E(e+uh|x)f_E(e+vh|x)+\dot f_E(e+uh|x)f_E(e+vh|x)(uh)\\
&\quad + f_E(e+uh|x)\dot f_E(e+vh|x)(vh)+\frac{1}{2}\big\{\ddot f_E(e+uh|x)f_E(e+vh|x)(uh)^2\\
&\quad + 2\dot f_E(e+uh|x)\dot f_E(e+vh|x)(uvh^2)+f_E(e+uh|x)\ddot f_E(e+vh|x)(vh)^2\big\}\Big]dudv\\
&\quad = f_E(e|x)^2+\ddot f_E(e|x)f_E(e|x)\mu_{K,2}h^2(1+o_p(1))
\end{aligned}
$$

This proves $E\big\{K_h(E_i-e)K_h(E_j-e)\Big|X=x\big\}-A^2=o_p(1)$, for a sufficiently small choice of $h>0$. Finally, we can verify the equation (12) as follows:

$$
\begin{aligned}
&\sum_{i\neq j}E\big\{v_i(x)v_j(x)K_h(E_i-e)K_h(E_j-e)\big\}-A^2\\
&\quad = \sum_{i\neq j}E\left[E\big\{v_i(x)v_j(x)K_h(E_i-e)K_h(E_j-e)\Big|X=x\big\}\right]-A^2\\
&\quad = \sum_{i\neq j}E\left[E\big\{v_i(x)v_j(x)\Big|X=x\big\}E\big\{K_h(E_i-e)K_h(E_j-e)\Big|X=x\big\}\right]-A^2\\
&\quad = \sum_{i\neq j}E\left[E\big\{v_i(x)v_j(x)\Big|X=x\big\}\Big\{E\big\{K_h(E_i-e)K_h(E_j-e)\Big|X=x\big\}-A^2+A^2\Big\}\right]-A^2\\
&\quad = \sum_{i\neq j}E\left[E\big\{v_i(x)v_j(x)\Big|X=x\big\}\big(o_p(1)+A^2\big)\right]-A^2\\
&\quad \to \sum_{i\neq j}A^2E\big[E\big\{v_i(x)v_j(x)\Big|X=x\big\}\big]-A^2\\
&\quad = A^2\Big(\sum_{i\neq j}E\big\{v_i(x)v_j(x)\big\}-1\Big)=A^2\Big(\sum_{i=1}^{n}E\big\{v_i(x)(1-v_i(x))\big\}-1\Big)\to 0
\end{aligned}
$$

as $n\to\infty$.

Lemma A.3 *Under Assumptions 1-5, the GLS-style Random Forest error $E_i^* := Y_i-\hat m_{\hat\Sigma}(X_i)$ is L_2-consistent to the true underlying Random Forest error $\tilde E_i^* := Y_i-m(X_i)$, that is, $\lim_{n\to\infty}E\int(E_i^*-\tilde E_i^*)dX_i=0$, and E_i^* is asymptotically independent on the out-of-bag weight $v_i(x)$ conditioning on X_i, for $i=1,\cdots,n$, respectively.*

[**Proof of Lemma A.3**]

If we consider the RF-GLS predictor $\hat m_{\hat\Sigma}$ is a $\mathcal{L}_2$-consistent estimator of the true underlying function m under assumptions 1-5 [21], we have

$$E_i^*-\tilde E_i^*=\big(Y_i-\hat m_{\hat\Sigma}(X_i)\big)-(Y_i-m(X_i))=m(X_i)-\hat m_{\hat\Sigma}(X_i)$$

This implies

$$\lim_{n\to\infty}E\int(E_i^*-\tilde E_i^*)dX_i=\lim_{n\to\infty}E\int(m(X_i)-\hat m_{\hat\Sigma}(X_i))dX_i=0$$

The out-of-bag weights $\{v_i(x)\}$ are built on a randomly chosen subset of bootstrapped samples that are not containing Y_i, so that conditioning on X_i yields sufficient independence on the true underlying random forest error $\tilde{E}_i^*$, i.e. $(v_i \perp\!\!\!\perp \tilde{E}_i^*)|X_i$. However, the weight $\{v_i(x)\}$ is not exactly independent on the estimated random forest error E_i^* conditioning on X_i, i.e. $(v_i \not\perp\!\!\!\perp E_i^*)|X_i$, since the same observations are used for the out-of-bag weights and for the tree construction of the predictor $\hat{m}_{\hat{\Sigma}}$. Therefore, we suggest an asymptotic independence between $v_i(x)$ and E_i^* by considering the above $\mathcal{L}_2$-consistency that implies $E_i^* \xrightarrow{p} \tilde{E}_i^*$, for $i = 1, \cdots, n$. Once we assume sufficiently large training sample size n, we can replace E_i^* by $(\tilde{E}_i^* + o_p(1))$ for the rest of our proof.

Lemma A.4 *Under Assumptions 10-14, as $n \to \infty$,*

$$\sum_{i=1}^{n} Var\Big[v_i(x)\Big(K_{h_n}(E_i^* - e) - f_E(e|X_i)\Big)\Big] \to 0$$

[**Proof of Lemma A.4**]

We decompose the sum of variances via the law of total variance:

$$\begin{aligned}
&\sum_{i=1}^{n} Var\Big[v_i(x)\Big(K_{h_n}(E_i^* - e) - f_E(e|X_i)\Big)\Big] \\
&\quad= \sum_{i=1}^{n} Var\left\{E\Big(v_i(x)\Big(K_{h_n}(E_i^* - e) - f_E(e|X_i)\Big)\Big|\Omega \setminus \{Y_i\}\Big)\right\} \\
&\quad+ \sum_{i=1}^{n} E\left\{Var\Big(v_i(x)\Big(K_{h_n}(E_i^* - e) - f_E(e|X_i)\Big)\Big|\Omega \setminus \{Y_i\}\Big)\right\}
\end{aligned}$$

The first term of the right side converges to zero:

$$\begin{aligned}
&\sum_{i=1}^{n} Var\left\{E\Big(v_i(x)\Big(K_{h_n}(E_i^* - e) - f_E(e|X_i)\Big)\Big|\Omega \setminus \{Y_i\}\Big)\right\} \\
&= \sum_{i=1}^{n} Var\left\{v_i(x)E\Big(\big(K_{h_n}(E_i^* - e) - f_E(e|X_i)\big)\Big|\Omega \setminus \{Y_i\}\Big)\right\} \\
&= \sum_{i=1}^{n} Var\left\{v_i(x)E\Big(\big(K_{h_n}(E_i^* - e) - f_E(e|X_i)\big)\Big|X_i\Big)\right\} \\
&= \sum_{i=1}^{n} Var\left\{v_i(x) \cdot \frac{1}{2}\mu_{K,2}\ddot{f}_E(\epsilon|x)h^2\right\} \to 0,
\end{aligned}$$

where the last equality follows by (8). The second term of the right side converges to zero:

$$
\begin{aligned}
0 &\le \sum_{i=1}^{n} E\left\{Var\Big(v_i(x)\Big(K_{h_n}(E_i^*-e)-f_E(e|X_i)\Big)\Big|\Omega\setminus\{Y_i\}\Big)\right\} \\
&= \sum_{i=1}^{n} E\left\{v_i^2(x)Var\Big(\big(K_{h_n}(E_i^*-e)-f_E(e|X_i)\big)\Big|\Omega\setminus\{Y_i\}\Big)\right\} \\
&= \sum_{i=1}^{n} E\left\{v_i^2(x)Var\Big(\big(K_{h_n}(E_i^*-e)-f_E(e|X_i)\big)\Big|X_i\Big)\right\} \\
&= \sum_{i=1}^{n} E\left\{v_i^2(x)Var\Big(K_{h_n}(E_i^*-e)\Big|X_i\Big)\right\} \le \sum_{i=1}^{n} E\left\{v_i^2(x)\right\} \\
&\le M_n \sum_{i=1}^{n} E\left\{v_i(x)\right\} \xrightarrow{n\to\infty} 0
\end{aligned}
$$

Lemma A.5 *Under Assumptions 10-14, as* $n\to\infty$*,*

$$
\sum_{i\ne j} Cov\Big[v_i(x)\Big(K_{h_n}(E_i^*-e)-f_E(e|X_i)\Big), v_j(x)\Big(K_{h_n}(E_j^*-e)-f_E(e|X_j)\Big)\Big] \to 0
$$

[**Proof of Lemma A.5**]

$$
\begin{aligned}
&\sum_{i\ne j} Cov\Big[v_i(x)\Big(K_{h_n}(E_i^*-e)-f_E(e|X_i)\Big), v_j(x)\Big(K_{h_n}(E_j^*-e)-f_E(e|X_j)\Big)\Big] \\
&= \sum_{i\ne j} E\Big\{v_i(x)v_j(x)\big(K_{h_n}(E_i^*-e)-f_E(e|X_i)\big)\big(K_{h_n}(E_j^*-e)-f_E(e|X_j)\big)\Big\} \\
&- \sum_{i\ne j} E\Big\{v_i(x)\big(K_{h_n}(E_i^*-e)-f_E(e|X_i)\big)\Big\}E\Big\{\big(v_j(x)K_{h_n}(E_j^*-e)-f_E(e|X_j)\big)\Big\} \\
&= \sum_{i\ne j} E\Big\{v_i(x)v_j(x)\big(K_{h_n}(E_i^*-e)-f_E(e|X_i)\big)\big(K_{h_n}(E_j^*-e)-f_E(e|X_j)\big)\Big\}
\end{aligned}
$$

where the last equality follows by (9). Note that

$$
\begin{aligned}
&E\big\{\big(K_{h_n}(E_i^*-e)-f_E(e|X_i)\big)\big(K_{h_n}(E_j^*-e)-f_E(e|X_j)\big)\Big|X_i,X_j\big\} \\
&\quad= E\big\{K_{h_n}(E_i^*-e)K_{h_n}(E_j^*-e)\Big|X_i,X_j\big\} - E\big\{K_{h_n}(E_i^*-e)f_E(e|X_j)\Big|X_i,X_j\big\} \\
&\quad- E\big\{f_E(e|X_i)K_{h_n}(E_j^*-e)\Big|X_i,X_j\big\} + E\big\{f_E(e|X_i)f_E(e|X_j)\Big|X_i,X_j\big\} \\
&\quad= E\big\{K_{h_n}(E_i^*-e)K_{h_n}(E_j^*-e)-f_E(e|X_i)f_E(e|X_j)\Big|X_i,X_j\big\}
\end{aligned}
$$

This is because the equation (8) implies

$$
E\big\{K_{h_n}(E_i^*-e)f_E(e|X_j)\Big|X_i,X_j\big\} = f_E(e|X_i)f_E(e|X_j)
$$

$$
E\big\{f_E(e|X_i)K_{h_n}(E_j^*-e)\Big|X_i,X_j\big\} = f_E(e|X_i)f_E(e|X_j)
$$

Then, conditioning on (X_i, X_j), we apply the second Taylor's expansion to the above kernel products:

$$
\begin{aligned}
&E\big\{K_{h_n}(E_i^* - e)K_{h_n}(E_j^* - e)\Big|X_i, X_j\big\} \\
&= h^{-2}\int\int K((u-e)/h)K((v-e)/h)f_E(u|X_i)f_E(v|X_j)dudv \\
&= \int\int K(u)K(v)f_E(e+uh|X_i)f_E(e+vh|X_j)dudv \\
&= \int\int K(u)K(v)\Big[f_E(e+uh|X_i)f_E(e+vh|X_j) + \dot{f}_E(e+uh|X_i)f_E(e+vh|X_j)(uh) \\
&+ f_E(e+uh|X_i)\dot{f}_E(e+vh|X_j)(vh) + \frac{1}{2}\big\{\ddot{f}_E(e+uh|X_i)f_E(e+vh|X_j)(uh)^2 \\
&+ 2\dot{f}_E(e+uh|X_i)\dot{f}_E(e+vh|X_j)(uvh^2) + f_E(e+uh|X_i)\ddot{f}_E(e+vh|X_j)(vh)^2\big\}\Big]dudv \\
&= f_E(e|X_i)f_E(e|X_j) + \frac{1}{2}\{\ddot{f}_E(e|X_i)f_E(e|X_j) + f_E(e|X_i)\ddot{f}_E(e|X_j)\}\mu_{K,2}h^2(1+o_p(1))
\end{aligned}
$$

This proves $E\big\{K_{h_n}(E_i^* - e)K_{h_n}(E_j^* - e) - f_E(e|X_i)f_E(e|X_j)\Big|X_i, X_j\big\} = o_p(1)$, for a sufficiently small choice of $h > 0$. Therefore, we conclude

$$
\begin{aligned}
&\sum_{i\neq j} Cov\Big[v_i(x)\Big(K_{h_n}(E_i^* - e) - f_E(e|X_i)\Big), v_j(x)\Big(K_{h_n}(E_j^* - e) - f_E(e|X_j)\Big)\Big] \\
&= \sum_{i\neq j} E\Big\{v_i(x)v_j(x)\big(K_{h_n}(E_i^* - e) - f_E(e|X_i)\big)\big(K_{h_n}(E_j^* - e) - f_E(e|X_j)\big)\Big\} \\
&= \sum_{i\neq j} E\Big[E\Big\{v_i(x)v_j(x)\big(K_{h_n}(E_i^* - e) - f_E(e|X_i)\big)\big(K_{h_n}(E_j^* - e) - f_E(e|X_j)\big)\Big\}\big|X_i, X_j\Big] \\
&= \sum_{i\neq j} E\Big[E\Big\{v_i(x)v_j(x)\big|X_i, X_j\Big\}E\Big\{\big(K_{h_n}(E_i^* - e) - f_E(e|X_i)\big)\big(K_{h_n}(E_j^* - e) - f_E(e|X_j)\big)\big|X_i, X_j\Big\}\Big] \\
&= \sum_{i\neq j} E\Big[E\Big\{v_i(x)v_j(x)\big|X_i, X_j\Big\}E\big\{K_{h_n}(E_i^* - e)K_{h_n}(E_j^* - e) - f_E(e|X_i)f_E(e|X_j)\Big|X_i, X_j\big\}\Big] \\
&= \sum_{i\neq j} E\Big[E\Big\{v_i(x)v_j(x)\big|X_i, X_j\Big\}\cdot o_p(1)\Big] \to 0
\end{aligned}
$$

as $n \to \infty$.

A.2 Supplementary Simulation Results

Table 2: MARGINAL PERFORMANCES UNDER A FEW SPATIAL ERROR Average coverage rates of 90% prediction intervals, widths, and interval score across 100 simulations constructed by Quantile Regression Forests (QRF), split conformal prediction (SC), the unweighted out-of-bag method (OOB), Local Spatial Conformal Prediction (LSCP), the weighted out-of-bag method (OOBW), and the generalized out-of-bag kernel method (OOBGK). Bold quantities represents the case showing the lowest values in interval length or average interval score, respectively, among the candidates.

	LINEAR			STEP			FRIEDMAN		
	CPR	LEN	AIS90	CPR	LEN	AIS90	CPR	LEN	AIS90
QRF	0.94	6.92	7.77	0.95	7.96	8.65	0.92	7.47	8.83
SC	0.92	6.37	7.48	0.90	7.98	9.80	0.91	7.18	9.01
OOB	0.91	5.95	**7.20**	0.91	6.37	**8.00**	0.91	6.48	**8.02**
LSCP	0.91	5.92	**7.20**	0.90	6.33	8.04	0.91	6.51	**8.02**
OOBW	0.90	5.92	7.43	0.89	6.18	8.24	0.90	6.46	8.19
OOBGK	0.87	**5.47**	7.52	0.87	**5.86**	8.19	0.87	**6.04**	8.31
	SINUSOIDAL(HOMO)			SINUSOIDAL(HEAVY)			SINUSOIDAL(HETERO)		
	CPR	LEN	AIS90	CPR	LEN	AIS90	CPR	LEN	AIS90
QRF	0.91	6.48	7.80	0.92	4.48	5.97	0.91	9.09	11.24
SC	0.90	6.47	7.96	0.91	4.52	6.08	0.91	9.21	11.58
OOB	0.90	5.89	**7.28**	0.90	3.71	**5.45**	0.90	8.46	**10.93**
LSCP	0.90	5.89	7.29	0.91	3.69	5.47	0.90	8.49	10.94
OOBW	0.90	5.92	7.47	0.88	3.78	5.71	0.90	8.61	10.93
OOBGK	0.85	**5.32**	7.64	0.87	**3.44**	5.62	0.86	**7.71**	11.28

Figure 3: SENSITIVITY TO A MEASUREMENT ERROR DISTRIBUTION The top left panel represents boxplots of estimated marginal coverage probabilities with 90% confidence level under homoscadestic (HOMO), heavy-tailed (HEAVY), and heteroscadestic (HETERO) measurement error distribution, respectively, as illustrated in section 3. The rest of the panels are scatterplots of estimated versus nominal miscoverage rate ranged from 0.02 to 0.20 assuming HOMO, HEAVY, HETERO measurement error distribution, respectively. For all the panels, blue color represents OOBGK method, green represents OOBW, and red represents LSCP.

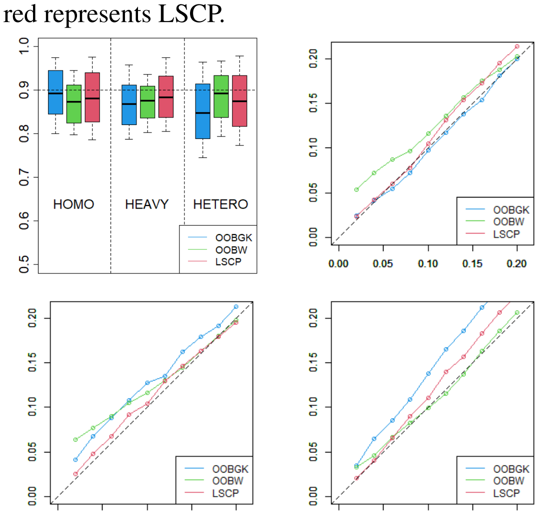

Figure 4: SENSITIVITY TO A TUNING PARAMETERS Based on the simulation settings in section 3, the first row of panels represent the relationship between tuning parameters and the estimated coverage rate on average, and the second row of panels represent the relationship between tuning parameters and the estimated prediction interval width on average. The first column of panels come from the kernel bandwidth(h), the second column come from the number of trees(n_{tree}), the third column come from the number of predictors for node-splitting (m_{try}), and the fourth colmn come from the minimum number of final nodes ($nodesize$)

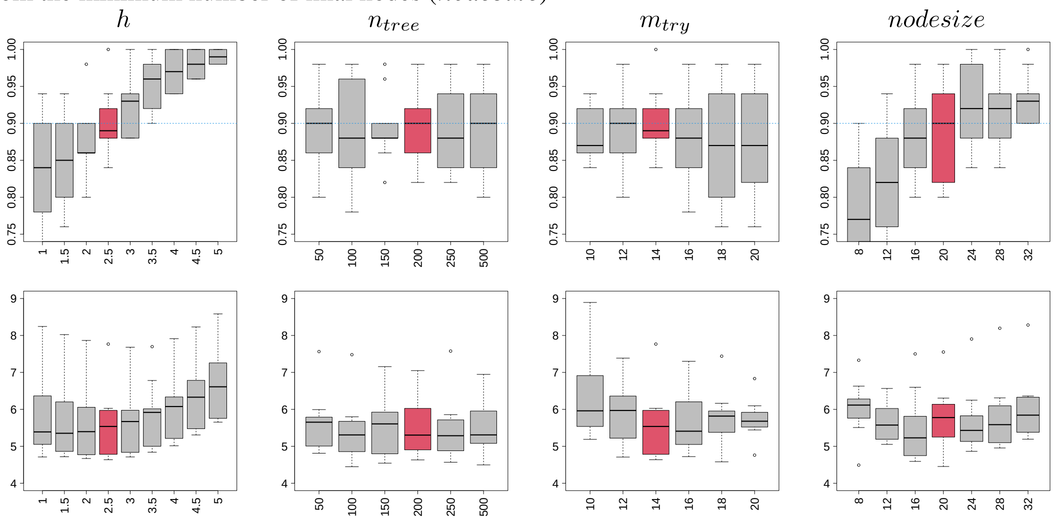

Table 3: SENSITIVITY TO TRUE UNDERLYING COVARIANCE Based on the simulation settings in section 3, we assume the Matérn covariance but three different cases of the true underlying parameters, denoted as Bumpy Matérn ($\nu = 0.1$), Exponential ($\nu = 0.5$), and Smooth Matérn ($\nu = 2.0$). For each cases, we provide average coverage rates of 90% prediction intervals, widths, and interval score across 100 simulations constructed by Quantile Regression Forests (QRF), split conformal prediction (SC), the unweighted out-of-bag method (OOB), Local Spatial Conformal Prediction (LSCP), the weighted out-of-bag method (OOBW), and the generalized out-of-bag kernel method (OOBGK). OOBGK(oracle) represents the proposed method using the true specified spatial covariance. OOBGK(ν) represents the proposed method but using the arbitrary specified parameters ν. Bold quantities represents the case showing the lowest values in interval length or average interval score, respectively, among the candidates.

	Bumpy Matérn ($\nu = 0.1$)			Exponential ($\nu = 0.5$)			Smooth Matérn ($\nu = 2.0$)		
	CPR	LEN	AIS90	CPR	LEN	AIS90	CPR	LEN	AIS90
QRF	0.90	7.09	8.77	0.94	6.60	7.37	0.95	6.00	6.42
SC	0.90	7.34	9.11	0.92	6.16	7.59	0.92	5.47	6.64
OOB	0.88	6.10	7.75	0.90	5.19	6.28	0.92	4.65	5.59
LSCP	0.91	6.06	7.42	0.91	4.98	6.06	0.92	4.55	5.45
OOBW	0.88	6.08	7.79	0.90	5.20	6.39	0.91	4.58	5.45
OOBGK($\nu = 0.1$)	0.91	7.98	9.21	0.90	6.12	7.70	0.93	5.74	7.01
OOBGK($\nu = 2.0$)	0.94	8.07	9.18	0.92	6.10	7.11	0.95	5.65	6.34
OOBGK(oracle)	0.92	6.32	**7.14**	0.91	4.89	5.71	0.94	4.71	5.44
OOBGK	0.86	**5.86**	7.56	0.89	**4.49**	**5.63**	0.90	**4.24**	**5.35**

Table 4: MARGINAL PERFORMANCES UNDER DOMINANT SPATIAL ERROR WITH DIFFERENT LEVEL OF THE NOMINAL MISCOVERAGE RATE For the different set of nominal miscoverage rates $\alpha \in \{0.05, 0.10, 0.20\}$, we provide average coverage rates of 90% prediction intervals, widths, and interval score across 100 simulations constructed by Quantile Regression Forests (QRF), split conformal prediction (SC), the unweighted out-of-bag method (OOB), Local Spatial Conformal Prediction (LSCP), the weighted out-of-bag method (OOBW), and the generalized out-of-bag kernel method (OOBGK). Bold quantities represents the case showing the lowest values in interval length or average interval score, respectively, among the candidates.

α		LINEAR			SINUSOIDAL			STEP			FRIEDMAN		
	Method	CPR	LEN	AIS90	CPR	LEN	AIS90	CPR	LEN	AIS90	CPR	LEN	AIS90
0.2	QRF	0.91	4.59	5.07	0.85	3.97	4.83	0.92	5.52	5.99	0.86	5.02	6.07
	SC	0.85	4.08	4.98	0.82	4.00	5.17	0.82	5.88	7.73	0.80	4.76	6.69
	OOB	0.83	3.50	4.54	0.80	3.35	4.53	0.81	3.87	5.41	0.80	4.02	5.57
	LSCP	0.83	3.24	4.18	0.80	3.05	4.13	0.81	3.60	5.07	0.80	3.80	5.24
	OOBW	0.81	3.42	4.54	0.80	3.28	4.47	0.78	3.63	5.36	0.80	3.98	5.37
	OOBGK	0.81	**3.00**	**4.06**	0.78	**2.89**	**4.01**	0.79	**3.16**	**4.83**	0.77	**3.56**	**5.05**
0.1	QRF	0.96	5.73	6.14	0.93	5.04	5.77	0.97	6.85	7.21	0.93	6.32	7.19
	SC	0.93	4.99	5.73	0.91	4.98	6.05	0.91	6.76	8.20	0.91	5.92	7.32
	OOB	0.91	4.41	5.30	0.90	4.18	5.25	0.90	4.85	6.52	0.90	5.05	6.35
	LSCP	0.91	4.11	4.89	0.90	3.86	4.79	0.90	4.49	6.11	0.90	4.77	5.95
	OOBW	0.90	4.29	5.35	0.89	4.12	5.31	0.87	4.63	6.70	0.90	4.98	6.27
	OOBGK	0.91	**3.86**	**4.73**	0.89	**3.74**	**4.78**	0.88	**4.24**	**6.04**	0.88	**4.64**	**5.93**
0.05	QRF	0.99	7.19	7.39	0.97	6.36	6.93	1.00	9.46	9.55	0.97	8.02	8.84
	SC	0.97	6.23	6.67	0.96	6.03	6.81	0.95	8.93	10.19	0.95	7.61	9.11
	OOB	0.96	5.31	6.02	0.95	5.03	6.00	0.95	6.04	7.82	0.95	6.13	7.46
	LSCP	0.96	4.97	5.60	0.95	4.61	**5.43**	0.95	5.61	7.41	0.95	5.88	7.10
	OOBW	0.95	5.19	6.26	0.94	4.93	6.19	0.93	5.72	8.17	0.94	6.06	7.36
	OOBGK	0.95	**4.60**	**5.45**	0.94	**4.50**	5.49	0.93	**5.10**	**7.35**	0.93	**5.50**	**6.95**

Figure 5: SENSITIVITY TO KERNEL BANDWIDTH SELECTION Based on the simulation settings in section 3, the first row of panels represent the relationship between tuning parameters and the estimated coverage rate on average, and the second row of panels represent the relationship between tuning parameters and the estimated prediction interval width on average. Each column of panels com form the linear, sinusoidal, step, and Friedman mean function, respectively.