Supplementary materials

Bayesian kernel machine meta-regression: an application in environmental epidemiology

List of contents

Section S1. Derivation of MCMC algorithm.

(1) Full conditional distribution of $\mathbf{h}$

The likelihood $p(\boldsymbol{\theta}|\mathbf{h},\sigma^2)$ is

$$p(\boldsymbol{\theta}|\mathbf{h},\sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{\theta}-\mathbf{h})^\top(\boldsymbol{\theta}-\mathbf{h})\right)$$

The prior $p(\mathbf{h}|\tau^2,\rho)$ is

$$p(\mathbf{h}|\tau^2,\rho) \propto (\tau^2)^{-\frac{n}{2}}\left|\mathbf{K}_\rho\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\tau^2}\mathbf{h}^\top\mathbf{K}_\rho^{-1}\mathbf{h}\right)$$

By combining the exponential terms, we obtain

$$p(\mathbf{h}|\boldsymbol{\theta},\sigma^2,\tau^2,\rho) \propto \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{\theta}-\mathbf{h})^\top(\boldsymbol{\theta}-\mathbf{h}) - \frac{1}{2\tau^2}\mathbf{h}^\top\mathbf{K}_\rho^{-1}\mathbf{h}\right)$$

Then

$$\mathbf{h}|\boldsymbol{\theta},\sigma^2,\tau^2,\rho \sim \mathrm{MVN}(\boldsymbol{\mu}_\mathbf{h},\boldsymbol{\Sigma}_\mathbf{h})$$

where

$$\boldsymbol{\Sigma}_\mathbf{h} = \left(\frac{1}{\sigma^2}\boldsymbol{I}_n + \frac{1}{\tau^2}\mathbf{K}_\rho^{-1}\right)^{-1}$$

$$\boldsymbol{\mu}_\mathbf{h} = \boldsymbol{\Sigma}_\mathbf{h}\left(\frac{1}{\sigma^2}\boldsymbol{\theta}\right)$$

(2) Full conditional distribution of $\boldsymbol{\theta}$

The likelihood $p(\hat{\theta}_i|\theta_i,\hat{s}_i^2)$ for $i = 1,\cdots,n$ is

$$p\left(\hat{\theta}_i\middle|\theta_i,\hat{s}_i^2\right) \propto \exp\left(-\frac{1}{2\hat{s}_i^2}\left(\hat{\theta}_i-\theta_i\right)^2\right)$$

The prior $p(\theta_i|h_i,\sigma^2)$ is

$$p(\theta_i|h_i,\sigma^2) \propto \exp\left(-\frac{1}{2\sigma^2}(\theta_i-h_i)^2\right)$$

By combining the exponential terms, we obtain

$$p\left(\theta_i\middle|\hat{\theta}_i,h_i,\sigma^2,\hat{s}_i^2\right) \propto \exp\left(-\frac{1}{2}\left(\left(\frac{1}{\hat{s}_i^2}+\frac{1}{\sigma^2}\right)\theta_i^2 - 2\left(\frac{\hat{\theta}_i}{\hat{s}_i^2}+\frac{h_i}{\sigma^2}\right)\theta_i\right)\right)$$

Then

$$\theta_i|\hat{\theta}_i,h_i,\sigma^2,\hat{s}_i^2 \sim \mathrm{N}\left(\mu_{\theta_i},\Sigma_{\theta_i}\right)$$

where

$$\Sigma_{\theta_i} = \left(\frac{1}{\hat{s}_i^2} + \frac{1}{\sigma^2}\right)^{-1}$$

$$\mu_{\theta_i} = \Sigma_{\theta_i}\left(\frac{\hat{\theta}_i}{\hat{s}_i^2} + \frac{h_i}{\sigma^2}\right)$$

(3) Full conditional distribution of σ^2

The likelihood $p(\boldsymbol{\theta}|\mathbf{h}, \sigma^2)$ is

$$p(\boldsymbol{\theta}|\mathbf{h}, \sigma^2) \propto (\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(\theta_i - h_i)^2\right)$$

The prior $p(\sigma^2)$ is

$$p(\sigma^2) \propto (\sigma^2)^{-(a_\sigma+1)} \exp\left(-\frac{b_\sigma}{\sigma^2}\right)$$

We obtain

$$p(\sigma^2|\mathbf{h}, \boldsymbol{\theta}) \propto (\sigma^2)^{-\left(a_\sigma+1+\frac{n}{2}\right)} \exp\left(-\frac{1}{\sigma^2}\left(b_\sigma + \frac{1}{2}\sum_{i=1}^{n}(\theta_i - h_i)^2\right)\right)$$

Then

$$\sigma^2|\mathbf{h}, \boldsymbol{\theta} \sim \mathrm{IG}\left(a_\sigma + \frac{n}{2}, b_\sigma + \frac{1}{2}\sum_{i=1}^{n}(\theta_i - h_i)^2\right)$$

(4) Full conditional distribution of τ^2

The likelihood $p(\mathbf{h}|\tau^2, \rho)$ is

$$p(\mathbf{h}|\tau^2, \rho) \propto (\tau^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\tau^2}\mathbf{h}^\top \mathbf{K}_\rho^{-1}\mathbf{h}\right)$$

The prior $p(\tau^2)$ is

$$p(\tau^2) \propto (\tau^2)^{-(a_\tau+1)} \exp\left(-\frac{b_\tau}{\tau^2}\right)$$

We obtain

$$p(\tau^2|\mathbf{h}, \rho) \propto (\tau^2)^{-\left(a_\tau+1+\frac{n}{2}\right)} \exp\left(-\frac{1}{\tau^2}\left(b_\tau + \frac{1}{2}\mathbf{h}^\top \mathbf{K}_\rho^{-1}\mathbf{h}\right)\right)$$

Then

$$\tau^2|\mathbf{h}, \rho \sim \mathrm{IG}\left(a_\tau + \frac{n}{2}, b_\tau + \frac{1}{2}\mathbf{h}^\top \mathbf{K}_\rho^{-1}\mathbf{h}\right)$$

(5) MH algorithm for ρ

If we sample $\rho^* \sim N(\rho^{(s)}, \delta_\rho^2)$ directly, the proposal ρ^* could be negative, especially when $\rho^{(s)}$ is close to zero. However, a negative value of ρ^* is undefined in BKMMR. By transforming to $\eta = \log(\rho)$, the parameter space changes from $(0, \infty)$ for ρ to $(-\infty, \infty)$ for η. Then we can sample η^* from the proposal distribution $N\left(\eta^{(s)}, \delta_\eta^2\right)$. If we transform back to $\rho^* = \exp(\eta^*)$, it ensures that ρ^* is always positive.

Let $\pi(\rho)$ denote our original target posterior distribution for ρ. It can be expressed as:

$$\begin{aligned} \pi(\rho) &= p(\rho|\mathbf{h}, \tau^2) \\ &\propto p(\mathbf{h}|\tau^2, \rho)p(\rho) \end{aligned}$$

Let $g(\eta)$ denote the density for η. Using the change of variables formula:

$$g(\eta) = \pi\big(\rho(\eta)\big)\left|\frac{d\rho}{d\eta}\right|$$

where $\left|\frac{d\rho}{d\eta}\right|$ is the Jacobian. Since $\rho = \exp(\eta)$, the Jacobian is $\left|\frac{d\rho}{d\eta}\right| = \rho$. Then

$$g(\eta) = \pi(\rho)\rho$$

Let $q\left(\eta^*\middle|\eta^{(s)}\right)$ be the proposal distribution of η. Since the proposal distribution is symmetric, the proposal ratio cancels out, and the acceptance ratio r is calculated as:

$$\begin{aligned} r &= \frac{g(\eta^*)}{g(\eta^{(s)})} \times \frac{q\left(\eta^{(s)}\middle|\eta^*\right)}{q\left(\eta^*\middle|\eta^{(s)}\right)} \\ &= \frac{g(\eta^*)}{g(\eta^{(s)})} \end{aligned}$$

Substituting our expression for $g(\eta)$, r is expressed as:

$$\begin{aligned} r &= \frac{\pi(\rho^*) \times \rho^*}{\pi(\rho^{(s)}) \times \rho^{(s)}} \\ &= \frac{p(\mathbf{h}|\tau^2, \rho^*)p(\rho^*)}{p\left(\mathbf{h}\middle|\tau^2, \rho^{(s)}\right)p(\rho^{(s)})} \times \frac{\rho^*}{\rho^{(s)}} \end{aligned}$$

Section S2. Derivation of posterior predictive distribution.

The conditional distribution for $\begin{pmatrix}\mathbf{h}_{obs}\\ \mathbf{h}_{mis}\end{pmatrix}$ (or $\begin{pmatrix}\mathbf{h}_{loc}\\ \mathbf{h}_{sub}\end{pmatrix}$) is derived as follows. Let $\mathbf{h} = \left(\mathbf{h}_{obs}^{\top}, \mathbf{h}_{mis}^{\top}\right)^{\top}$ be jointly multivariate normal where

$$\begin{aligned}
\begin{pmatrix}\mathbf{h}_{obs}\\ \mathbf{h}_{mis}\end{pmatrix} | \tau^2, \rho &\sim \mathrm{MVN}\left(\begin{pmatrix}\boldsymbol{\mu}_{obs}\\ \boldsymbol{\mu}_{mis}\end{pmatrix}, \begin{pmatrix}\tau^2\mathbf{K}_{obs,obs} & \tau^2\mathbf{K}_{obs,mis}\\ \tau^2\mathbf{K}_{mis,obs} & \tau^2\mathbf{K}_{mis,mis}\end{pmatrix}\right)\\
&= \mathrm{MVN}\left(\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\tau^2\mathbf{K}_{obs,obs} & \tau^2\mathbf{K}_{obs,mis}\\ \tau^2\mathbf{K}_{mis,obs} & \tau^2\mathbf{K}_{mis,mis}\end{pmatrix}\right)
\end{aligned}$$

Using the properties of the multivariate normal distribution, the conditional distribution of $\mathbf{h}_{mis}$ given $\mathbf{h}_{obs}$, τ^2, and ρ follows $\mathrm{MVN}\left(\boldsymbol{\mu}_{pred}, \boldsymbol{\Sigma}_{pred}\right)$ where

$$\begin{aligned}
\boldsymbol{\mu}_{pred} &= \boldsymbol{\mu}_{mis} + \left(\tau^2\mathbf{K}_{mis,obs}\right)\left(\tau^2\mathbf{K}_{obs,obs}\right)^{-1}\left(\mathbf{h}_{obs} - \boldsymbol{\mu}_{obs}\right)\\
&= \mathbf{0} + \left(\tau^2\mathbf{K}_{mis,obs}\right)\left(\tau^2\mathbf{K}_{obs,obs}\right)^{-1}\left(\mathbf{h}_{obs} - \mathbf{0}\right)\\
&= \tau^2\mathbf{K}_{mis,obs}\frac{1}{\tau^2}\left(\mathbf{K}_{obs,obs}\right)^{-1}\mathbf{h}_{obs}\\
&= \mathbf{K}_{mis,obs}\left(\mathbf{K}_{obs,obs}\right)^{-1}\mathbf{h}_{obs}
\end{aligned}$$

and

$$\begin{aligned}
\boldsymbol{\Sigma}_{pred} &= \tau^2\mathbf{K}_{mis,mis} - \left(\tau^2\mathbf{K}_{mis,obs}\right)\left(\tau^2\mathbf{K}_{obs,obs}\right)^{-1}\left(\tau^2\mathbf{K}_{obs,mis}\right)\\
&= \tau^2\mathbf{K}_{mis,mis} - \tau^2\mathbf{K}_{mis,obs}\left(\mathbf{K}_{obs,obs}\right)^{-1}\mathbf{K}_{obs,mis}\\
&= \tau^2\left(\mathbf{K}_{mis,mis} - \mathbf{K}_{mis,obs}\left(\mathbf{K}_{obs,obs}\right)^{-1}\mathbf{K}_{obs,mis}\right)
\end{aligned}$$

Section S3. Convergence diagnostics for MCMC sampling in BKMMR.

Convergence of the MCMC algorithm for the BKMMR model is assessed in both the simulation study and data application. Convergence is evaluated using the Gelman-Rubin statistic ($\hat{R}$), effective sample size (ESS), and trace plots.[1]

For each monitored parameter ψ, suppose that m independent chains are run and n posterior samples are retained from each chain. Let $\bar{\psi}_j$ denote the posterior mean from chain j, $\bar{\psi}$ denote the posterior mean across chains, and v_j^2 denote the within-chain sample variances of chain j. The within-chain variance W and between-chain variance B are defined as:

$$W = \frac{1}{m}\sum_{j=1}^{m} v_j^2 \, , B = \frac{n}{m-1}\sum_{j=1}^{m}\left(\bar{\psi}_j - \bar{\psi}\right)^2$$

The estimated marginal posterior variance is computed as:

$$\hat{V} = \frac{n-1}{n}W + \frac{1}{n}B$$

Finally, the statistic $\hat{R}$ is defined as:[1]

$$\hat{R} = \sqrt{\frac{\hat{V}}{W}}$$

Values of $\hat{R}$ close to 1 indicate that the between-chain variability is similar to the within-chain variability, suggesting that the chains converge to a common posterior distribution. In general, values of $\hat{R} < 1.1$ are considered to indicate satisfactory convergence.

ESS is used to quantify the amount of independent information contained in the autocorrelated MCMC samples. For each parameter ψ, the ESS is defined as:[1]

$$\text{ESS} = \frac{N}{1 + 2\sum_{k=1}^{\infty}\rho_k}$$

where N is the total number of posterior samples, and ρ_k is the autocorrelation at lag k. A larger ESS indicates lower autocorrelation and more independent information in the posterior samples. In general, ESS greater than 400 is often considered adequate for stable posterior inference.

For both the simulation study and data application, convergence diagnostics are computed for σ^2, τ^2, ρ, all location-specific parameters θ_i, and all latent function values h_i.

In the simulation study, the assessment is conducted for one representative setting: Scenario C under the estimation setting using the first generated dataset. Three independent MCMC chains are run with different initial values for the key variance parameters. The numerical convergence diagnostics for the covariance parameters are summarized in Table S1.The Gelman-Rubin statistics range from 1.002 to 1.007 across the monitored covariance parameters, and the ESS values are sufficiently large relative to the total number of posterior samples. For the 100 θ_i parameters, the maximum $\hat{R}$ is 1.005 and the minimum ESS is 2736.77. For the 100 h_i values, the maximum $\hat{R}$ is 1.003 and the minimum ESS is 2456.50. These results indicate stable convergence for the covariance parameters, location-specific parameters, and latent function values.

In the data application, the assessment is conducted for the estimation analysis. Three independent MCMC chains are run with different initial values for the key variance parameters. The numerical convergence diagnostics for the covariance parameters are summarized in Table S1. The Gelman-Rubin statistics range from 1.000 to 1.017 across the monitored covariance parameters, and the ESS values are sufficiently large relative to the total number of posterior samples. For the 77 θ_i parameters, the maximum $\hat{R}$ is 1.003 and the minimum ESS is 1824.44. For the 77 h_i values, the maximum $\hat{R}$ is 1.003 and the minimum ESS is 1634.63. These results indicate stable convergence for the covariance parameters, location-specific parameters, and latent function values.

Trace plots for the simulation study and data application are presented in Figures S1 and S2, respectively. Each figure shows trace plots from the three independent MCMC chains for σ^2, τ^2, ρ, one representative location-specific parameter θ_i, and one representative latent function value h_i. The trace plots show stable mixing across chains without apparent separation.

Section S4. Sensitivity analysis to prior specifications in BKMMR.

We conduct a sensitivity analysis to evaluate whether the BKMMR results are sensitive to the prior specification. We compare multiple prior specifications by refitting the model under each specification and evaluating model fit using the leave-one-out information criterion (LOOIC).[2] The LOOIC is computed using the loo() function from the loo package in R.[2,3]

For each prior specification, the pointwise log-likelihood is computed for each posterior sample $s = 1, \cdots, S$ and location $i = 1, \cdots, n$ as:

$$l_i^{(s)} = \log p\left(\hat{\theta}_i \middle| \theta_i^{(s)}, \hat{s}_i^2\right)$$

which is evaluated under the Gaussian likelihood as:

$$l_i^{(s)} = -\frac{1}{2}\log\left(2\pi\hat{s}_i^2\right) - \frac{\left(\hat{\theta}_i - \theta_i^{(s)}\right)^2}{2\hat{s}_i^2}$$

where $\theta_i^{(s)}$ is the posterior sample of the location-specific parameter.

The collection of $l_i^{(s)}$ values over posterior samples and locations forms the pointwise log-likelihood matrix used as the input to the loo() function. Using this matrix, Pareto-smoothed importance sampling leave-one-out cross-validation approximates the leave-one-out predictive density for each location, $\log p\left(\hat{\theta}_i \middle| \hat{\theta}_{-i}\right)$, where $\hat{\theta}_{-i}$ denotes the set of first-stage estimates excluding location i. This quantity represents how well the model predicts the first-stage estimate at location i when that location is left out of the model fitting.

The leave-one-out expected log pointwise predictive density is obtained by summing the leave-one-out log predictive densities across all locations:

$$\text{elpd}_{loo} = \sum_{i=1}^{n} \log p\left(\hat{\theta}_i \middle| \hat{\theta}_{-i}\right)$$

Finally, LOOIC is defined as -2 times this quantity:[2]

$$\text{LOOIC} = -2 \times \text{elpd}_{loo} = -2\sum_{i=1}^{n} \log p\left(\hat{\theta}_i \middle| \hat{\theta}_{-i}\right)$$

Lower LOOIC values indicate better expected out-of-sample predictive fit.

The sensitivity analysis focuses on the prior distributions for the Gaussian process covariance parameters τ^2 and ρ. The prior for σ^2 is fixed across all sensitivity analyses. For τ^2 and ρ, we consider the same 3×3 grid of prior specifications in both the simulation study and data application, $a_\tau = b_\tau \in \{0.001, 2, 10\}$ and $a_\rho = b_\rho \in \{0.001, 2, 10\}$, resulting in a total of nine prior specifications.

In the simulation study, the sensitivity analysis is conducted for one representative setting: Scenario C under the estimation setting using the first generated dataset. The prior for σ^2 is fixed as $a_\sigma = 2$ and $b_\sigma = 0.01$. For each of the nine prior specifications, the BKMMR model is refitted and evaluated using LOOIC.

In the data application, the sensitivity analysis is conducted for the estimation analysis. The prior for σ^2 is fixed as $a_\sigma = 2$ and $b_\sigma = 10^{-5}$. For each of the nine prior specifications, the BKMMR model is refitted and evaluated using LOOIC.

The LOOIC values under the nine prior specifications are summarized in Table S2. In the simulation study, the lowest LOOIC is obtained under case (a), with $a_\tau = b_\tau = 0.001$ and $a_\rho = b_\rho = 0.001$, whereas in the data application, the lowest LOOIC is obtained under case (d), with $a_\tau = b_\tau = 0.001$ and $a_\rho = b_\rho = 2$. The LOOIC value for case (a) is -201.016 in the simulation study, whereas the LOOIC value for case (d) is -496.445 in the data application. In the simulation study, the LOOIC values are similar across prior specifications, ranging from -201.016 to -193.845. In the data application, the LOOIC values show a clearer difference across prior specifications, ranging from -496.445 to -468.948, suggesting greater sensitivity to the prior specification than in the simulation study. Based on these results, case (a) is selected as the main prior specification for the simulation study, while case (d) is selected for the data application.

Section S5. Details for implementing the competing methods.

Conventional linear meta-regression (LMR) is fitted using the mixmeta() function from the mixmeta package in statistical software R, with restricted maximum likelihood estimation.[4]

Random forest (RF) is implemented using the ranger() function from the ranger package in statistical software R.[5] To ensure a fair comparison, hyperparameters are selected using five-fold cross-validation in both the simulation study and data application. During cross-validation, model performance is evaluated using the weighted root mean squared error, where the inverse of the estimated variance is used as the sample weight to account for heteroscedasticity in the meta-analysis. In the simulation study, tuning is performed independently for each generated dataset, scenario, and analytical setting (estimation, prediction, and downscaling), while in the data application tuning is performed independently for each analytical setting. The final number of trees is set to 1,000 in the simulation study and 2,000 in the data application, while 500 trees are used during cross-validation to reduce computational burden. The tuning grid includes the number of variables randomly sampled as candidates at each split, with values of 1, 2, and 3, and the minimum terminal node size, with values of 1, 3, and 5.

Extreme gradient boosting (XGBoost) is implemented using the xgb.train() function from the xgboost package in statistical software R.[6] Hyperparameters are selected using five-fold cross-validation based on the weighted root mean squared error, with the inverse of the estimated variance used as the sample weight. In the simulation study, tuning is performed independently for each generated dataset, scenario, and analytical setting, while in the data application tuning is performed separately for each analytical setting. The 95% prediction intervals are constructed using 100 bootstrap iterations. In the simulation study, the tuning grid includes learning rates of 0.01 and 0.05; maximum tree depths of 3 and 5; numbers of boosting rounds of 100, 200, and 300; minimum sums of instance weights required in a child node of 1 and 3; and column subsampling ratios of 0.7 and 1.0. In the data application, the tuning grid includes learning rates of 0.01, 0.05, and 0.1; maximum tree depths of 3 and 5; and numbers of boosting rounds of 100, 200, and 300.

Table S1. Summary of convergence diagnostics for MCMC sampling in BKMMR.

Parameter	$\hat{\boldsymbol{R}}$		**ESS**	
	Simulation study	**Data application**	**Simulation study**	**Data application**
σ^2	1.002	1.000	3000.00	3101.83
τ^2	1.007	1.017	2455.57	1505.10
ρ	1.002	1.001	2728.76	484.19

Note. $\hat{R}$ = Gelman-Rubin statistic; ESS = effective sample size.

Table S2. Comparison of LOOIC values under nine prior specifications.

Panel	Prior hyperparameters		LOOIC[a]	
	$a_\tau = b_\tau$	$a_\rho = b_\rho$	**Simulation study**	**Data application**
(a)	0.001	0.001	**-201.016**	-491.522
(b)	2	0.001	-195.583	-490.051
(c)	10	0.001	-199.537	-489.601
(d)	0.001	2	-198.661	**-496.445**
(e)	2	2	-196.062	-480.874
(f)	10	2	-194.743	-479.832
(g)	0.001	10	-194.237	-481.931
(h)	2	10	-198.265	-473.059
(i)	10	10	-193.845	-468.948

[a] LOOIC = leave-one-out information criterion.

Table S3. Statistics (RMSE, Corr, CP, and AIW) across scenarios for estimation performance in simulation study.

Scenario	Metric	BKMMR	LMR	RF	XGBoost
A	RMSE	0.048	0.048	0.060	0.064
	Corr	0.954	0.954	0.927	0.924
	CP	0.951	0.976	0.968	0.632
	AIW	0.186	0.214	0.409	0.170
B	RMSE	0.052	0.055	0.082	0.057
	Corr	0.972	0.970	0.943	0.968
	CP	0.950	0.992	0.964	0.735
	AIW	0.200	0.298	0.581	0.244
C	RMSE	0.052	0.055	0.102	0.071
	Corr	0.981	0.979	0.943	0.967
	CP	0.946	0.994	0.963	0.694
	AIW	0.198	0.317	0.704	0.311

Note. RMSE = root mean squared error; Corr = Pearson correlation coefficient; CP = coverage probability; AIW = average interval width; BKMMR = Bayesian kernel machine meta-regression; LMR = linear meta-regression; RF = random forest; XGBoost = extreme gradient boosting.

Table S4. Statistics (RMSE, Corr, CP, and AIW) across scenarios for predictive performance at unobserved locations in simulation study.

Scenario	Metric	BKMMR	LMR	RF	XGBoost
A	RMSE	0.106	0.105	0.118	0.120
	Corr	0.746	0.753	0.675	0.659
	CP	0.944	0.424	0.956	0.587
	AIW	0.411	0.130	0.493	0.195
B	RMSE	0.187	0.234	0.175	0.167
	Corr	0.558	0.109	0.615	0.664
	CP	0.947	0.402	0.930	0.636
	AIW	0.703	0.257	0.693	0.292
C	RMSE	0.224	0.267	0.237	0.235
	Corr	0.587	0.325	0.539	0.533
	CP	0.935	0.423	0.926	0.596
	AIW	0.728	0.287	0.855	0.345

Note. RMSE = root mean squared error; Corr = Pearson correlation coefficient; CP = coverage probability; AIW = average interval width; BKMMR = Bayesian kernel machine meta-regression; LMR = linear meta-regression; RF = random forest; XGBoost = extreme gradient boosting.

Table S5. Statistics (RMSE, Corr, CP, and AIW) across scenarios for downscaling performance at sublocations in simulation study.

Scenario	Metric	BKMMR	LMR	RF	XGBoost
A	RMSE	0.106	0.105	0.113	0.114
	Corr	0.794	0.794	0.761	0.755
	CP	0.774	0.264	0.877	0.476
	AIW	0.257	0.069	0.389	0.141
B	RMSE	0.129	0.211	0.139	0.137
	Corr	0.810	0.287	0.780	0.785
	CP	0.875	0.328	0.891	0.543
	AIW	0.410	0.185	0.544	0.225
C	RMSE	0.142	0.260	0.184	0.182
	Corr	0.875	0.469	0.785	0.786
	CP	0.938	0.355	0.938	0.631
	AIW	0.528	0.240	0.863	0.354

Note. RMSE = root mean squared error; Corr = Pearson correlation coefficient; CP = coverage probability; AIW = average interval width; BKMMR = Bayesian kernel machine meta-regression; LMR = linear meta-regression; RF = random forest; XGBoost = extreme gradient boosting.

Table S6. Summary statistics for daily $PM_{2.5}$ concentration and mortality in 77 Si/gun/gu-level locations in data application.

Si/gun/gu code	Name of Si/gun/gu	Name of Si/do	Total number of deaths	Average of daily mean $PM_{2.5}$ ($\mu g/m^3$)	Observed[a]
11010	Jongno-gu	Seoul	6,078	23.50	No
11020	Jung-gu	Seoul	4,946	23.19	Yes
11030	Yongsan-gu	Seoul	8,120	23.39	Yes
11040	Seongdong-gu	Seoul	9,482	23.24	Yes
11050	Gwangjin-gu	Seoul	10,061	22.92	Yes
11060	Dongdaemun-gu	Seoul	13,252	22.63	No
11070	Jungnang-gu	Seoul	14,922	22.42	No
11080	Seongbuk-gu	Seoul	15,374	21.98	Yes
11090	Gangbuk-gu	Seoul	13,410	22.22	Yes
11100	Dobong-gu	Seoul	12,995	21.90	Yes
11110	Nowon-gu	Seoul	19,481	23.21	Yes
11120	Eunpyeong-gu	Seoul	17,363	22.61	Yes
11130	Seodaemun-gu	Seoul	11,165	22.85	Yes
11140	Mapo-gu	Seoul	11,310	24.05	No
11150	Yangcheon-gu	Seoul	13,284	23.74	No
11160	Gangseo-gu	Seoul	18,769	24.67	Yes
11170	Guro-gu	Seoul	13,036	23.93	Yes
11180	Geumcheon-gu	Seoul	8,371	24.25	Yes
11190	Yeongdeungpo-gu	Seoul	11,975	24.64	Yes
11200	Dongjak-gu	Seoul	12,068	23.30	No
11210	Gwanak-gu	Seoul	15,189	23.52	No
11220	Seocho-gu	Seoul	10,152	22.74	Yes
11230	Gangnam-gu	Seoul	12,708	23.07	Yes
11240	Songpa-gu	Seoul	16,031	22.79	Yes
11250	Gangdong-gu	Seoul	13,351	23.41	No

23010	Jung-gu	Incheon	5,009	21.72	Yes
23020	Dong-gu	Incheon	3,672	24.38	Yes
23040	Yeonsu-gu	Incheon	9,092	23.12	Yes
23050	Namdong-gu	Incheon	17,891	24.02	Yes
23060	Bupyeong-gu	Incheon	19,069	23.24	Yes
23070	Gyeyang-gu	Incheon	10,489	24.16	Yes
23080	Seo-gu	Incheon	14,899	22.94	Yes
23090	Michuhol-gu	Incheon	10,358	24.24	No
23310	Ganghwa-gun	Incheon	5,622	22.63	No
23320	Ongjin-gun	Incheon	1,311	21.49	No
31011	Jangan-gu	Gyeonggi-do	9,491	23.06	Yes
31012	Gwonseon-gu	Gyeonggi-do	10,462	24.28	No
31013	Paldal-gu	Gyeonggi-do	7,293	23.73	Yes
31014	Yeongtong-gu	Gyeonggi-do	6,387	23.00	No
31021	Sujeong-gu	Gyeonggi-do	8,719	22.73	Yes
31022	Jungwon-gu	Gyeonggi-do	8,738	23.19	Yes
31023	Bundang-gu	Gyeonggi-do	12,404	22.93	No
31030	Uijeongbu-si	Gyeonggi-do	16,658	24.00	No
31041	Manan-gu	Gyeonggi-do	8,596	23.43	No
31042	Dongan-gu	Gyeonggi-do	8,245	23.48	Yes
31050	Bucheon-si	Gyeonggi-do	23,218	26.38	Yes
31060	Gwangmyeong-si	Gyeonggi-do	9,933	24.56	Yes
31070	Pyeongtaek-si	Gyeonggi-do	17,208	28.33	Yes
31080	Dongducheon-si	Gyeonggi-do	5,404	25.01	Yes
31091	Sangrok-gu	Gyeonggi-do	11,282	24.47	Yes
31092	Danwon-gu	Gyeonggi-do	8,954	24.06	No
31101	Deogyang-gu	Gyeonggi-do	15,720	24.34	Yes
31103	Ilsandong-gu	Gyeonggi-do	9,055	25.28	Yes
31104	Ilsanseo-gu	Gyeonggi-do	8,739	25.99	No
31110	Gwacheon-si	Gyeonggi-do	1,831	22.69	Yes

31120	Guri-si	Gyeonggi-do	5,979	24.58	Yes
31130	Namyangju-si	Gyeonggi-do	22,144	23.51	Yes
31140	Osan-si	Gyeonggi-do	5,456	24.85	Yes
31150	Siheung-si	Gyeonggi-do	11,677	27.02	No
31160	Gunpo-si	Gyeonggi-do	8,083	24.19	Yes
31170	Uiwang-si	Gyeonggi-do	4,771	23.69	Yes
31180	Hanam-si	Gyeonggi-do	6,869	22.48	No
31191	Cheoin-gu	Gyeonggi-do	9,453	26.33	Yes
31192	Giheung-gu	Gyeonggi-do	10,445	24.45	Yes
31193	Suji-gu	Gyeonggi-do	8,117	23.72	Yes
31200	Paju-si	Gyeonggi-do	16,402	26.54	No
31210	Icheon-si	Gyeonggi-do	8,230	27.86	Yes
31220	Anseong-si	Gyeonggi-do	8,703	27.09	No
31230	Gimpo-si	Gyeonggi-do	11,470	26.29	Yes
31240	Hwaseong-si	Gyeonggi-do	17,185	26.51	Yes
31250	Gwangju-si	Gyeonggi-do	10,694	25.79	Yes
31260	Yangju-si	Gyeonggi-do	8,909	25.97	Yes
31270	Pocheon-si	Gyeonggi-do	8,397	22.85	Yes
31280	Yeoju-si	Gyeonggi-do	6,424	26.37	No
31350	Yeoncheon-gun	Gyeonggi-do	3,214	23.10	Yes
31370	Gapyeong-gun	Gyeonggi-do	4,355	21.99	Yes
31380	Yangpyeong-gun	Gyeonggi-do	7,038	23.63	Yes

[a] “Observed” indicates whether the location is included in the observed set for fitting the second-stage model in the prediction analysis.

Table S7. Detailed information for meta-predictors used in data application.

Meta-predictor	**Data source**	**Spatiotemporal resolution**
NDVI	MODIS Terra MOD13Q1 product, accessed through Google Earth Engine	250 m, 16-day
Proportion of single-person households	Korean Statistical Information Service (KOSIS)	Si/gun/gu-level, Eup/myeon/dong-level, 2020
Proportion of the population aged 65 years or older	Korean Statistical Information Service (KOSIS)	Si/gun/gu-level, Eup/myeon/dong-level, 2020
Population density	Seoul Open Data Plaza for Seoul; Korean Statistical Information Service (KOSIS) for Incheon and Gyeonggi-do	Si/gun/gu-level, Eup/myeon/dong-level, 2020

NDVI is derived from the MODIS Terra MOD13Q1 product accessed through Google Earth Engine. Values from May to October 2020 are used to represent the peak greenness season in South Korea. Negative NDVI pixels are masked, and the median NDVI value is calculated for each pixel before averaging within administrative boundaries.[7]

The proportion of single-person households is calculated as the number of single-person households divided by the number of general households, multiplied by 100.

The proportion of the population aged 65 years or older was calculated as the population aged 65 years or older divided by the total population, multiplied by 100.

The population density is calculated as total population divided by land area in km^2.

Table S8. Original-scale summary statistics for standardized meta-predictors used in data application.

Meta-predictor	Original unit	Mean	SD	-1 SD	+1 SD
NDVI	Index	0.508	0.132	0.376	0.640
Proportion of single-person households	%	30.773	6.366	24.407	37.139
Proportion of the population aged 65 years or older	%	15.247	4.238	11.008	19.485
Population density	persons/km^2	9072.558	7373.334	1699.225	16445.892

Mean and standard deviation (SD) are calculated from the original values. Values of -1 and +1 correspond to one standard deviation below and above the mean on the standardized scale used in Figure 5.

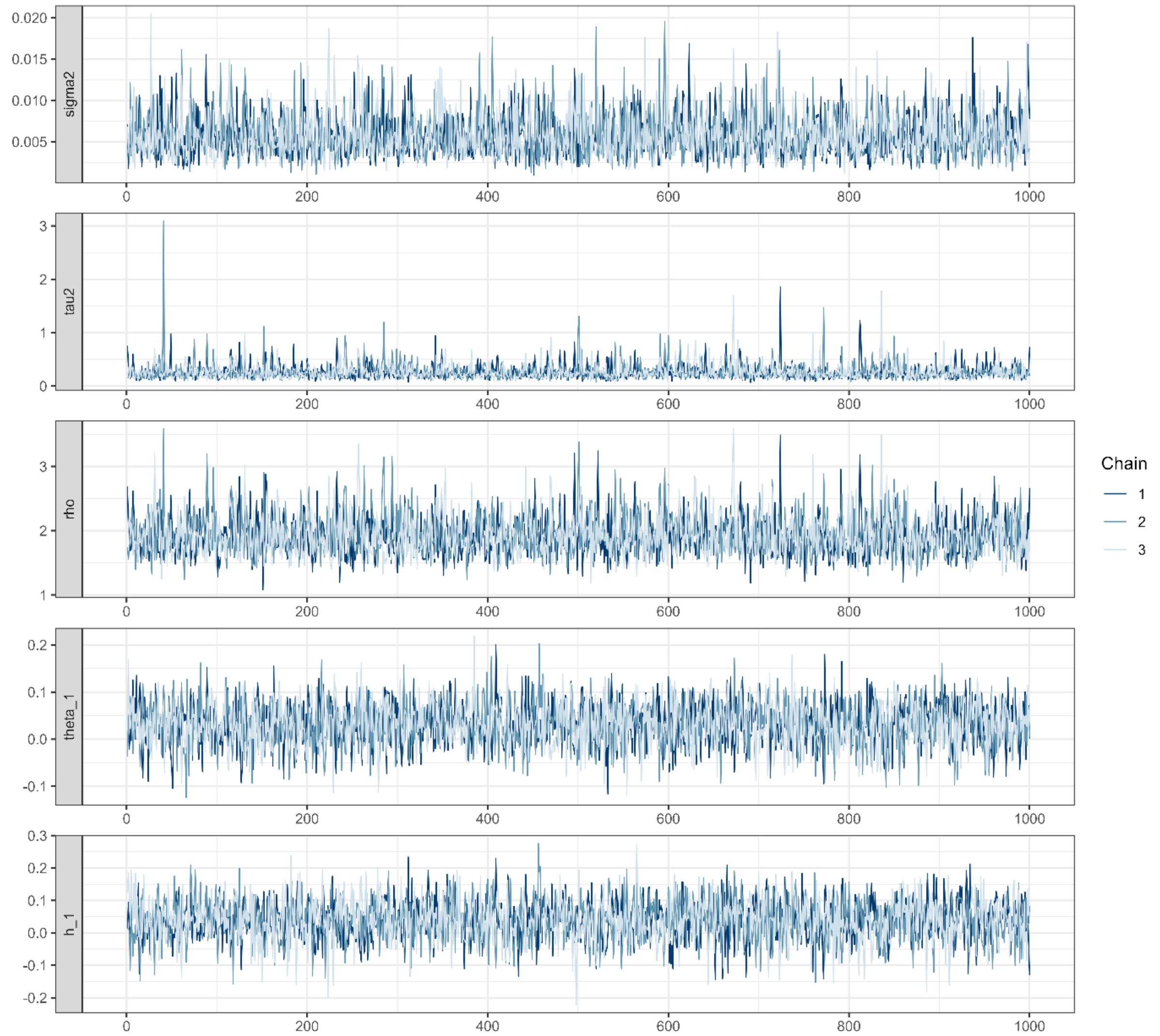


Figure S1. Trace plots for BKMMR parameters in simulation study.

Each figure shows trace plots for σ^2, τ^2, ρ, and representative θ_i and h_i from the three independent MCMC chains.

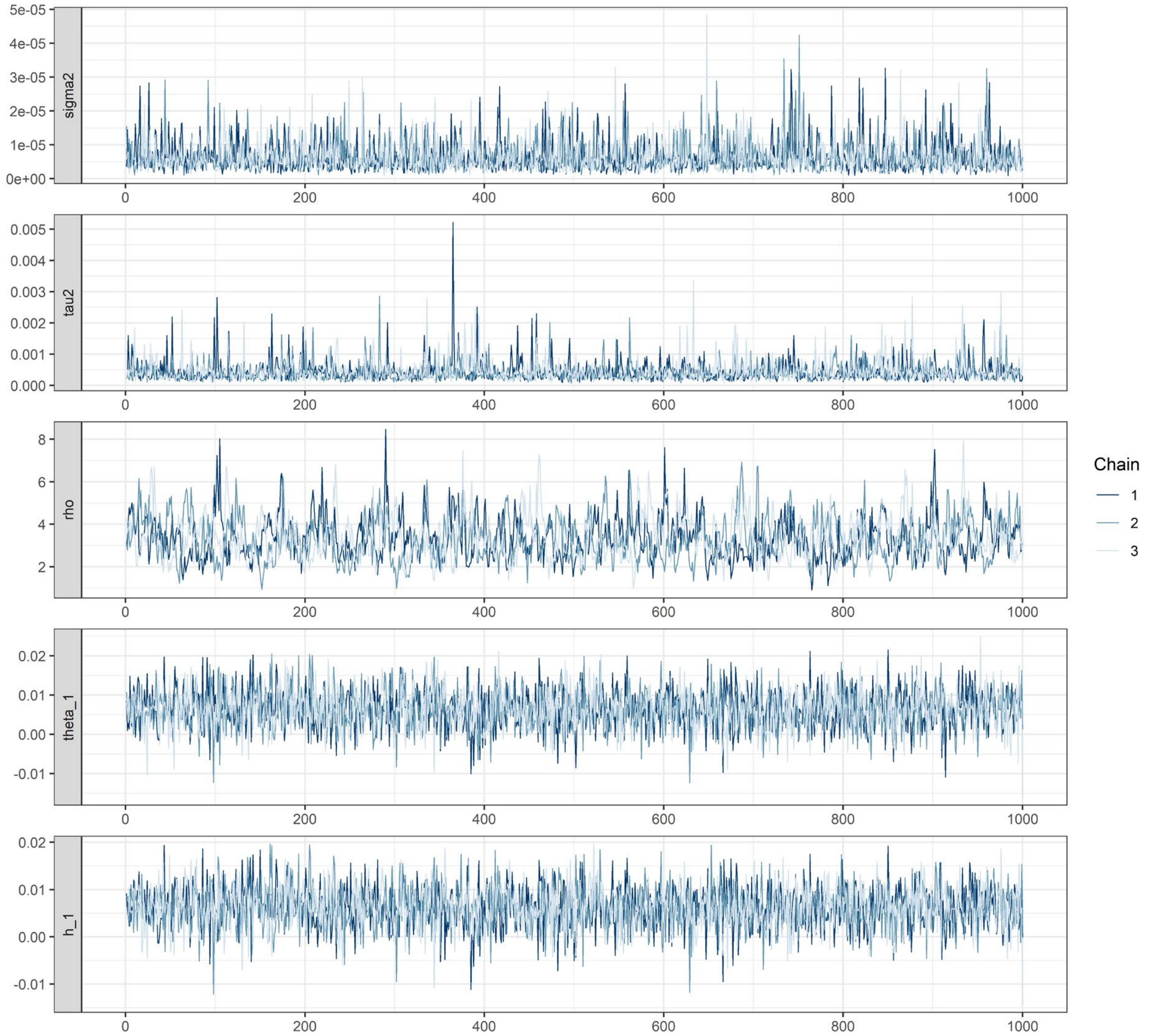


Figure S2. Trace plots for BKMMR parameters in data application.

Each figure shows trace plots for σ^2, τ^2, ρ, and representative θ_i and h_i from the three independent MCMC chains.

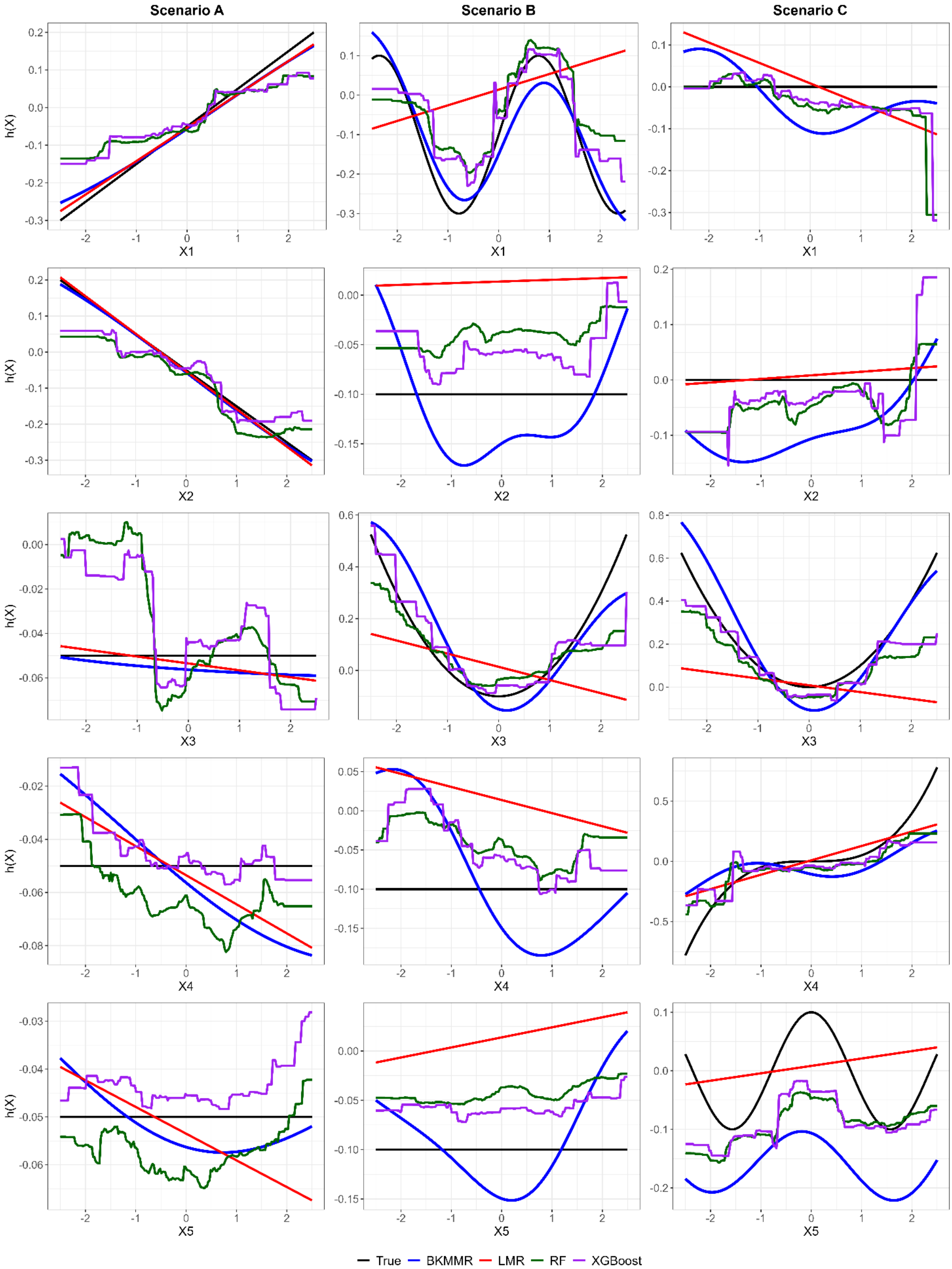


Figure S3. True and estimated meta-regression functions obtained from different methods across scenarios in simulation study.

Each panel shows the true and estimated functions as the displayed meta-predictor varies while the remaining meta-predictors are fixed at their mean values. Colors denote the true function (black), BKMMR (blue), LMR (red), RF (green), and XGBoost (purple). Estimated functions obtained from each method represent the mean estimates across 100 simulated datasets.

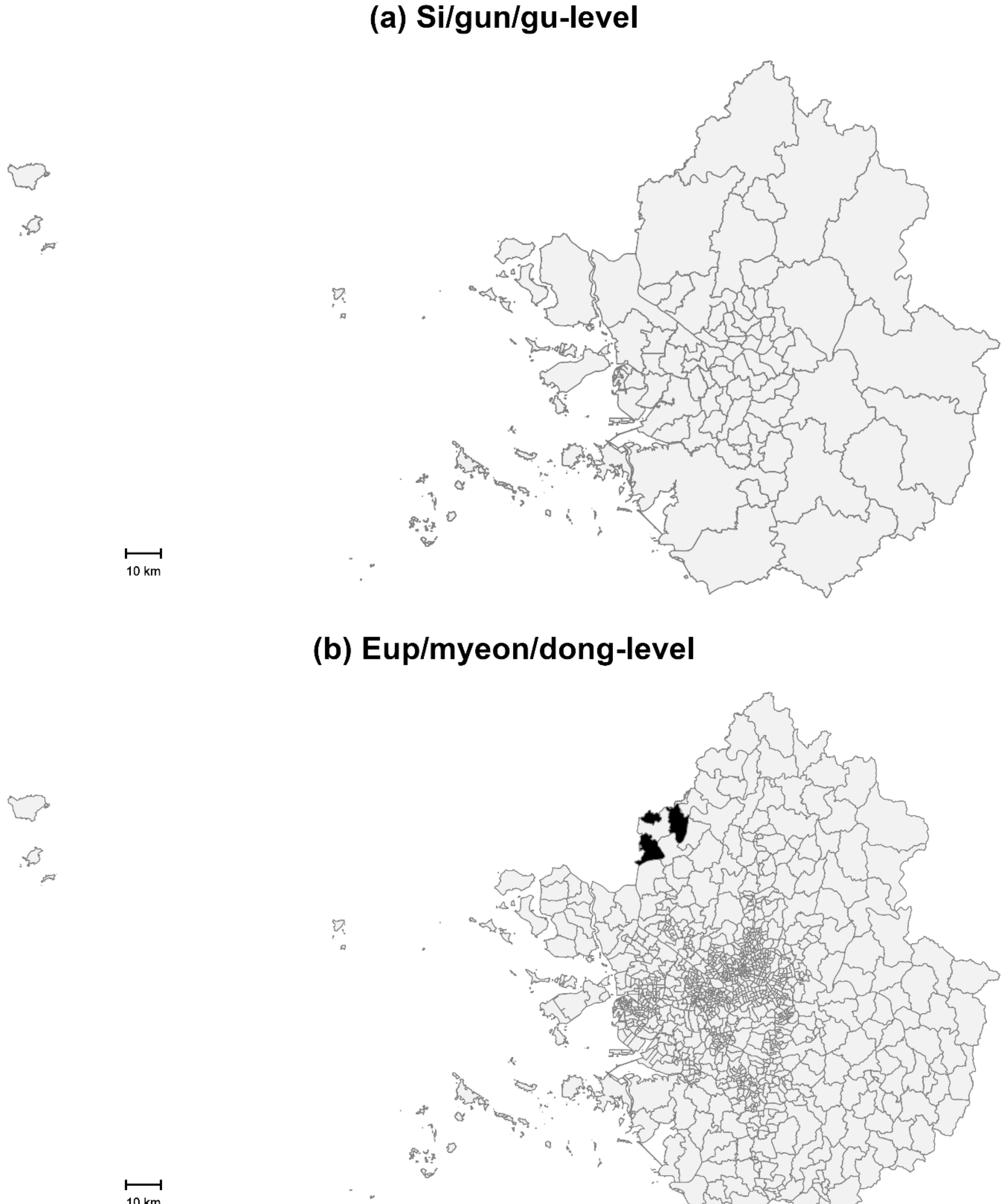


Figure S4. Spatial distribution of Si/gun/gu-level locations and Eup/myeon/dong-level sublocations. Black polygons indicate the excluded Eup/myeon/dong units.

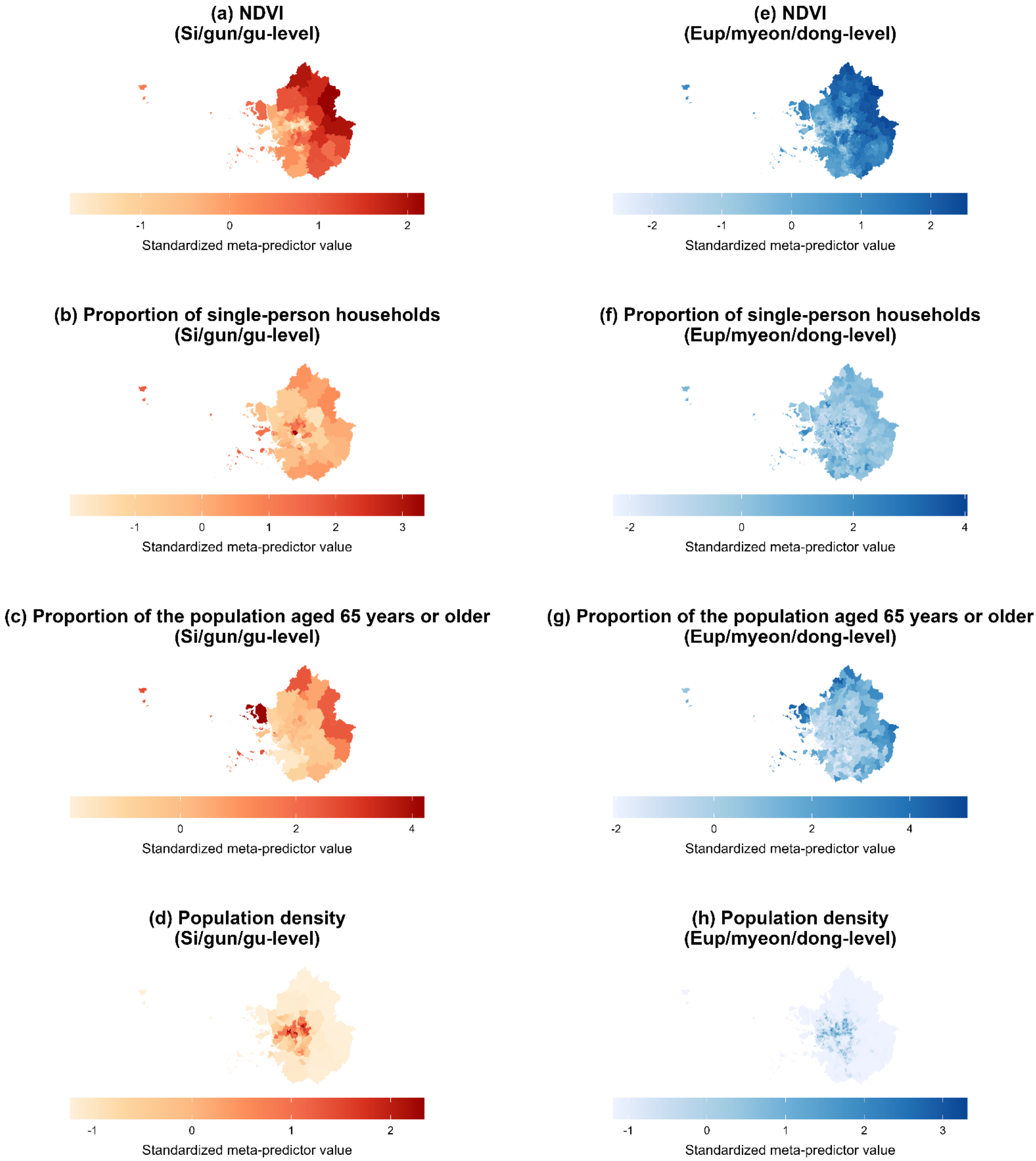


Figure S5. Spatial distribution of meta-predictors across Si/gun/gu and Eup/myeon/dong levels. Colors indicate the standardized value of each meta-predictor.

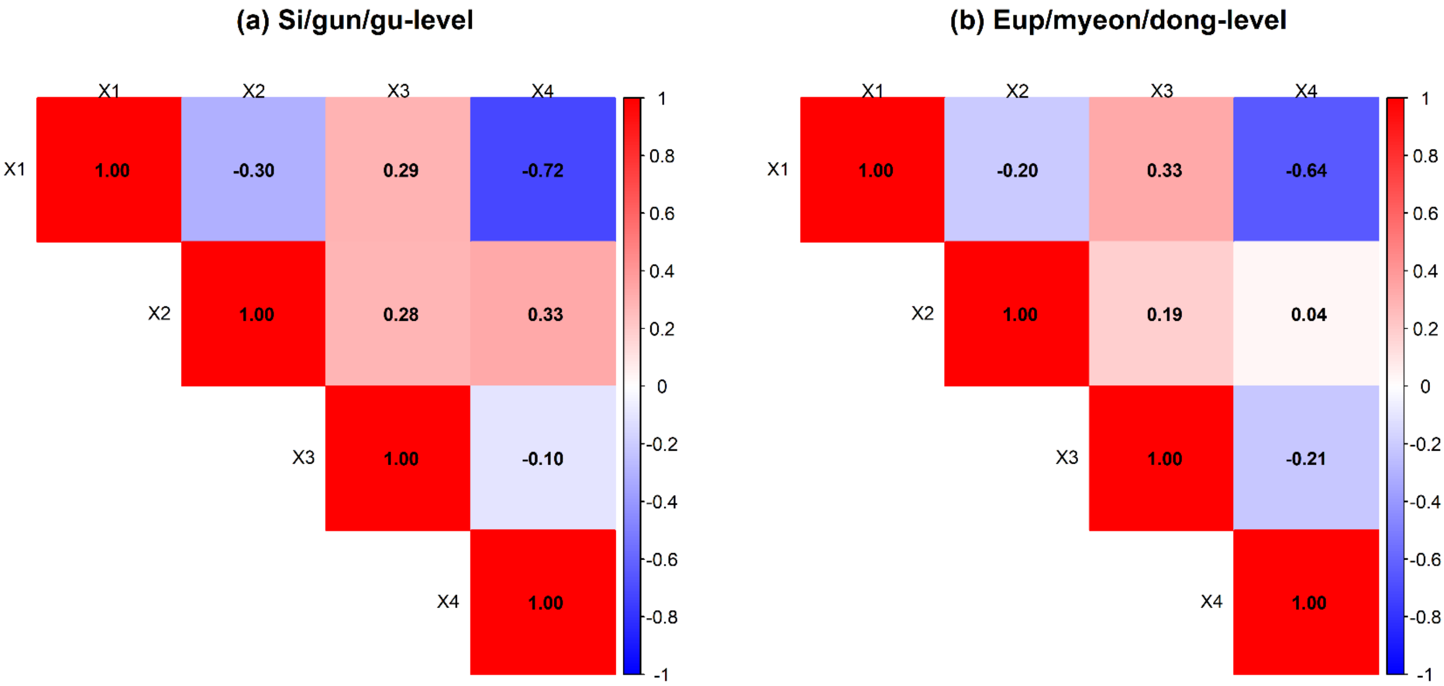
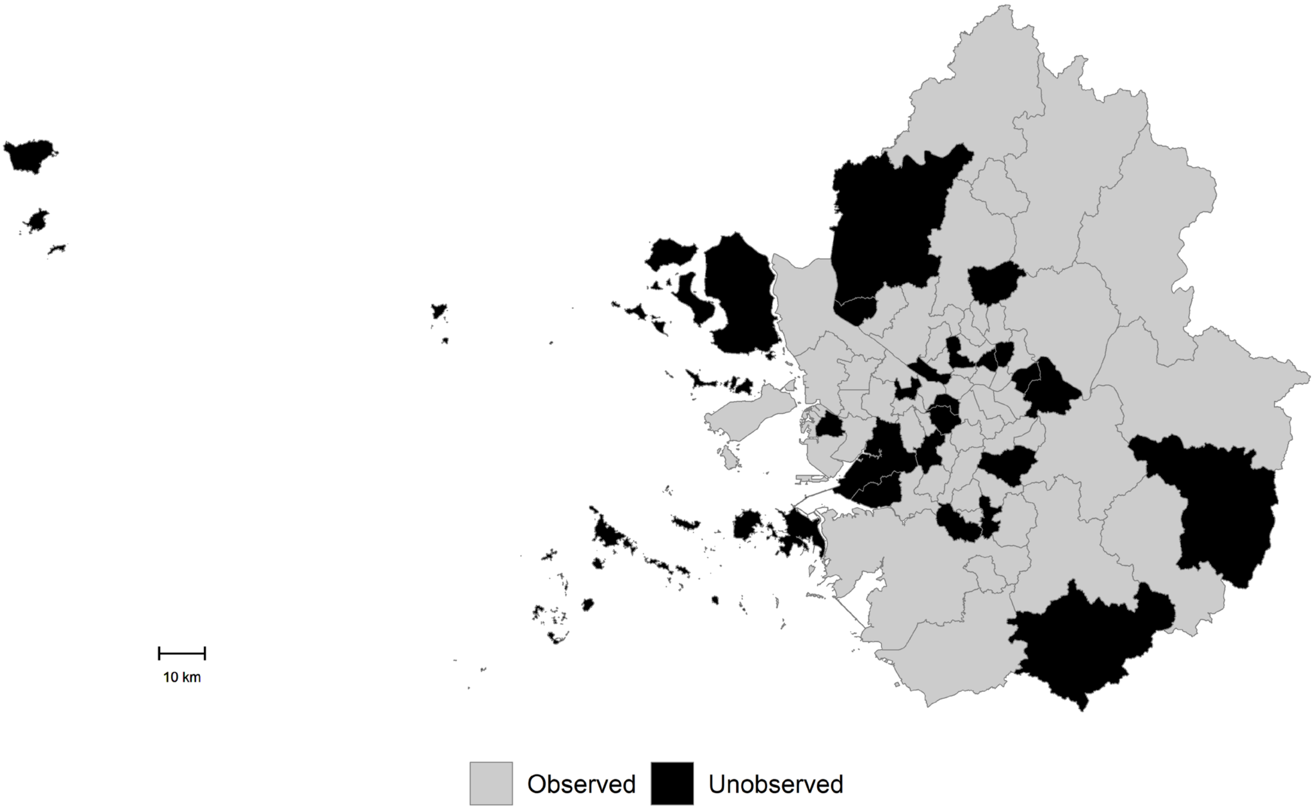


Figure S6. Pairwise correlations among meta-predictors at Si/gun/gu and Eup/myeon/dong levels.

X1, X2, X3, and X4 correspond to NDVI, the proportion of single-person households, the proportion of the population aged 65 years or older, and population density, respectively. Colors indicate the magnitude and direction of the Pearson correlation coefficients.



298 **Figure S7. Spatial distribution of observed and unobserved Si/gun/gu-level locations for the**
299 **prediction analysis.**

300 Grey polygons indicate observed locations, and black polygons indicate unobserved locations.

References

1. Vehtari A, Gelman A, Simpson D, Carpenter B and Bürkner P-C. Rank-normalization, folding, and localization: an improved R for assessing convergence of MCMC. Bayesian Anal 2021; 16(2): 667–718.
2. Vehtari A, Gelman A and Gabry J. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. Stat Comput 2017; 27: 1413–1432.
3. Gabry J, Vehtari A, Magnusson M, Yao Y, Bürkner P-C, Paananen T and Gelman A. loo: Efficient Leave-One-Out Cross-Validation and WAIC for Bayesian Models (Version 2.9.0). Vienna, Austria: Comprehensive R Archive Network (CRAN), 2025.
4. Gasparrini A and Sera F. mixmeta: An extended mixed-effects framework for meta-analysis (Version 1.2.2). Vienna, Austria: Comprehensive R Archive Network (CRAN), 2025.
5. Wright MN, Wager S and Probst P. ranger: A fast implementation of random forests (Version 0.18.0). Vienna, Austria: Comprehensive R Archive Network (CRAN), 2026.
6. Yuan J, Chen T, He T, et al. xgboost: Extreme Gradient Boosting (Version 3.2.0.1). Vienna, Austria: Comprehensive R Archive Network (CRAN), 2026.
7. Byun G, Kim S, Choi Y, et al. Long-term exposure to PM2.5 and mortality in a national cohort in South Korea: effect modification by community deprivation, medical infrastructure, and greenness. BMC Public Health 2024; 24: 1266.