

# Supplemental Materials for “Imputing Satellite-derived Aerosol Optical Depth Using a Multi-Resolution Spatial Model and Random Forest for PM<sub>2.5</sub> Prediction”

Behzad Kianian, Yang Liu, and Howard H. Chang

Animations are best viewed in Acrobat Reader.

## **S1 Additional AOD Figures and Tables**

Figure S1: Daily split between training and testing data.

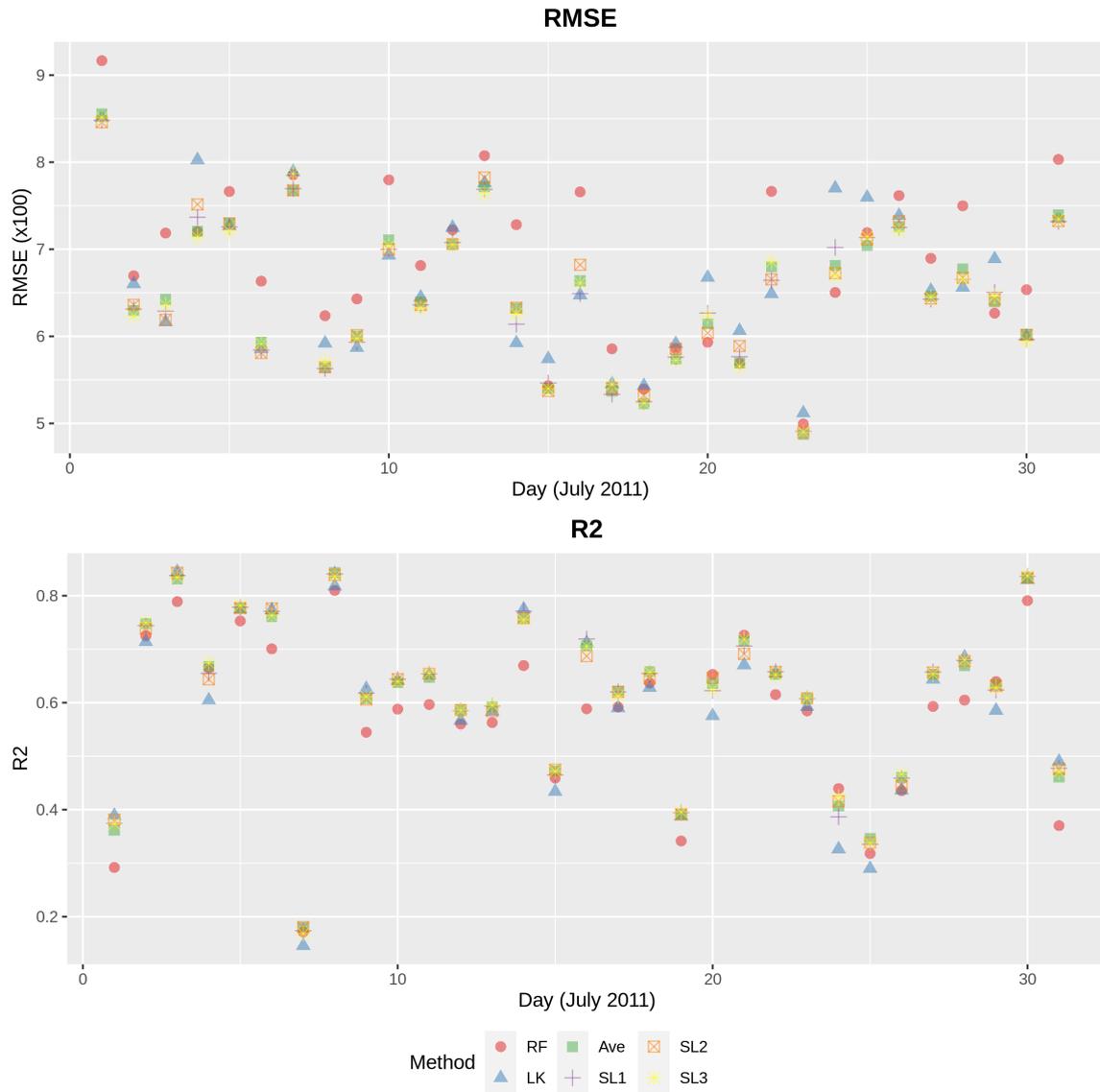


Figure S2: Root mean-squared error (RMSE) and  $R^2$  across days for LK, RF, average of RF and LK (Ave), SL: Overall (SL1), SL: Daily (SL2), and SL: Distance-based (SL3) methods.

Figure S3: Daily observed and predicted AOD values. Values outside of range are truncated for display.

Figure S4: Daily differences between LatticeKrig and Random Forest

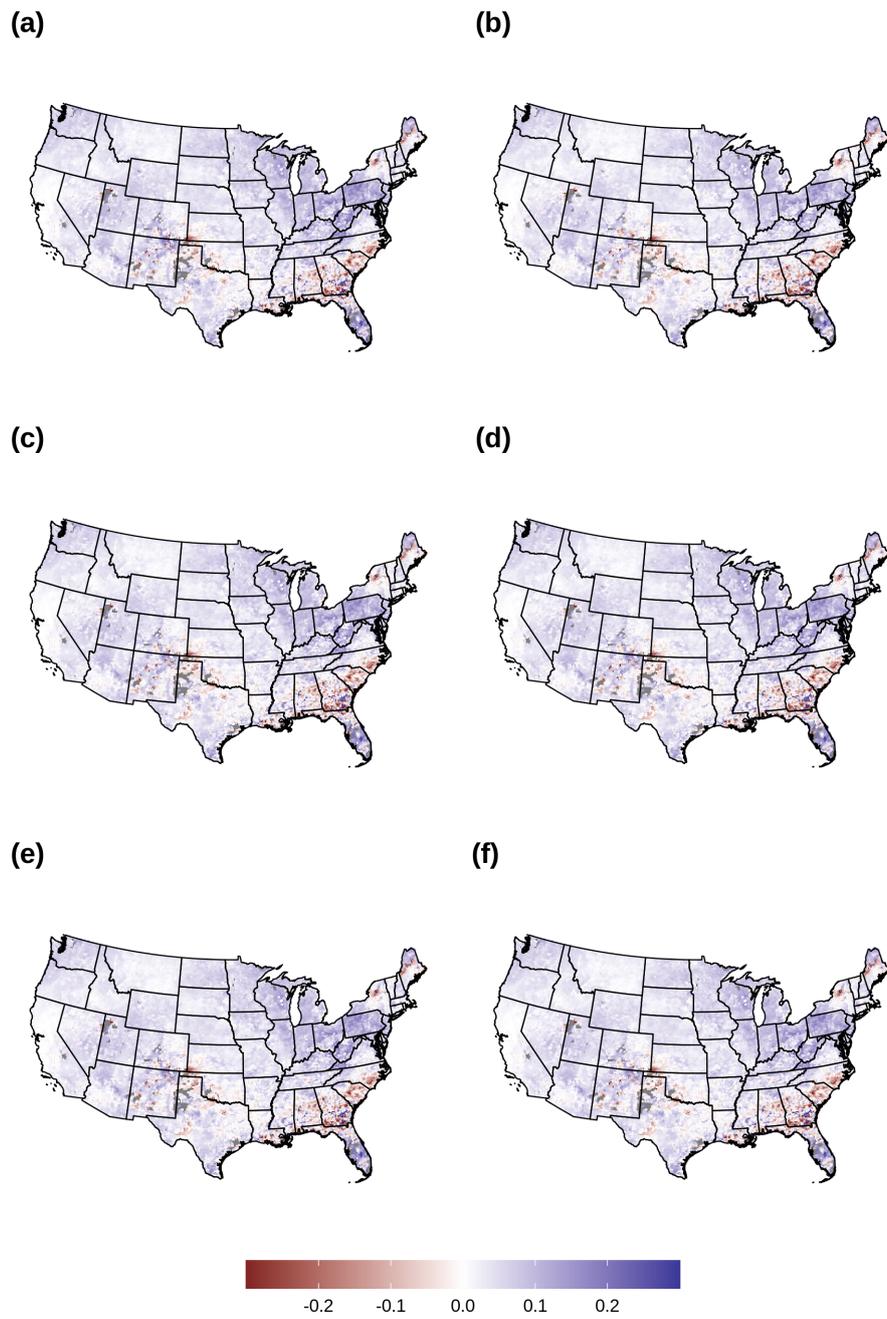


Figure S5: Difference in average predictions and observed daily values for July 2011: (a) LK; (b) RF; (c) Average of LK and RF; (d) SL: Overall; (e) SL: Daily; (f) SL: Distance-based. Differences outside of range of  $(-0.3, 0.3)$  are truncated for figure appearance.

Figure S6: Daily 10-fold spatially clustered CV. Each color represents a distinct fold, generated by the R package `blockCV`.

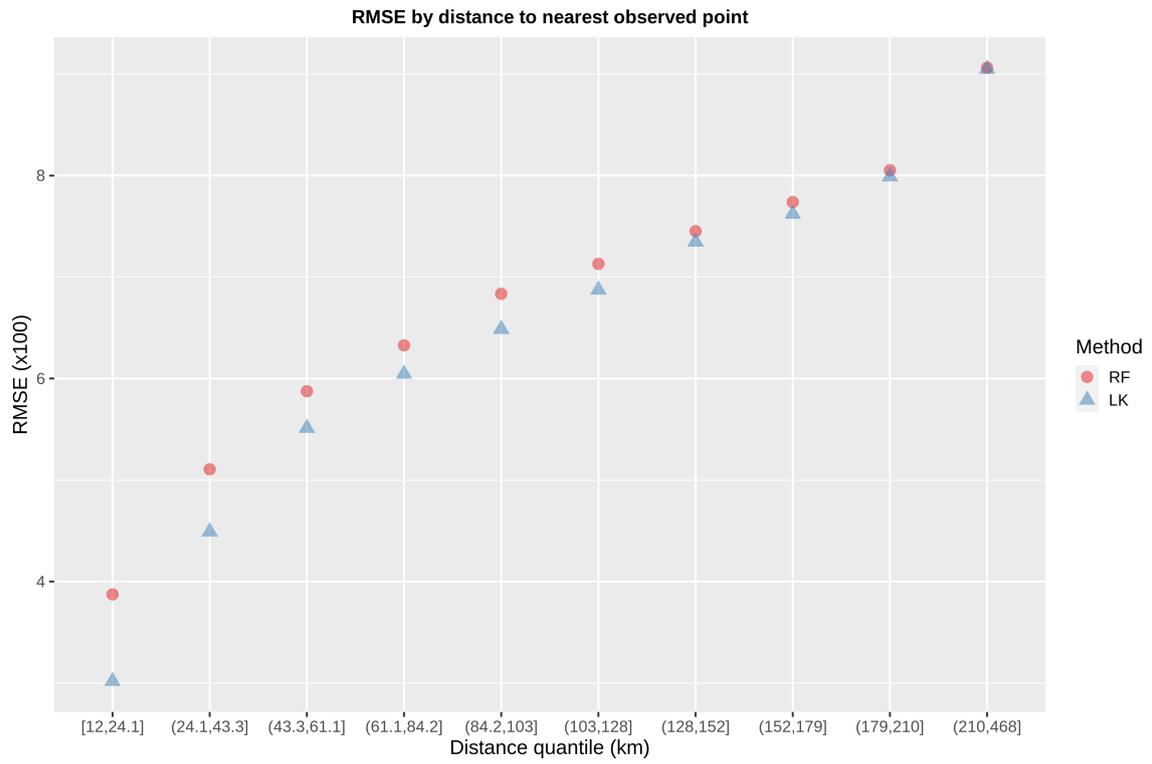


Figure S7: Comparison of LatticeKrig and Random Forest at different distances between test data and training data across all days.

## S2 Additional PM<sub>2.5</sub> Figures and Results

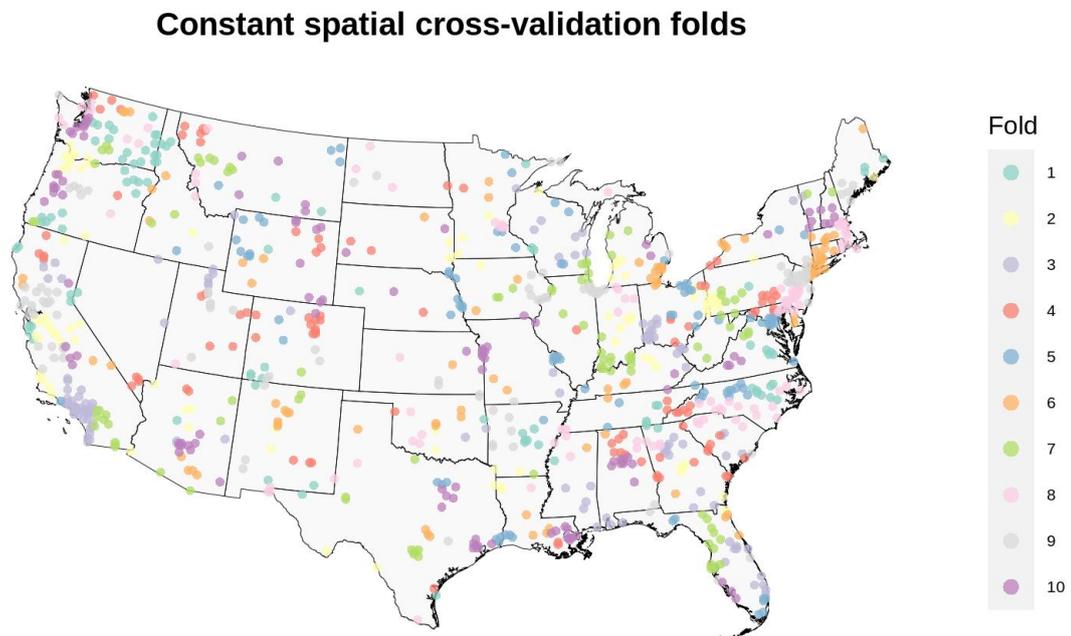


Figure S8: Constant spatial clustering cross-validation map for PM<sub>2.5</sub> analyses.

Features	M1	M2	M3	M4	M5
Convolution layer PM <sub>2.5</sub>	35.46	33.19	30.65	33.54	26.32
CMAQ-X Coordinate	18.77	13.52	13.12	15.88	12.18
GEOS-Chem		6.69	6.31		5.79
CMAQ-Y Coordinate	7.83	6.58	6.05	6.92	4.91
Convective available potential energy	8.71	6.64	5.66	7.20	4.60
Pressure at surface	6.24	5.95	5.30	6.40	4.76
Surface DW longwave radiation flux	7.46	6.35	5.14	6.62	5.61
Temperature	6.24	5.72	4.94	5.93	4.49
Imputed AOD			4.43		
Elevation	5.95	4.81	4.40	5.02	4.69
AOD/GEOS-Chem combination				3.47	
Observed AOD					2.99
Potential evaporation	3.42	3.11	3.07	3.18	2.86
Population density	3.47	3.05	2.81	3.03	2.79
Relative humidity	3.30	2.83	2.67	2.91	1.74
Day	2.52	2.49	2.24	2.31	1.46
Impervious surface (%)	2.48	2.01	1.89	2.07	1.88
Surface DW shortwave radiation flux	1.92	1.83	1.83	1.81	1.58
Percent forest cover	1.64	1.38	1.23	1.37	1.53
u-direction wind-speed	1.27	1.20	1.09	1.10	0.67
v-direction wind speed	1.28	1.13	0.98	1.10	1.18
Precipitation	0.85	0.59	0.68	0.72	0.15
Total length of local road	0.92	0.76	0.65	0.78	0.83
Faction of total precipitation that is convective	0.64	0.51	0.58	0.53	0.03
Day of the Week	0.50	0.50	0.43	0.46	0.37
AOD Missing Indicator	0.17	0.17	0.18	0.32	
Total length of limited-access road	0.10	0.10	0.09	0.09	0.11
Total length of highway	0.13	0.11	0.08	0.10	0.11
EPA 2011 emission inventory	0.07	0.06	0.06	0.07	0.07

Table S1: Feature importance (permutation-based, mean decrease in accuracy) from spatio-temporal random forest model based on `mtry = 4`.

Description_Features	M1	M2	M3	M4	M5
Convolution layer PM <sub>2.5</sub>	50.44	47.58	45.01	48.13	39.51
CMAQ-X Coordinate	19.59	13.80	12.02	17.62	10.36
GEOS-Chem		5.90	5.69		5.38
CMAQ-Y Coordinate	6.13	5.32	4.82	5.48	4.06
Pressure at surface	5.47	5.21	4.59	4.93	4.11
Surface DW longwave radiation flux	5.92	5.13	4.56	5.65	4.14
Convective available potential energy	6.13	5.15	3.99	5.59	2.82
Temperature	4.63	4.61	3.96	4.67	2.99
Imputed AOD			3.50		
Elevation	4.17	3.71	3.33	3.77	3.68
AOD/GEOS-Chem combination				2.70	
Observed AOD					2.30
Population density	3.15	2.84	2.67	2.84	2.52
Relative humidity	2.49	2.16	2.18	2.40	1.15
Potential evaporation	2.21	2.04	1.99	2.06	1.93
Impervious surface (%)	2.03	1.77	1.65	1.74	1.55
Day	1.27	1.27	1.33	1.17	0.78
Surface DW shortwave radiation flux	1.25	1.16	1.18	1.14	0.93
Percent forest cover	1.30	1.12	1.16	1.16	1.22
u-direction wind-speed	0.85	0.87	0.77	0.75	0.56
v-direction wind speed	0.84	0.78	0.67	0.79	0.96
Precipitation	0.63	0.53	0.51	0.61	0.17
Total length of local road	0.63	0.55	0.48	0.56	0.70
Faction of total precipitation that is convective	0.47	0.41	0.42	0.40	0.03
Day of the Week	0.22	0.25	0.22	0.23	0.17
AOD Missing Indicator	0.08	0.08	0.08	0.12	
Total length of limited-access road	0.08	0.06	0.07	0.07	0.09
Total length of highway	0.09	0.08	0.07	0.08	0.08
EPA 2011 emission inventory	0.06	0.05	0.05	0.05	0.06

Table S2: Feature importance (mean decrease in accuracy) from pooled random forest model based on `mtry = 8`.

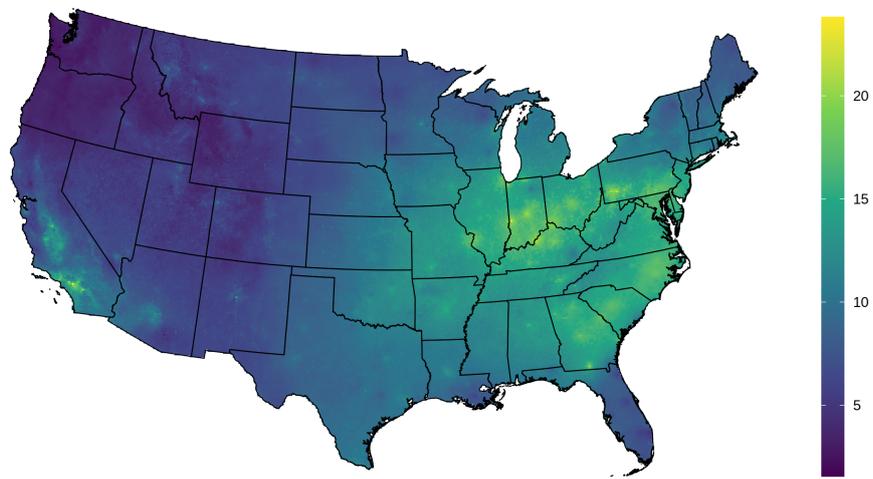


Figure S9: Average July 2011 PM<sub>2.5</sub> predicted map using imputed AOD random forest model for  $m_{\text{try}} = 8$ .

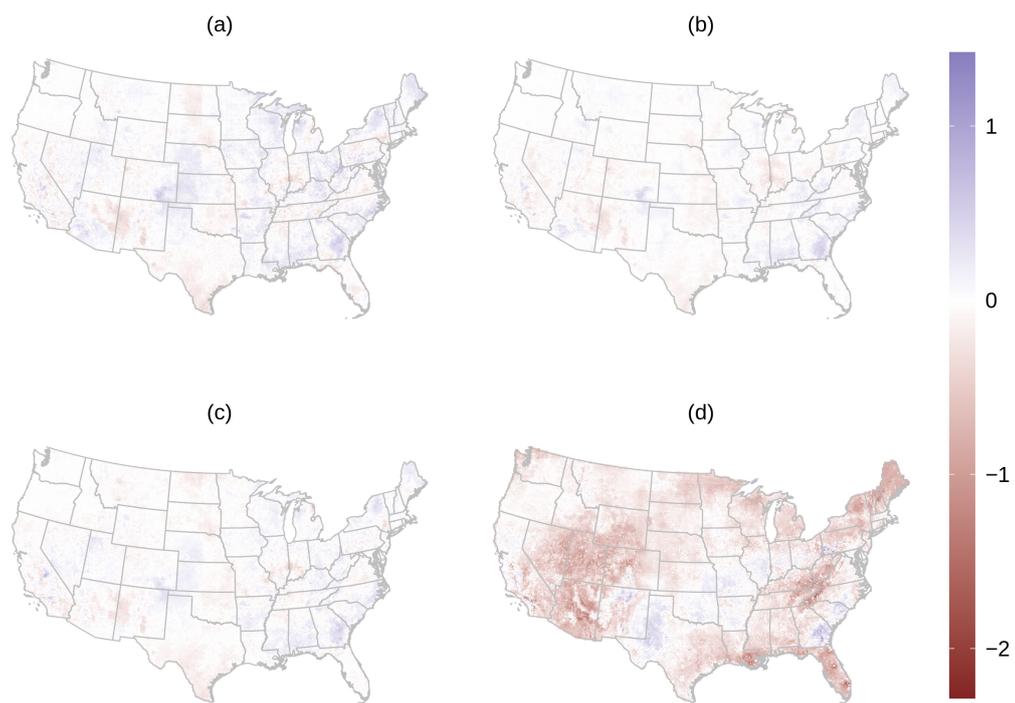
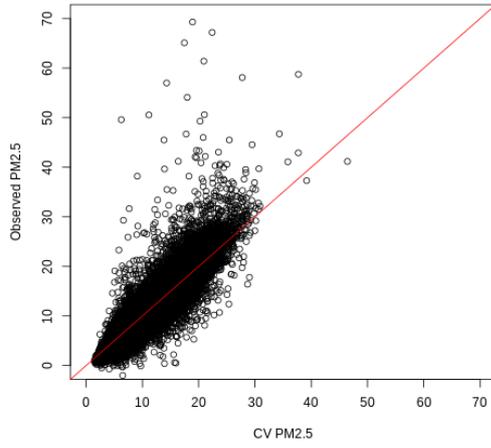
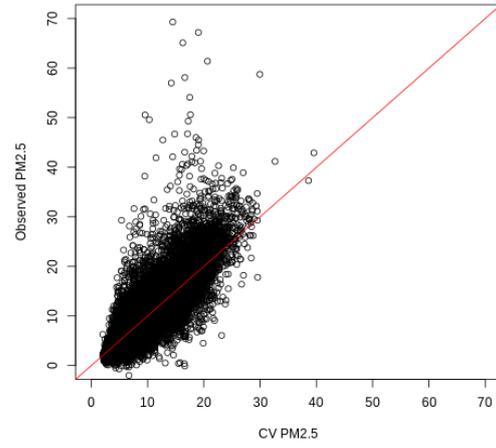


Figure S10: Difference between the imputed AOD RF model (M3) and other RF models in average July 2011 PM<sub>2.5</sub> predictions for mtry = 8: (a) M1: model with no AOD or GEOS-Chem; (b) M2: GEOS-Chem; (c) M4: Replacing missing values of AOD with GEOS-Chem; (d) M5: Train on observed AOD, and predict by replacing missing AOD values with imputed values.

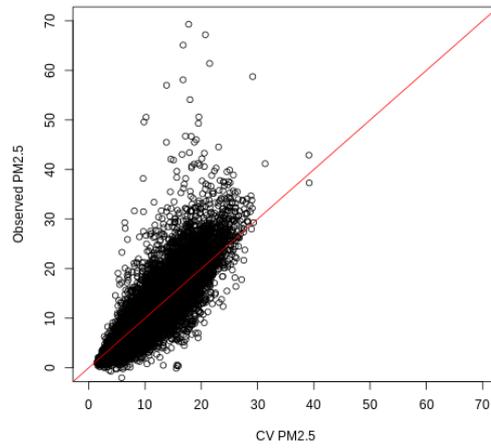
Figure S11: Difference between the imputed AOD RF model (M3) and other RF models in daily  $\text{PM}_{2.5}$  predictions for `mtry = 4`. M1: model with no AOD or GEOS-Chem; M2: GEOS-Chem; M4: Replacing missing values of AOD with GEOS-Chem; M5: Train on observed AOD, and predict by replacing missing AOD values with imputed values. Green points denote cells with observed  $\text{PM}_{2.5}$  monitors. Values outside of range truncated for display.



(a) Random

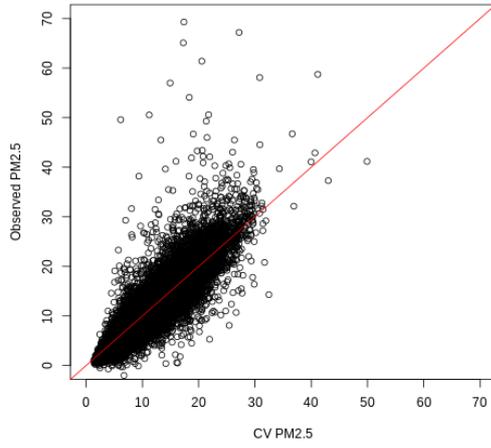


(b) Constant spatial cluster

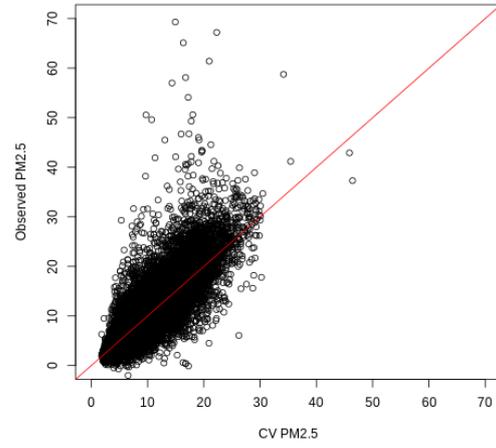


(c) Varying spatial cluster

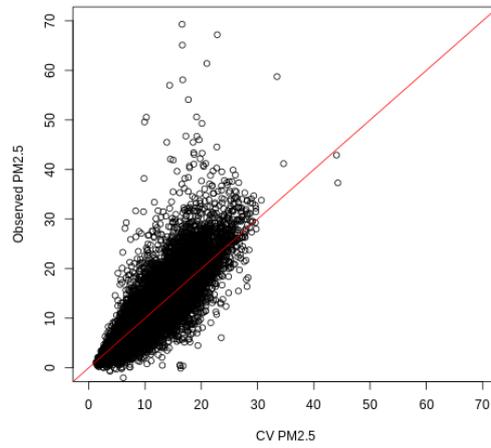
Figure S12: Scatter plots comparing observed  $\text{PM}_{2.5}$  values with cross-validation predictions from spatio-temporal random forest models including imputed AOD (M3) with `mtry = 4` for (a) random cross-validation, (b) constant spatially clustered cross-validation, and (c) varying spatially clustered cross-validation. Axes limited to  $(0, 70)$  for clarity of visual presentation. Red line is the  $y = x$  line.



(a) Random

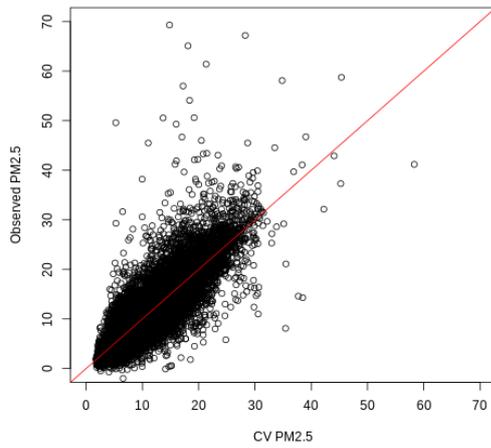


(b) Constant spatial cluster

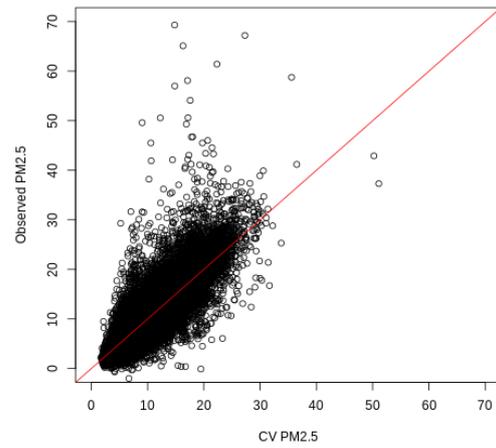


(c) Varying spatial cluster

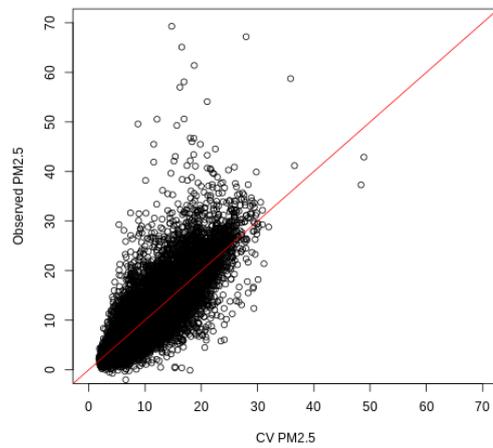
Figure S13: Scatter plots comparing observed  $\text{PM}_{2.5}$  values with cross-validation predictions from spatio-temporal random forest models including imputed AOD (M3) with  $\text{mtry} = 8$  for (a) random cross-validation, (b) constant spatially clustered cross-validation, and (c) varying spatially clustered cross-validation. Axes limited to  $(0, 70)$  for clarity of visual presentation. Red line is the  $y = x$  line.



(a) Random



(b) Constant spatial cluster



(c) Varying spatial cluster

Figure S14: Scatter plots comparing observed  $\text{PM}_{2.5}$  values with cross-validation predictions from daily random forest models including imputed AOD (M3) with `mtry = 8` for (a) random cross-validation, (b) constant spatially clustered cross-validation, and (c) varying spatially clustered cross-validation. Axes limited to (0, 70) for clarity of visual presentation. Red line is the  $y = x$  line.

Setting	AOD Status	Daily					Spatio-temporal				
		M1a	M2a	M3a	M4a	M5a	M1b	M2b	M3b	M4b	M5b
Intercept											
Random	All	-0.19	-0.21	-0.23	-0.23	-0.37	-0.71	-0.73	-0.41	-0.39	-0.75
Random	Missing	-0.26	-0.29	-0.31	-0.27	-0.78	-0.76	-0.79	-0.44	-0.40	-1.01
Random	Observed	-0.09	-0.09	-0.14	-0.16	0.06	-0.64	-0.66	-0.37	-0.35	-0.47
Constant cluster	All	-0.32	-0.33	-0.35	-0.37	-0.51	-0.42	-0.44	-0.47	-0.47	-0.92
Constant cluster	Missing	-0.60	-0.64	-0.65	-0.61	-1.12	-0.73	-0.74	-0.78	-0.73	-1.48
Constant cluster	Observed	0.03	0.05	0.02	-0.04	0.05	-0.02	-0.05	-0.10	-0.12	-0.33
Varying cluster	All	-0.40	-0.41	-0.44	-0.44	-0.53	-1.13	-1.15	-1.15	-1.17	-1.26
Varying cluster	Missing	-0.64	-0.67	-0.69	-0.64	-1.05	-1.33	-1.35	-1.33	-1.32	-1.66
Varying cluster	Observed	-0.07	-0.06	-0.12	-0.14	-0.02	-0.88	-0.90	-0.94	-0.95	-0.84
Slope											
Random	All	1.01	1.01	1.02	1.02	1.00	1.06	1.06	1.03	1.03	1.05
Random	Missing	1.02	1.02	1.02	1.02	1.02	1.07	1.07	1.03	1.04	1.06
Random	Observed	1.00	1.01	1.01	1.01	0.99	1.06	1.06	1.03	1.02	1.04
Constant cluster	All	1.04	1.04	1.04	1.05	1.03	1.06	1.06	1.06	1.06	1.08
Constant cluster	Missing	1.06	1.06	1.06	1.07	1.06	1.08	1.07	1.08	1.08	1.11
Constant cluster	Observed	1.02	1.02	1.03	1.02	1.02	1.03	1.03	1.04	1.03	1.06
Varying cluster	All	1.05	1.05	1.05	1.05	1.03	1.11	1.12	1.11	1.12	1.11
Varying cluster	Missing	1.06	1.06	1.06	1.07	1.05	1.13	1.13	1.12	1.13	1.13
Varying cluster	Observed	1.02	1.02	1.03	1.02	1.01	1.10	1.10	1.11	1.10	1.10

Table S3: Intercept and slope estimates from daily and spatio-temporal random forest model for different 10-fold cross-validation settings. Intercept and slope estimated from linear regression model Observed =  $\beta_0 + \beta_1$ Predicted.

Setting	AOD Status	Daily					Spatio-temporal				
		M1a	M2a	M3a	M4a	M5a	M1b	M2b	M3b	M4b	M5b
Random	Central	0.07	-0.04	-0.12	0.01	1.19	-1.73	-1.77	-0.83	-0.76	-0.92
	East North Central	-0.66	-0.67	-0.68	-0.65	-0.62	-1.82	-1.91	-1.17	-1.13	-1.49
	Northeast	-0.17	-0.24	-0.28	-0.23	0.24	-1.03	-0.98	-0.49	-0.47	-0.99
	Northwest	0.32	0.35	0.34	0.33	0.70	-0.70	-0.71	-0.35	-0.35	-0.43
	South	0.28	0.23	0.10	0.22	-0.32	-0.84	-0.88	-0.31	-0.17	-1.14
	Southeast	-0.38	-0.39	-0.35	-0.38	-1.61	-0.99	-1.04	-0.44	-0.43	-1.51
	Southwest	-0.00	-0.12	-0.18	-0.20	0.49	-1.47	-1.59	-0.99	-0.89	-1.05
	West	0.10	0.14	0.03	-0.01	0.47	-1.49	-1.61	-1.02	-1.01	-1.21
Constant cluster	West North Central	0.71	0.75	0.72	0.76	1.11	-0.90	-0.87	-0.68	-0.65	-0.74
	Central	-0.11	-0.31	-0.35	-0.43	1.17	-0.88	-0.97	-1.17	-1.09	-1.33
	East North Central	-0.94	-0.93	-0.95	-0.97	-1.04	-1.59	-1.72	-1.79	-1.72	-2.26
	Northeast	-0.35	-0.41	-0.45	-0.41	-0.08	-0.48	-0.57	-0.64	-0.55	-1.32
	Northwest	0.06	0.08	0.09	0.09	0.42	-0.34	-0.31	-0.30	-0.33	-0.57
	South	-0.25	-0.36	-0.54	-0.37	-0.83	-0.83	-0.98	-1.08	-0.93	-2.36
	Southeast	-1.09	-1.15	-1.13	-1.08	-2.45	-1.35	-1.34	-1.35	-1.33	-2.60
	Southwest	-0.22	-0.33	-0.44	-0.45	0.15	0.10	0.02	-0.18	-0.19	-0.60
Varying cluster	West	0.87	1.01	1.12	0.80	1.27	0.21	0.22	0.16	0.11	-0.13
	West North Central	0.84	0.85	0.86	0.86	1.29	0.86	0.81	0.76	0.78	0.59
	Central	0.41	0.20	0.12	0.08	1.75	-1.93	-2.03	-2.28	-2.12	-1.27
	East North Central	-1.11	-1.07	-1.07	-1.09	-0.98	-2.43	-2.55	-2.55	-2.54	-2.41
	Northeast	-0.62	-0.68	-0.69	-0.68	-0.25	-1.71	-1.64	-1.63	-1.69	-1.74
	Northwest	0.22	0.22	0.24	0.24	0.59	-0.89	-0.87	-0.88	-0.89	-0.76
	South	-0.28	-0.38	-0.51	-0.36	-0.67	-1.83	-1.95	-2.01	-1.91	-2.43
	Southeast	-1.14	-1.18	-1.17	-1.11	-2.06	-2.00	-2.08	-1.96	-2.02	-2.70
Southwest	-0.18	-0.21	-0.40	-0.41	0.33	-2.17	-2.26	-2.43	-2.66	-2.09	
West	0.81	0.90	0.78	0.75	1.25	-2.05	-2.16	-2.27	-2.19	-1.56	
West North Central	0.64	0.65	0.65	0.66	1.05	-1.29	-1.26	-1.32	-1.34	-1.25	

Table S4: Regional intercept estimates for daily and spatio-temporal random forest model for different 10-fold cross-validation settings.

Setting	AOD Status	Daily					Spatio-temporal				
		M1a	M2a	M3a	M4a	M5a	M1b	M2b	M3b	M4b	M5b
Random	Central	1.01	1.02	1.02	1.01	0.93	1.12	1.12	1.06	1.05	1.06
	East North Central	1.02	1.02	1.02	1.02	0.99	1.13	1.13	1.07	1.07	1.08
	Northeast	1.01	1.02	1.02	1.02	0.98	1.08	1.07	1.03	1.03	1.06
	Northwest	0.87	0.86	0.86	0.86	0.71	1.15	1.16	1.07	1.07	1.05
	South	0.96	0.96	0.98	0.96	0.97	1.07	1.07	1.02	1.00	1.07
	Southeast	1.02	1.03	1.02	1.02	1.07	1.07	1.08	1.03	1.03	1.08
	Southwest	0.95	0.96	0.97	0.97	0.82	1.19	1.20	1.12	1.11	1.07
	West	1.01	1.01	1.02	1.02	0.96	1.15	1.17	1.10	1.10	1.11
	West North Central	0.91	0.90	0.91	0.90	0.82	1.14	1.13	1.11	1.11	1.10
Constant cluster	Central	1.07	1.08	1.08	1.09	0.98	1.13	1.13	1.14	1.14	1.15
	East North Central	1.04	1.03	1.03	1.03	1.01	1.09	1.10	1.11	1.10	1.13
	Northeast	1.04	1.04	1.05	1.04	1.01	1.04	1.05	1.05	1.05	1.09
	Northwest	0.91	0.91	0.90	0.90	0.75	1.02	1.02	1.01	1.02	1.03
	South	1.01	1.02	1.04	1.02	1.01	1.06	1.07	1.08	1.07	1.18
	Southeast	1.08	1.08	1.08	1.08	1.13	1.10	1.10	1.10	1.10	1.17
	Southwest	1.00	1.01	1.03	1.03	0.89	0.98	0.99	1.02	1.02	1.02
	West	1.03	1.02	1.00	1.04	0.97	1.09	1.09	1.09	1.10	1.12
	West North Central	0.90	0.90	0.90	0.90	0.80	0.92	0.92	0.93	0.93	0.94
Varying cluster	Central	1.03	1.04	1.05	1.05	0.94	1.18	1.19	1.20	1.19	1.14
	East North Central	1.05	1.04	1.04	1.05	1.01	1.17	1.17	1.17	1.17	1.15
	Northeast	1.07	1.07	1.07	1.07	1.03	1.14	1.13	1.13	1.14	1.13
	Northwest	0.88	0.88	0.88	0.87	0.72	1.18	1.17	1.17	1.17	1.10
	South	1.01	1.01	1.03	1.01	1.00	1.15	1.16	1.17	1.16	1.19
	Southeast	1.08	1.08	1.08	1.08	1.10	1.15	1.16	1.14	1.15	1.18
	Southwest	1.00	1.00	1.03	1.03	0.87	1.30	1.31	1.33	1.37	1.22
	West	1.00	0.99	1.00	1.00	0.93	1.26	1.27	1.29	1.28	1.20
	West North Central	0.92	0.92	0.92	0.92	0.83	1.19	1.19	1.19	1.20	1.16

Table S5: Regional slope estimates for daily and spatio-temporal random forest model for different 10-fold cross-validation settings.

### S3 Additional LatticeKrig modeling details

We follow the model description of lattice kriging (LatticeKrig or LK) laid out by Nychka et al. [2015]. At a high-level, LK models the spatial process using several levels of two-dimensional basis functions, which are laid out on a grid and approximately double with each successive layer. These basis functions are compact, which means that for a particular point only a small number of basis function are used to make the prediction. The coefficients associated with the basis functions are assumed to be correlated, and this structure can flexibly model observed spatial covariance structures. Estimation proceeds through a likelihood-based approach after specifying various tuning parameters.

Following the notation of Nychka et al. [2015], we observe  $\{y_i\}$  at locations  $\{\mathbf{x}_i\}$  for  $i = 1, \dots, n$ . We assume  $\{y_i\}$  follow an additive model consisting of a mean function based on covariates, a spatial process, and a measurement error term:

$$y_i = \mathbf{Z}_i^T \mathbf{d} + g(\mathbf{x}_i) + \epsilon_i, \quad (1)$$

where  $\mathbf{d}$  is a  $p \times 1$  vector of fixed coefficients associated with the covariates  $\mathbf{Z}_i$ , and  $g(\mathbf{x}_i)$  denotes the spatial process. The mean-zero error terms  $\epsilon_i$  are presumed to be independent and identically distributed, i.e.,  $\boldsymbol{\epsilon} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$ , where  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T$ .

The overall spatial process  $g(\mathbf{x}_i)$  can be written as a sum of  $L$  independent spatial processes  $g_l(\mathbf{x}_i)$ :

$$g(\mathbf{x}_i) = \sum_{l=1}^L g_l(\mathbf{x}_i) = \sum_{l=1}^L \sum_{j=1}^{m(l)} c_j^l \phi_{j,l}(\mathbf{x}_i), \quad (2)$$

where  $\phi_{j,l}$  denotes the the  $l$ th level of resolution's  $j$ th basis function, and  $c_j^l$  denotes the coefficient associated with this basis function. Although the basis functions and number of levels are fixed (i.e., chosen), the coefficients for each level  $l$ ,  $\mathbf{c}^l = (c_1^l, \dots, c_{m(l)}^l)^T$  are assumed to follow a multivariate normal with mean zero and covariance  $\rho \mathbf{Q}_l^{-1}$ :

$$\mathbf{c}^l \sim N(\mathbf{0}, \rho \mathbf{Q}_l^{-1}). \quad (3)$$

Each level's spatial process is independent with marginal variance  $\rho \alpha_l$  subject to the constraint  $\sum_{l=1}^L \alpha_l = 1$ , so that the marginal variance of the overall spatial process  $g(\mathbf{x}_i)$  is  $\rho$ .

Let  $m$  denote the total number of basis functions, and for simplicity consider a single level  $L = 1$ , so that  $g(\mathbf{x}) = \sum_{j=1}^m c_j \phi_j(\mathbf{x})$ . Then, for any two locations  $\mathbf{x}$  and  $\mathbf{x}'$ , the covariance is given as:

$$\text{Cov}(g(\mathbf{x}), g(\mathbf{x}')) = \rho \sum_{j=1}^m \sum_{k=1}^m \mathbf{Q}_{j,k}^{-1} \phi_j(\mathbf{x}) \phi_k(\mathbf{x}'). \quad (4)$$

Denote  $\boldsymbol{\Phi}$  as the  $n \times m$  matrix of basis functions evaluated at the observed locations. The full marginal distribution  $\mathbf{y}$  is then given as

$$\mathbf{y} \sim N(\mathbf{Z}\mathbf{d}, \rho \boldsymbol{\Phi} \mathbf{Q}^{-1} \boldsymbol{\Phi}^T + \sigma^2 \mathbf{I}). \quad (5)$$

By setting  $\lambda = \sigma^2/\rho$  (a noise to signal ratio), and  $M_\lambda = \boldsymbol{\Phi} \mathbf{Q}^{-1} \boldsymbol{\Phi}^T + \lambda \mathbf{I}$ , this may be further re-written as

$$\mathbf{y} \sim N(\mathbf{Z}\mathbf{d}, \rho M_\lambda). \quad (6)$$

Nychka et al. [2015] provide further details on estimation of the key parameters using the profile log-likelihood such that the likelihood only depends on  $\lambda$  and parameters determining  $\mathbf{Q}$ .

Nychka et al. [2015] propose using two-dimensional radial basis functions (RBF) using the Wendland functions that have a compact support. These basis functions take the following form for scaled distance  $0 \leq d \leq 1$ :

$$\phi(d) = (1 - d)^6(35d^2 + 18d + 3)/3. \quad (7)$$

By default, the distance is scaled to be 2.5 times the grid spacing for each level of resolution. The basis functions are thus defined as:

$$\phi_j^*(\mathbf{x}) = \phi(\|\mathbf{x} - \mu_j\|/\theta), \quad (8)$$

where  $\mu_j$  is the location of the basis function, and  $\theta$  is set to determine the amount of overlap. Nychka et al. [2015] additionally recommend and implement basis function normalization by default as part of their estimation in order to obtain a constant marginal variance.

### S3.1 Parameters

Several parameters can impact LK's predictions and inference. We review them here and discuss their impact on the implied spatial covariance, along with the associated parameter name in the R package `LatticeKrig` (version 8.4) [Nychka et al., 2016] implementing the method:

- *Number of basis functions and levels*: The number of basis functions follows from: (1) specifying the number of levels of resolutions, denoted by `nlevel` in the package, and (2) specifying the number of basis functions along the longest dimension at the first (coarsest) level of resolution, parameterized by `NC`. Each successive level of resolution has roughly double the basis functions, so this determines the entire grid. Nychka et al. [2015] suggests choosing these so that the coarsest level of resolution can capture the overall correlation range, and so that the finest level of resolution can capture fine scale changes in the spatial process. Holding all else constant (including the levels of resolution), increasing the number of basis functions at the coarsest level decreases the implied covariance for a given distance. A parameter for adding extra basis functions to the edges to reduce artifacts in prediction is determined by `NC.buffer`, which is set to 5 by default.
- *Relative weight of each spatial level's process*: Recall that each level's spatial process  $g_l(\mathbf{x}_i)$  has a marginal variance of  $\rho\alpha_l$  where  $\sum_{l=1}^L \alpha_l = 1$ . In the package implementation,  $\sqrt{\alpha_l}$  multiplies the basis functions (after normalization), such that

$$g(\mathbf{x}) = \sum_{l=1}^L \sqrt{\alpha_l} g_l(\mathbf{x}) = \sum_{l=1}^L \sum_{j=1}^{m(l)} c_j^l (\sqrt{\alpha_l} \phi_{j,l}(\mathbf{x}_i)).$$

Choosing  $\alpha$  parameters (relative weights) can be simplified into a single tuning parameter  $\nu$  (`nu` in the R package), where  $\alpha_l \propto 2^{-2l\nu}$ . Small values of  $\nu$  (e.g., 0.1) weight each level of resolution more equally, while larger values of  $\nu$  (e.g., 1.25) result in more heavily weighting the coarsest level of resolution.

- *Scale/range parameter*: Briefly, the coefficient vector  $\mathbf{c}^l$  for level  $l$  follows a Gaussian Markov random field, and in particular, a spatial autoregression (SAR). In the `LatticeKrig` package, one specifies  $a = 4 + \kappa^4$  (or `a.wght`). Holding other parameters constant, large values of  $a$  suggest less effective correlation range, i.e., for a given distance the implied correlation of the LK model will be lower as  $a$  is increased. A small value of  $a$ , e.g. 4.01 (the default setting in `LatticeKrig`) is similar to a thin-plate spline where there is a very large range and strong

spatial dependence. Some greater detail is provided in the Supplemental Materials as well as in the originating paper and package documentation [Nychka et al., 2015, 2016].

## References

- Douglas Nychka, Soutir Bandyopadhyay, Dorit Hammerling, Finn Lindgren, and Stephan Sain. A multiresolution gaussian process model for the analysis of large spatial datasets. *Journal of Computational and Graphical Statistics*, 24(2):579–599, 2015.
- Douglas Nychka, Dorit Hammerling, Stephan Sain, and Nathan Lenssen. Latticekrig: Multiresolution kriging based on markov random fields, 2016. URL <https://github.com/NCAR/LatticeKrig>. R package version 8.4.