Supplementary Material: Additive Dynamic Models for Correcting Numerical Model Outputs

Yewen Chen, Xiaohui Chang, Fangzhi Luo, Hui Huang

In this document, our focus is on several aspects. Firstly, we examine the distribution of observed data in the BTH region. Secondly, we provide an overview of the different methods discussed in the manuscript. Additionally, we conduct additional comparisons using simulated data. Lastly, we perform a sensitivity analysis for the ADCM by exploring different specifications of the precision matrices of IDE models.

S1. Data transformation

Following Berrocal et al. (2010), histograms are employed to investigate the distributions of the data. In addition to analyzing the original observations in the BTH region, we also consider the square root transformed data as suggested by Sahu et al. (2006). As shown in Figure S1 (left), the observed $PM_{2.5}$ concentrations exhibit clear right skewness. Consequently, we perform a square root transformation on the observations to achieve variance stabilization, as depicted in Figure S1 (right).

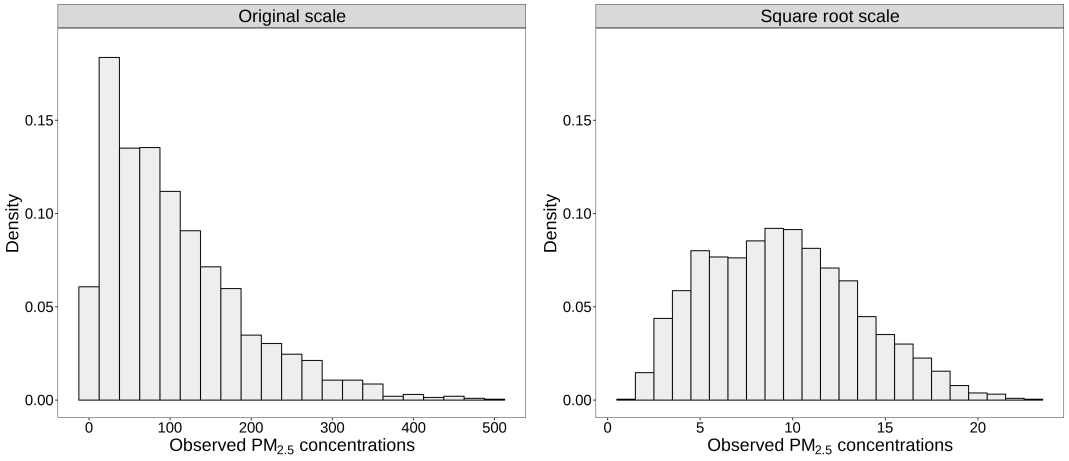

Figure S1: Histograms of observed $PM_{2.5}$ concentrations at two different scales in the BTH region in the Winter of 2015, including the original scale (left) and square root scale (right).

S2. Competing models

S2.1. Spatially varying coefficients (SVC) model

We use the following version of the SVC model:

$$y_t(s) = \beta_0 + \beta_1 x_t(C_s) + \sum_{k=1}^{4} \alpha_k z_{t,k}(C_s) + w_t(s) + \varepsilon_t(s), \tag{S1}$$

where $y_t(s)$ represents the square root of the reading of the monitoring station at location $\boldsymbol{s} \in D$, $x_t(C_s)$ represents the square root of the weighted average CMAQ outputs at location $\boldsymbol{s}$, t is a time variable, and $z_{t,1}(C_{\boldsymbol{s}}), \cdots, z_{t,4}(C_{\boldsymbol{s}})$ represent surface temperature, surface pressure, northern and eastern cumulative wind powers, respectively (if not

stated, all variables have been standardized, the same for all models), $w_t(s)$ is the spatial random process used to capture an additive bias, and $\varepsilon_t(\boldsymbol{s})$ is a zero-mean Gaussian white noise with variance σ^2_ϵ.

We assume that $w_t(s)$ and $\varepsilon_t(\boldsymbol{s})$ are mutually and serially independent. At time t, we model $w_t(s)$ as spatial stationary Gaussian random processes with mean $\mathbf{0}$ and an exponential covariance function from the Matérn family (Cressie and Wikle, 2011), i.e., $\text{Cov}\big(w_t(s_i), w_t(s_j)\big) = \tau^2_w \exp\big(-\|s_i - s_j\|/\phi_w\big)$. The unknown parameters including τ^2_w, ϕ_w, and others appeared in equation (S1) can be estimated using the Markov chain Monte Carlo (MCMC) algorithm or MLE methods. To ensure a fair comparison between the SVC (S1) and the proposed ADCM, our focus is mainly on the SVC under maximum likelihood estimation methods (Dambon et al., 2021a), which can be implemented through the R package `varycoef` (Dambon et al., 2021b).

S2.2. Spatiotemporally varying coefficient (STVC) model

Following section 3.2 of Berrocal et al. (2010), we extend the spatial SVC (S1) to a spatio-temporal downscaler model that allows for spatiotemporally varying coefficient (STVC), given by

$$\begin{cases} y_t(\boldsymbol{s}) = \beta_0 + \beta_{0,t} + (\beta_1 + \beta_{1,t})x_t(C_s) + \beta_2 t + \sum_{k=1}^4 \alpha_k z_{t,k}(C_s) + w_t(s) + \varepsilon_t(s) \\ \beta_{l,t} = \rho_l \beta_{l,t-1} + \eta_{l,t}, \text{ for } l = 0, 1, \end{cases} \tag{S2}$$

where $w_t(\boldsymbol{s})$ is a spatial stationary Gaussian random process with mean $\mathbf{0}$, and $\boldsymbol{\eta}_{l,t} = \big(\eta_{l,t}, \cdots, \eta_{l,t}\big)^T \sim \mathcal{N}\big(0, \delta^2_l \boldsymbol{I}_{N_t}\big)$. In the STVC, for each l, $\eta_{l,t}$ and $w_t(\boldsymbol{s})$ are mutually independent for all t and $\boldsymbol{s}$, and the same for $\eta_{l,t}$ and $\varepsilon_t(s)$.

In practice, adding constraints of the STVC may be important to make the latent effects, $\beta_{l,t}$ and $w_t(s)$, identifiable, and thus, we set a linear constrain to $\sum_{t=1}^{N_t} \beta_{l,t} = 0$. Based on the cross-validation, we found that the linear constraint can significantly improve the prediction performance of the STVC for the BTH data. In this work, we implement the STVC by using the `R` package `INLA` (`https://www.r-inla.org/`, Lindgren and Rue (2015)). See Section 6.6 of Gómez-Rubio (2020) for a more detailed description on linear constraint methods in `INLA`.

Additionally, through cross-validation, we found that the performance of (S2) for the BTH data is improved when $\beta_{0,t} = 0$ for all time point t. Hence, the results presented in Table 2 of the manuscript and Table S1 of this document are obtained using the setting of $\beta_{0,t} = 0$ for all time point t.

S2.3. First-order spatiotemporal autoregression (STAR) model

We use the following STAR model (Blangiardo et al., 2013):

$$\begin{cases} y(\boldsymbol{s}) = \beta_0 + \beta_1 x_t(C_s) + \beta_2 t + \sum_{k=1}^4 \alpha_k z_{t,k}(C_s) + w_t(\boldsymbol{s}) + \varepsilon_t(\boldsymbol{s}) \\ w_t(\boldsymbol{s}) = \rho w_{t-1}(\boldsymbol{s}) + \eta_t(\boldsymbol{s}), \end{cases} \tag{S3}$$

where $\eta_t(s)$ follows a Gaussian process. Let $\boldsymbol{w}_t = \big(w_t(s_1), \cdots, w_t(s_n)\big)^T$, $\boldsymbol{\varepsilon}_t = \big(\varepsilon_t(s_1), \cdots, \varepsilon_t(s_n)\big)^T$, and $\boldsymbol{\eta}_t = \big(\eta_t(s_1), \cdots, \eta_t(s_n)\big)^T$. We assume that $\boldsymbol{w}_t$ and ε_t are mutually and serially independent, which also holds for $\varepsilon_t(\boldsymbol{s})$ and $\boldsymbol{\eta}_t$, and for $\boldsymbol{w}_{t-1}$ and $\boldsymbol{\eta}_t$. The initial process $\boldsymbol{w}_0 \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, and the random process $\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_\eta)$. The STAR is also implemented using the `R` package `INLA`.

In two INLA-based methods, STVC and STAR, we set the number of mesh nodes to be 882, which is comparable to the grid size of 788 in the proposed ADCM; see Figure S2 for the triangulated mesh of domain.

In INLA-based methods, the integrated nested Laplace approximation (INLA; see Rue et al. (2009)) is implemented in the R package `INLA` to solve the stochastic partial differential equation (SPDE; see Lindgren et al. (2011)). In the SPDE, the structure of the covariance function is closely related to a parameter α. Specifically, $\alpha = \nu + d/2$, where ν represents the smoothness parameter of the Matérn covariance family and d is the dimension of the domain of interest, as defined in Equation (2) of Lindgren et al. (2011). Although there are no Matérn covariance functions for $\nu = 0$, which corresponds to $\alpha = 1$, the solution for this case still exists in the SPDE approach, as discussed by Lindgren et al. (2011).

Based on the data analysis for the $PM_{2.5}$ datasets, the choice of covariance functions has a greater impact on the predictive performance of STVC and STAR models, compared to other components such as approximation strategies, integration methods, priors, etc. In this analysis, the covariance function is determined through a cross-validation procedure, while the remaining components are fixed using default settings in the `INLA` package. The cross-validation procedure is performed for each value of $\alpha \in \{0.5, 1, 1.5, 2\}$. Based on the cross-validation results, $\alpha = 1$ is predetermined as the optimal choice for both STVC and STAR.

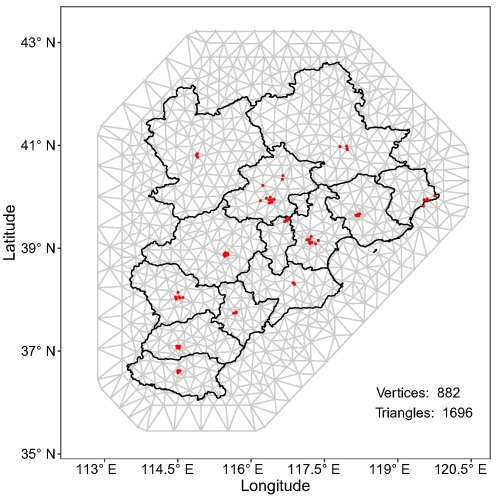

Figure S2: Triangulated mesh of the BTH region and extended area which is used to avoid boundary effects.

S2.4. Additive model (ADM)

We use the following addictive model:

$$y_t(\boldsymbol{s}) = \beta_0 + \beta_1 x_{t,1}(\boldsymbol{C_s}) + g_0(t) + \sum_{k=1}^{4} g_k\Big(z_{t,k}(\boldsymbol{C_s})\Big) + g_5\big(s^x, s^y, t\big) + \epsilon_t(s), \tag{S4}$$

where s^x and s^y represent longitude and latitude, respectively. The functions $\{g_k(\cdot)\}_{k=0}^{5}$ are estimated by basis expansion $g_k\big(z_{t,k}\big) = \sum_{j=1}^{J_k} \alpha_{k,j}\phi_{k,j}$ with $J_k = 5$ for univariate functions $\{g_k(\cdot)\}_{k=0}^{4}$ and $J_5 = 180$ for the three-dimensional smooth function $g_5(\cdot)$. We use cyclic cubic-spline basis functions for $\{\phi_{k,j}\}_{k=0}^{4}$ and thin plate splines for $\phi_{5,j}$; see Wood (2017) for more discussions on additive models. The ADM is implemented using the `R` package `mgcv` (Wood, 2022).

S3. Model comparisons with the weighted CMAQ outputs as a single predictor

Following Berrocal et al. (2010), downscaler models commonly employ the CMAQ outputs as a single predictor for real-world applications such as the SVC (Berrocal et al., 2020). As a result, all models in this section consider the weighted CMAQ outputs as a single predictor. Notably, the time variable t and four meteorological variables, $z_{t,1}(\boldsymbol{C_s}), \cdots, z_{t,4}(\boldsymbol{C_s})$, have been excluded from the mean structure of all models. Additionally, for the SVC, a Bayesian SVC (BSVC) is also taken into consideration by Berrocal et al. (2010), along with maximum likelihood estimation (MLE) methods. In the case of the BSVC, 50,000 MCMC samples were generated from the posterior distribution, with a burn-in period of 50,000 iterations.

The results of the leave-one-city-out cross-validation (LOCOCV) are presented in Table S1. Compared to all other models, the ADCM with the single predictor improves the performance by at least 7.82% in average RMSE and 7.13% in average CRPS.

Table S1
Averaged RMSE and CRPS of $PM_{2.5}$ concentration predictions ($\mu g/m^3$) at 13 cities in the BTH region using LOCOCV from five models: (1) Bayesian spatially-varying coefficient downscaling model (BSVC[1]), (2) spatially-varying coefficient downscaling model (SVC) with the maximum likelihood estimation, (3) spatiotemporally-varying coefficient downscaling model (STVC), (4) first-order spatiotemporal autoregression (STAR), (5) additive model (ADM), and (6) proposed additive dynamic calibration model (ADCM). The smallest RMSE and CRPS are in bold, and the second smallest ones are underlined. All models with the weighted CMAQ outputs as a single predictor. Daily data from November 1, 2015 to January 31, 2016 are considered.

City	RMSE						CRPS					
	BSVC	SVC	STVC	STAR	ADM	ADCM	BSVC	SVC	STVC	STAR	ADM	ADCM
Baoding	75.53	77.05	77.24	77.50	**73.13**	74.26	41.25	41.85	41.66	41.82	40.87	**40.18**
Beijing	61.74	62.68	60.23	63.17	69.39	**58.19**	30.36	30.69	30.21	30.95	33.24	**30.04**
Cangzhou	35.63	34.47	**30.68**	32.79	43.89	36.51	25.67	23.93	22.81	23.59	29.13	**21.78**
Chengde	34.87	32.04	28.54	**26.98**	32.32	29.66	19.49	18.43	16.99	**16.37**	18.46	16.86
Handan	58.24	54.64	55.06	56.79	54.27	**44.64**	37.24	35.23	35.62	36.24	33.20	**29.57**
Hengshui	57.85	64.93	58.43	58.11	61.20	**51.85**	32.26	33.87	31.48	31.58	34.59	**30.94**
Langfang	**34.81**	36.40	35.33	38.47	42.44	35.60	26.96	26.92	27.27	**26.40**	27.60	28.39
Qinhuangdao	35.29	38.99	33.00	41.07	36.15	**25.99**	20.79	22.59	19.31	23.79	19.18	**14.09**
Shijiazhuang	49.45	49.19	47.80	**46.28**	68.10	53.08	33.10	32.86	31.46	**30.28**	41.51	34.19
Tangshan	32.79	30.47	36.70	**29.58**	34.43	33.58	21.50	21.32	21.40	**19.82**	22.42	20.95
Tianjin	32.52	30.76	**29.95**	32.93	46.67	31.83	24.56	**23.50**	23.51	24.01	29.71	24.41
Xingtai	53.17	53.98	54.36	54.03	55.57	**50.85**	31.39	31.57	31.66	**30.98**	32.16	31.67
Zhangjiakou	79.00	66.76	60.88	73.86	83.54	**34.52**	43.31	36.54	35.08	41.42	46.47	**19.11**
Average	49.30	48.64	46.78	48.58	53.93	**43.12**	29.84	29.18	28.34	29.02	31.43	**26.32**

[1]BSVC is implemented through the `R` package `spBayes` (https://cran.r-project.org/web/packages/spBayes/; see Finley and Banerjee (2020)).

S4. Additional comparisons with simulated data

To assess the computational efficiency of our method, we compare the running time of the ADCM with the benchmark SVC based on the MLE and STAR based on `INLA` by using simulated data. Especially, another SVC with the sparse constraint for the tapering covariance (SVCt) is also included, where the tapering distance in space is set as $0.6 \max \|\boldsymbol{s}_i - \boldsymbol{s}_j\|$, here $i, j = 1, 2, \cdots, n$. We use two resolutions for the ADCM and set $N_v = 188 + 680 = 868$ for the simulation. To a fair comparison, we set the dimension for a triangulation mesh of the STAR to be 882, which is comparable to the grid size of the ADCM. The simulated data are generated from the following model:

$$y_t(\boldsymbol{s}) = \beta_0 + \beta_1 x_{1,t}(\boldsymbol{s}) + \beta_2 x_{2,t}(\boldsymbol{s}) + w_t(\boldsymbol{s}) + \varepsilon_t(\boldsymbol{s}), t = 0, 1/29, \cdots, 1. \tag{S5}$$

All settings of (S5) are the same as (14) of the manuscript, except that (S5) removes the nonlinear functions $g(\cdot)$ in Equation (14) of the manuscript. Under each setting, we repeat the procedure 50 times and report the results in Table S2. To summarize:

- The ADCM is much more computationally efficient than all other methods when the number of spatial points (i.e., n) is large. For example, when $n = 1{,}000$, the ADCM is almost 20 times faster than the SVC, 10 times faster than the INLA, and 3 times faster than the SVCt though SVCt's covariance matrix is tapered for the sparsity.
- The MSE of the regression coefficients $\boldsymbol{\beta}$ and the predictions of the ADCM are the best among all methods, in almost all cases.
- The computation of the SVC is much more sensitive to the number of spatial points (i.e., n) than the STAR and ADCM. When n increases from 300 to 1,000, the SVC takes 22.4 times more time while STAR and ADCM take about the same amount of time.

Table S2
The biases, standard deviations (SD), and mean squared errors (MSEs) of the regression coefficients, and predictions when the training spatial sample size $n = 300$ and 1,000, along with the total running time (seconds). The results are based on 50 independent simulations and obtained using three models: (1) the spatially-varying coefficient downscaling model (SVC) through an MLE method, (2) SVC with a tapering method (SVCt), (3) the first-order spatiotemporal autoregression (STAR) through an INLA method, and (4) the proposed additive dynamic calibration model (ADCM) through an EM method. The lowest running time and MSE are in bold.

n	Model	Running time	Criterion	$\beta_0(=15)$	$\beta_1(=1)$	$\beta_2(=1)$	Predction
300	SVC	17,960	Bias	-0.19967	-0.00034	-0.00735	
			SD	0.03124	0.00417	0.00714	
			MSE	0.04084	0.00002	0.00011	0.13195
	SVCt	**4,387**	Bias	-0.01476	0.00657	-0.05421	
			SD	0.03487	0.00950	0.01537	
			MSE	0.00143	0.00013	0.00318	1.07103
	STAR	157,032	Bias	-0.10150	-0.00034	-0.00705	
			SD	0.01250	0.00415	0.00623	
			MSE	0.01046	0.00002	0.00009	0.12527
	ADCM	19,818	Bias	0.00309	0.00047	0.00006	
			SD	0.03302	0.00326	0.00802	
			MSE	**0.00110**	**0.00001**	**0.00006**	**0.12488**
1,000	SVC	403,069	Bias	-0.16581	-0.00007	-0.00635	
			SD	0.01651	0.00124	0.00299	
			MSE	0.02776	**0.00000**	0.00005	0.11431
	SVCt	64,641	Bias	-0.01591	0.00737	-0.05188	
			SD	0.01262	0.00462	0.00607	
			MSE	0.00041	0.00008	0.00273	1.06329
	STAR	166,852	Bias	-0.09423	-0.00019	-0.00625	
			SD	0.00565	0.00124	0.00261	
			MSE	0.00891	0.00000	0.00005	0.11072
	ADCM	**21,779**	Bias	0.00292	0.00016	0.00025	
			SD	0.01343	0.00204	0.00554	
			MSE	**0.00019**	0.00000	**0.00003**	**0.11064**

S5. A sensitivity analysis for the ADCM for different specifications of the precision matrices of IDE models

In this section, we carry out a sensitivity analysis based on the cross-validation of the BTH data by using six different specifications of the precision matrices $\boldsymbol{Q}(\cdot)$ in the ADCM:

C0 $\boldsymbol{Q}_0 = \tau_0^2\left(\boldsymbol{G} + \zeta_0^2\boldsymbol{I}\right)$ and $\boldsymbol{Q} = \tau^2\left(\boldsymbol{G} + \zeta^2\boldsymbol{I}\right)$, the precision matrix adopted in Section 4.1 of the manuscript.

C1 $\boldsymbol{Q}_0 = \tau_0^2\boldsymbol{I}$ and $\boldsymbol{Q} = \tau^2\boldsymbol{I}$.

C2 $\boldsymbol{Q}_0 = \tau_0^2\left(\boldsymbol{I} - \frac{\zeta_0}{\lambda_{\max}}\boldsymbol{G}\right)$ and $\boldsymbol{Q} = \tau^2\left(\boldsymbol{I} - \frac{\zeta}{\lambda_{\max}}\boldsymbol{G}\right)$ for a conditional autoregressive (CAR) specification (see Table 1 of Bivand et al. (2015) or Section 8.3.2 of Blangiardo and Cameletti (2015)), where $\zeta_0, \zeta \in [0, 1)$ and $\lambda_{\max}$ is the maximum eigenvalue of $\boldsymbol{G}$.

C3 Considering $\boldsymbol{Q}_0 = \boldsymbol{I} - \frac{\zeta_0}{\lambda_{\max}}\boldsymbol{G}$ and $\boldsymbol{Q} = \boldsymbol{I} - \frac{\zeta}{\lambda_{\max}}\boldsymbol{G}$ by setting two scale parameters of Case 2 to $\tau_0^2 = 1$ and $\tau^2 = 1$, corresponding to Equation (13) of Bivand et al. (2015); see also Section 3.3.5 of Gómez-Rubio (2020) and references therein.

Table S3
Averaged RMSE and total running time of $PM_{2.5}$ concentration predictions ($\mu g/m^3$) at 13 cities in the BTH region using LOCOCV through the ADCM with six specifications of the precision matrices: **C0** The specification employed in the manuscript, **C1** Diagonal matrix, **C2** Conditional autoregressive (CAR) with two parameters, **C3** Conditional autoregressive (CAR) with the scale parameter of 1, **C4** Leroux model, and **C5** Simultaneous autoregressive (SAR). The smallest RMSE and running time are in bold, and the second smallest ones are underlined. Daily data from November 1, 2015 to January 31, 2016 are considered.

City	RMSE						City	Running time (Seconds)					
	C0	**C1**	**C2**	**C3**	**C4**	**C5**		**C0**	**C1**	**C2**	**C3**	**C4**	**C5**
Baoding	**73.96**	77.21	77.36	81.16	79.30	77.43	Baoding	948	528	584	589	**436**	824
Beijing	**54.23**	55.63	55.44	59.21	55.50	55.39	Beijing	824	493	527	563	**475**	744
Cangzhou	34.91	35.17	**33.43**	37.41	33.79	33.55	Cangzhou	849	579	706	617	**511**	933
Chengde	**26.83**	29.46	29.23	30.10	28.52	29.44	Chengde	837	**462**	539	576	496	714
Handan	48.11	46.14	45.34	46.61	45.90	**45.26**	Handan	886	590	572	602	**471**	797
Hengshui	**56.11**	63.17	62.73	69.46	64.17	62.68	Hengshui	776	693	587	598	**514**	768
Langfang	**36.21**	36.87	38.30	40.64	37.47	38.33	Langfang	601	585	582	616	**486**	797
Qinhuangdao	23.60	**22.53**	23.57	24.81	23.61	22.54	Qinhuangdao	662	**473**	640	582	478	708
Shijiazhuang	53.07	51.52	50.36	53.95	51.78	**50.33**	Shijiazhuang	640	758	575	584	**460**	766
Tangshan	32.94	33.92	**32.33**	35.83	33.68	33.54	Tangshan	535	707	600	575	**447**	865
Tianjin	32.84	**30.68**	31.09	32.73	31.25	30.73	Tianjin	507	513	532	545	**466**	664
Xingtai	**52.98**	55.88	54.93	57.51	56.44	54.93	Xingtai	552	570	495	532	**456**	626
Zhangjiakou	**27.46**	27.99	27.99	29.73	30.36	27.94	Zhangjiakou	668	525	585	517	**494**	757
Average	**42.56**	43.55	43.24	46.09	43.98	43.24	**Total**	9,285	7,477	7,524	7,496	**6,189**	9,964

C4 $\boldsymbol{Q}_0 = \tau_0^2\Big((1-\zeta_0)\boldsymbol{I} - \zeta_0\boldsymbol{G}\Big)$ and $\boldsymbol{Q} = \tau^2\Big((1-\zeta)\boldsymbol{I} - \zeta\boldsymbol{G}\Big)$ for the Leroux model specification (Leroux et al., 2000), where $\zeta_0, \zeta \in [0, 1]$. See also Section 5.1 of Blangiardo and Cameletti (2015).

C5 $\boldsymbol{Q}_0 = \tau_0^2\big(\boldsymbol{I} - \zeta_0\boldsymbol{G}\big)^T\big(\boldsymbol{I} - \zeta_0\boldsymbol{G}\big)$ and $\boldsymbol{Q} = \tau^2\big(\boldsymbol{I} - \zeta\boldsymbol{G}\big)^T\big(\boldsymbol{I} - \zeta\boldsymbol{G}\big)$ for a simultaneous autoregressive (SAR) specification; see Section 5.2 of Bivand et al. (2015).

Table S3 demonstrates the ADCM for all specifications of the precision matrix has the lowest average RMSE compared to other methods presented in Table 2 of the manuscript. Table S3 also shows that the ADCM is robust for $\boldsymbol{Q}(\cdot)$ in all cases. Compared to other cases, the ADCM for a precision matrix of **C3** is worse for the predictions of the BTH $PM_{2.5}$ data, which demonstrates the importance of the scale parameters, τ_0^2 and τ^2. Compared to **C1**−**C5**, **C0** has the best prediction performance for the BTH data using the average RMSE.

Moreover, although **C1** and **C3** have only one free parameter while the other cases have two parameters, the computations of **C1** and **C3** are not the most efficient as shown in the running time in Table S3, mostly due to the number of iterations involved in the computation. The calculations of τ^2 and ζ for the BTH data are not difficult because of the sparsity and the small dimensions of $\boldsymbol{Q}$ and $\boldsymbol{M}$ at each resolution.

References

Berrocal, V.J., Gelfand, A.E., Holland, D.M., 2010. A spatio-temporal downscaler for output from numerical models. Journal of Agricultural, Biological and Environmental Statistics 15, 176–197.

Berrocal, V.J., Guan, Y., Muyskens, A., Wang, H., Reich, B.J., Mulholland, J.A., Chang, H.H., 2020. A comparison of statistical and machine learning methods for creating national daily maps of ambient $PM_{2.5}$ concentration. Atmospheric Environment 222, 117–130.

Bivand, R., Gómez-Rubio, V., Rue, H., 2015. Spatial data analysis with R-INLA with some extensions. Journal of Statistical Software 63, 1–31.

Blangiardo, M., Cameletti, M., 2015. Spatial and Spatio-temporal Bayesian Models with R-INLA. John Wiley & Sons, New York.

Blangiardo, M., Cameletti, M., Baio, G., Rue, H., 2013. Spatial and spatio-temporal models with R-INLA. Spatial and Spatio-temporal Epidemiology 4, 33–49.

Cressie, N., Wikle, C.K., 2011. Statistics for Spatio-temporal Data. John Wiley & Sons, Hoboken, N.J.

Dambon, J.A., Sigrist, F., Furrer, R., 2021a. Maximum likelihood estimation of spatially varying coefficient models for large data with an application to real estate price prediction. Spatial Statistics 41, 100470.

Dambon, J.A., Sigrist, F., Furrer, R., 2021b. varycoef: An R package for Gaussian process-based spatially varying coefficient models. arXiv preprint arXiv:2106.02364 .

Finley, A.O., Banerjee, S., 2020. Bayesian spatially varying coefficient models in the spbayes r package. Environmental Modelling & Software 125, 104608.

Gómez-Rubio, V., 2020. Bayesian inference with INLA. CRC Press.

Leroux, B.G., Lei, X., Breslow, N., 2000. Estimation of disease rates in small areas: A new mixed model for spatial dependence. In Statistical Models in Epidemiology, the Environment, and Clinical Trials, pages 179–191. Springer.

Lindgren, F., Rue, H., 2015. Bayesian spatial modelling with R-INLA. Journal of Statistical Software 63, 1–25.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 73, 423–498.

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 71, 319–392.

Sahu, S.K., Gelfand, A.E., Holland, D.M., 2006. Spatio-temporal modeling of fine particulate matter. Journal of Agricultural, Biological and Environmental Statistics 11, 61–86.

Wood, S.N., 2017. Generalized Additive Models: An Introduction with R. Chapman & Hall/CRC, New York.

Wood, S.N., 2022. Package 'mgcv' R package version 1.8-41. https://CRAN.R-project.org/package=mgcv.