Supplementary Document for A Bayesian Spatio-Temporal Geostatistical Model with an Auxiliary Lattice for Large Datasets

Ganggang Xu[1], Faming Liang[1] and Marc G. Genton[2]

[1] *Texas A&M University and* [2]*King Abdullah University of Science and Technology*

1. Proof of Proposition 1

Using (4), if all eigenvalues of $\mathbf{\Phi}(\boldsymbol{\beta})$ have absolute value less than 1, then one has the representation

$$\vec{U}_t = \sum_{j=0}^{\infty} \mathbf{\Phi}(\boldsymbol{\beta})^j \vec{\boldsymbol{\zeta}}_{t-j},$$

which implies that the variance of $\vec{U}_t$ is

$$\mathbf{\Sigma}_U(\boldsymbol{\beta}) = \sigma_q^2 \sum_{j=0}^{\infty} \mathbf{\Phi}(\boldsymbol{\beta})^j \mathbf{\Lambda}^{-1}(\boldsymbol{\beta}) \mathbf{\Phi}(\boldsymbol{\beta})^j.$$

By definition, $\mathbf{\Phi}(\boldsymbol{\beta})$ and $\mathbf{\Lambda}(\boldsymbol{\beta})$ are symmetric block circulant matrices, that have the same set of eigenvectors with all eigenvalues real (Rue and Held (2005), Chapter 2). Block circulant matrices also have the property that $\mathbf{\Phi}(\boldsymbol{\beta})\mathbf{\Lambda}(\boldsymbol{\beta}) = \mathbf{\Lambda}(\boldsymbol{\beta})\mathbf{\Phi}(\boldsymbol{\beta})$. Then, straightforward algebra yields

$$\mathbf{\Sigma}_U(\boldsymbol{\beta}) = \sigma_q^2 \{(\mathbf{I} - \mathbf{\Phi}^2(\boldsymbol{\beta}))\mathbf{\Lambda}(\boldsymbol{\beta})\}^{-1} = \sigma_q^2 \mathbf{Q}^{-1}(\boldsymbol{\beta}),$$

which gives $\mathbf{Q}(\boldsymbol{\beta}) = (\mathbf{I} - \mathbf{\Phi}^2(\boldsymbol{\beta}))\mathbf{\Lambda}(\boldsymbol{\beta})$. □

2. MCMC algorithm details

Given an odd number K, let $\mathbf{\Phi}_K(\boldsymbol{\beta})$ and $\mathbf{\Lambda}_K(\boldsymbol{\beta})$ be the matrices obtained by setting $m_1 = m_2 = K$ in (6) and (7), respectively. Take $d_K = (K-1)/2$,

$$\vec{U}_{(-kl)t,\partial K} = (U_{(k-d_K)(l-d_K),t}, U_{(k-d_K)(l-d_K+1),t}, \ldots, \underbrace{0}_{kl,t}, \ldots, U_{(k+d_K)(l+d_K),t})^{\mathrm{T}},$$

$$\vec{U}_t^{(1)} = \mathbf{\Lambda}(\boldsymbol{\beta})\vec{U}_t, \quad \vec{U}_t^{(2)} = \mathbf{\Phi}(\boldsymbol{\beta})\vec{U}_t, \quad \vec{U}_t^{(3)} = \mathbf{\Lambda}(\boldsymbol{\beta})\mathbf{\Phi}(\boldsymbol{\beta})\vec{U}_t,$$

where we can show that the entry corresponding to the (k,l)th location in the lattice W is

$$\begin{aligned}
\vec{U}_{kl,t}^{(1)} &= U_{kl,t} + \beta_{010}(U_{k(l+1),t} + U_{k(l-1),t}) + \beta_{100}(U_{(k+1)l,t} + U_{(k-1)l,t}) \\
&\quad + \beta_{110}(U_{(k+1)(l+1),t} + U_{(k-1)(l-1),t} + U_{(k-1)(l+1),t} + U_{(k+1)(l-1),t}), \\
\vec{U}_{kl,t}^{(2)} &= \beta_{001}U_{kl,t} + \beta_{011}(U_{k(l+1),t} + U_{k(l-1),t}) + \beta_{101}(U_{(k+1)l,t} + U_{(k-1)l,t}) \\
&\quad + \beta_{111}(U_{(k+1)(l+1),t} + U_{(k-1)(l-1),t} + U_{(k-1)(l+1),t} + U_{(k+1)(l-1),t}), \\
\vec{U}_{kl,t}^{(3)} &= U_{kl,t}^{(2)} + \beta_{010}(U_{k(l+1),t}^{(2)} + U_{k(l-1),t}^{(2)}) + \beta_{100}(U_{(k+1)l,t}^{(2)} + U_{(k-1)l,t}^{(2)}) \\
&\quad + \beta_{110}(U_{(k+1)(l+1),t}^{(2)} + U_{(k-1)(l-1),t}^{(2)} + U_{(k-1)(l+1),t}^{(2)} + U_{(k+1)(l-1),t}^{(2)}),
\end{aligned}$$

for $k = 1, \ldots, m_1$, $l = 1, \ldots, m_2$, $t = 1, \ldots, T$.

To simulate samples from (19), we treat $\mathbf{U}_1, \ldots, \mathbf{U}_T$ as missing data and use a Gibbs sampler as a data augmentation tool. The rest of the parameters are updated one by one using the Metropolis-Hasting (MH) algorithm. To facilitate the sampling, the parameters are grouped into subgroups after appropriate transformations: $\boldsymbol{\theta}_1 = (\boldsymbol{\beta}, \log\sigma_q^2)$, $\boldsymbol{\theta}_2 = (\log\phi, \boldsymbol{\xi}, \log\sigma_e^2)$. With samples at the bth iteration, say $\boldsymbol{\theta}_1^b$, $\boldsymbol{\theta}_2^b$ and $\mathbf{U}_1^b, \ldots, \mathbf{U}_T^b$, let $\boldsymbol{\theta}_1^*, \boldsymbol{\theta}_2^*$ and $\mathbf{U}_1^*, \ldots, \mathbf{U}_T^*$ denote the samples of $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$, and $\mathbf{U}_1, \ldots \mathbf{U}_T$, at the $(b+1)$th iteration. If we suppress the subscript b, the sampling process can be described as follows.

1. Update $(\mathbf{U}_1, \ldots, \mathbf{U}_t)$: Generate $U_{kl,t}^*$ using the Gibbs sampler from the conditional density $f(U_{kl,t}|\mathbf{U}_{-(kl,t)})$, where $\mathbf{U}_{-(kl,t)}$ denotes the set of all elements of $(\mathbf{U}_1, \ldots, \mathbf{U}_t)$ except for $U_{kl,t}$. With extensive algebra, we can show that$U_{kl,t}^*|\mathbf{U}_{-(kl,t)}, \boldsymbol{\theta}, \mathbf{Y} \sim N(E_\partial^2 F_\partial, E_\partial^2)$, where

$$E_\partial^2 = \left\{ \frac{1}{\sigma_q^2}(1+h) + \sum_{i=1}^{n_t} I(w_{kl} \in \partial\boldsymbol{s}_{it}) \frac{c^2(\boldsymbol{\beta})[(\boldsymbol{r}_{\partial\boldsymbol{s}_{it}}^{\mathrm{T}}\mathbf{\Sigma}_\partial^{-1}(\boldsymbol{\beta}))]_{i(kl)}^2}{\sigma_e^2 + \sigma_{it\partial}^2} \right\}^{-1},$$

$$
\begin{aligned}
F_\partial = &\left\{ \frac{1}{\sigma_q^2} \left(-\boldsymbol{b}_1^{\mathrm{T}} \vec{U}_{(-kl)t,\partial 3} - \boldsymbol{b}_2^{\mathrm{T}} \vec{U}_{(-kl)t,\partial 7} + \vec{U}^{(3)}_{kl,(t+1)} + \vec{U}^{(3)}_{kl,(t-1)} \right) \right\} \\
&+ \sum_{i=1}^{n_t} I(w_{klt} \in \partial \boldsymbol{s}_{it}) \frac{c(\boldsymbol{\beta})[(\boldsymbol{r}^{\mathrm{T}}_{\partial \boldsymbol{s}_{it}} \boldsymbol{\Sigma}_\partial^{-1}(\boldsymbol{\beta}))]_{i(kl)}}{\sigma_e^2 + \sigma_{it\partial}^2} \\
&\qquad \times \{ Y(\boldsymbol{s}_{it}, t) - \mu(\boldsymbol{s}_{it}, t) - c(\boldsymbol{\beta}) \sum_{ab \neq kl} [\boldsymbol{r}^{\mathrm{T}}_{\partial \boldsymbol{s}_{it}} \boldsymbol{\Sigma}_\partial^{-1}(\boldsymbol{\beta})]_{i(ab)} U_{ab,t} \}.
\end{aligned}
$$

Here $i(kl)$ signifies that $U^*_{kl,t}$ is the $i(kl)$th element of $\vec{U}_{\partial \boldsymbol{s}_{it},t}$, $[(\boldsymbol{r}^{\mathrm{T}}_{\partial \boldsymbol{s}_{it}} \boldsymbol{\Sigma}_\partial^{-1}(\boldsymbol{\beta}))]_{i(kl)}$ is the $i(kl)$th entry of the vector $\boldsymbol{r}^{\mathrm{T}}_{\partial \boldsymbol{s}_{it}} \boldsymbol{\Sigma}_\partial^{-1}(\boldsymbol{\beta})$, h is the $(1,1)$th entry of $\mathbf{B}_7(\boldsymbol{\beta}) = \boldsymbol{\Phi}_7(\boldsymbol{\beta}_1)\boldsymbol{\Lambda}_7(\boldsymbol{\beta})\boldsymbol{\Phi}_7(\boldsymbol{\beta})$, and $\boldsymbol{b}_1$ and $\boldsymbol{b}_2$ are the 5th and 25th columns of $\boldsymbol{\Lambda}_3(\boldsymbol{\beta})$ and $\mathbf{B}_7(\boldsymbol{\beta})$, respectively.

2. Update $\boldsymbol{\theta}_1$: Generate $\boldsymbol{\theta}_1^*$ using the MH algorithm from the density

$$
f(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2, \mathbf{U}, \mathbf{Y}) \propto \pi(\boldsymbol{\theta}_1 | \boldsymbol{\theta}_2) \prod_{t=1}^{\mathrm{T}} f(\mathbf{U}_t | \boldsymbol{\theta}_1) f(\mathbf{Y}_t | \mathbf{U}_t, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2).
$$

3. Update $\boldsymbol{\theta}_2$: Generate $\boldsymbol{\theta}_2^*$ using the MH algorithm from the density

$$
f(\boldsymbol{\theta}_2 | \boldsymbol{\theta}_1, \mathbf{U}, \mathbf{Y}) \propto \pi(\boldsymbol{\theta}_2 | \boldsymbol{\theta}_1) \prod_{t=1}^{\mathrm{T}} f(\mathbf{Y}_t | \mathbf{U}_t, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2).
$$

Remark. Since the number of parameters involved in the MCMC algorithm is not small (10 in the simplest case), it is important to start the sampling process with good initial values so that the Markov chain converges quickly and generates well-mixed samples. We recommend setting $\boldsymbol{\Phi}(\boldsymbol{\beta}) = \mathbf{0}$ and using data from only one time point to get rough estimates of the remaining parameters, that are then used as the initial values for the spatio-temporal hierarchical model. While the computation time of this step is negligible compared with that of the spatio-temporal kriging, our numerical examples indicate that initial values obtained in this way can greatly reduce the convergence time of the Markov chain while stabilizing the prediction performance.

3. Additional Numerical Results

Neighborhoods	Case I ($\phi = 8$) 2×2	Case I ($\phi = 8$) 3×3	Case II ($\phi = 10$) 2×2	Case II ($\phi = 10$) 3×3
	Bias(SE)	Bias(SE)	Bias(SE)	Bias(SE)
$\boldsymbol{\beta}_{010}$	-0.012(0.035)	-0.027(0.058)	-0.037(0.032)	-0.049(0.054)
$\boldsymbol{\beta}_{100}$	-0.019(0.050)	-0.037(0.058)	-0.025(0.035)	-0.061(0.049)
$\boldsymbol{\beta}_{110}$	0.018(0.032)	0.051(0.051)	0.028(0.026)	0.053(0.039)
$\boldsymbol{\beta}_{001}$	-0.015(0.048)	0.014(0.051)	-0.001(0.071)	0.013(0.089)
$\boldsymbol{\beta}_{011}$	0.011(0.035)	-0.015(0.037)	0.001(0.042)	-0.021(0.048)
$\boldsymbol{\beta}_{101}$	0.012(0.027)	-0.010(0.037)	0.003(0.047)	-0.019(0.049)
$\boldsymbol{\beta}_{111}$	-0.006(0.023)	-0.014(0.026)	-0.003(0.030)	0.013(0.027)
ϕ	0.108(0.500)	-0.184(0.335)	0.013(0.288)	-0.092(0.574)
σ_q^2	-0.127(0.370)	-0.169(0.333)	-0.262(0.157)	-0.374(0.342)
σ_e^2	-0.002(0.246)	0.035(0.173)	0.018(0.160)	-0.022(0.145)

Table S1: The bias and standard error of parameter estimates over 20 simulated datasets.

	$\phi = 20$			
	$\tau = 2$		$\tau = 5$	
Neighborhoods	2×2	3×3	2×2	3×3
$\boldsymbol{\beta}_{010}$	-0.09(0.05)	-0.24(0.07)	-0.10(0.07)	-0.27(0.05)
$\boldsymbol{\beta}_{100}$	-0.10(0.06)	-0.25(0.05)	-0.08(0.08)	-0.25(0.06)
$\boldsymbol{\beta}_{110}$	-0.15(0.03)	0.00(0.05)	-0.16(0.04)	0.01(0.05)
$\boldsymbol{\beta}_{001}$	0.56(0.04)	0.57(0.04)	0.77(0.04)	0.79(0.02)
$\boldsymbol{\beta}_{011}$	0.02(0.03)	0.05(0.03)	0.02(0.02)	0.02(0.02)
$\boldsymbol{\beta}_{101}$	0.03(0.03)	0.05(0.02)	0.02(0.03)	0.03(0.02)
$\boldsymbol{\beta}_{111}$	0.00(0.02)	-0.03(0.02)	0.00(0.01)	-0.02(0.01)
ϕ	12.93(1.39)	13.61(1.25)	13.83(3.19)	13.18(1.47)
σ_0^2	7.4(1.02)	7.07(0.91)	7.64(1.80)	7.07(0.75)
σ_e^2	0.79(0.10)	0.99(0.09)	0.76(0.10)	0.82(0.07)
MSPE(24×24)	2.30(0.12)	2.29(0.12)	2.08(0.12)	2.07(0.11)
MSPE(28×28)	2.25(0.11)	2.24(0.11)	1.99(0.10)	2.00(0.11)
MSPE(32×32)	2.22(0.11)	2.22(0.11)	1.95(0.11)	1.95(0.11)
MSPE(35×35)	2.21(0.11)	2.20(0.11)	1.92(0.11)	1.92(0.10)
MSPE(SP)	2.38(0.12)	2.38(0.12)	2.37(0.14)	2.37(0.14)
MSPE(SPT)	2.07(0.09)	2.07(0.09)	1.77(0.08)	1.77(0.08)
CPU (m)	18.63	24.15	20.76	26.86
	$\phi = 40$			
	$\tau = 2$		$\tau = 5$	
Neighborhoods	2×2	3×3	2×2	3×3
$\boldsymbol{\beta}_{010}$	-0.09(0.05)	-0.27(0.05)	-0.09(0.05)	-0.25(0.05)
$\boldsymbol{\beta}_{100}$	-0.09(0.05)	-0.25(0.06)	-0.08(0.05)	-0.26(0.06)
$\boldsymbol{\beta}_{110}$	-0.16(0.04)	0.01(0.04)	-0.16(0.03)	0.01(0.04)
$\boldsymbol{\beta}_{001}$	0.46(0.04)	0.50(0.05)	0.66(0.07)	0.69(0.05)
$\boldsymbol{\beta}_{011}$	0.07(0.03)	0.06(0.03)	0.04(0.03)	0.06(0.03)
$\boldsymbol{\beta}_{101}$	0.06(0.04)	0.07(0.03)	0.05(0.03)	0.05(0.03)
$\boldsymbol{\beta}_{111}$	0.00(0.02)	-0.02(0.02)	0.00(0.02)	-0.02(0.02)
ϕ	19.70(4.33)	21.41(3.93)	25.75(9.17)	21.32(3.48)
σ_0^2	6.73(1.37)	6.41(0.87)	8.18(2.45)	6.88(1.28)
σ_e^2	0.81(0.08)	0.93(0.10)	0.79(0.07)	0.74(0.08)
MSPE(24×24)	1.76(0.09)	1.75(0.09)	1.61(0.08)	1.60(0.08)
MSPE(28×28)	1.75(0.10)	1.75(0.09)	1.59(0.07)	1.58(0.08)
MSPE(32×32)	1.74(0.09)	1.74(0.09)	1.57(0.07)	1.57(0.08)
MSPE(35×35)	1.75(0.10)	1.74(0.09)	1.57(0.07)	1.57(0.07)
MSPE(SP)	1.80(0.07)	1.80(0.07)	1.77(0.11)	1.77(0.11)
MSPE(SPT)	1.66(0.07)	1.66(0.07)	1.48(0.07)	1.48(0.07)
CPU (m)	18.76	23.41	18.45	26.86

Table S2: The mean of the estimated parameters averaged over 20 simulated datasets when the covariance function is correctly specified. The numbers in the parentheses are the standard errors of the estimates. The size of the auxiliary lattice is 32×32.

	$\phi = 20$			
	$\tau = 2$		$\tau = 5$	
Neighborhoods	2×2	3×3	2×2	3×3
$\boldsymbol{\beta}_{010}$	-0.17(0.06)	-0.26(0.05)	-0.18(0.10)	-0.28(0.05)
$\boldsymbol{\beta}_{100}$	-0.18(0.07)	-0.26(0.05)	-0.17(0.10)	-0.27(0.06)
$\boldsymbol{\beta}_{110}$	-0.06(0.07)	0.03(0.04)	-0.06(0.09)	0.04(0.04)
$\boldsymbol{\beta}_{001}$	0.62(0.03)	0.62(0.03)	0.82(0.03)	0.82(0.05)
$\boldsymbol{\beta}_{011}$	0.00(0.03)	0.02(0.02)	0.00(0.02)	0.01(0.03)
$\boldsymbol{\beta}_{101}$	0.00(0.03)	0.02(0.02)	0.00(0.02)	0.01(0.03)
$\boldsymbol{\beta}_{111}$	0.01(0.01)	-0.02(0.01)	0.00(0.01)	-0.02(0.02)
ϕ	9.75(0.48)	9.32(0.43)	9.75(0.65)	9.07(3.48)
σ_0^2	7.65(0.37)	7.22(0.37)	7.53(0.66)	6.86(1.28)
σ_e^2	0.55(0.16)	0.72(0.15)	0.58(0.12)	0.61(0.08)
MSPE(24×24)	2.70(0.13)	2.70(0.13)	2.35(0.10)	2.34(0.10)
MSPE(28×28)	2.59(0.12)	2.60(0.14)	2.20(0.08)	2.21(0.08)
MSPE(32×32)	2.54(0.12)	2.54(0.13)	2.12(0.07)	2.13(0.08)
MSPE(35×35)	2.52(0.12)	2.50(0.12)	2.09(0.07)	2.10(0.07)
MSPE(SP)	2.92(0.13)	2.92(0.13)	2.91(0.13)	2.91(0.13)
MSPE(SPT)	2.40(0.12)	2.40(0.12)	1.97(0.06)	1.97(0.07)
CPU (m)	21.15	26.87	23.03	24.28
	$\phi = 40$			
	$\tau = 2$		$\tau = 5$	
Neighborhoods	2×2	3×3	2×2	3×3
$\boldsymbol{\beta}_{010}$	-0.07(0.05)	-0.26(0.07)	-0.10(0.04)	-0.25(0.05)
$\boldsymbol{\beta}_{100}$	-0.06(0.05)	-0.27(0.07)	-0.08(0.04)	-0.24(0.07)
$\boldsymbol{\beta}_{110}$	-0.18(0.03)	0.02(0.05)	-0.16(0.02)	0.00(0.05)
$\boldsymbol{\beta}_{001}$	0.41(0.04)	0.47(0.06)	0.64(0.05)	0.69(0.10)
$\boldsymbol{\beta}_{011}$	0.09(0.03)	0.06(0.05)	0.06(0.03)	0.03(0.02)
$\boldsymbol{\beta}_{101}$	0.09(0.03)	0.06(0.05)	0.06(0.03)	0.04(0.02)
$\boldsymbol{\beta}_{111}$	-0.02(0.02)	-0.01(0.03)	0.00(0.02)	0.00(0.05)
ϕ	15.73(2.61)	16.69(2.49)	21.88(5.17)	21.25(8.32)
σ_0^2	7.24(1.07)	6.93(1.15)	8.98(2.29)	8.88(3.69)
σ_e^2	0.61(0.06)	0.73(0.04)	0.63(0.10)	0.68(0.11)
MSPE(24×24)	1.73(0.08)	1.71(0.08)	1.65(0.09)	1.63(0.09)
MSPE(28×28)	1.73(0.09)	1.70(0.08)	1.61(0.09)	1.59(0.09)
MSPE(32×32)	1.74(0.09)	1.70(0.09)	1.60(0.09)	1.59(0.08)
MSPE(35×35)	1.74(0.10)	1.69(0.09)	1.61(0.09)	1.58(0.09)
MSPE(SP)	1.80(0.10)	1.80(0.10)	1.83(0.11)	1.83(0.11)
MSPE(SPT)	1.63(0.08)	1.63(0.08)	1.53(0.08)	1.53(0.08)
CPU (m)	18.22	21.30	15.23	21.50

Table S3: The mean of the estimated parameters averaged over 20 simulated datasets when the covariance function is mis-specified. The numbers in the parentheses are the standard errors of the estimates. The size of the auxiliary lattice is 32×32.

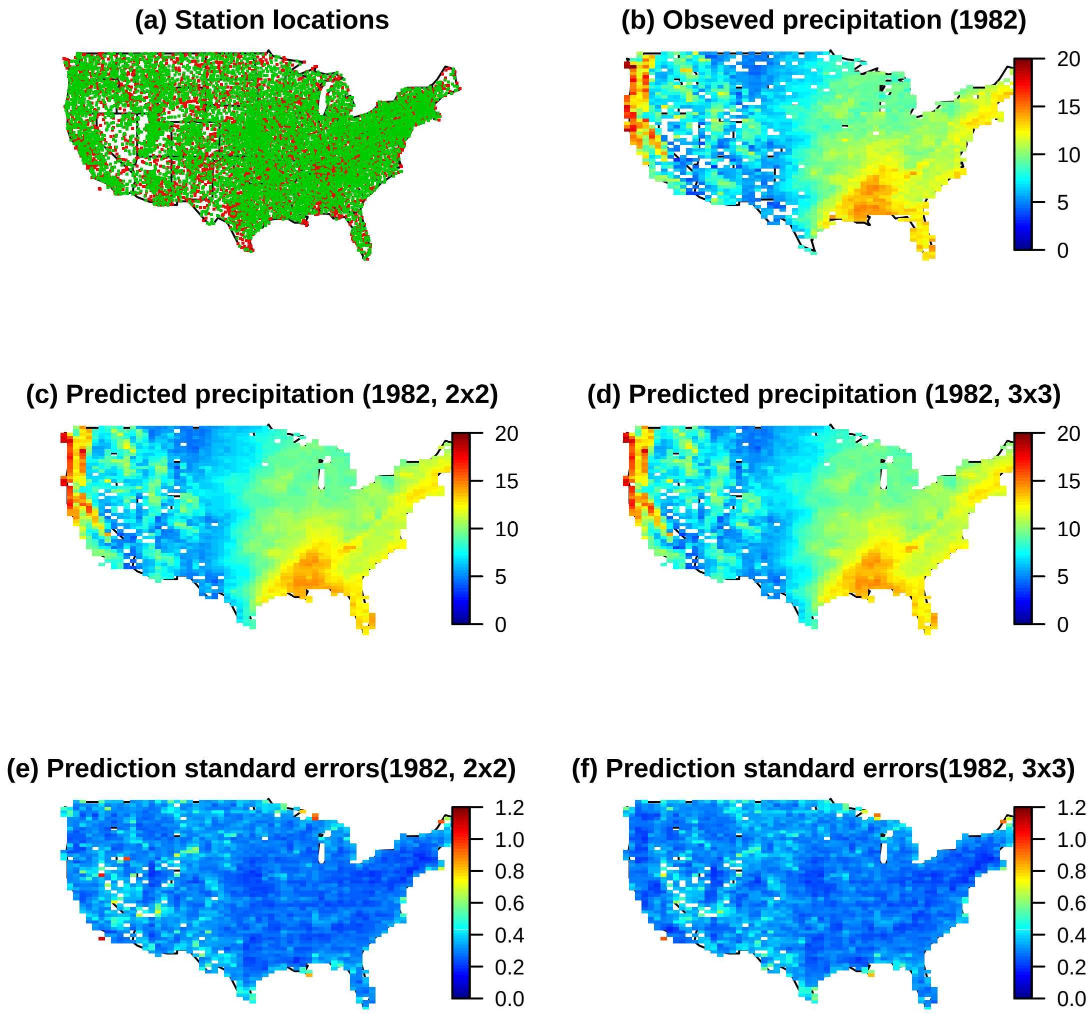

Figure S1: (a) Station locations: red dots are stations without observations; green dots are stations with observations. (b) Annual total precipitation reported from 6595 stations (stations with greed dots in Fig. 3(a)). (c) Predicted annual total precipitation for 11, 918 stations using a 2×2 neighborhood structure. (d) Predicted annual total precipitation for 11, 918 stations using a 3×3 neighborhood structure. (e) Estimated prediction standard errors for 11, 918 stations using a 2×2 neighborhood structure. (f) Estimated prediction standard errors for 11, 918 stations using a 3×3 neighborhood structure.

Institute for Applied Mathematics and Computational Science (IAMCS), Texas A&M University, College Station, TX 77843

E-mail: gang@stat.tamu.edu

Department of Statistics, Texas A&M University, College Station, TX 77843

E-mail: fliang@stat.tamu.edu

CEMSE Division, King Abdullah University of Science and Technology, Thuwal,23955-6900, Saudi Arabia.

E-mail: marc.genton@kaust.edu.sa