Full-Scale Approximations of Spatio-Temporal Covariance Models for Large Datasets

Bohai Zhang, Huiyan Sang, Jianhua Z. Huang

Texas A&M University

Supplementary Material

S1 Variations of the FSA method

S1.1 FSA for separable spatio-temporal covariance functions

In some cases, by taking advantage of the specific dependence structures, one can modify the general-purpose spatio-temporal FSA approach to achieve better covariance approximation and/or further reduce computational cost. For example, suppose $\Gamma_w(\mathbf{x}, \mathbf{x}') = \Gamma_u(\mathbf{s}, \mathbf{s}')\Gamma_v(t, t')$ where Γ_u and Γ_v are valid covariance functions in space and time domains, respectively, and that $n = N \times T$ observations are collected at spatial sites $\mathcal{S} = \{\mathbf{s}_1, \ldots, \mathbf{s}_N\}$ through time points set $\mathcal{T} = \{t_1, \cdots, t_T\}$. Then if we permute the observations by sorting the time in an increasing order, the covariance matrix of $w(\mathbf{s}, t)$ at $\mathcal{S} \times \mathcal{T}$ can be written as $\Sigma_w = \Sigma_v \otimes \Sigma_u$, where $\Sigma_u = [\Gamma_u(\mathbf{s}_i, \mathbf{s}_j)]_{i,j=1:N}$, $\Sigma_v = [\Gamma_v(t_i, t_j)]_{i,j=1:T}$, and $\otimes$ is the Kronecker product. If $\epsilon(\mathbf{s}, t) \equiv 0$, the resultant data covariance matrix $\Sigma_{\mathbb{Y}} = \Sigma_t \otimes \Sigma_{\mathbf{s}}$. The separability structure of $\Gamma_w(\mathbf{x}, \mathbf{x}')$ can alleviate the computational demand because it reduces the dimension of the data covariance matrices that need to be inverted. Specifically, $|\Sigma_{\mathbb{Y}}| = |\Sigma_v|^T |\Sigma_u|^N$ and $\Sigma_{\mathbb{Y}}^{-1} = \Sigma_v^{-1} \otimes \Sigma_u^{-1}$. However, computational challenge still exists in the presence of large spatial locations and/or time points. In this case, the FSA approach can be applied to approximate the spatial covariance Σ_u and the temporal covariance Σ_v respectively.

S1.2 Further improvement of computational efficiency by pre-tapering

Spatio-temporal datasets can be massive if they are observed at a large spatial domain during a long time period. Although the FSA approach is conceptually applicable to such datasets, direct application is impractical. A dense knot set is desirable to achieve a reasonable approximation, and the cross-covariance matrix. $\mathcal{C}(\mathscr{X}, \mathscr{X}^*)$ can bring challenges in matrix operations and storage. The computational issue is further complicated

by a full MCMC implementation requiring a large number of iterations. Here we propose to pre-taper the original spatio-temporal covariance function, and then apply our FSA approach to the tapered covariance.

Consider a separable tapering function $\mathcal{K}(\mathbf{x}, \mathbf{x}') = \mathcal{K}_u(\mathbf{s}, \mathbf{s}'; \gamma_u)\mathcal{K}_v(t, t'; \gamma_v)$, where $\mathcal{K}_u$ and $\mathcal{K}_v$ are, respectively, spatial and temporal tapering functions. We create a sparse approximation to the original covariance Γ_w as

$$\Gamma_{\tilde{w}}(\mathbf{x}, \mathbf{x}') = \Gamma_w(\mathbf{x}, \mathbf{x}')\mathcal{K}(\mathbf{x}, \mathbf{x}').$$

This leads to a spatio-temporal covariance matrix sparser than the original covariance matrix. If the tapering is done conservatively, the tapered covariance is expected to retain most of the spatio-temporal dependence information. Since this matrix still keeps a large number of non-zero entries due to the massive size of the original covariance, we further reduce the computational cost by applying the FSA to $\Gamma_{\tilde{w}}(\mathbf{x}, \mathbf{x}')$.

Let $\Sigma_{\tilde{w}_l} + \Sigma_{\tilde{w}_s}$ be the FSA covariance matrix applied to the pre-tapered covariance matrix. Due to the pre-tapering, $\Sigma_{\tilde{w}_l}$ takes a quadratic form that involves a sparse $n \times m$ cross-covariance matrix $\mathcal{C}(\mathscr{X}, \mathscr{X}^*)$ and the inverse of a sparse $m \times m$ matrix $\mathcal{C}^*$. This greatly alleviates the computational burden in matrix operations and storage by applying sparse matrix techniques.

S2 More simulation results

S2.1 Gneiting's model without nugget effect

We compared the methods assuming no nugget effect in the covariance model, but do not include the results of the predictive process model because, without nugget it leads to a low-rank covariance matrix that can't be inverted. The results are presented in Table 1. Again, the FSA-Block outperforms the independent blocks method in terms of prediction. Besides, comparing with the case with the nugget effect, the FSA-Block provides more accurate parameter estimation when there is no nugget effect. In particular for set-up 2, the estimates of a, η, and σ^2 by the FSA-Block are very close to those from the full covariance model.

S2.2 The Matérn covariance model

We experimented with $\nu = 0.5, 1, 2$ to investigate the performance of the FSA-Block method in terms of parameter estimations and predictions under different level of smoothness. Other particulars were the same as in the simulation study for Gneiting's nonseparable covariance model. The MLEs were obtained for all model parameters and we took $\nu \in (0, 3]$ when estimating it. The parameter estimations and prediction results are shown in Table 2.

Table 1: The means and MSEs (in parenthesis) of each parameter and MSPE results for the covariance model without nugget. The results are based on 100 runs of simulations.

Settings	Method	Mean and MSEs				MSPE
		a	c	η	σ^2	
Set-up 1		10	20	0.5	1	
	FM	10.16 (1.14)	20.32 (2.43)	0.53 (0.0373)	1.01 (0.0048)	0.33
	FSA-Block	10.52 (1.58)	22.55 (10.03)	0.46 (0.0433)	1.05 (0.0076)	0.37
	Block	9.76 (1.32)	19.81 (1.95)	0.48 (0.0649)	0.99 (0.0042)	0.42
Set-up 2		5	10	0.5	1	
	FM	5.01 (0.27)	10.01 (0.34)	0.51 (0.0311)	1.00 (0.0021)	0.59
	FSA-Block	5.11 (0.28)	10.67 (0.94)	0.53 (0.0379)	1.01 (0.0021)	0.62
	Block	4.90 (0.28)	9.98 (0.47)	0.47 (0.0501)	0.99 (0.0024)	0.66

Table 2: The means and MSEs (in parenthesis) of each parameter and MSPE results for Matérn's covariance model. The results are based on 100 runs of simulations.

Settings	Method	Mean and MSEs					MSPE
		ϕ_t	ϕ_s	ν	σ^2	τ^2	
Set-up 1		5	10	0.5	1	0.01	
	FM	4.65 (0.68)	9.23 (2.45)	0.54 (0.0041)	0.96 (0.0102)	0.02 ($4.1 \cdot 10^{-4}$)	0.37
	FSA-Block	4.42 (0.86)	8.87 (3.24)	0.61 (0.0196)	0.95 (0.0113)	0.05 ($2.0 \cdot 10^{-3}$)	0.39
	Block	4.46 (0.73)	8.87 (2.81)	0.55 (0.0044)	0.94 (0.0107)	0.02 ($3.8 \cdot 10^{-4}$)	0.42
	PP	2.13 (8.27)	4.19 (33.89)	3.00 (6.2494)	1.13 (0.0408)	0.34 ($1.1 \cdot 10^{-1}$)	0.46
	Modified PP	2.15 (8.17)	4.24 (33.30)	3.00 (6.2499)	0.90 (0.0235)	0.25 ($5.7 \cdot 10^{-2}$)	0.46
Set-up 2		5	10	1	1	0.01	
	FM	4.90 (0.54)	9.71 (1.99)	1.02 (0.0051)	0.97 (0.0213)	0.01 ($6.6 \cdot 10^{-6}$)	0.11
	FSA-Block	4.73 (0.59)	9.44 (2.21)	1.08 (0.0134)	0.98 (0.0217)	0.01 ($2.0 \cdot 10^{-5}$)	0.13
	Block	4.69 (0.57)	9.31 (2.25)	1.04 (0.0068)	0.94 (0.0245)	0.01 ($7.3 \cdot 10^{-6}$)	0.16
	PP	2.69 (5.43)	5.33 (21.99)	2.98 (3.9449)	1.78 (0.7571)	0.11 ($1.1 \cdot 10^{-2}$)	0.18
	Modified PP	2.73 (5.24)	5.42 (21.19)	2.99 (3.9710)	1.19 (0.1121)	0.07 ($3.3 \cdot 10^{-3}$)	0.18
Set up 3		5	10	2	1	0.01	
	FM	4.83 (0.39)	9.61 (1.54)	2.06 (0.0281)	0.95 (0.0422)	0.01 ($2.2 \cdot 10^{-7}$)	0.02
	FSA-Block	4.81 (0.56)	9.56 (2.04)	2.21 (0.0963)	1.13 (0.1020)	0.01 ($4.7 \cdot 10^{-7}$)	0.02
	Block	4.56 (0.66)	9.06 (2.64)	2.15 (0.0709)	0.90 (0.0551)	0.01 ($2.7 \cdot 10^{-7}$)	0.13
	PP	4.61 (0.34)	9.14 (1.21)	2.60 (0.3807)	2.23 (1.6538)	0.02 ($1.1 \cdot 10^{-4}$)	0.03
	Modified PP	4.81 (0.23)	9.59 (0.76)	2.56 (0.3500)	1.42 (0.2555)	0.01 ($2.6 \cdot 10^{-6}$)	0.03

Overall, the FSA-Block method and independent blocks method give reasonable estimates of the model parameters for different true values of ν, while the predictive process/modified predictive process tends to overestimate ν and gives notably biased estimates of other parameters, especially when ν is relatively small. In terms of prediction performance, it appears that the FSA-Block method achieves comparable results to the full covariance model under all the three parameter settings. The prediction performance of the other three competing methods depends on the true value of smoothness parame-

ter ν. Specifically, the independent blocks method achieves good prediction results and outperforms the predictive process/modified predictive process models when ν is relatively small, but its prediction performance is inferior to the predictive process/modified predictive process models when ν is large.

S2.3 RJMCMC algorithm

We applied the RJMCMC algorithm described in Section 2.4 to automatically select knots when applying the FSA. We used the same set of space-time locations as in simulation study and we simulated $Y(\mathbf{s}, t)$ at these locations following the model in (2.1), with $\mu = 0$ and $\tau^2 = 0$. The spatio-temporal random effect w was assumed to have a separable nonstationary correlation function $\Gamma_w(\mathbf{s}_i, t_i; \mathbf{s}_j, t_j) = \Gamma_u(\mathbf{s}_i, \mathbf{s}_j) \cdot \Gamma_v(t_i, t_j)$, where Γ_u and Γ_v were nonstationary spatial and temporal covariances constructed following Paciorek and Schervish (2006). Here the square spatial domain $[0,\ 20] \times [0,\ 20]$ was divided equally into 4 subregions with $D(\mathbf{s}_i)$ such that $D(\mathbf{s}_i) = l$ if $\mathbf{s}_i$ belong to the lth subregion,

$$\Gamma_u(\mathbf{s}_i, \mathbf{s}_j) = |\mathbf{H}_{D(\mathbf{s}_i)}|^{\frac{1}{4}} |\mathbf{H}_{D(\mathbf{s}_j)}|^{\frac{1}{4}} \left| \frac{\mathbf{H}_{D(\mathbf{s}_i)} + \mathbf{H}_{D(\mathbf{s}_j)}}{2} \right|^{-\frac{1}{2}} \exp(\sqrt{Q_{ij}}),$$

where $Q_{ij} = (\mathbf{s}_i - \mathbf{s}_j)^T \left(\frac{\mathbf{H}_{D(\mathbf{s}_i)} + \mathbf{H}_{D(\mathbf{s}_j)}}{2} \right)^{-1} (\mathbf{s}_i - \mathbf{s}_j)$ is the Mahalanobis distance. $\mathbf{H}_{D(\mathbf{s})}$ is referred to as the kernel covariance matrix. The eigenvalue decomposition of $\mathbf{H}_{D(\mathbf{s})}$ has clear geometric interpretations: the square roots of the eigenvalues of $\mathbf{H}_{D(\mathbf{s})}$ control the range of the spatial dependence and the eigenvector matrix corresponds to a rotation matrix. We reparameterize $\mathbf{H}_{D(\mathbf{s})}$ as

$$\mathbf{H}_{D(\mathbf{s})} = \mathbf{R}(\theta_{D(\mathbf{s})}) \begin{pmatrix} \lambda_{D(\mathbf{s}),1} & 0 \\ 0 & \lambda_{D(\mathbf{s}),2} \end{pmatrix} \mathbf{R}^T(\theta_{D(\mathbf{s})}),$$

where $\lambda_{D(\mathbf{s}),1}, \lambda_{D(\mathbf{s}),2}$ are eigenvalues of $\mathbf{H}_{D(\mathbf{s})}$, and $\mathbf{R}(\theta_{D(\mathbf{s})})$ is a rotation matrix. Anisotropy is introduced to the covariance function by allowing different values of $\lambda_{D(\mathbf{s}),1}$ and $\lambda_{D(\mathbf{s}),2}$. Nonstationarity is achieved by assuming a spatially-varying kernel covariance across different subregions. The time domain was divided equally into 2 intervals and Γ_v was constructed in a similar way as Γ_u. In this experiment, for simplicity, we set $\mathbf{R}$'s to be identity matrices. We used the same 35 blocks as in simulation study 1 for the FSA-Block approach. The true values of model parameters are shown in the second column of Table 3.

For Bayesian posterior inferences, flat priors were adopted for all λ's, and truncated normal distributions on $(0,\ 50)$ were used as their proposal distributions. The prior for knot number m was set to be Poisson(50), truncated at $\lambda_0 = 700$. Conditional on the knot number, we assign uniform priors from the set containing all observed space-time points for knot locations. We then followed the RJMCMC algorithm in Section 2.4 to draw samples of knots. We compared this method with the FSA-Block approach using the fixed knot design, where the knot set was predetermined by choosing a random

sample from the observed location set. We ran 7000 iterations after a burning period of 1000 iterations. 3500 posterior samples were collected with thinning, using every 3rd iteration.

Table 3: Parameter estimation and prediction results for FSA-Block approach with knots selected by RJMCMC algorithm.

Parameters	True value	Full Model	Random knots	$m = 50$	$m = 100$	$m = 300$
$\lambda_{\mathbf{s}_{11}}, \lambda_{\mathbf{s}_{12}}$	40	39.81 (4.28)	42.95 (3.79)	42.22 (4.10)	39.45 (4.37)	40.86 (4.10)
$\lambda_{\mathbf{s}_{21}}, \lambda_{\mathbf{s}_{22}}$	25	20.91 (2.53)	22.75 (2.70)	19.76 (2.32)	20.86 (2.38)	21.95 (2.48)
$\lambda_{\mathbf{s}_{31}}, \lambda_{\mathbf{s}_{32}}$	20	23.37 (2.82)	23.93 (2.95)	24.64 (3.23)	22.97 (2.92)	22.27 (2.84)
$\lambda_{\mathbf{s}_{41}}, \lambda_{\mathbf{s}_{42}}$	10	10.08 (1.07)	10.13 (1.14)	9.98 (1.11)	9.87 (1.18)	9.42 (1.07)
λ_{t_1}	40	37.06 (4.98)	43.75 (4.28)	41.93 (4.80)	41.52 (4.92)	40.40 (5.03)
λ_{t_2}	10	11.59 (1.53)	12.23 (1.64)	11.86 (1.65)	11.51 (1.50)	11.81 (1.64)
MSPE	-	0.231(0.001)	0.265(0.008)	0.287(0.001)	0.271 (0.001)	0.262(0.001)
posterior mean of m	-	-	82.90	-	-	-
Time(hour)	-	25.49	8.05	6.35	7.75	8.90

In Table 3, the fixed and random knot designs give fairly close estimates of the covariance parameters. These estimates are also close to those from the full model, suggesting that the FSA-Block is capable of providing a good approximation to a nonstationary spatio-temporal covariance model. The prediction performance of the FSA-Block approach with the fixed knot design seems to depend on the knot number. Under the random knot scenario, the posterior mean of the knot number m given by the RJMCMC algorithm is close to 83, but its MSPE is just slightly larger than that of the FSA with 300 fixed knots, indicating that the RJMCMC algorithm can be effective in determining a reasonable knot number and selecting the "most useful" knots from a candidate set.

S3 Analysis of the eastern US ozone data

Figure 1 shows the locations of the 513 monitoring stations and the seasonal effect in 1999.

S3.1 Analysis of monthly data chunks

Table 4 shows the parameter estimation and prediction results for each monthly dataset in June and July in 1998 and 1999. In general, parameter estimates of the FSA-Block method are close to those from the full model, implying that the FSA-Block method can approximate the original model reasonably well. The WCL method overestimates the spatial range parameter c. This may be due to the fact that it only includes pairs within certain distance for inference and thus may fail to incorporate large-scale dependence information. As the WCL estimate of η is always on the boundary of its parameter space, the estimate of α has the same problem in half of the cases, indicating possible convergence problems when using the WCL for parameter estimation.

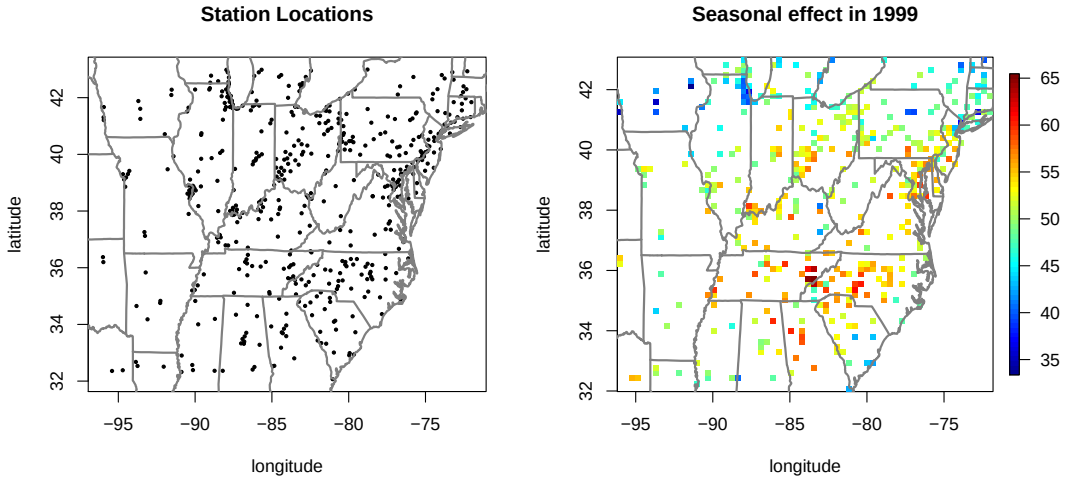

Figure 1: 513 ozone monitoring station locations and the estimated averaged seasonal effect in 1999.

We focused on comparing the prediction performance using the parameter estimates from the WCL, the FSA-Block, and the full model. Since the training set is large, computation of the BLUP using all training data is expensive, we used the training data within 4 days from each test data point when computing the BLUP. Although our FSA-Block method can be applied to the full training set and provides efficient computation for the BLUP, we used the partial training data and the full covariance model for fair comparison. Table 4 shows that the prediction results of the FSA-Block method and the full model are close to each other, and that both methods provide slightly better prediction performance than the WCL method.

S3.2 FSA with the random knot selection on the summer ozone datasets

We applied the RJMCMC algorithm described in Section 2.4 to automatically select knots when applying the FSA-Block on the summer ozone datasets in 1998 and 1999. We considered only Gneiting's model since it achieves better prediction performance than the Matérn covariance model. As the MLEs of η are close to zero for both datasets, and there is no significant difference between the separable and nonseparable models in terms of the parameter estimates of other parameters and prediction, we only applied model A here. For Bayesian inference, a uniform prior with support specified according to the prior belief on the practical range was assigned to the spatial/temproal range parameter. A uniform prior on $(0, 1]$ was taken for α, and a Poisson(100) prior truncated at $\lambda_0 = 700$ was assigned to the knot number m. We ran 8000 iterations to collect 2000 posterior samples after a burn-in period of 4000 iterations, thinning by using every third iteration.

Table 4: Parameter estimation and prediction results of monthly data. The root mean squared predictive errors (RMSPE) were made based on the within 4 days' training data.

Month	Methods	a	c	η	α	RMSPE
1998, June	FSA	22.70	414.40	0.023	0.212	0.335
	WCL	24.73	1299.63	1	0.503	0.339
	FullModel	20.95	414.22	0.033	0.149	0.334
1998, July	FSA	21.94	358.58	0.092	0.167	0.359
	WCL	25.77	1119.10	1	0.3467	0.359
	FullModel	21.28	356.99	0.076	0.138	0.359
1999, June	FSA	20.63	454.71	0.061	0.305	0.317
	WCL	17.69	1027.91	1	1	0.330
	FullModel	18.45	467.64	0.042	0.233	0.317
1999, July	FSA	23.91	338.94	0.036	0.260	0.430
	WCL	16.90	827.70	1	1	0.450
	FullModel	20.36	353.93	0.010	0.181	0.430

The posterior prediction was calculated using the partial training data by plugging in the maximum a posteriori (MAP) estimates of the knot set and the other model parameters. We used the same number of blocks as in the MLE cases.

Table 5: Parameter estimation and prediction results of summer ozone datasets by FSA-Block with random knot selection.

Method	Year	a	c	α	m	RMSPE
FSA	1998	18.80(0.46)	380.82(4.05)	0.164(0.009)	247.74	0.373
	1999	18.83(0.40)	417.70(3.90)	0.191(0.009)	246.92	0.380

Table 5 shows the Bayesian posterior sample means and standard deviations (in the parenthesis) of model parameters from FSA-Block with random knot selection for the summer ozone datasets in 1998 and 1999. For both datasets, the posterior sample means of range parameters are close to the corresponding MLE estimates with 400 fixed knots, while the posterior sample mean of α is slightly less than its MLE counterpart. For each summer ozone dataset, the posterior mean of the knot number m is about 250, but its RMSPE is about the same as that from the MLE result with 400 fixed knots, implying that the RJMCMC algorithm can be effective in determining a reasonable knot number and selecting useful knots from the candidate set.

Bibliography

Paciorek, C. J. and Schervish, M. J. (2006) Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics*, **17**, 483–506.