Spatial Multilevel Fused Lasso model for large datasets

Yewen Chen

October 13, 2020

CONTENTS

LIST OF FIGURES

LIST OF TABLES

1 MODEL SPECIFICATION

$$y(s) = \mathbf{x}(s)\boldsymbol{\beta} + w(s) + \epsilon(s). \tag{1}$$

Let $w(s) = \sum_{l=1}^{L} \sum_{k=1}^{m_l} h_k^l(s)\phi_k^l$, where $\mathbf{h}^l$ is a pre-defined kernel function on the spatial lth grid-level whose coefficients are denoted as $\boldsymbol{\phi}^l = (\phi_1^l, \cdots, \phi_{m_l}^l)$.

Specifically, we propose to estimate all parameters by minimizing the following objective function:

$$\frac{1}{n}\sum_{i=1}^{n}\left\{y(s_i) - \sum_{j=1}^{p}\beta_j x_j(s_i) - \sum_{l=1}^{L}\sum_{k=1}^{m_l} h_k^l(s_i)\phi_k\right\}^2 + \sum_{l=1}^{L}\sum_{k=1}^{m_l} P_\lambda\left(\phi_k^l - \sum_{j\in N(k)} \nu_j \phi_j^l\right) \tag{2}$$

where $N(k)$ denotes the set of neighbors index of the vertice k.

2 SOME EXAMPLES

2.1 Synthetic example from Heaton et al. [2019]

The size of training data: 105, 569
The size of testing data: 44, 431

Table 1: RMSE for each competing method on the simulated data.

Method	RMSE	Run time(min)
FRK	1.31	
Gapfill	1.00	
LatticKrig	0.87	
LAGP	1.11	
Metakriging	0.97	
MRA	0.83	
NNGP	0.88	70
Partition	0.86	
Pre. Proc.	1.43	
SPDE	0.86	
Tapering	0.97	
Periodic Embedding	0.91	
Model (2)	0.86	35

2.2 Simulation study

The $n = 500$ spatial nonregularly points in $[0, 50] \times [0, 50] \subset \mathcal{D}$ was randomly generated by

$$\mathbf{Y} = \boldsymbol{W} + \varepsilon, \tag{3}$$

where $\varepsilon_t \sim N(0, 0.5^2\mathbf{I})$ and $\boldsymbol{w} \sim GP\left(1 + \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2, 15\exp\left(-\frac{D}{10}\right)\right)$, the two covariates $(\mathbf{X}_1, \mathbf{X}_2)$ were, respectively, generated from $\texttt{beta}(20, 1)$ and $\texttt{beta}(2, 10)$ or $\texttt{N}(20, 1)$ and $\texttt{N}(2, 10)$.

The 90% of the samples were randomly selected as training set, the rest as test. The testing results for model (2) and NNGP are as follows:

Table 2: Mean of RMSE and CRPS for the simulated data using 50 different seeds.

nu	model	RMSE	CRPS	Covariate X
0.5	NNGP	2.00	1.35	beta dis.
0.5	Model (2)	1.90	1.17	
2.0	NNGP	1.02	0.89	
2.0	Model (2)	0.86	0.74	
2.0	NNGP	1.23	11.87	Gaussian dis.
2.0	Model (2)	0.99	11.86	

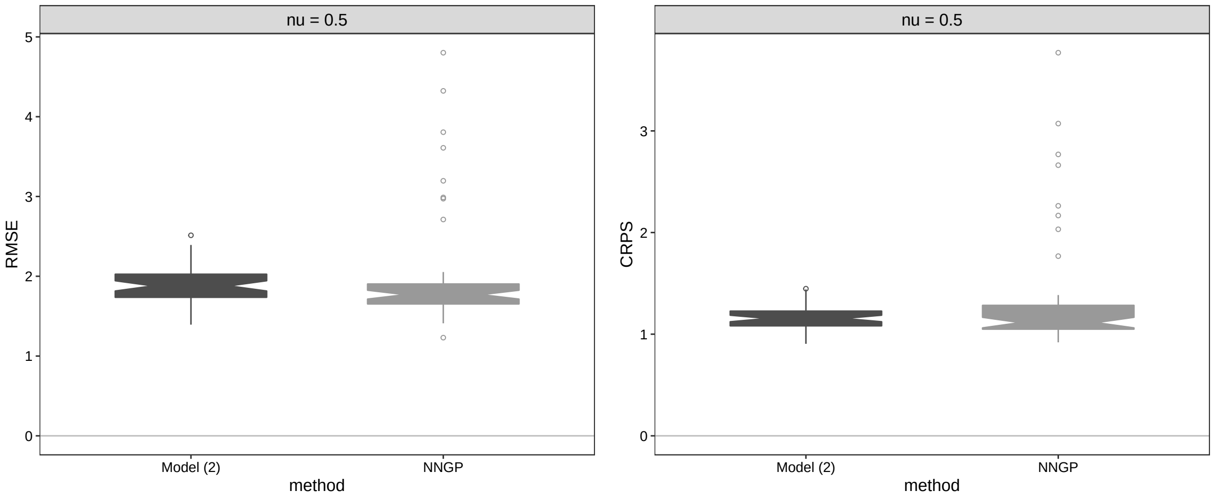

Figure 1: Note that the covariance model(i.e. Matern function with the smoothness parameter 0.5) is correctly specified for NNGP model. Left: RMSE for the two models. Right: CRPS for the two models.

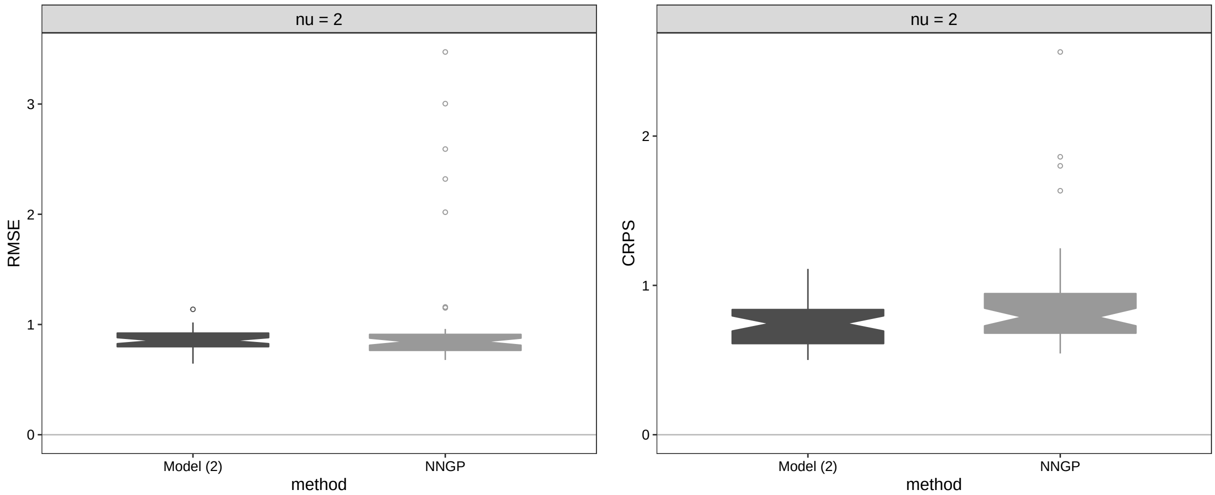

Figure 2: Note that the NNGP covariance model is incorrectly specified and the specification is Matern function with the smoothness parameter 2. Left: RMSE for the two models. Right: CRPS for the two models.

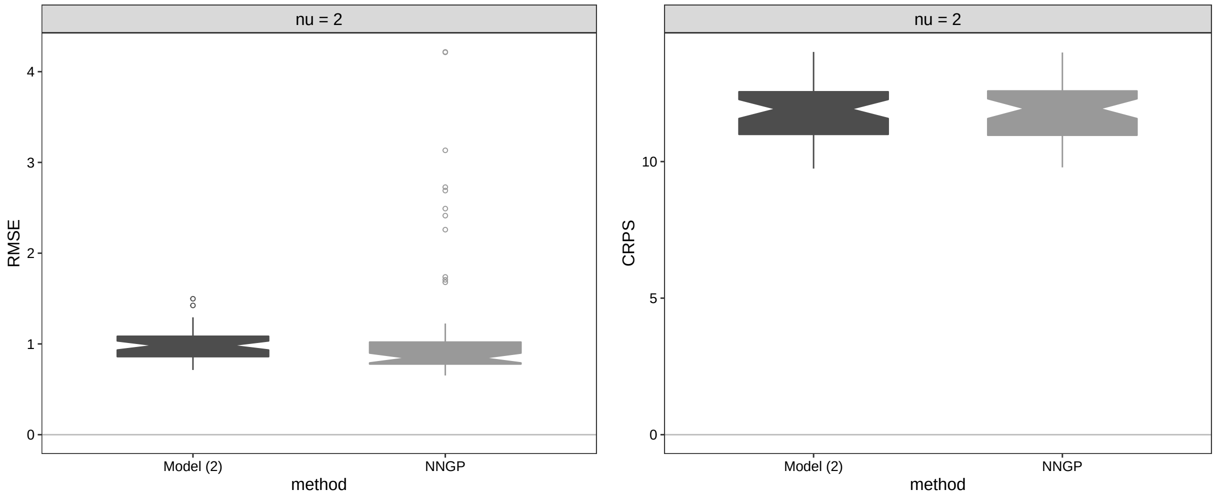

Figure 3: The covariate X are from Gaussian distribution.

2.3 Real data from Beijing-Tianjin-Hebei region

City	RMSE							CRPS						
	CMAQ	UK	RF	SVC	fLasso	fLassoM	HDCM	CMAQ	UK	RF	SVC	fLasso	fLassoM	HDCM
Baoding	36.25	28.09	25.29	34.58	**23.10**	22.37	22.29	21.73	15.07	13.95	16.77	**12.56**	12.14	11.96
Beijing	36.87	25.75	30.02	23.39	**20.09**	21.55	19.60	20.74	14.15	17.11	12.73	**11.31**	11.83	11.02
Cangzhou	34.41	22.94	21.57	20.99	**20.70**	19.96	19.19	23.12	12.78	12.13	11.84	**11.18**	11.19	10.45
Chengde	25.04	21.50	20.34	19.44	**18.78**	18.38	16.69	12.56	11.84	10.79	10.62	**9.55**	10.13	8.43
Handan	34.31	28.28	28.11	22.46	**20.16**	19.98	27.55	20.26	15.29	15.99	12.85	**11.78**	11.77	16.02
Hengshui	44.14	25.61	26.32	20.86	**19.11**	18.60	20.29	30.33	14.28	14.68	11.82	**11.12**	10.70	11.36
Langfang	27.52	14.79	22.53	13.45	**14.50**	13.18	12.93	15.99	9.72	12.40	9.22	**9.90**	9.13	9.07
Qinhuangdao	24.15	30.52	26.35	25.76	**24.71**	31.19	22.88	12.78	16.40	15.80	14.26	**13.05**	17.73	11.97
Shijiazhuang	34.22	35.00	31.55	28.41	**26.21**	26.71	26.38	19.50	19.53	18.37	15.86	**14.47**	14.70	14.52
Tangshan	43.65	32.56	27.84	28.81	**26.59**	29.66	24.63	26.34	18.37	16.68	16.31	**14.85**	17.29	14.00
Tianjin	24.24	21.24	23.60	18.85	**14.33**	18.30	15.37	13.89	12.21	13.42	11.08	**8.74**	10.72	9.21
Xingtai	40.07	35.35	32.94	32.56	**20.32**	23.43	24.28	23.71	19.69	19.10	18.22	**11.49**	12.97	13.47
Zhangjiakou	24.02	33.32	18.03	36.24	**20.10**	22.50	15.04	13.42	18.26	10.11	19.37	**11.26**	12.55	8.38
Mean	32.99	27.30	25.73	25.06	**20.67**	21.99	20.55	19.57	15.20	14.66	13.92	**11.63**	12.53	11.53

(a)

City	RMSE							CRPS						
	CMAQ	UK	RF	SVC	fLasso	fLassoM	HDCM	CMAQ	UK	RF	SVC	fLasso	fLassoM	HDCM
Baoding	94.64	77.78	75.45	73.08	**73.41**	73.63	68.43	50.36	43.49	40.61	40.92	**41.03**	40.85	38.16
Beijing	92.52	75.11	63.58	66.41	**66.11**	62.27	52.31	46.19	38.80	32.64	34.01	**33.06**	31.47	28.96
Cangzhou	60.99	72.15	41.4	46.93	**34.63**	34.82	35.68	31.87	41.41	23.86	30.21	**23.50**	24.61	22.20
Chengde	41.62	48.72	65.89	47.15	**33.43**	32.68	26.49	22.69	23.51	34.98	24.11	**18.53**	18.96	14.95
Handan	66.09	105.45	54.94	68.69	**58.78**	55.15	47.10	37.60	53.12	31.81	41.53	**37.17**	35.61	27.65
Hengshui	91.19	74.45	80.29	63.09	**64.51**	57.86	53.06	46.21	39.74	38.82	33.30	**34.99**	32.18	30.48
Langfang	85.44	39.03	52.44	35.81	**39.28**	36.13	35.26	46.14	27.26	29.86	26.79	**29.04**	27.19	27.15
Qinhuangdao	46.19	93.31	45.67	80.97	**30.75**	42.72	28.69	24.91	46.07	25.68	42.66	**16.12**	23.86	15.04
Shijiazhuang	87.72	88.08	60.96	44.91	**55.33**	54.19	49.41	47.99	50.56	35.87	31.62	**36.31**	35.31	32.40
Tangshan	95.75	51.63	60.68	47.2	**34.47**	40.34	26.07	51.84	32.03	34.60	29.80	**23.79**	24.03	19.49
Tianjin	70.74	51.91	49.84	44.78	**30.83**	33.37	31.62	38.08	32.13	28.46	29.40	**24.26**	25.22	24.43
Xingtai	85.79	73.25	74.17	64.72	**54.34**	57.15	54.60	45.57	41.69	39.12	36.37	**32.48**	33.34	32.11
Zhangjiakou	64.42	113.43	54.55	90.37	**29.41**	52.16	38.53	33.03	48.02	28.15	45.39	**15.62**	28.24	21.01
Mean	75.62	74.18	59.99	59.55	**46.56**	48.65	42.10	40.19	39.83	32.65	34.32	**28.15**	29.30	25.69

(b)

Figure 4: (a):The cross validation for the warm season with the 6 different models. (b): The cross validation for the cool season with the 6 different model. The label "`fLasso`" corresponds to model (2) and the label "`fLassoM`" is a multi-variable version.

REFERENCES

M. J. Heaton, A. Datta, A. O. Finley, R. Furrer, J. Guinness, R. Guhaniyogi, F. Gerber, R. B. Gramacy, D. Hammerling, M. Katzfuss, et al. A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological and Environmental Statistics*, 24 (3):398–425, 2019. URL https://chenyw68.github.io/Literature/[2018]Acasestudycompetitionamongmethodsforanalyzinglargespatialdata.pdf.