Small scale spatio-temporal data analysis

Jaehong Jeong and Mikyoung Jun

Texas A&M University

June 20, 2014

Data description

The log 2-week average time series at 42 monitoring sites (from 1999-01-06 to 2011-12-21).

- $42 \times 339 = 14238$ observations (including 1061 missing values).
- Its mean is roughly 0 and standard deviation is 0.2064.

10 of the 52 monitoring sites for validation.

Data description

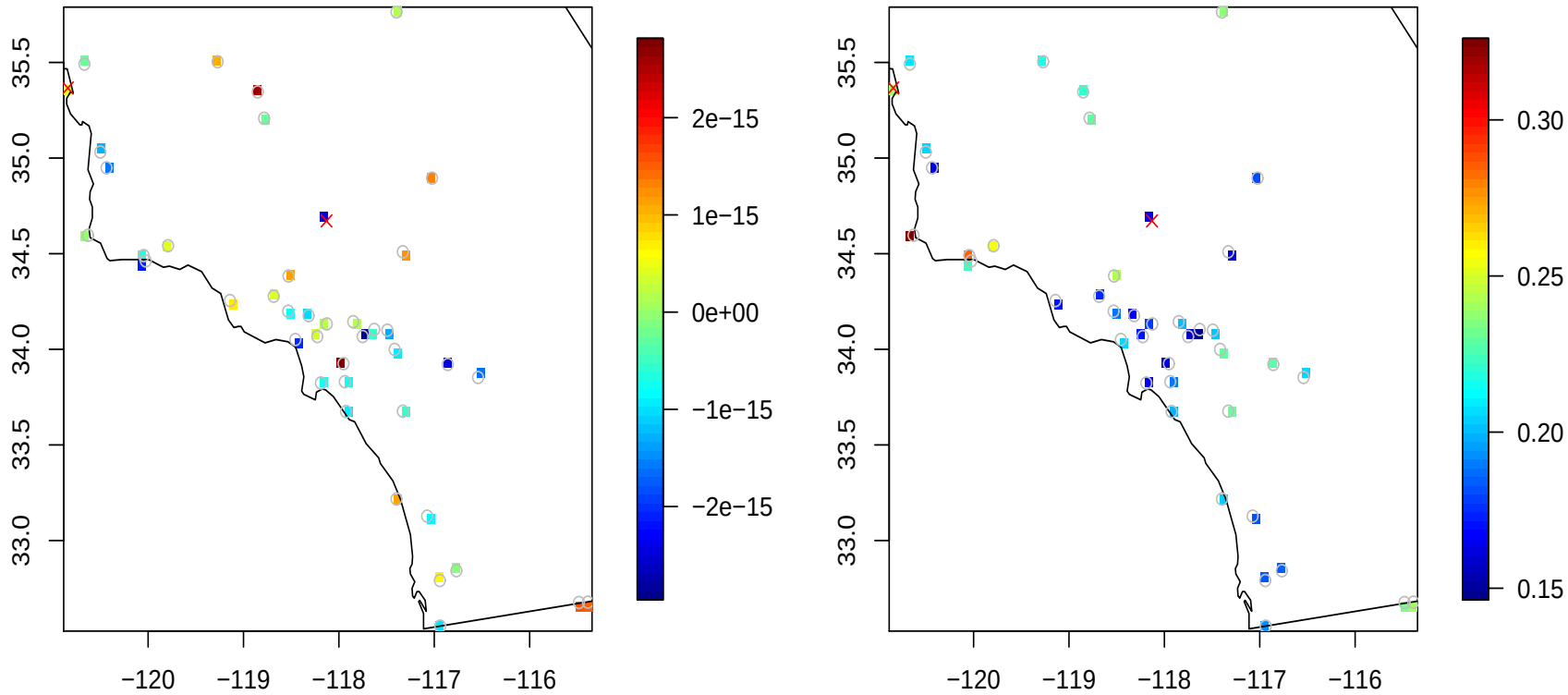

Figure 1: Locations. ∘ : 40 locations for estimation and × : 2 locations (station 60793001 and 60379033) for prediction for preliminary study. For given location, (L) mean of residuals over the whole time and (R) standard deviation of residuals over the whole time.

Data description

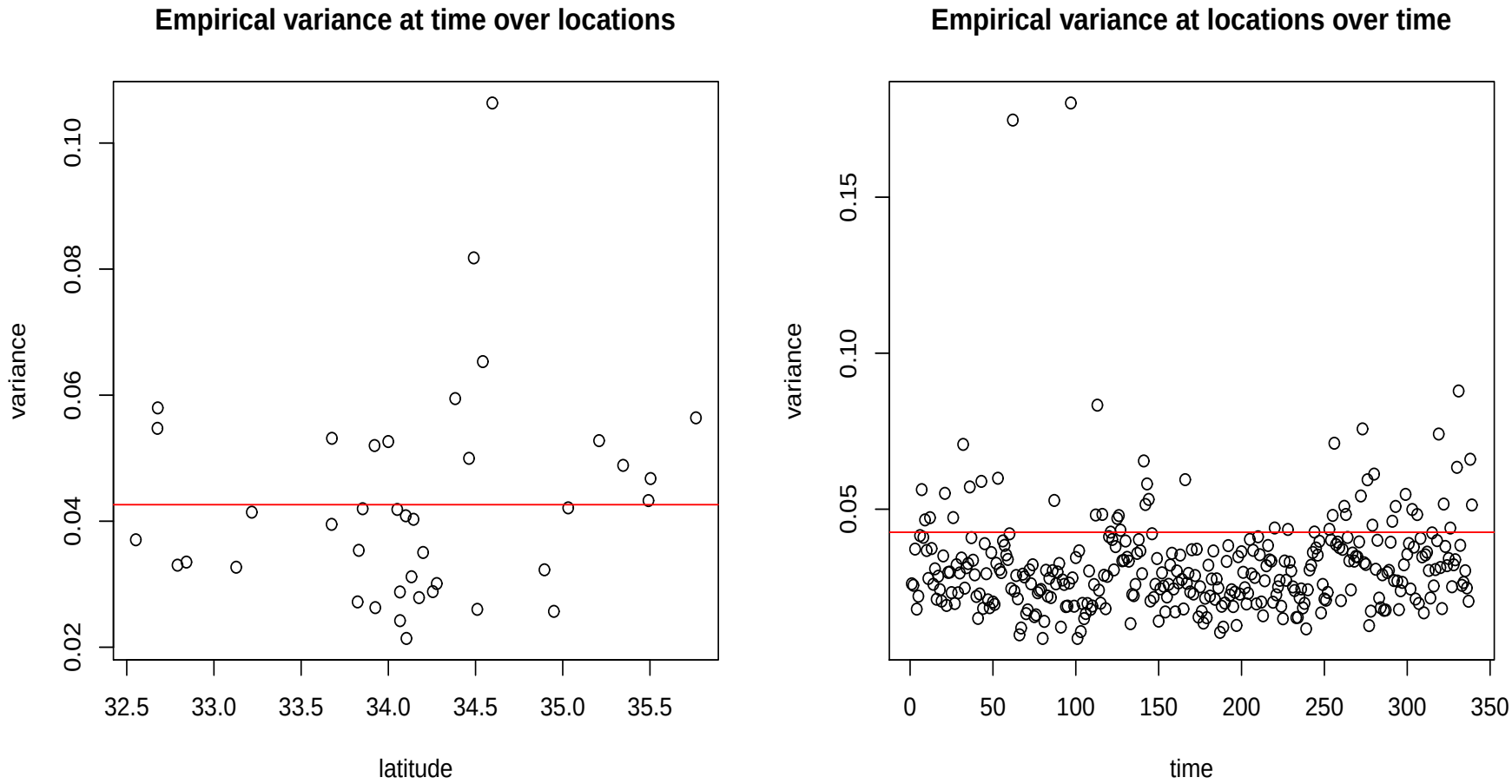

Figure 2: Empirical variance at time over locations (L) and at locations over time (R). Red line represents overall empirical variance, 0.0426.

Variograms

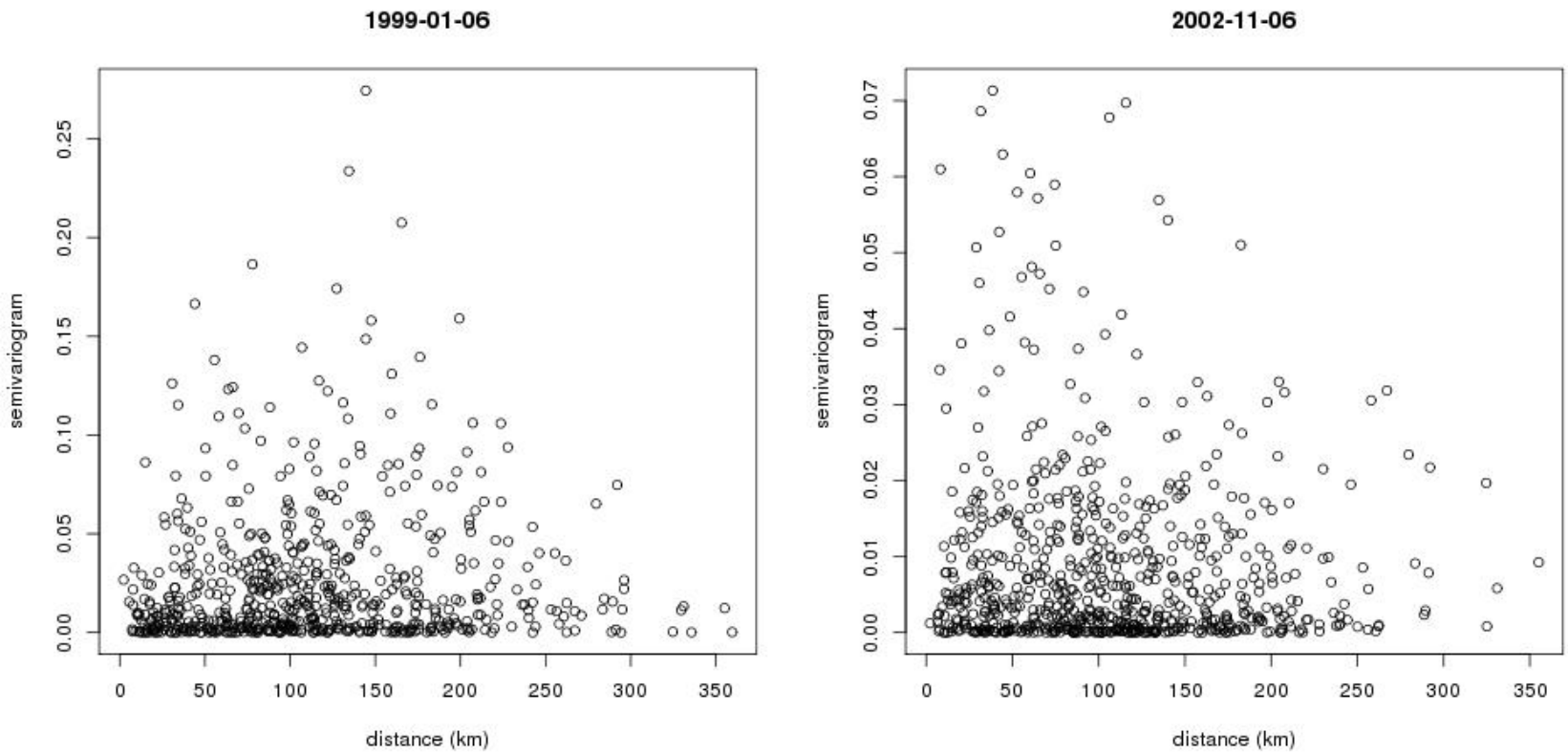

Figure 3: Empirical semivariogram against distance (km) at 1999-01-05 and 2002-11-06.

Variograms

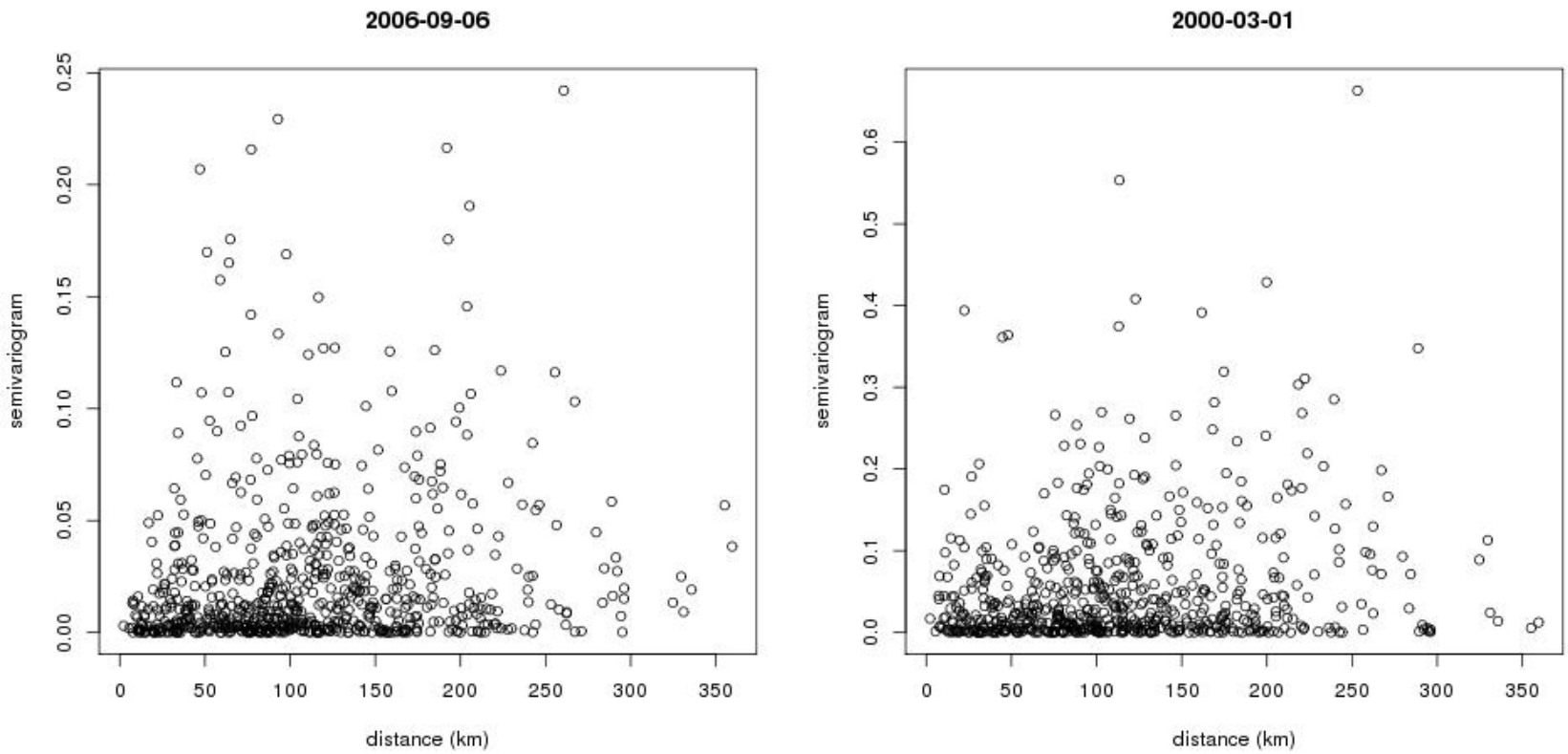

Figure 4: Empirical semivariogram against distance (km) at 2006-09-06 and 2000-03-01.

Variograms

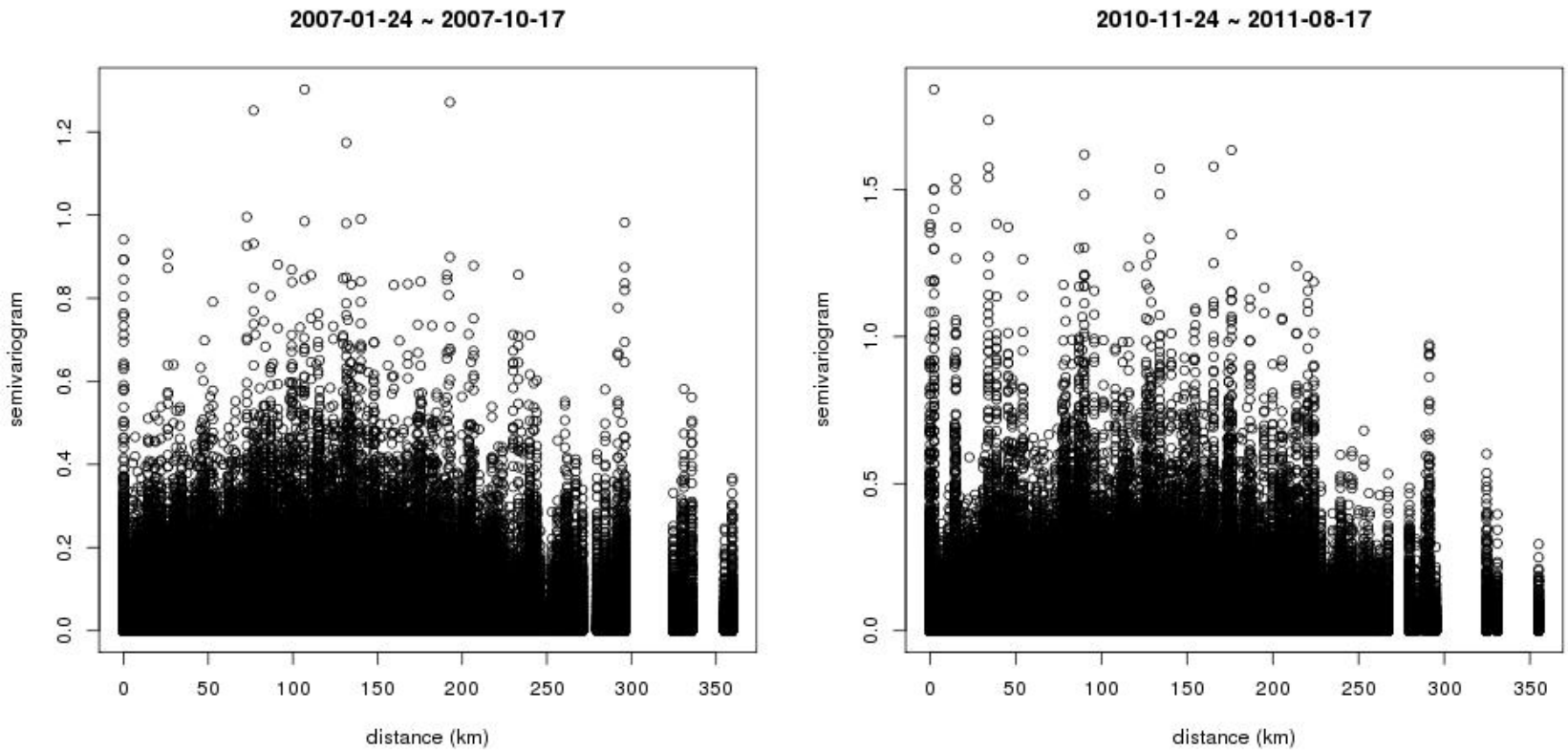

Figure 5: Empirical semivariogram against distance (km) at some periods.

Variograms

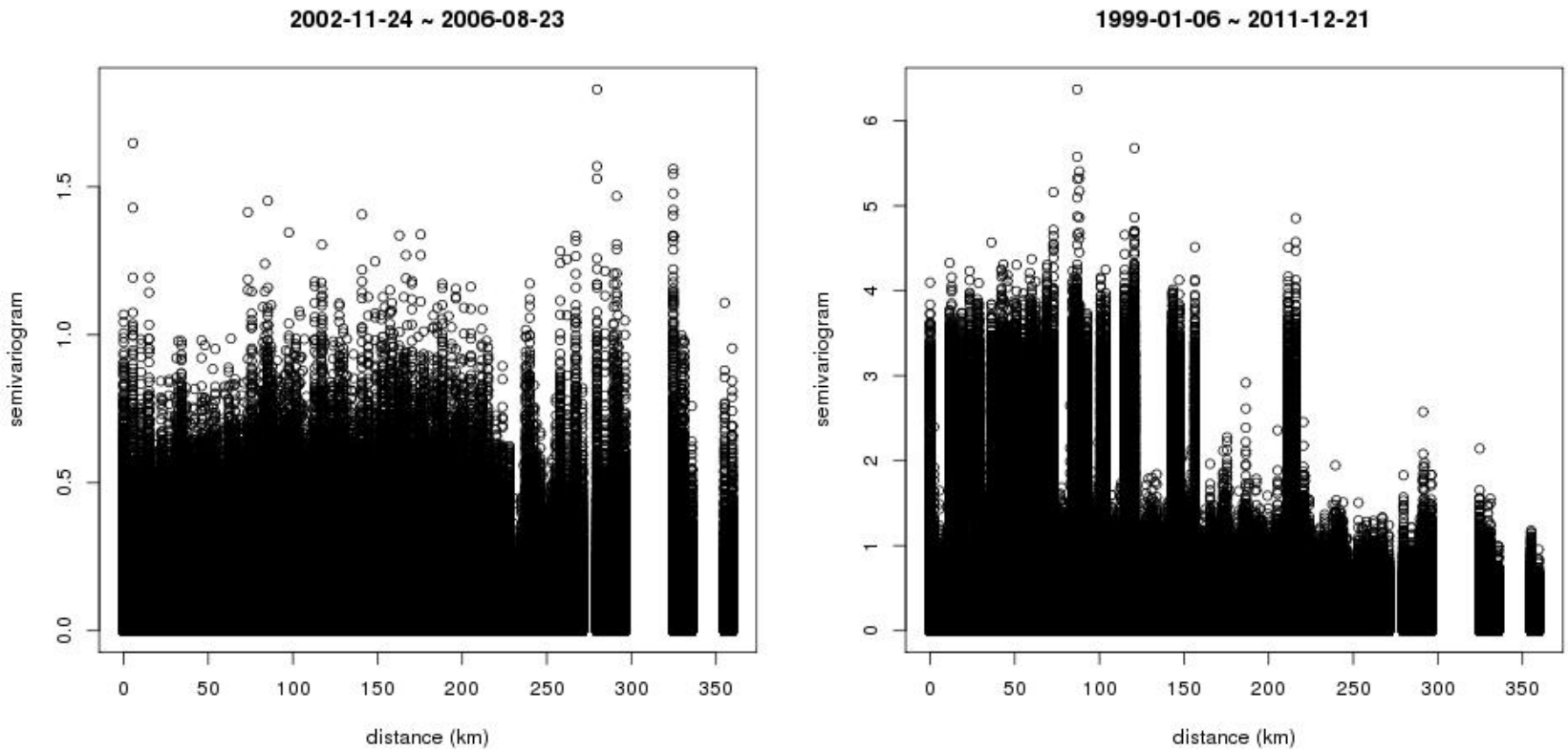

Figure 6: Empirical semivariogram against distance (km) at some periods.

Variograms

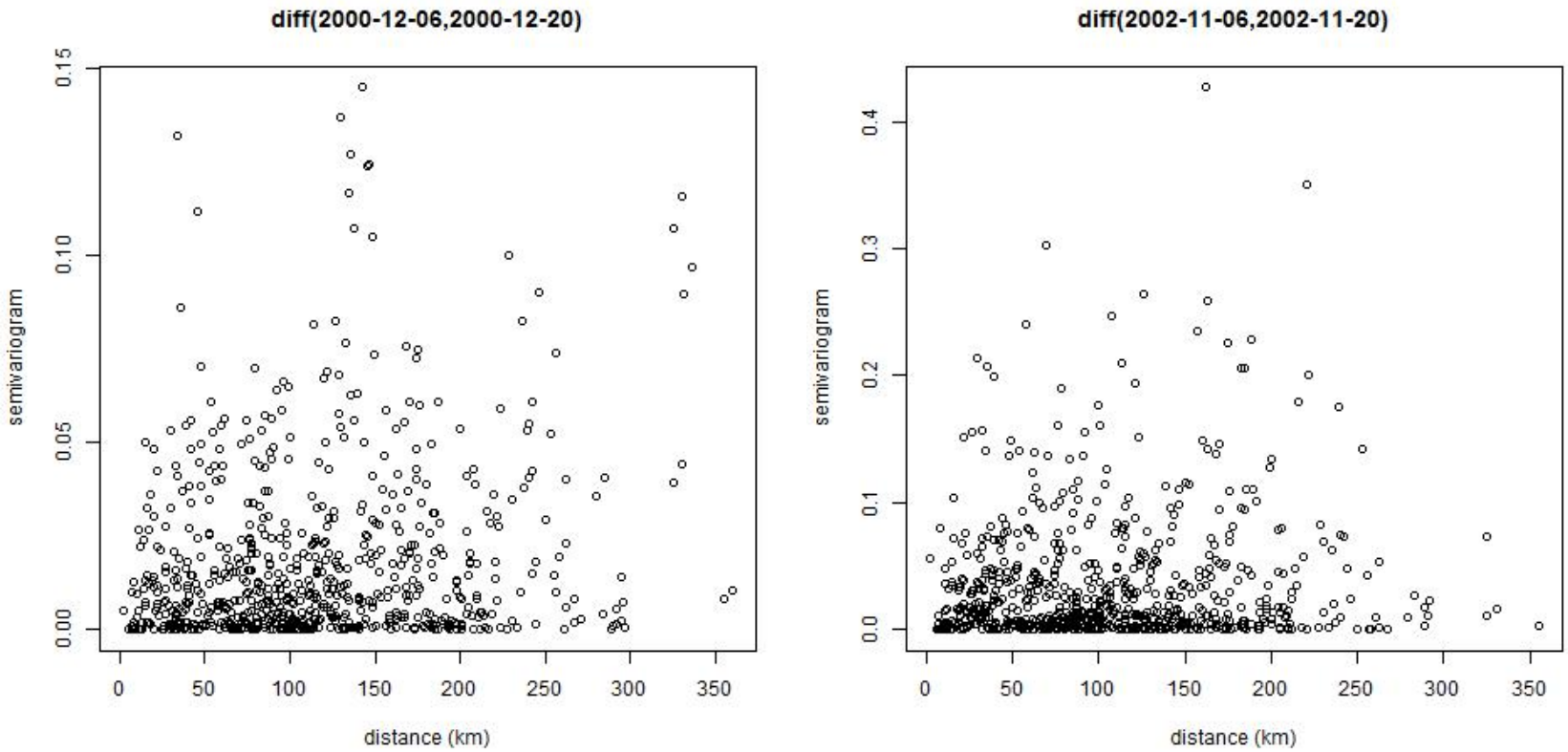

Figure 7: Empirical semivariogram of $Z(s,t) - Z(s,t-1)$ against distance (km) at some selected times.

Variograms

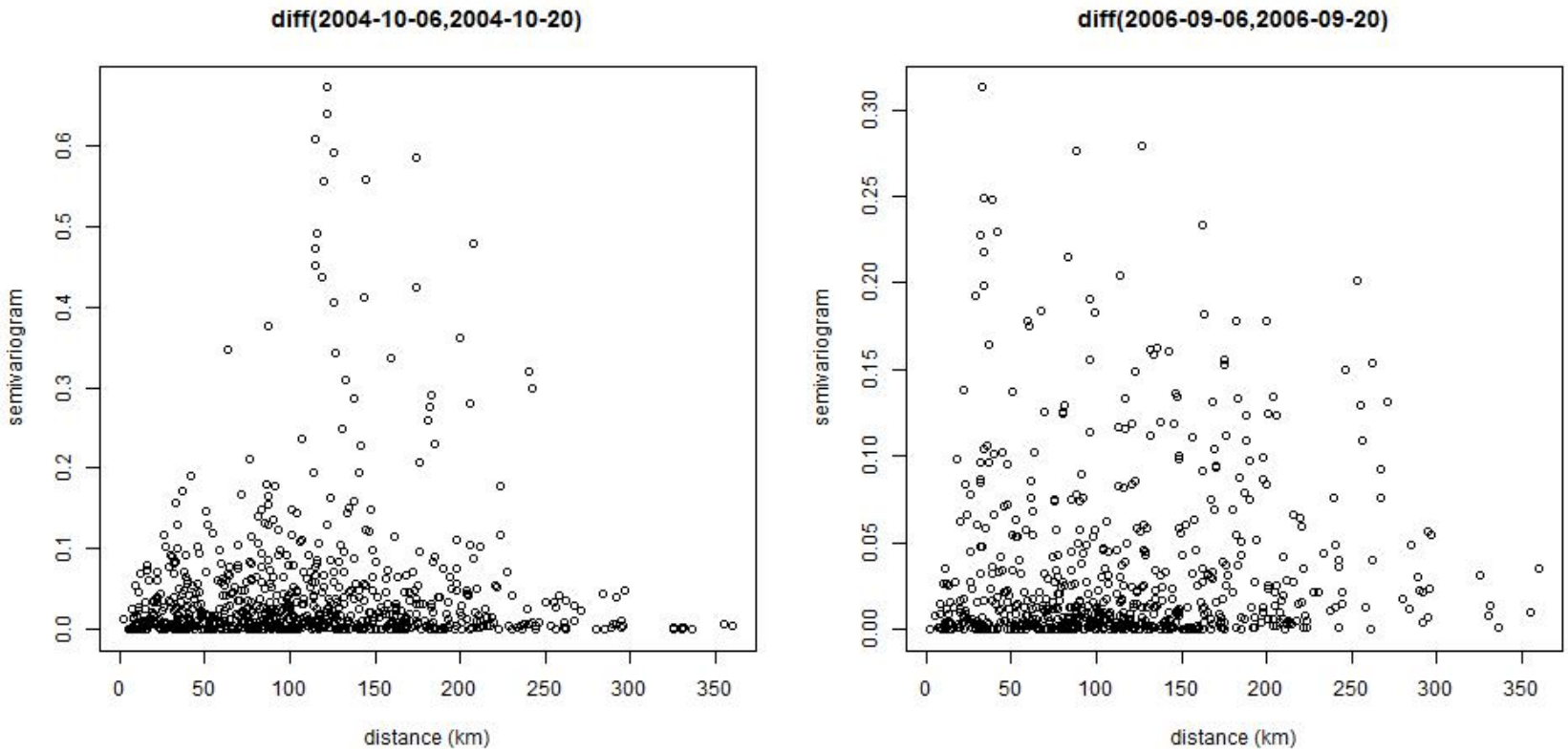

Figure 8: Empirical semivariogram of $Z(s,t) - Z(s,t-1)$ against distance (km) at some selected times.

Autocorrelation plot

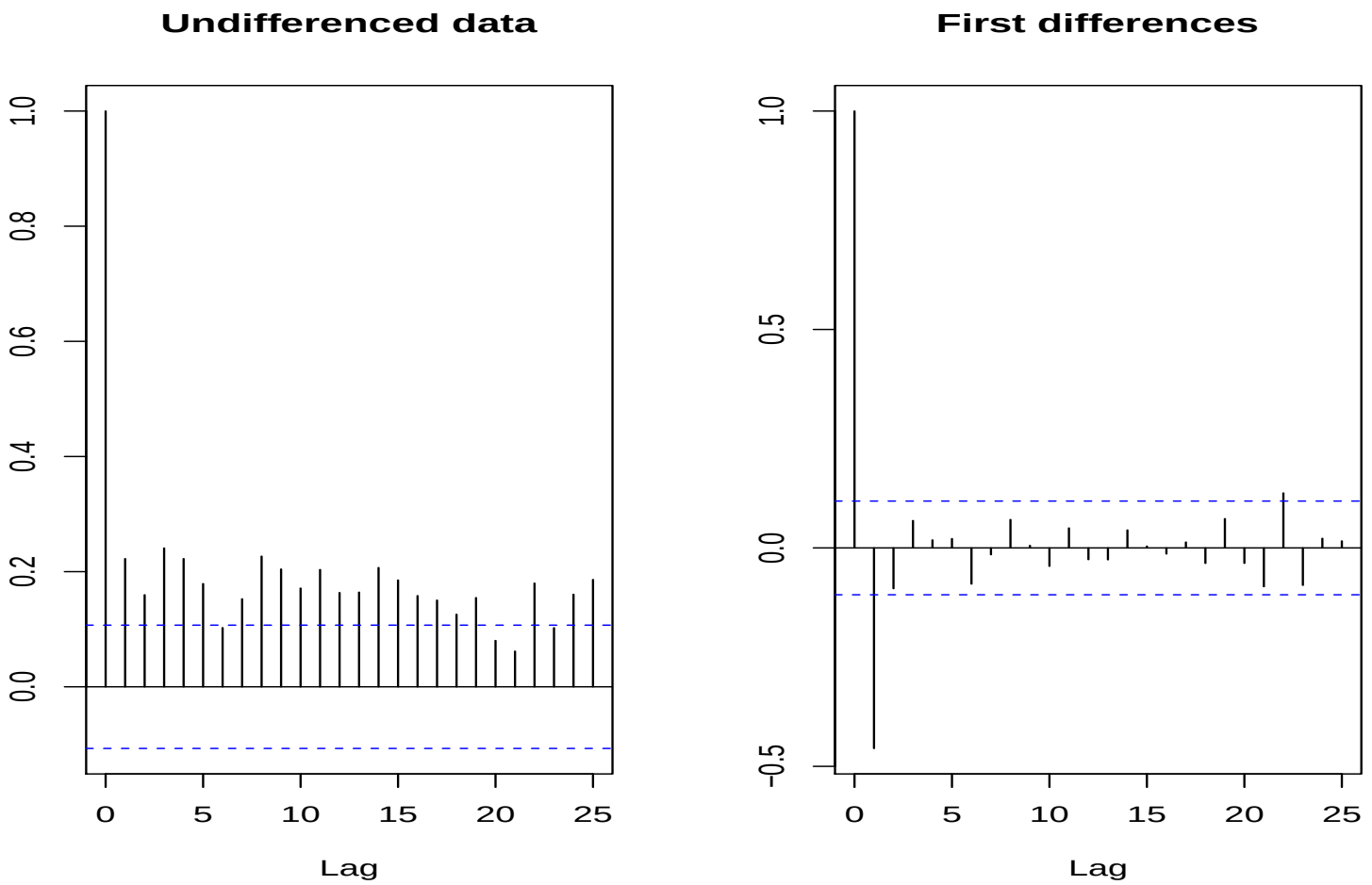

Figure 9: ACF plots at 60831025 station.

Autocorrelation plot

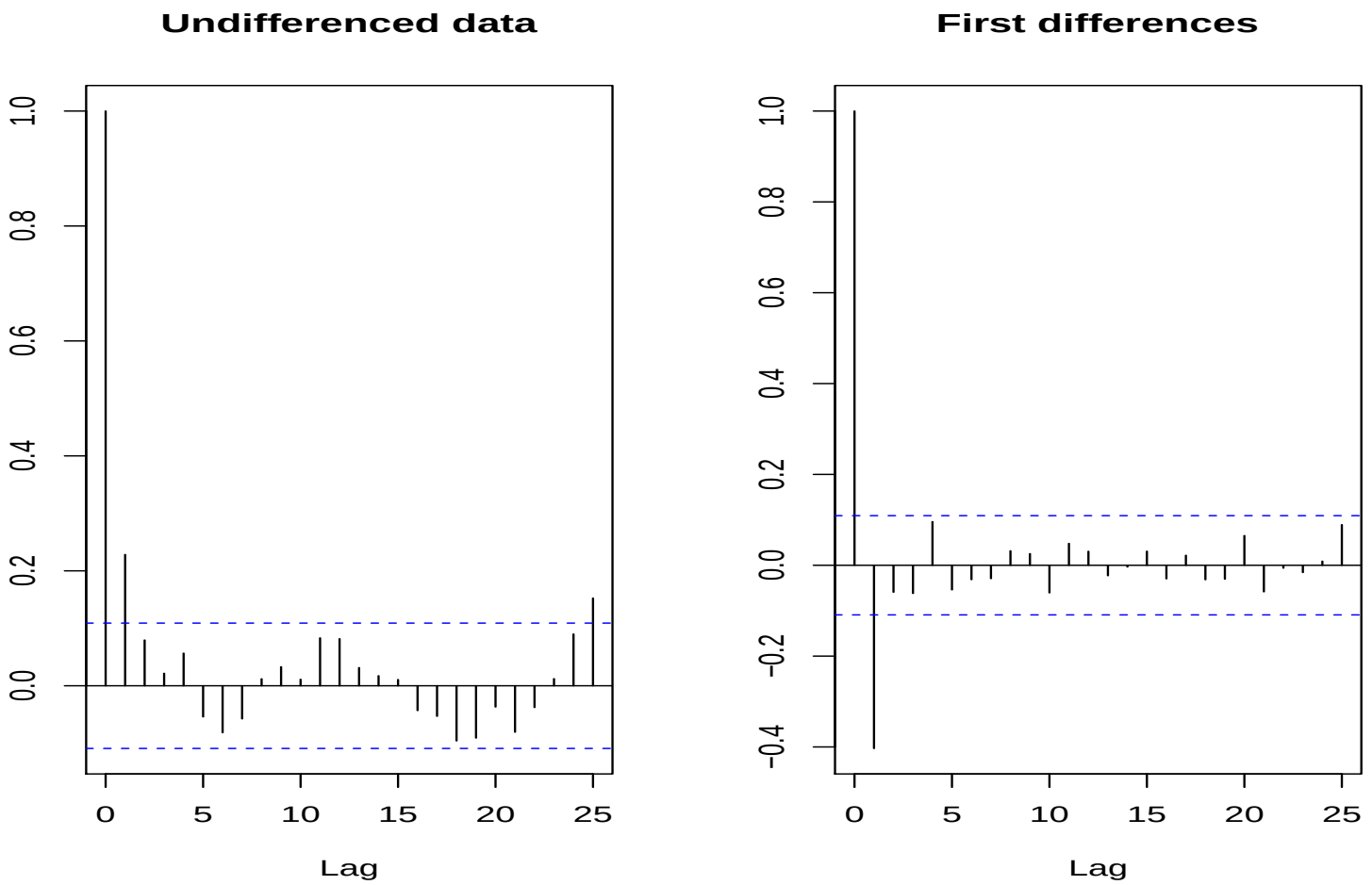

Figure 10: ACF plots for 60793001 station.

Model

We splitted whole time domain into 4 parts (1 : 1999-01-06 $\sim$ 2002-03-27, 2 : 2002-04-10 $\sim$ 2005-06-29, 3 : 2005-07-13 $\sim$ 2008-10-01, and 4 : 2008-10-15 $\sim$ 2011-12-21). Following models were considered.

- MC - Matérn covariance function.
- MCS - Separable space-time Matérn covariance function.
- DF - Differential operator terms + Matérn covariance function (Jun and Stein, 2007).
- EX - Exponential covariance function.

For priliminary study, we used 40 stations for estimation and 2 stations for prediction.

Differential Operator Model (Jun and Stein, 2007)

Consider

$$Z(L,l,t) = \left\{ A(L)\frac{\partial}{\partial L} + B(L)\frac{\partial}{\partial l} + C(L)\frac{\partial}{\partial t} \right\} Z_0(L,l,t) + Z_0'(L,l,t)$$

where $A(L) = P(L; a_0, \ldots, a_p) \times cos(L)$, $B(L) = P(L; b_0, \ldots, b_q)$, and $C(L) = P(L; c_0, \ldots, c_r)$. $P(L;\cdot)$ is the linear combinations of Legendre polynomials, and L and l are latitude and longitude, respectively.
Then, $\mathcal{M}_\nu(x) = \alpha x^\nu \mathcal{K}_\nu(x)$,

$$K_Z(L_1, L_2, l_1 - l_2, t_1 - t_2) = C_1 \mathcal{M}_{\nu-1}(\sqrt{h}) + C_2 \mathcal{M}_\nu(\sqrt{h}) + \alpha' \mathcal{M}_{\nu'}(\sqrt{h})$$

where
$h(L_1, L_2, l_1 - l_2, t_1 - t_2) = \mathrm{ch}^2(L_1, L_2, l_1 - l_2)/\beta_1^2 + (t_1 - t_2)^2/\beta_2^2$
and C functions are expressed as sums of products of the A, B, C, α, and the partial derivatives of h with respect to latitude lag, longitude lag, and time lag.

Model1 - Matérn

Table 1: Maximum loglikelihood value and prediction measures for each model.

Model	Max.loglik	RMSE	MAE	CRPS
MC1	3765.011	0.2097	0.1732	0.1203
MC2	4020.027	0.1790	0.1349	0.0985
MC3	4318.228	0.1432	0.1118	0.0793
MC4	3576.529	0.2101	0.1394	0.1042

Model1 - Matérn

Table 2: Parameter estimates for each model. β_s : spatial range (km), β_t temporal range (2 weeks), α : scale and ν : smoothness parameter.

Estimate	MC1	MC2	MC3	MC4
β_s	758.32	833.70	609.29	524.63
β_t	0.32	0.11	0.13	0.11
α	0.05	0.05	0.04	0.05
ν	0.08	0.08	0.11	0.09

Model1 - MC1

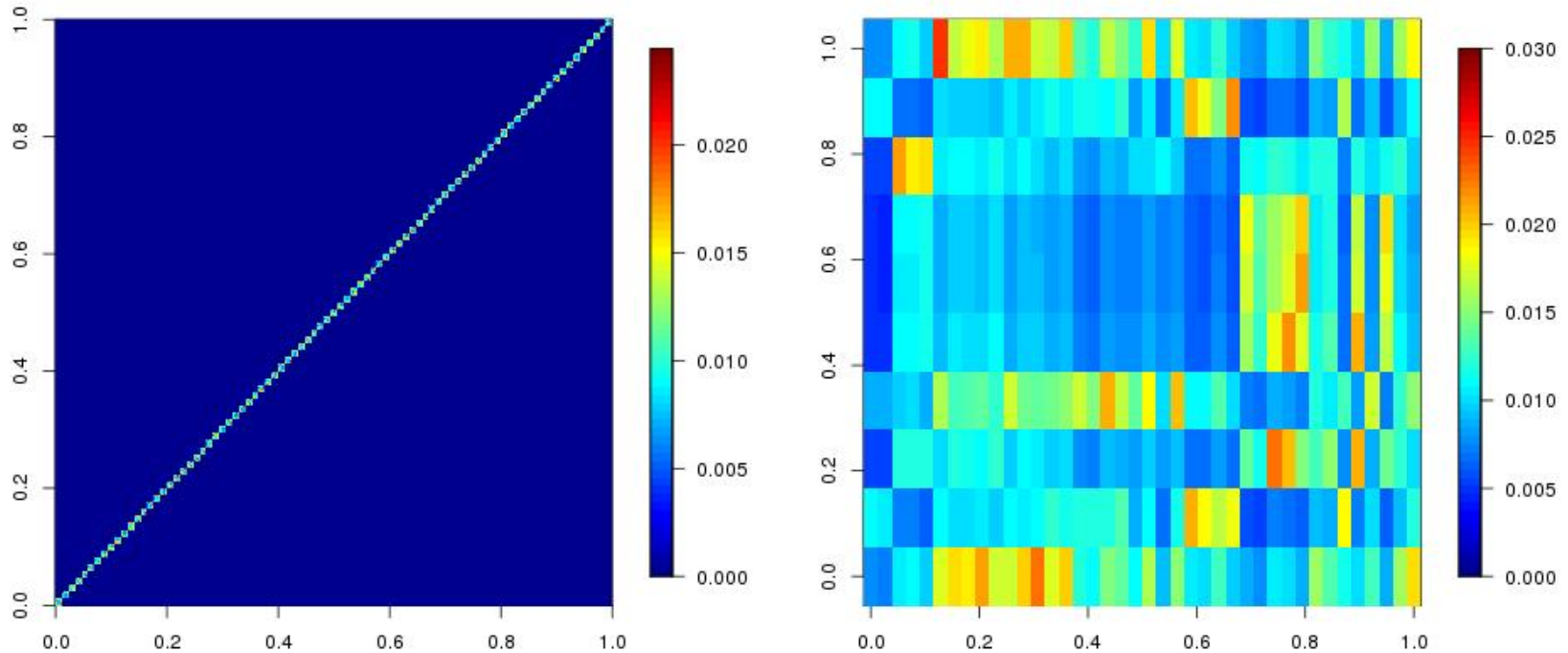

Figure 11: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (1999-01-06 $\sim$ 2002-03-27).

Model1 - MC2

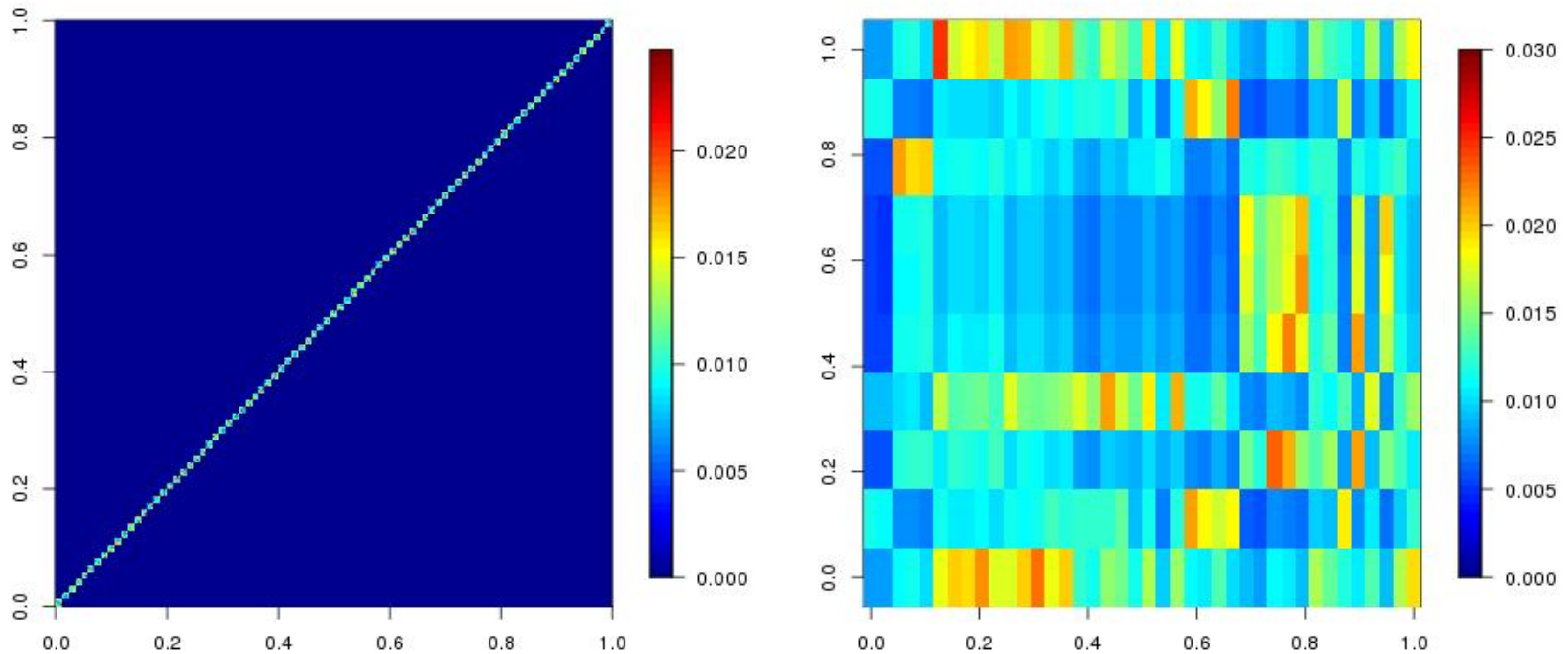

Figure 12: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (2002-04-10 $\sim$ 2005-06-29).

Model1 - MC3

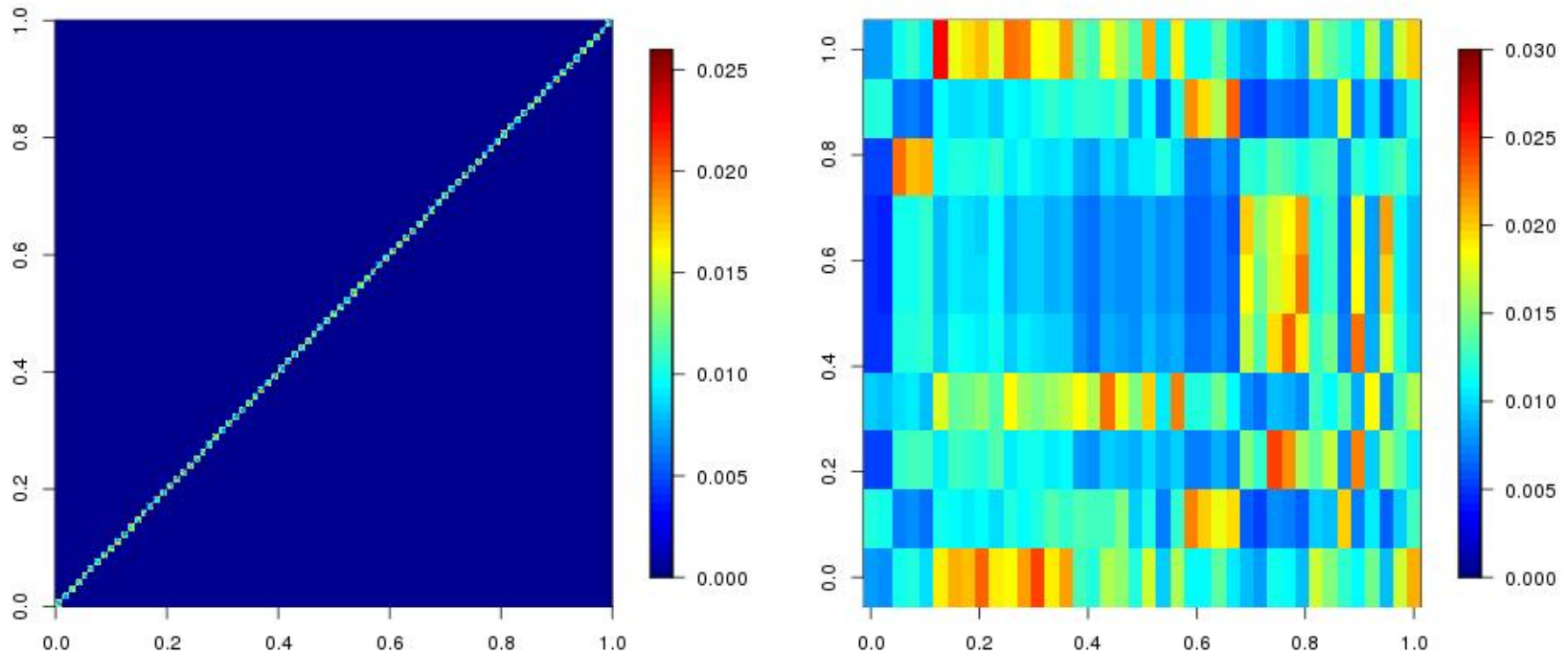

Figure 13: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (2005-07-13 $\sim$ 2008-10-01).

Model1 - MC4

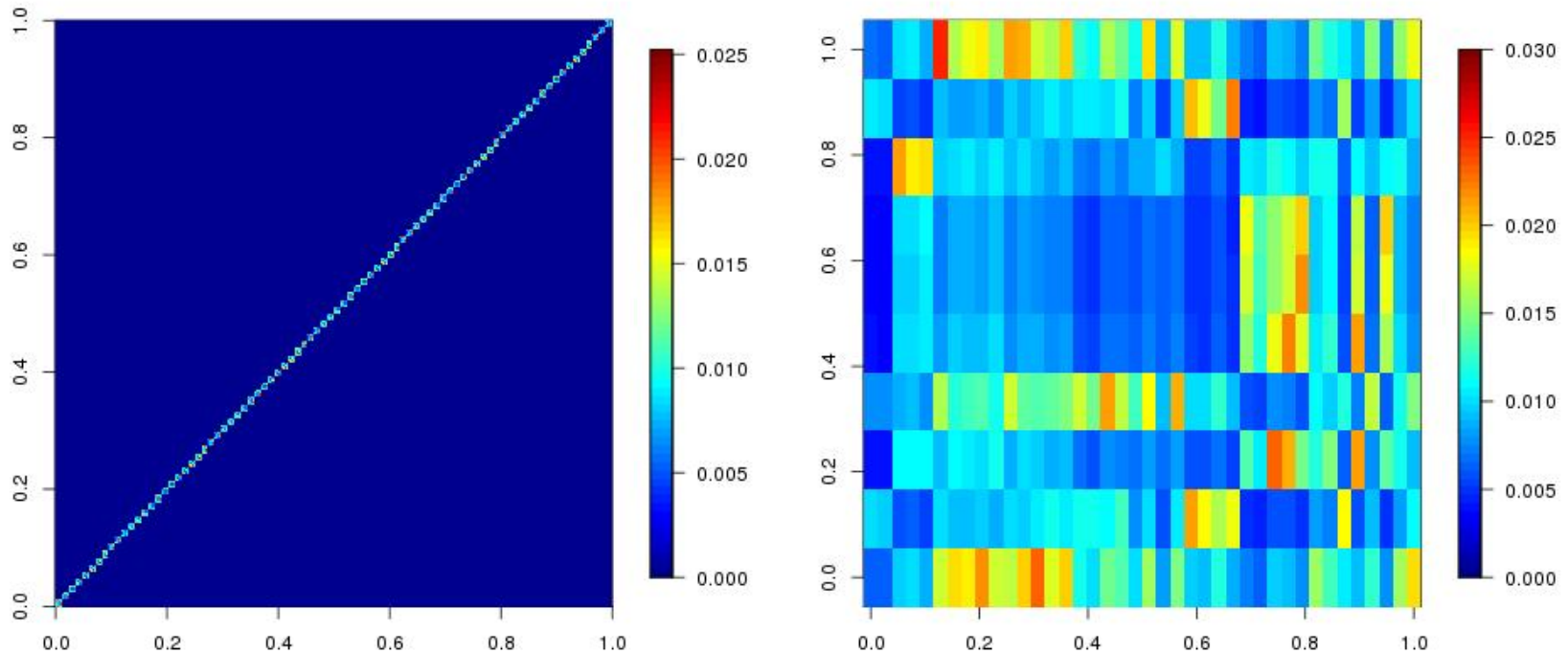

Figure 14: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (2008-10-15 $\sim$ 2011-12-21).

Model1 - MC - Prediction

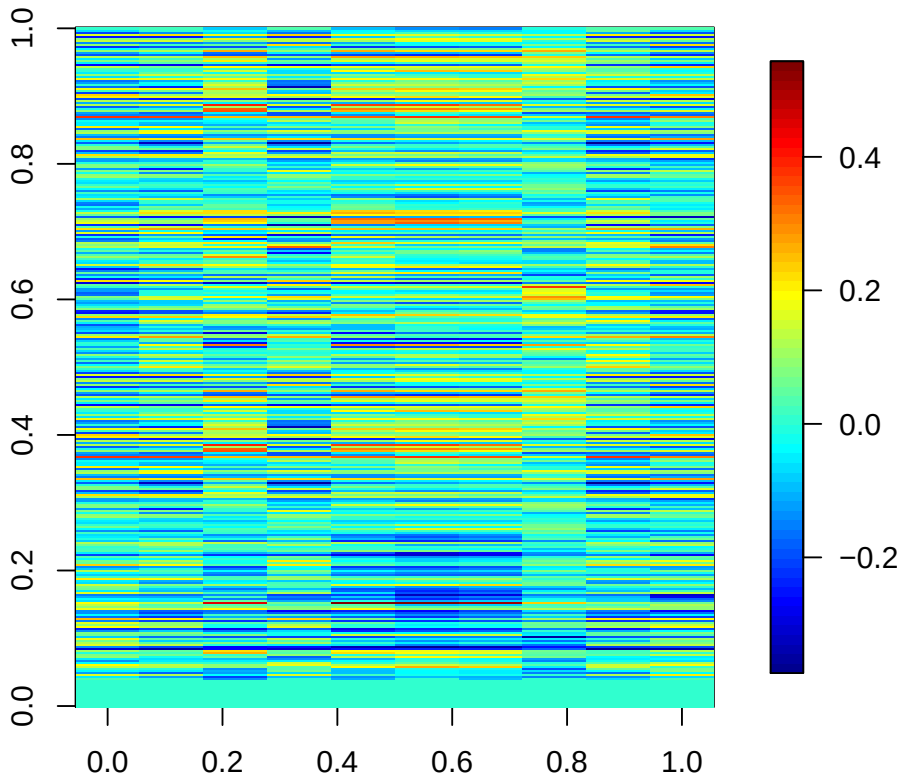

Figure 15: Image of 10×339 prediction values matrix.

Model2 - Separable space-time Matérn

Table 3: Maximum loglikelihood value and prediction measures for each model.

Model	Max.loglik	RMSE	MAE	CRPS
MCS1	4185.360	0.2132	0.1753	0.1223
MCS2	4469.583	0.1794	0.1356	0.0988
MCS3	4740.078	0.1433	0.1118	0.0794
MCS4	3954.572	0.2108	0.1405	0.1046

Model2 - Separable space-time Matérn

Table 4: Parameter estimates for each model. β_s : spatial range (km) and β_t temporal range (2 weeks).

Estimate	MCS1	MCS2	MCS3	MCS4
β_s	1473.23	1502.75	820.07	1009.62
β_t	6.71	5.97	8.11	5.39
α_s	0.21	0.21	0.13	0.19
α_t	0.25	0.27	0.39	0.30
ν_s	0.09	0.10	0.14	0.10
ν_t	0.16	0.17	0.14	0.17

Model2 - MCS1

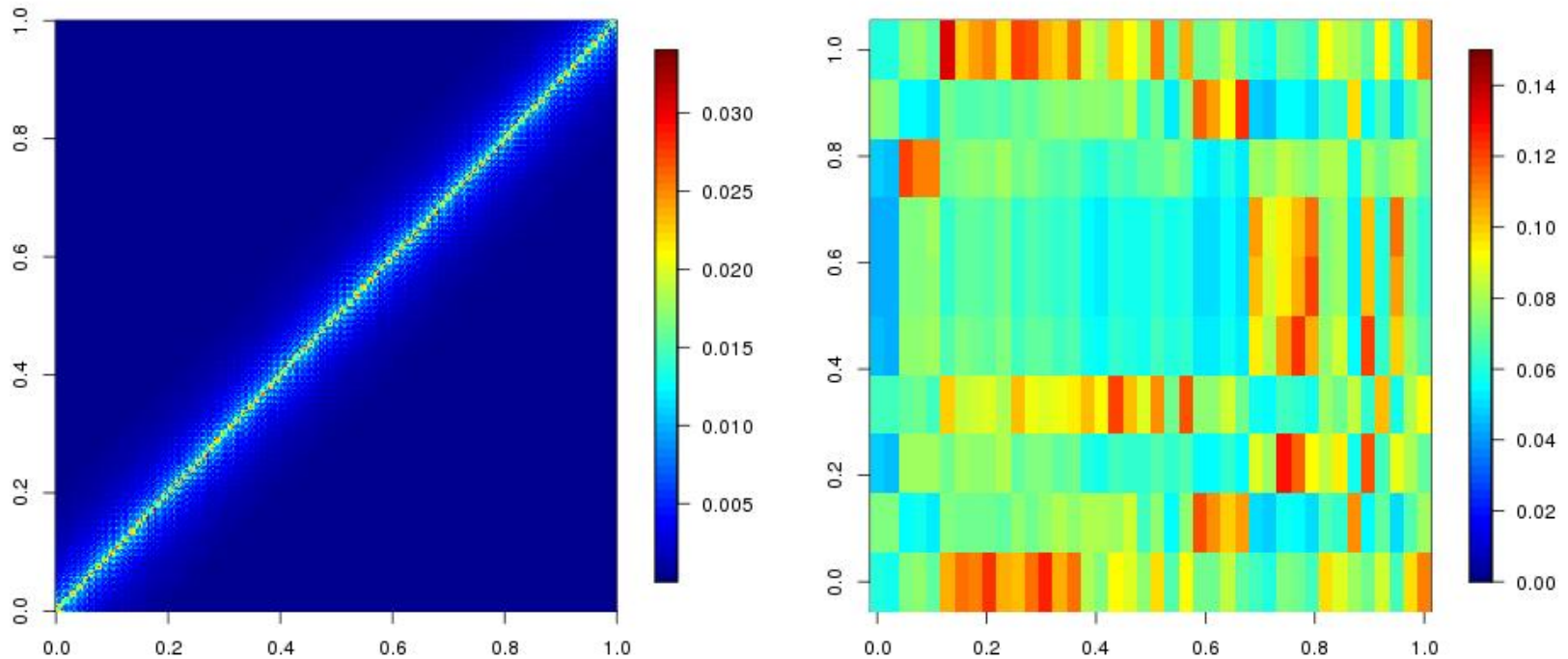

Figure 16: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (1999-01-06 $\sim$ 2002-03-27).

Model2 - MCS2

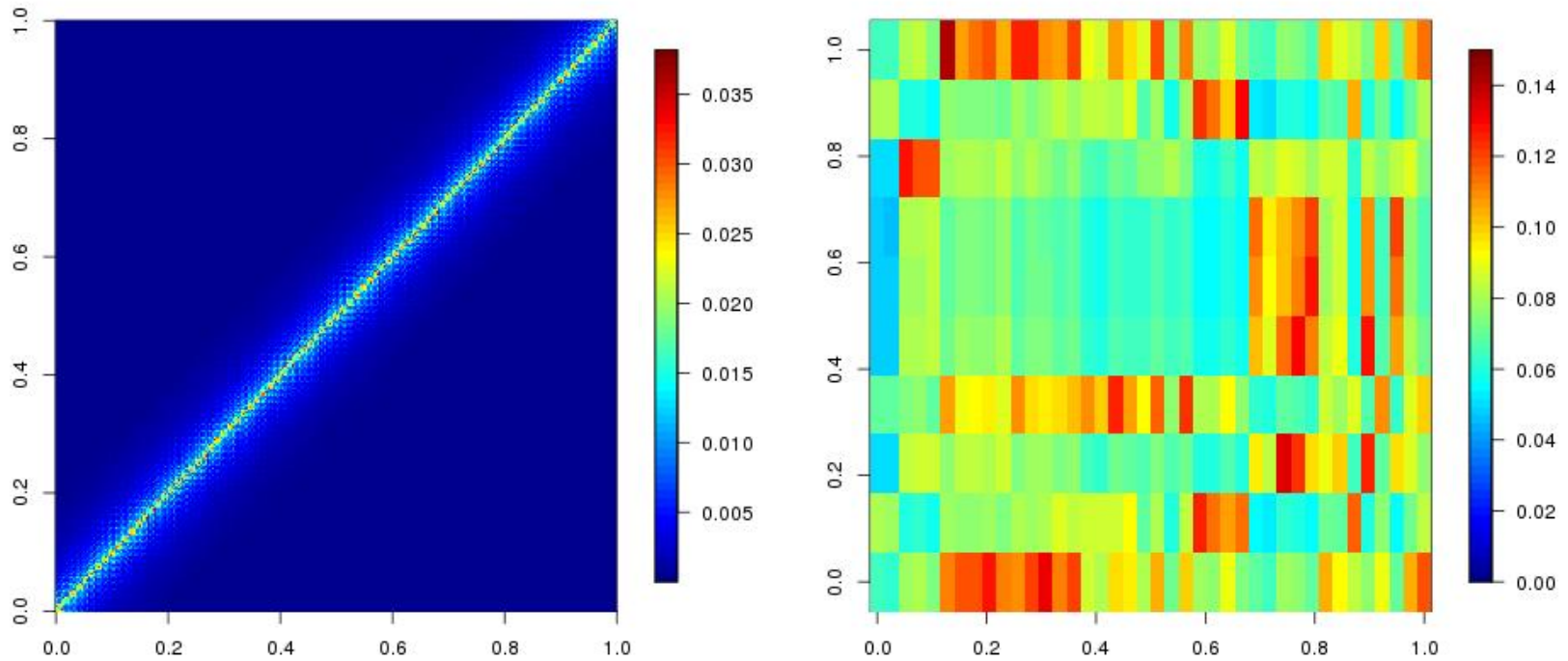

Figure 17: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (2002-04-10 $\sim$ 2005-06-29).

Model2 - MCS3

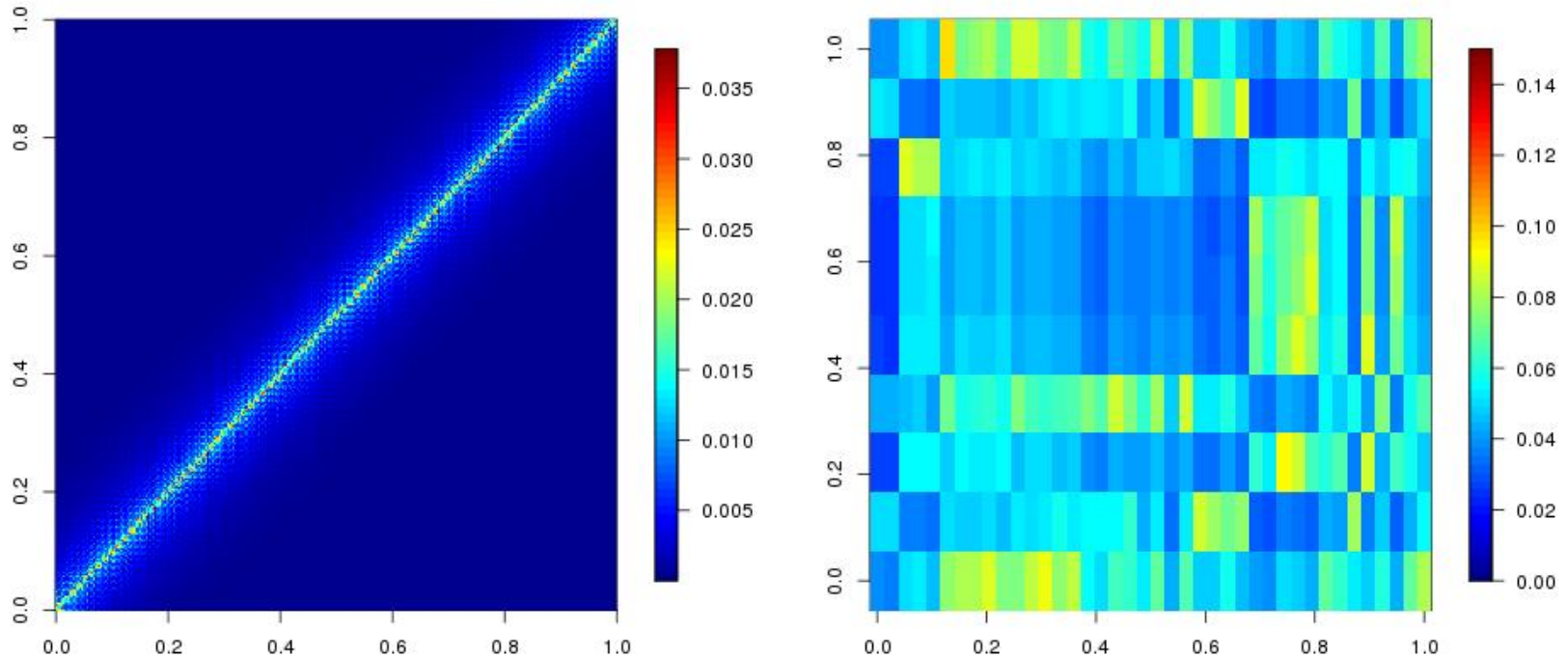

Figure 18: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (2005-07-13 $\sim$ 2008-10-01).

Model2 - MCS4

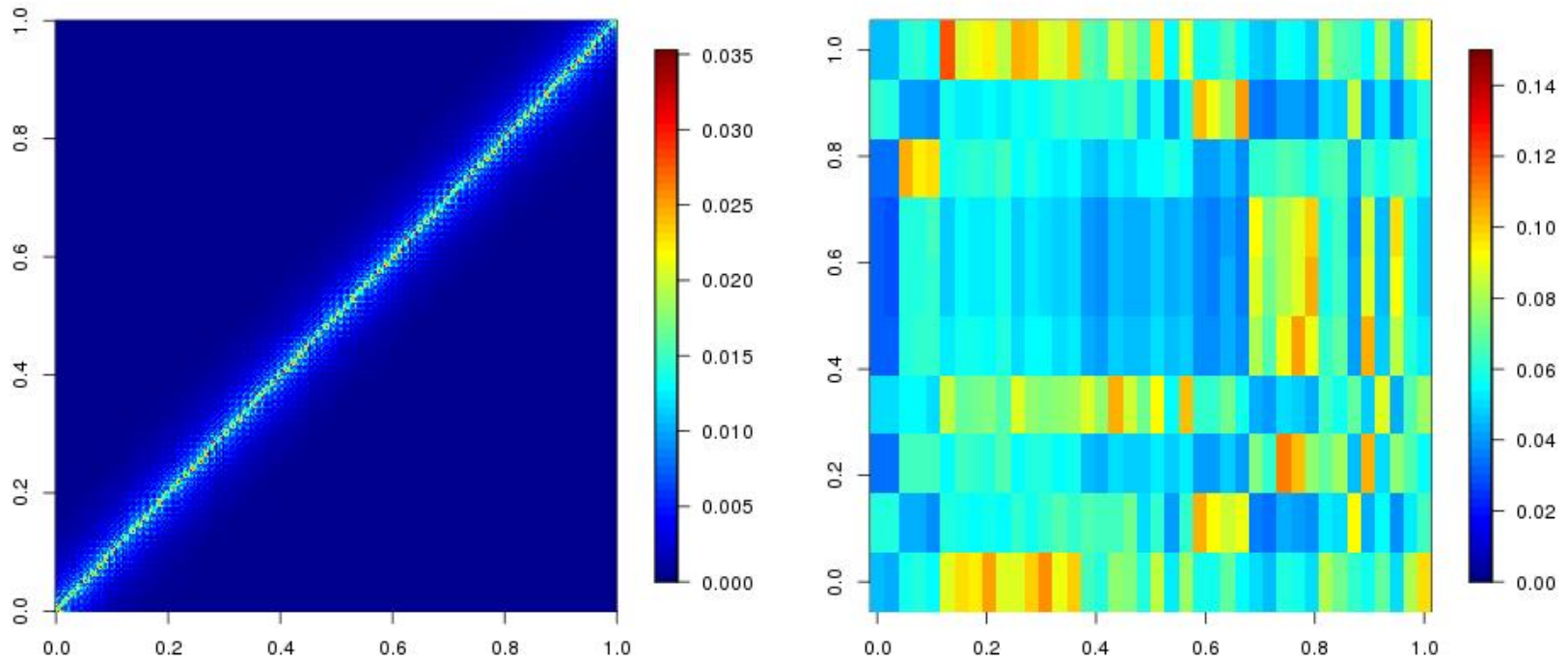

Figure 19: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (2008-10-15 $\sim$ 2011-12-21).

Model2 - DF - Prediction

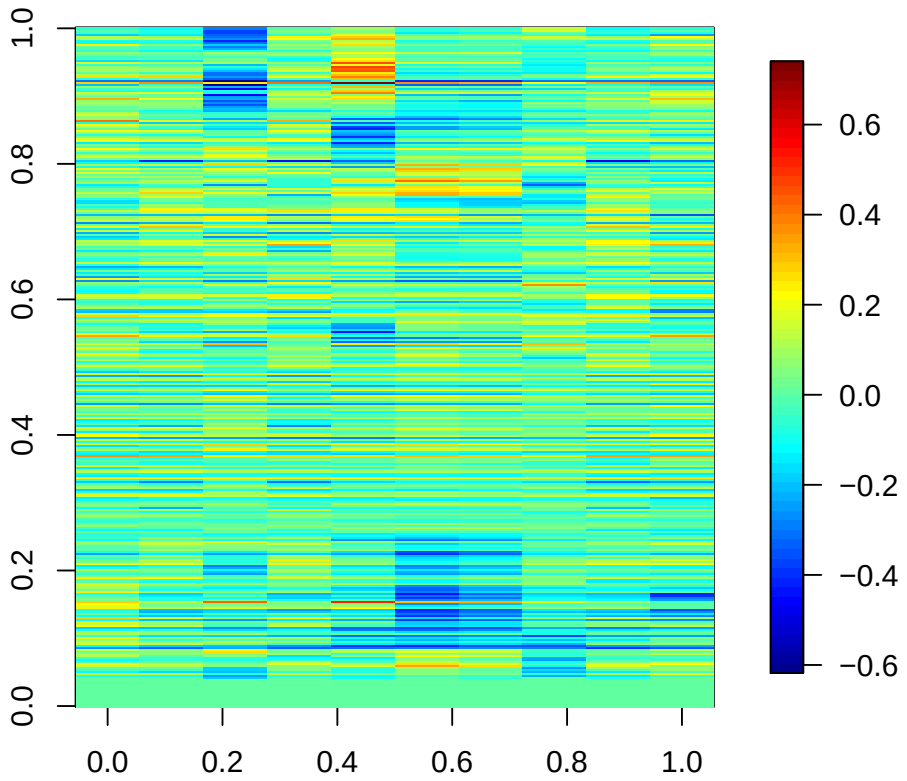

Figure 20: Image of 10×339 prediction values matrix.

Model3 - DF

Table 5: Maximum loglikelihood value and prediction measures for each model.

Model	Max.loglik	RMSE	MAE	CRPS
DF1	4119.963	0.1966	0.1613	0.1111
DF1*	3768.563	0.2109	0.1736	0.1209
DF2	4379.737	0.1827	0.1372	0.1007
DF2*	4035.548	0.1768	0.1332	0.0971
DF3	4606.413	0.1466	0.1201	0.0841
DF3*	4322.475	0.1430	0.1119	0.0792
DF4	3919.680	0.2180	0.1504	0.1191
DF4*	3581.579	0.2099	0.1399	0.1046

DF* results were obtained with different starting points.

Model3 - DF

Table 6: Parameter estimates for each model. β_s (km) and β_t (2 weeks).

Estimate	DF1/DF1*	DF2/DF2*	DF3/DF3*	DF4/DF4*
a_0	-0.01/0.01	-0.01/0.10	0.00/0.01	-0.01/0.04
a_1	0.02/0.00	0.01/-0.13	0.00/-0.02	0.01/-0.08
a_2	0.00/-0.23	0.00/-1.76	0.00/0.00	0.00/0.05
b_0	-0.01/0.01	0.00/0.01	0.00/-0.01	0.02/0.08
b_1	0.01/0.00	0.00/0.04	0.00/0.01	-0.03/-0.04
b_2	0.00/0.35	0.00/-0.82	0.00/-0.30	0.01/1.84
c_0	0.12/-0.03	0.12/-0.03	0.09/-0.01	-0.32/-0.05
c_1	-0.04/-0.06	0.53/-0.02	-0.09/0.06	-0.11/0.10
c_2	-5.95/0.77	2.09/0.67	0.95/-0.42	-4.08/0.39
β_s	16.32/1563.47	4.41/2338.87	9.43/976.21	13.28/4705.90
β_t	5.69/0.78	4.64/0.32	2.40/0.56	5.64/4.73
α	0.81/0.17	0.17/0.08	8.79/0.46	4.79/0.60
ν	1.00/1.98	1.28/2.48	1.81/1.99	1.00/2.34
β_s'	5791.49/773.75	4366.19/284.05	1424.56/568.62	3482.06/833.98
β_t'	1.88/0.29	0.30/0.25	3.76/0.17	0.94/0.05
α'	0.04/0.04	0.04/0.04	0.04/0.04	0.03/0.05
ν'	0.05/0.07	0.04/0.07	0.12/0.11	0.08/0.08

Model3 - DF1

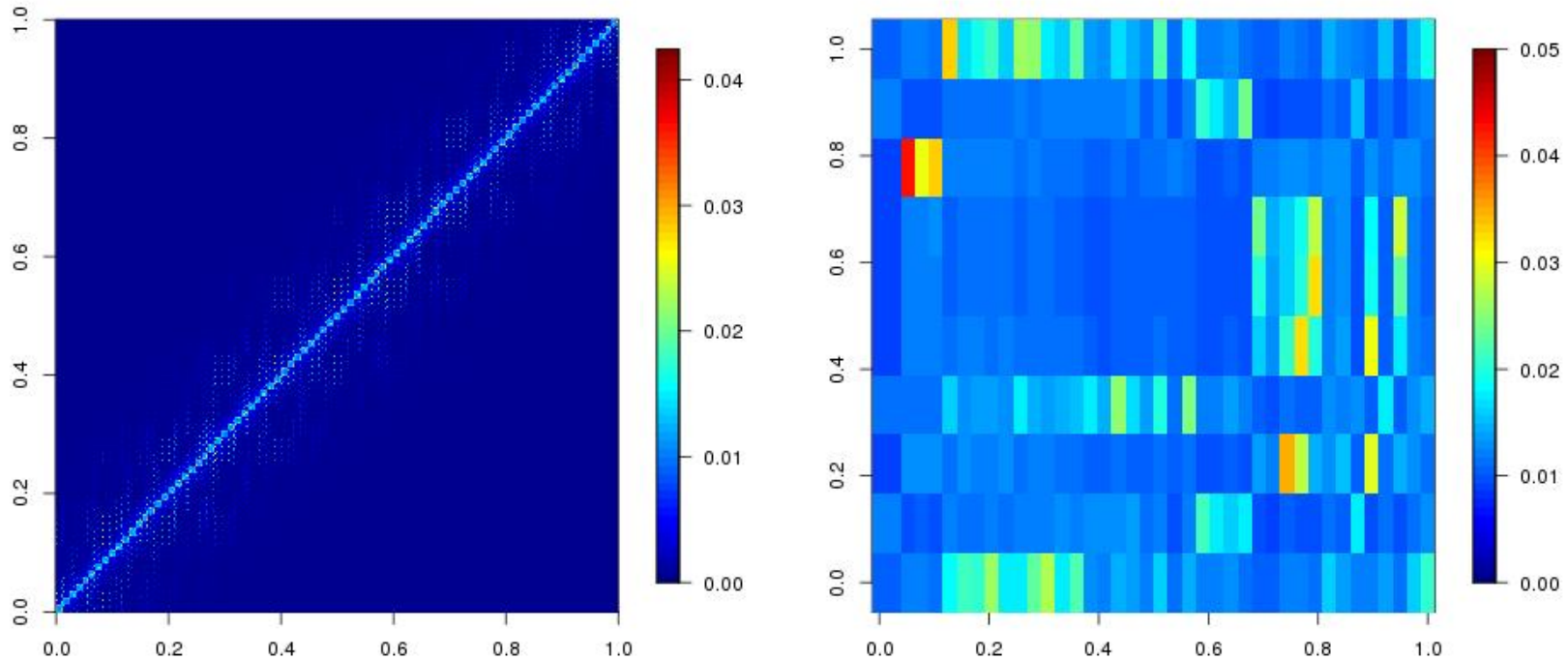

Figure 21: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (1999-01-06 $\sim$ 2002-03-27).

Model3 - DF2

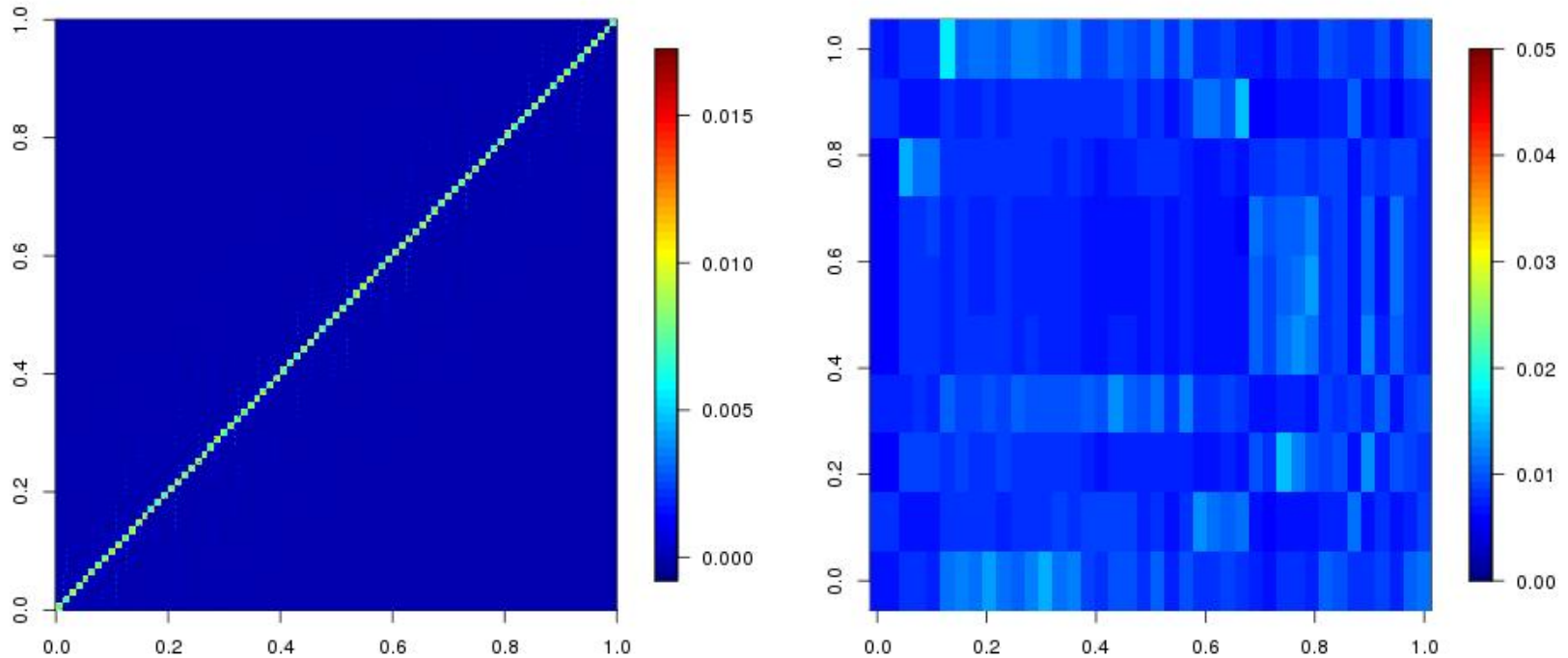

Figure 22: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (2002-04-10 $\sim$ 2005-06-29).

Model3 - DF3

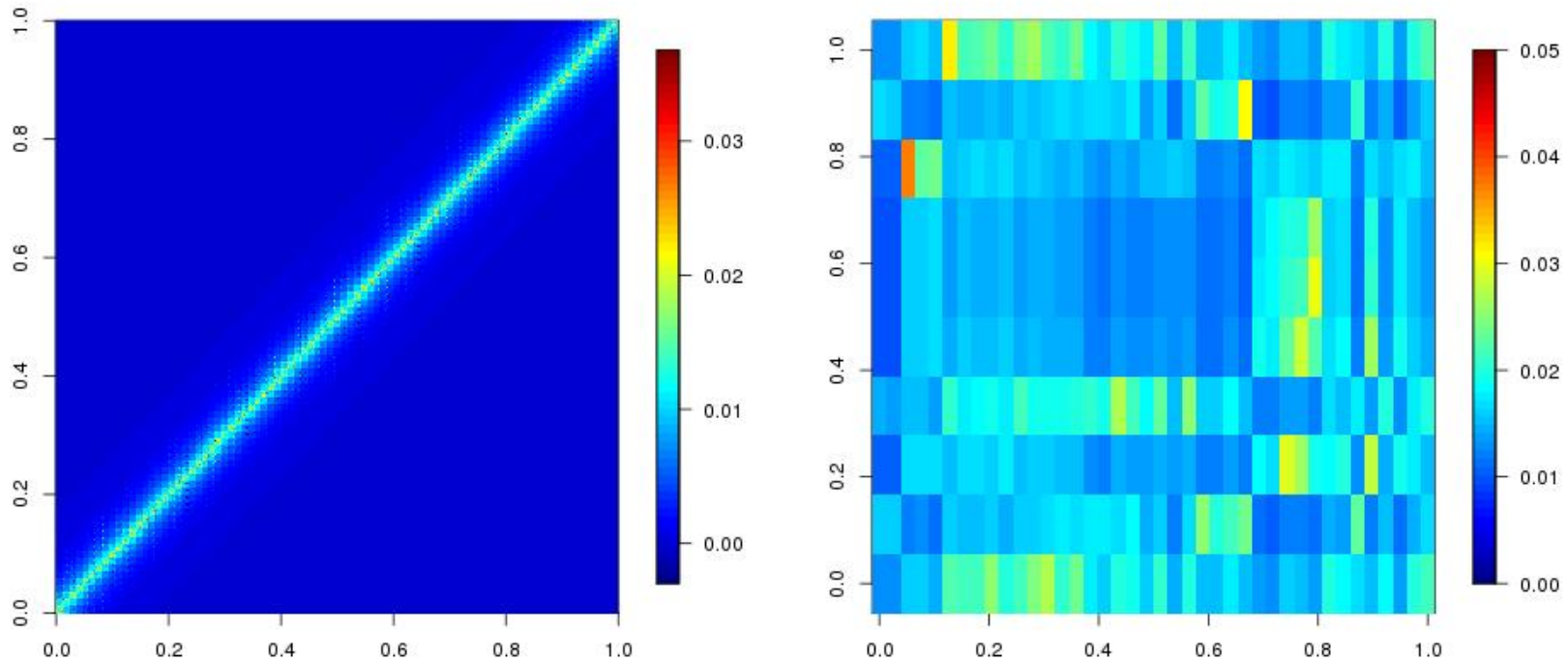

Figure 23: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (2005-07-13 $\sim$ 2008-10-01).

Model3 - DF4

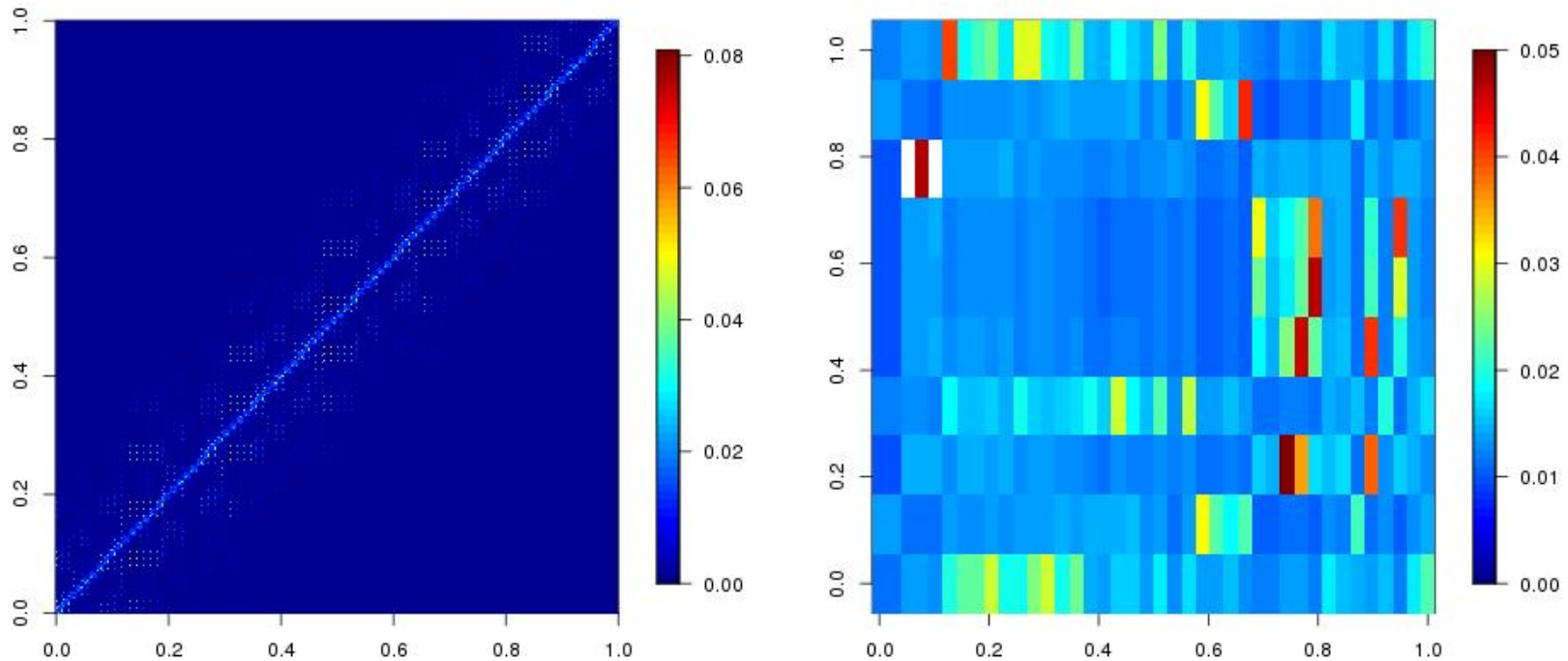

Figure 24: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (2008-10-15 $\sim$ 2011-12-21).

Model3 - DF - Prediction

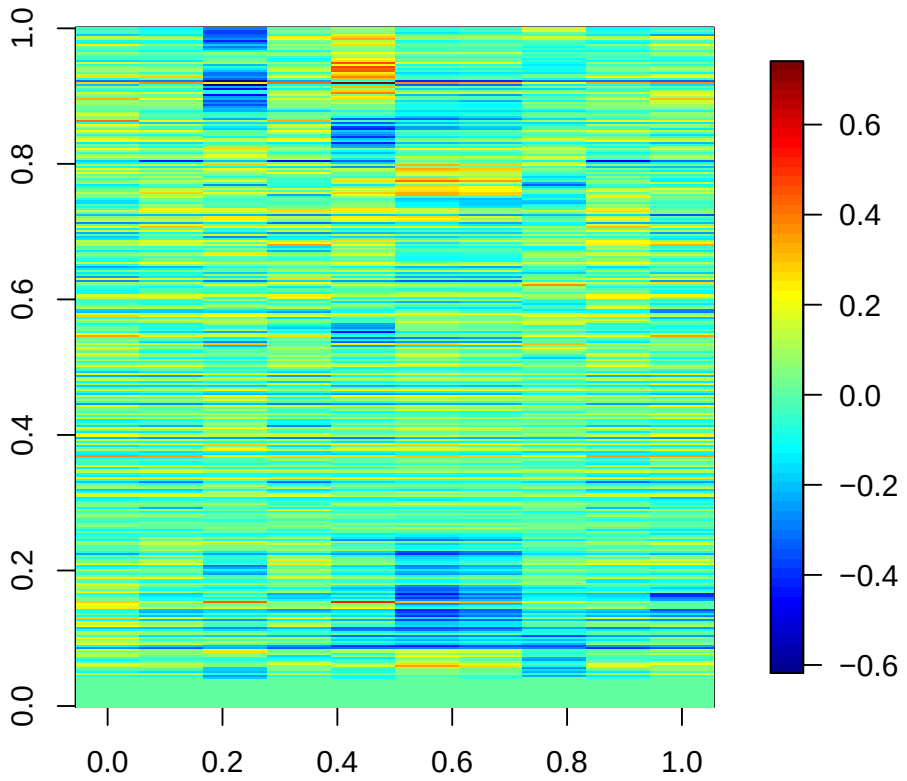

Figure 25: Image of 10×339 prediction values matrix.

Model3 - DF1*

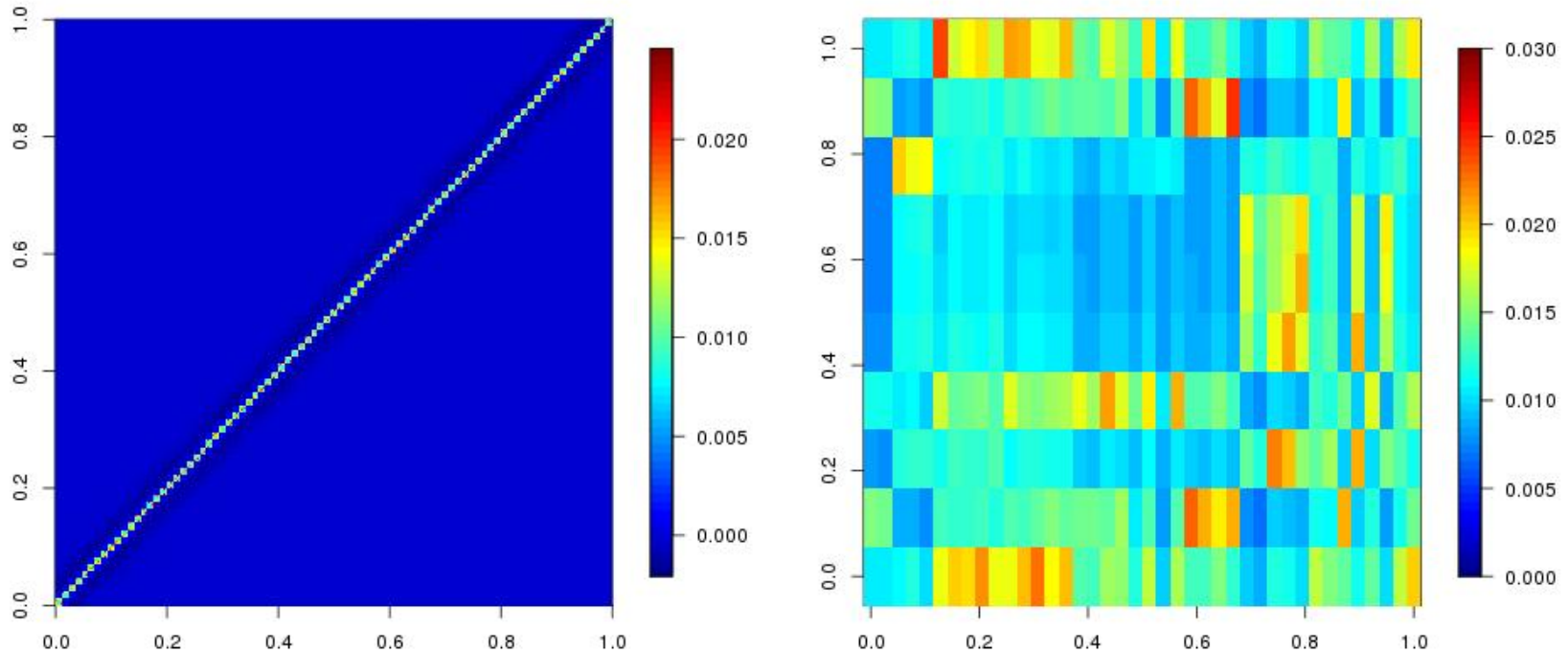

Figure 26: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (1999-01-06 $\sim$ 2002-03-27).

Model3 - DF2*

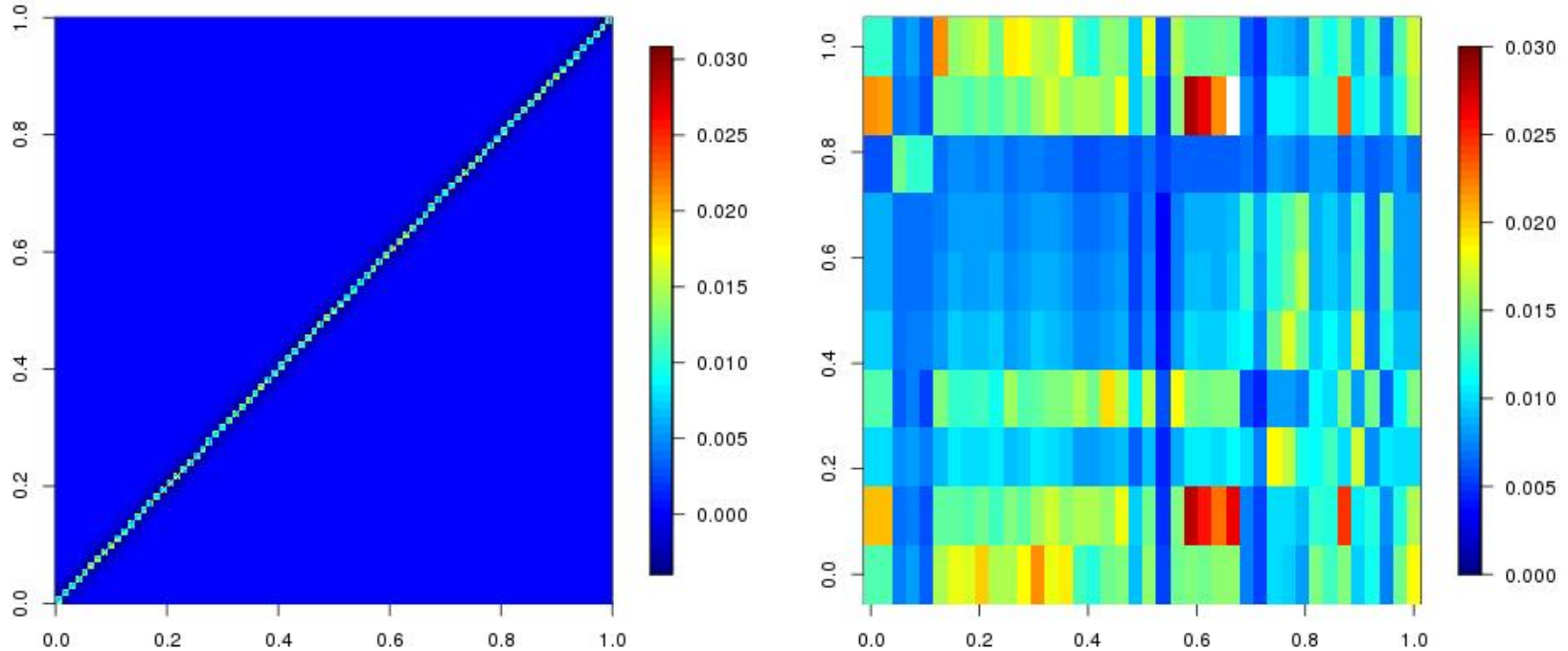

Figure 27: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (2002-04-10 $\sim$ 2005-06-29).

Model3 - DF3*

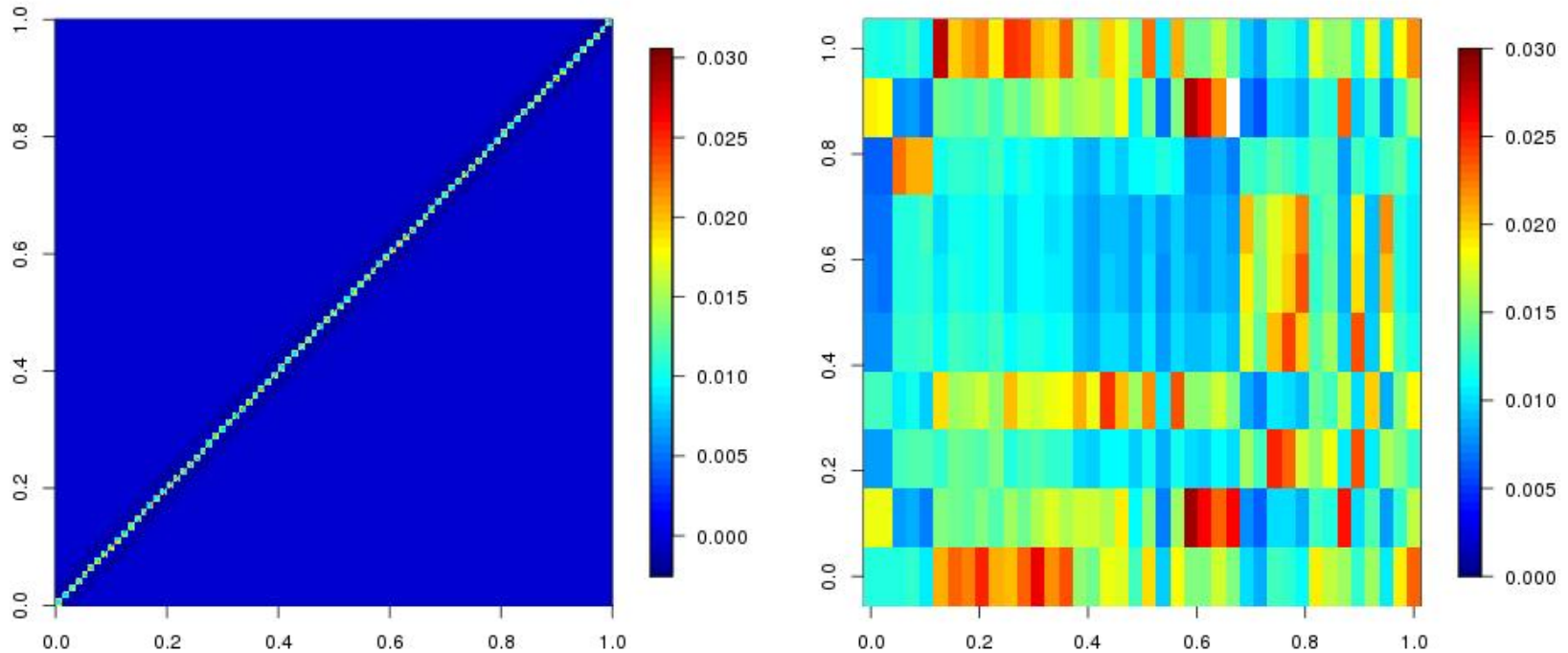

Figure 28: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (2005-07-13 $\sim$ 2008-10-01).

Model3 - DF4*

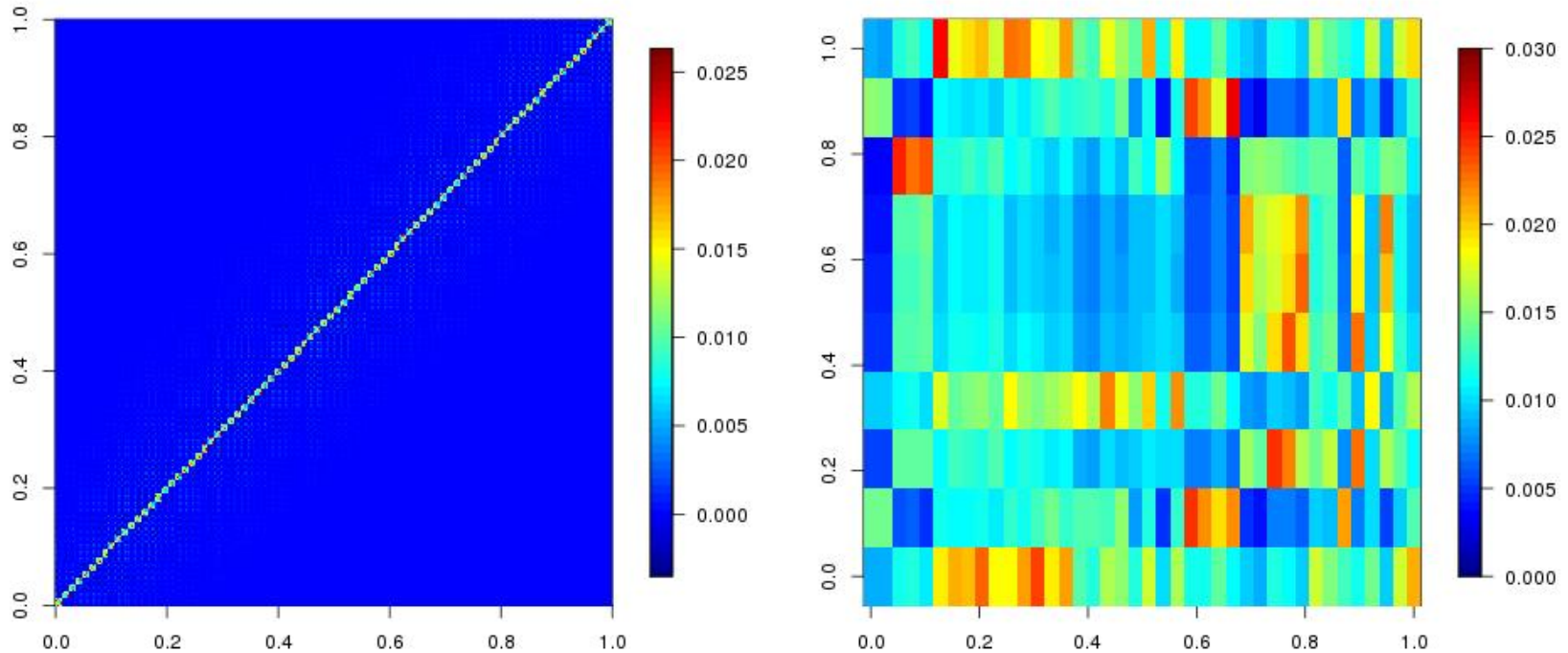

Figure 29: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (2008-10-15 $\sim$ 2011-12-21).

Model3 - DF* - Prediction

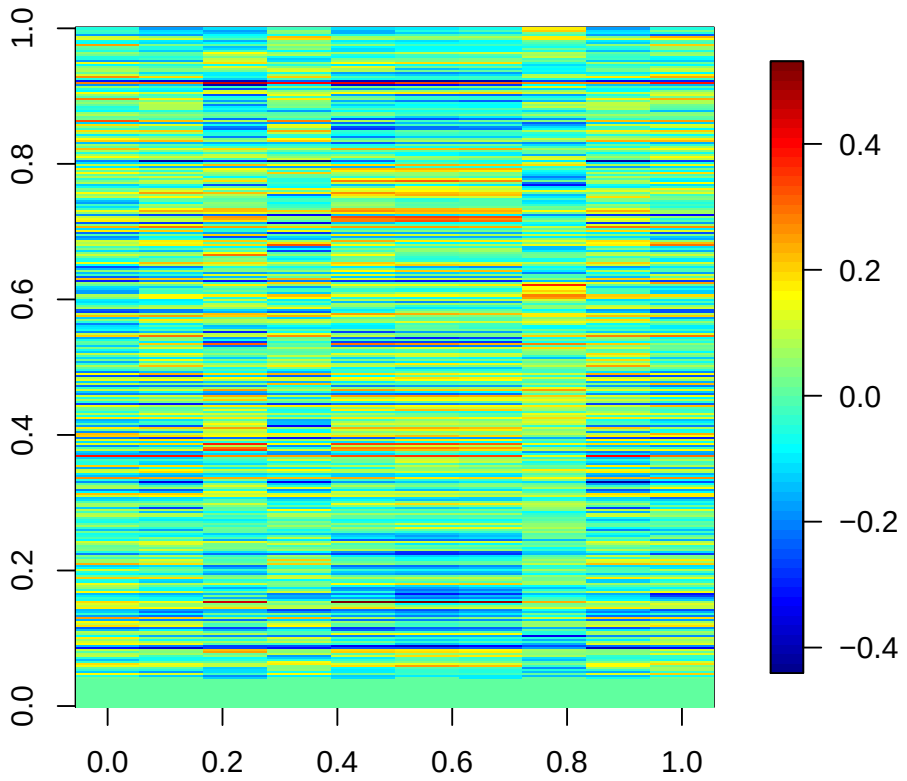

Figure 30: Image of 10×339 prediction values matrix.

Model4 - Exponential

Table 7: Maximum loglikelihood value and prediction measures for each model.

Model	Max.loglik	RMSE	MAE	CRPS
EX1	3673.584	0.2092	0.1724	0.1207
EX2	3762.053	0.1999	0.1578	0.1128
EX3	4100.743	0.1514	0.1204	0.0857
EX4	3299.550	0.2355	0.1644	0.1208

Model4 - Exponential

Table 8: Parameter estimates for each model. β_s : spatial range (km), β_t temporal range (2 weeks), and α : scale parameter.

Estimate	EX1	EX2	EX3	EX4
β_s	28.07	25.46	32.31	17.82
β_t	0.82	0.57	0.57	0.64
α	0.05	0.05	0.04	0.05

Model4 - EX1

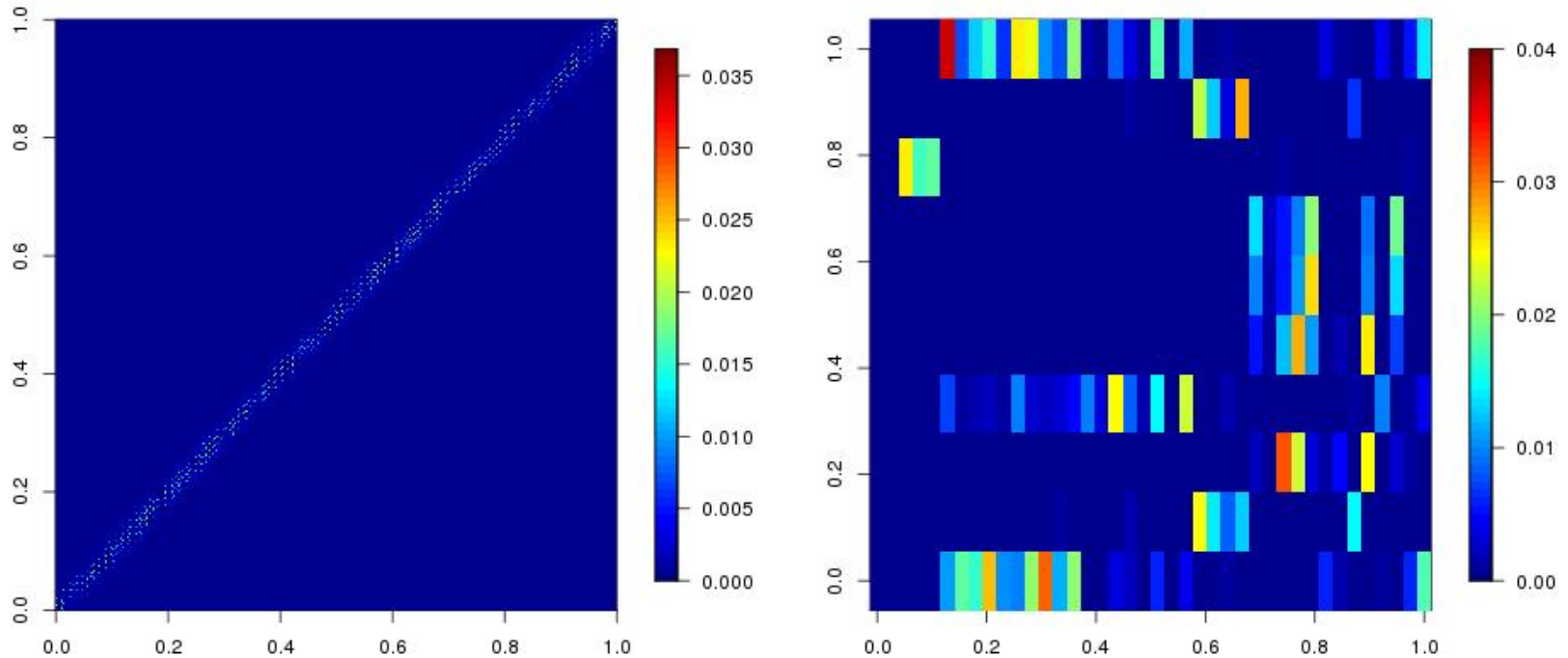

Figure 31: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (1999-01-06 $\sim$ 2002-03-27).

Model4 - EX2

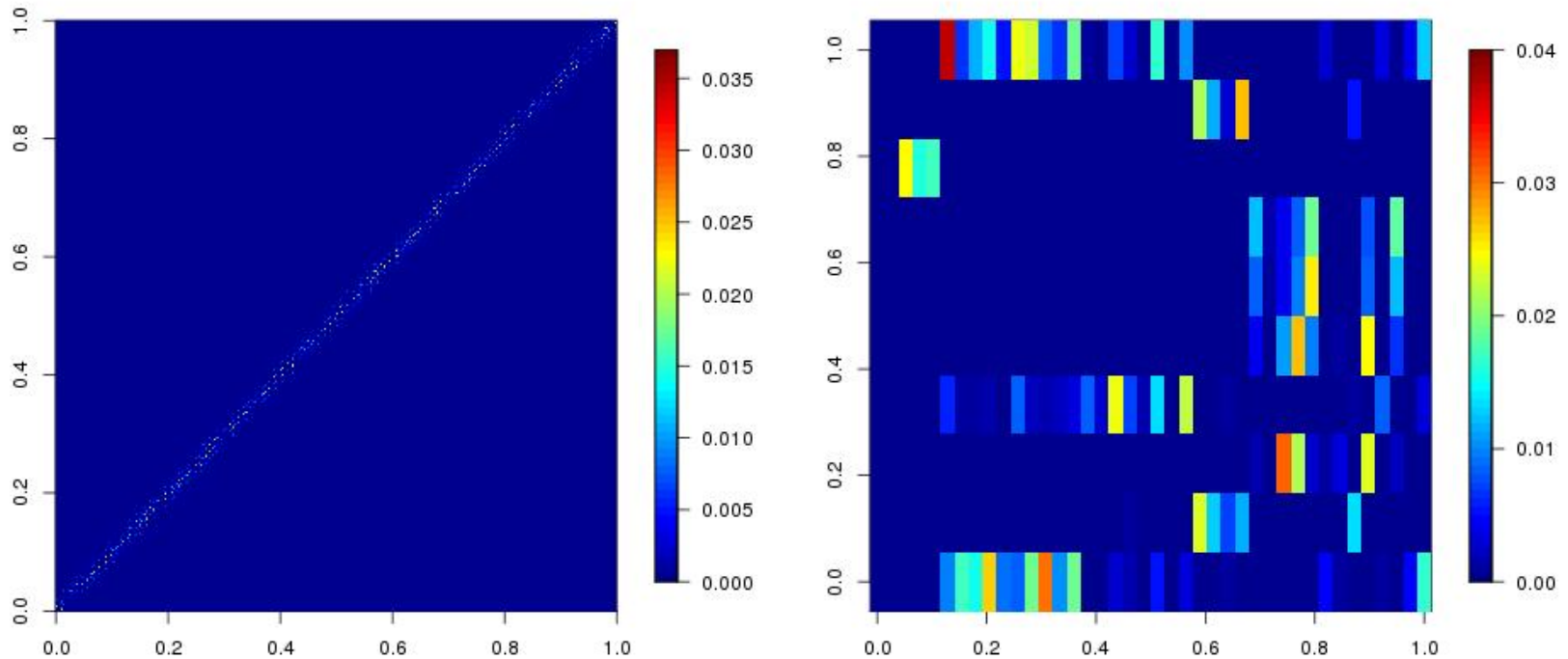

Figure 32: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (2002-04-10 $\sim$ 2005-06-29).

Model4 - EX3

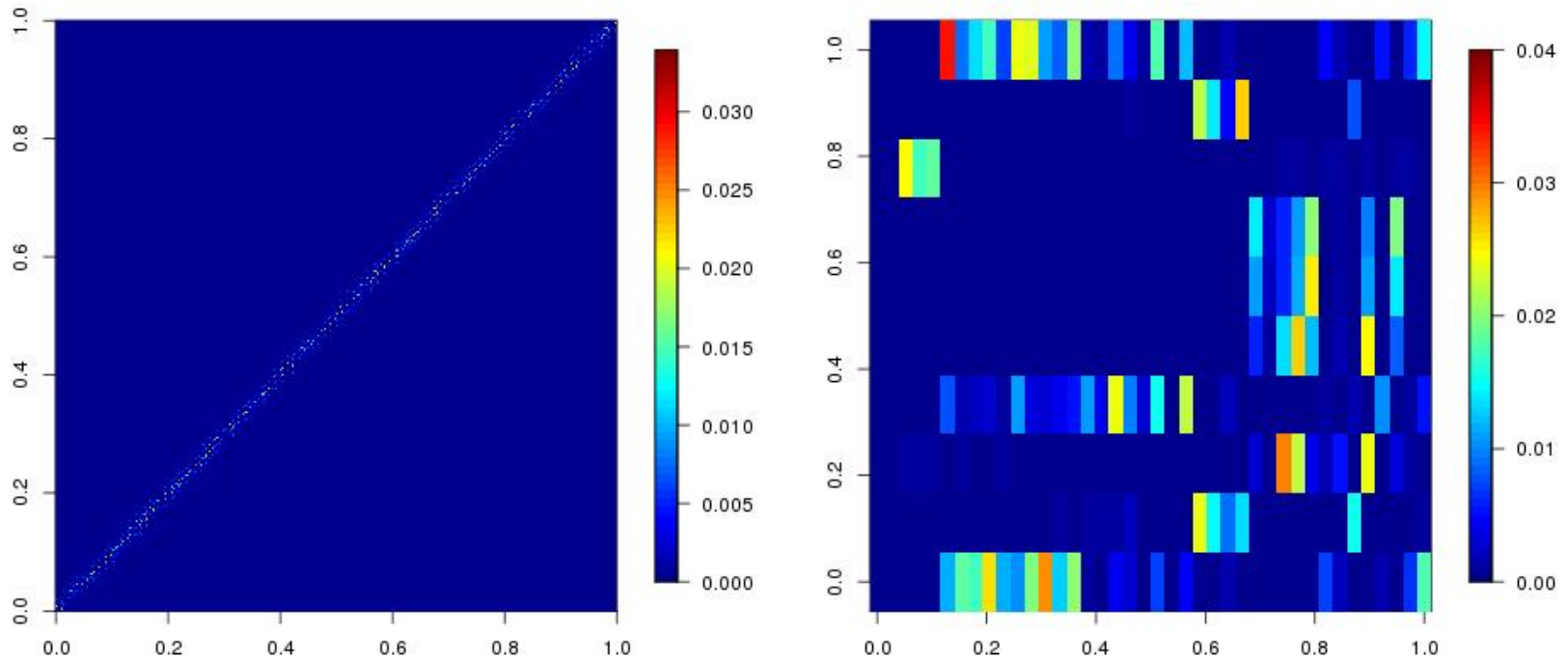

Figure 33: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (2005-07-13 $\sim$ 2008-10-01).

Model4 - EX4

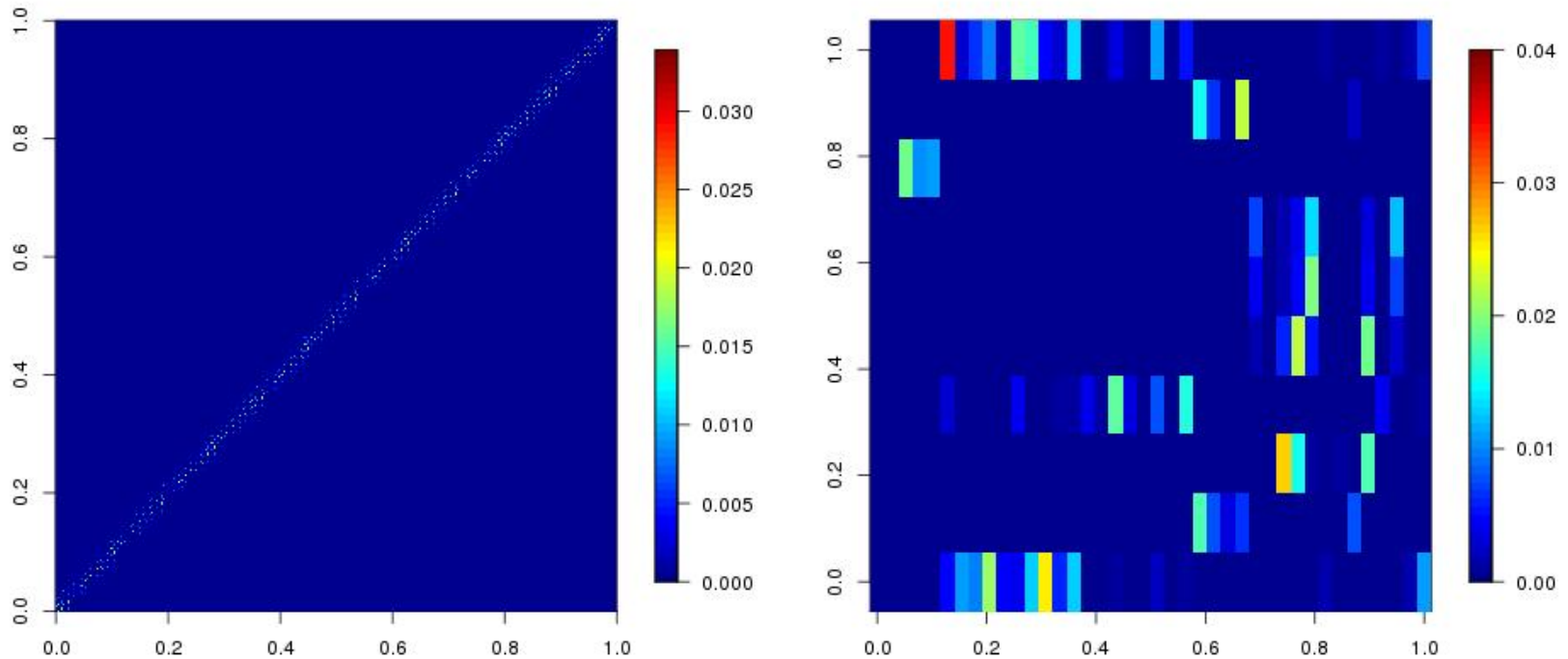

Figure 34: 3400×850 covariance matrix and 40×10 covariance matrix for prediction (2008-10-15 $\sim$ 2011-12-21).

Model4 - EX - Prediction

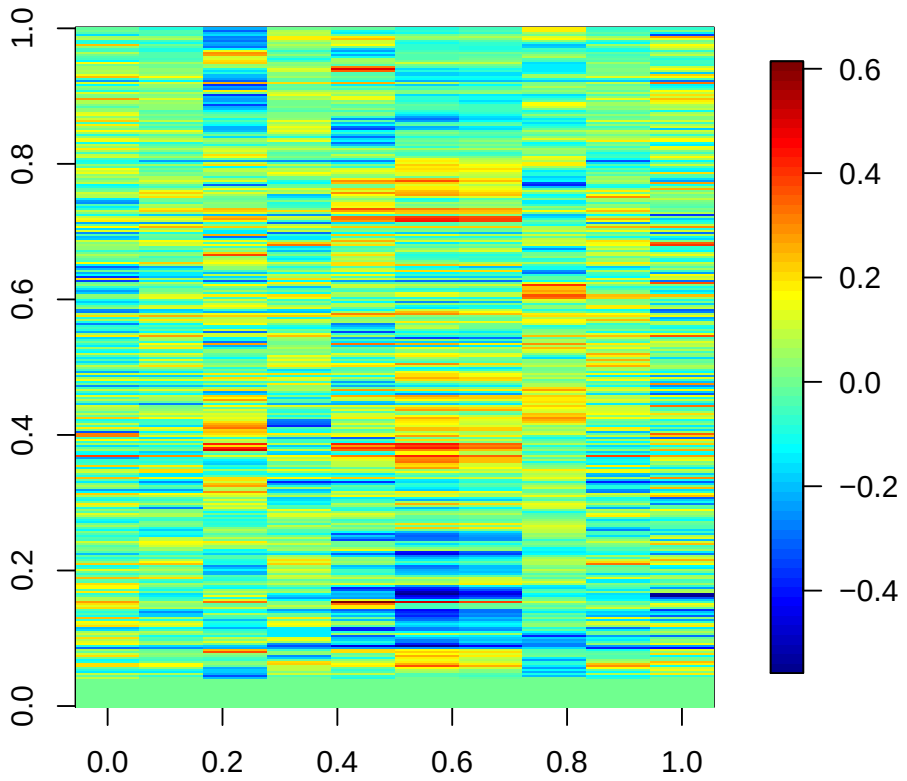

Figure 35: Image of 10×339 prediction values matrix.