Spatio-temporal modelling of environmental monitoring data

Luís Manuel S. M. Margalho

Coimbra Institute of Engineering

II Xornada de Usuarios R en Galicia

October 1, 2015

Motivation example

Spatio-temporal data results from observations collected not only in space but also along time.

Motivation example

Spatio-temporal data results from observations collected not only in space but also along time.

Some environmental applications of this kind of data may be found in the literature:

- Cocchi *et al.* (2007) **Hierarchical space-time modelling of PM_{10} pollution**, under the Bayesian framework, use data from 11 spatial locations collected over 1096 days
- Mitchell *et al.* (2005) **Testing for separability of space-time covariances**, aiming to study the effect of high levels of CO_2 on rice, use data from 13 spatial locations and 112 time points and test the separability of the proposed spatio-temporal model by rearranging the data as in the context of multivariate repeated measures
- Bruno *et al.* (2003) **Non-separability of space-time covariance models in environmental studies**, use a data set consisting of daily ozone measurements made at 32 monitoring locations, for the period 1998-2002, enabling the identification of the temporal variability which, when removed, leaves separable space and time correlation components

These examples share one common feature: the number of time observations is (much) larger than the number of spatial locations.

Motivation example

Spatio-temporal data results from observations collected not only in space but also along time.

Some environmental applications of this kind of data may be found in the literature:

- Cocchi *et al.* (2007) **Hierarchical space-time modelling of PM_{10} pollution**, under the Bayesian framework, use data from 11 spatial locations collected over 1096 days
- Mitchell *et al.* (2005) **Testing for separability of space-time covariances**, aiming to study the effect of high levels of CO_2 on rice, use data from 13 spatial locations and 112 time points and test the separability of the proposed spatio-temporal model by rearranging the data as in the context of multivariate repeated measures
- Bruno *et al.* (2003) **Non-separability of space-time covariance models in environmental studies**, use a data set consisting of daily ozone measurements made at 32 monitoring locations, for the period 1998-2002, enabling the identification of the temporal variability which, when removed, leaves separable space and time correlation components

These examples share one common feature: the number of time observations is (much) larger than the number of spatial locations.

There are cases, however, where data are collected over several locations but only few times, disabling the use of time series techniques.

Geoestatistics / Spatial Statistics

- Geostatistical data results from the measurement of something over some spatial region

Geoestatistics / Spatial Statistics

- Geostatistical data results from the measurement of something over some spatial region
- Sampling locations continuously spread over the spatial region

Geoestatistics / Spatial Statistics

- Geostatistical data results from the measurement of something over some spatial region
- Sampling locations continuously spread over the spatial region
- Tobler's Law: "Everything is related to everything else, but near things are more related than distant things"

Geoestatistics / Spatial Statistics

- Geostatistical data results from the measurement of something over some spatial region
- Sampling locations continuously spread over the spatial region
- Tobler's Law: "Everything is related to everything else, but near things are more related than distant things"
- Geostatistical models allow to interpolate the response variable at non-sampled locations, weighting values at nearby locations

Applications in

- mining industry (beginning of geostatistics, D. Krige, 1951)

Applications in

- mining industry (beginning of geostatistics, D. Krige, 1951)
- meteorology

Applications in

- mining industry (beginning of geostatistics, D. Krige, 1951)
- meteorology
- health (Broad Street cholera outbreak in 1854)

Applications in

- mining industry (beginning of geostatistics, D. Krige, 1951)
- meteorology
- health (Broad Street cholera outbreak in 1854)
- environment
- $\cdots$

Spatio-temporal Geostatistics

- Modern technologies make easy to gather data exhibiting both a spatial and a temporal dimension

Spatio-temporal Geostatistics

- Modern technologies make easy to gather data exhibiting both a spatial and a temporal dimension
- Spatio-temporal geostatistical models take into account not only *where* data was collected, but also *when* they were collected

Spatio-temporal Geostatistics

- Modern technologies make easy to gather data exhibiting both a spatial and a temporal dimension
- Spatio-temporal geostatistical models take into account not only *where* data was collected, but also *when* they were collected
- Extending the spatial framework to the spatio-temporal framework is more than just the inclusion of one more dimension. It must reflect the diference in nature of the two dimensions involved

Goals of this work

- Work mainly motivated by environmental monitoring studies

Goals of this work

- Work mainly motivated by environmental monitoring studies
- Understand the importance of including in the prediction task
    - information from the past

Goals of this work

- Work mainly motivated by environmental monitoring studies
- Understand the importance of including in the prediction task
  - information from the past
  - covariates explaining the process under observation

Goals of this work

- Work mainly motivated by environmental monitoring studies
- Understand the importance of including in the prediction task
    - information from the past
    - covariates explaining the process under observation
- Developed in two stages
    - extending an existing model by including covariates

Goals of this work

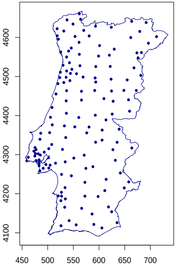
- Work mainly motivated by environmental monitoring studies
- Understand the importance of including in the prediction task
  - information from the past
  - covariates explaining the process under observation
- Developed in two stages
  - extending an existing model by including covariates
  - proposing a model suitable for studies with sparse time dimension

Application: environmental monitoring in mainland Portugal

- Data concerning heavy metal (Mn and Pb) deposition in mosses

Application: environmental monitoring in mainland Portugal

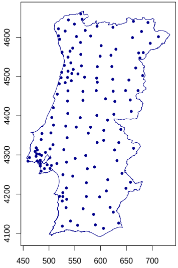
- Data concerning heavy metal (Mn and Pb) deposition in mosses



- 146 sampling locations

Application: environmental monitoring in mainland Portugal

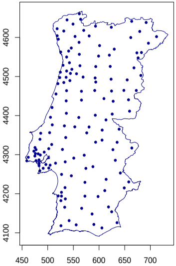
- Data concerning heavy metal (Mn and Pb) deposition in mosses



- 146 sampling locations
- intensified sampling at regions with high industrial or urban density (covariate to be considered)

Application: environmental monitoring in mainland Portugal

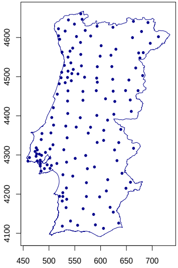
- Data concerning heavy metal (Mn and Pb) deposition in mosses



- 146 sampling locations
- intensified sampling at regions with high industrial or urban density (covariate to be considered)
- 3 surveys (1992, 1996, 2002)

Application: environmental monitoring in mainland Portugal

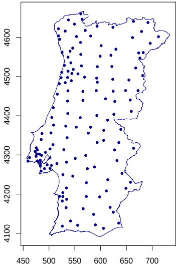
- Data concerning heavy metal (Mn and Pb) deposition in mosses



- 146 sampling locations
- intensified sampling at regions with high industrial or urban density (covariate to be considered)
- 3 surveys (1992, 1996, 2002)
- aim: predict heavy metal concentration at unsampled locations, for the most recent survey

Application: environmental monitoring in mainland Portugal

- Data concerning heavy metal (Mn and Pb) deposition in mosses

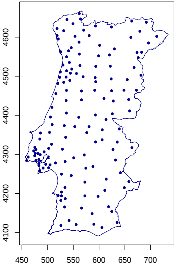


- 146 sampling locations
- intensified sampling at regions with high industrial or urban density (covariate to be considered)
- 3 surveys (1992, 1996, 2002)
- aim: predict heavy metal concentration at unsampled locations, for the most recent survey

- To assess the accuracy of predictions for the most recent survey: 4 different prediction models under comparison

Application: environmental monitoring in mainland Portugal

- Data concerning heavy metal (Mn and Pb) deposition in mosses

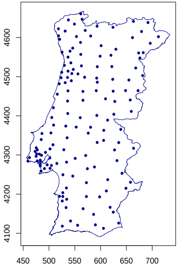


- 146 sampling locations
- intensified sampling at regions with high industrial or urban density (covariate to be considered)
- 3 surveys (1992, 1996, 2002)
- aim: predict heavy metal concentration at unsampled locations, for the most recent survey

- To assess the accuracy of predictions for the most recent survey: 4 different prediction models under comparison
    - spatial model with / without covariate included

Application: environmental monitoring in mainland Portugal

- Data concerning heavy metal (Mn and Pb) deposition in mosses



- 146 sampling locations
- intensified sampling at regions with high industrial or urban density (covariate to be considered)
- 3 surveys (1992, 1996, 2002)
- aim: predict heavy metal concentration at unsampled locations, for the most recent survey

- To assess the accuracy of predictions for the most recent survey: 4 different prediction models under comparison
  - spatial model with / without covariate included
  - spatio-temporal model with / without covariate included

Data summary of Mn concentration (observed and Box-Cox transformed values)

	Survey					
	1992		1996		2002	
	Observ.	Transf.	Observ.	Transf.	Observ.	Transf.
Min	16.00	3.39	4.03	1.54	23.14	3.96
Median	123.50	6.90	149.18	7.28	123.20	6.89
Max	970.00	11.63	685.55	10.73	503.11	9.97
Mean	161.62	6.88	178.62	7.13	147.12	6.85
St. dev.	147.97	1.69	136.01	1.65	99.28	1.41

Data summary of Pb concentration (observed and Box-Cox transformed values)

	Survey					
	1992		1996		2002	
	Observ.	Transf.	Observ.	Transf.	Observ.	Transf.
Min	0.50	-0.67	2.00	0.71	0.68	-0.38
Median	13.00	2.78	15.70	2.99	3.11	1.17
Max	172.00	6.04	191.17	6.18	109.93	5.44
Mean	16.40	2.75	21.72	3.03	5.40	1.29
St. dev.	16.91	0.86	24.05	0.92	10.39	0.90

One spatio-temporal model

- Model proposed by *Høst et al. (1995)*

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \omega(\mathbf{s}, t)\varepsilon(\mathbf{s}, t)$$

One spatio-temporal model

- Model proposed by *Høst et al. (1995)*

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \omega(\mathbf{s}, t)\varepsilon(\mathbf{s}, t)$$

  - $\mu(\mathbf{s}, t) = M_{\mathbf{1}}(\mathbf{s}) + m_{\mathbf{2}}(t)$

One spatio-temporal model

- Model proposed by *Høst et al. (1995)*

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \omega(\mathbf{s}, t)\varepsilon(\mathbf{s}, t)$$

  - $\mu(\mathbf{s}, t) = M_{\mathbf{1}}(\mathbf{s}) + m_{\mathbf{2}}(t)$
  - $\omega(\mathbf{s}, t) = S_{\mathbf{1}}(\mathbf{s})s_{\mathbf{2}}(t)$

One spatio-temporal model

- Model proposed by *Høst et al. (1995)*

$$Y(\mathbf{s},t) = \mu(\mathbf{s},t) + \omega(\mathbf{s},t)\varepsilon(\mathbf{s},t)$$

  - $\mu(\mathbf{s},t) = M_1(\mathbf{s}) + m_2(t)$
  - $\omega(\mathbf{s},t) = S_1(\mathbf{s})s_2(t)$
  - $\varepsilon(\mathbf{s},t)$ residuals

One spatio-temporal model

- Model proposed by *Høst et al. (1995)*

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \omega(\mathbf{s}, t)\varepsilon(\mathbf{s}, t)$$

  - $\mu(\mathbf{s}, t) = M_{\mathbf{1}}(\mathbf{s}) + m_{\mathbf{2}}(t)$
  - $\omega(\mathbf{s}, t) = S_{\mathbf{1}}(\mathbf{s})s_{\mathbf{2}}(t)$
  - $\varepsilon(\mathbf{s}, t)$ residuals

- Mean random field mainly decomposed by a spatial component added with a time correction: suitable when the temporal dimension is lower than the spatial

One spatio-temporal model

- Model proposed by *Høst et al. (1995)*

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + \omega(\mathbf{s}, t)\varepsilon(\mathbf{s}, t)$$

  - $\mu(\mathbf{s}, t) = M_{\mathbf{1}}(\mathbf{s}) + m_{\mathbf{2}}(t)$
  - $\omega(\mathbf{s}, t) = S_{\mathbf{1}}(\mathbf{s})s_{\mathbf{2}}(t)$
  - $\varepsilon(\mathbf{s}, t)$ residuals

- Mean random field mainly decomposed by a spatial component added with a time correction: suitable when the temporal dimension is lower than the spatial

- Extending the model allowing for the inclusion of spatial covariates

$$M_1(\mathbf{s}) = \sum_{i=1}^{p} \beta_i f_i(\mathbf{s})$$

Summary of predicted concentration for the 2002 survey (with interpolation errors)

		Spatial model				Spatio-temporal model			
		without covar.		with covar.		without covar.		with covar	
Mn									
	Min	4.88	(1.04)	4.39	(1.03)	4.87	(0.92)	4.30	(0.80)
	Median	6.99	(1.26)	7.57	(1.25)	7.21	(1.16)	7.53	(1.07)
	Max	8.59	(1.47)	8.58	(1.37)	8.72	(1.45)	8.93	(1.18)
	Mean	6.97	(1.26)	7.40	(1.24)	7.03	(1.17)	7.37	(1.06)
	St. dev.	0.63	(0.12)	0.63	(0.09)	0.73	(0.14)	0.66	(0.12)
Pb									
	Min	0.19	(0.70)	0.13	(0.66)	0.67	(0.59)	0.61	(0.59)
	Median	1.19	(0.88)	0.89	(0.85)	1.15	(0.66)	0.87	(0.63)
	Max	3.01	(0.94)	3.25	(0.89)	2.16	(0.70)	2.00	(0.65)
	Mean	1.20	(0.86)	0.99	(0.83)	1.16	(0.66)	0.84	(0.63)
	St. dev.	0.28	(0.07)	0.33	(0.05)	0.19	(0.03)	0.25	(0.02)

Summary of predicted concentration for the 2002 survey (with interpolation errors)

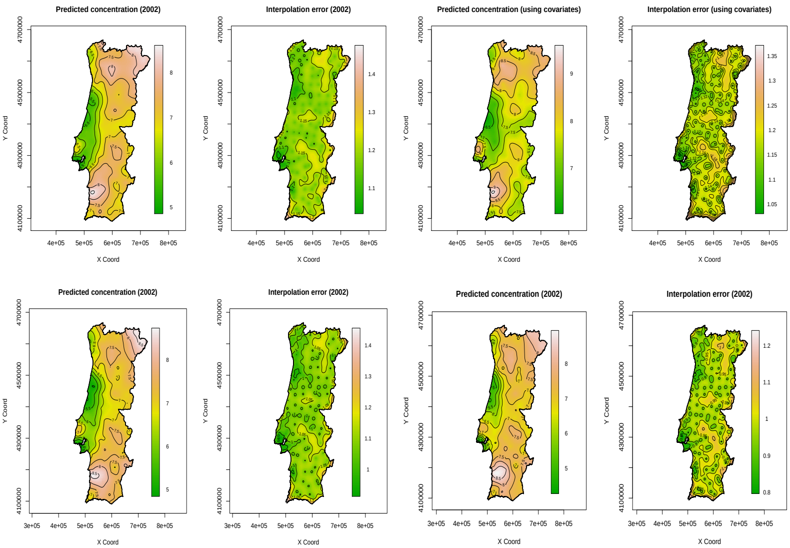
		Spatial model				Spatio-temporal model			
		without covar.		with covar.		without covar.		with covar	
Mn									
	Min	4.88	(1.04)	4.39	(1.03)	4.87	(0.92)	4.30	(0.80)
	Median	6.99	(1.26)	7.57	(1.25)	7.21	(1.16)	7.53	(1.07)
	Max	8.59	(1.47)	8.58	(1.37)	8.72	(1.45)	8.93	(1.18)
	Mean	6.97	(1.26)	7.40	(1.24)	7.03	(1.17)	7.37	(1.06)
	St. dev.	0.63	(0.12)	0.63	(0.09)	0.73	(0.14)	0.66	(0.12)
Pb									
	Min	0.19	(0.70)	0.13	(0.66)	0.67	(0.59)	0.61	(0.59)
	Median	1.19	(0.88)	0.89	(0.85)	1.15	(0.66)	0.87	(0.63)
	Max	3.01	(0.94)	3.25	(0.89)	2.16	(0.70)	2.00	(0.65)
	Mean	1.20	(0.86)	0.99	(0.83)	1.16	(0.66)	0.84	(0.63)
	St. dev.	0.28	(0.07)	0.33	(0.05)	0.19	(0.03)	0.25	(0.02)

- The inclusion of covariate corrects underestimation of Mn and overestimation of Pb

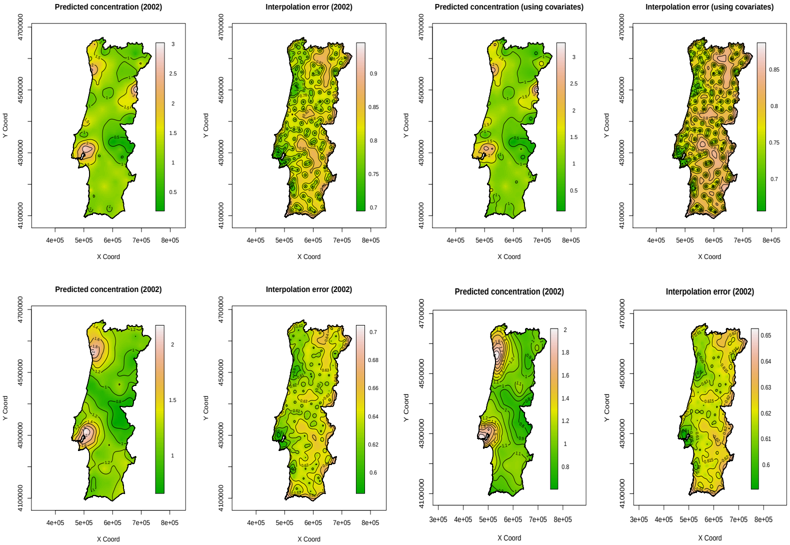
Summary of predicted concentration for the 2002 survey (with interpolation errors)

		Spatial model				Spatio-temporal model			
		without covar.		with covar.		without covar.		with covar	
Mn									
	Min	4.88	(1.04)	4.39	(1.03)	4.87	(0.92)	4.30	(0.80)
	Median	6.99	(1.26)	7.57	(1.25)	7.21	(1.16)	7.53	(1.07)
	Max	8.59	(1.47)	8.58	(1.37)	8.72	(1.45)	8.93	(1.18)
	Mean	6.97	(1.26)	7.40	(1.24)	7.03	(1.17)	7.37	(1.06)
	St. dev.	0.63	(0.12)	0.63	(0.09)	0.73	(0.14)	0.66	(0.12)
Pb									
	Min	0.19	(0.70)	0.13	(0.66)	0.67	(0.59)	0.61	(0.59)
	Median	1.19	(0.88)	0.89	(0.85)	1.15	(0.66)	0.87	(0.63)
	Max	3.01	(0.94)	3.25	(0.89)	2.16	(0.70)	2.00	(0.65)
	Mean	1.20	(0.86)	0.99	(0.83)	1.16	(0.66)	0.84	(0.63)
	St. dev.	0.28	(0.07)	0.33	(0.05)	0.19	(0.03)	0.25	(0.02)

- The inclusion of covariate corrects underestimation of Mn and overestimation of Pb
- Lower amount of interpolation error attained by the most informative model



Prediction and interpolation error maps for Mn (top row: spatial models, bottom row: spatio-temporal models; 2 left panel: without covariates, 2 right panel: with covariates)



Prediction and interpolation error maps for Pb (top row: spatial models, bottom row: spatio-temporal models, 2 left panel: without covariates, 2 right panel: with covariates)

- Different prediction methodologies lead to different predicted results

- Different prediction methodologies lead to different predicted results
- To understand the accuracy of predicted values, an exercise of cross-validation was performed

- Different prediction methodologies lead to different predicted results
- To understand the accuracy of predicted values, an exercise of cross-validation was performed
- **A**bsolute **P**rediction **E**rror as a discrepancy measure between the observed and the predicted value

Mean and standard deviation of the APE

		Spatial Model		Spatio-temporal Model	
		without covar.	with covar.	without covar.	with covar.
Mn					
	MAPE	0.95	0.93	0.94	0.92
	Sd(APE)	0.65	0.65	0.64	0.65
Pb					
	MAPE	0.61	0.59	0.57	0.56
	Sd(APE)	0.63	0.62	0.63	0.63

- Different prediction methodologies lead to different predicted results
- To understand the accuracy of predicted values, an exercise of cross-validation was performed
- **A**bsolute **P**rediction **E**rror as a discrepancy measure between the observed and the predicted value

Mean and standard deviation of the APE

		Spatial Model		Spatio-temporal Model	
		without covar.	with covar.	without covar.	with covar.
Mn					
	MAPE	0.95	0.93	0.94	0.92
	Sd(APE)	0.65	0.65	0.64	0.65
Pb					
	MAPE	0.61	0.59	0.57	0.56
	Sd(APE)	0.63	0.62	0.63	0.63

- Details in Margalho L., Menezes R., Sousa I. (2014). *Assessing interpolation error for space–time monitoring data*, Stochastic Environmental Research and Risk Assessment, 28:1307–1321

New model proposal

- Our proposal

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + Z(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t)$$

New model proposal

- Our proposal

$$Y(\mathbf{s}, t) = \mu(\mathbf{s}, t) + Z(\mathbf{s}, t) + \varepsilon(\mathbf{s}, t)$$

- Covariates indexed in space and/or time, incorporated in the mean component

$$\mu(\mathbf{s}, t) = \sum_{i=1}^{p} \beta_i f_i(\mathbf{s}, t)$$

such that $E[Y(\mathbf{s}, t)] = \mu(\mathbf{s}, t)$

New model proposal

- Our proposal

$$Y(\mathbf{s},t) = \mu(\mathbf{s},t) + Z(\mathbf{s},t) + \varepsilon(\mathbf{s},t)$$

- Covariates indexed in space and/or time, incorporated in the mean component

$$\mu(\mathbf{s},t) = \sum_{i=1}^{p} \beta_i f_i(\mathbf{s},t)$$

such that $E[Y(\mathbf{s},t)] = \mu(\mathbf{s},t)$

- The non-observed spatio-temporal process $Z(\mathbf{s},t)$ defined by

$$\boldsymbol{Z} \sim MVN(\mathbf{0}, \Sigma)$$

New model proposal

- Our proposal

$$Y(\mathbf{s},t) = \mu(\mathbf{s},t) + Z(\mathbf{s},t) + \varepsilon(\mathbf{s},t)$$

- Covariates indexed in space and/or time, incorporated in the mean component

$$\mu(\mathbf{s},t) = \sum_{i=1}^{p} \beta_i f_i(\mathbf{s},t)$$

such that $E[Y(\mathbf{s},t)] = \mu(\mathbf{s},t)$

- The non-observed spatio-temporal process $Z(\mathbf{s},t)$ defined by

$$\boldsymbol{Z} \sim MVN\left(\mathbf{0}, \Sigma\right)$$

- Gaussian space-time measurement errors,

$$\boldsymbol{\varepsilon} \sim MVN\left(\mathbf{0}, \tau^2 I_{NT}\right)$$

- Assuming a separable covariance structure,

$$
\begin{aligned}
\Sigma^{i,j,k,l} &= \mathrm{Cov}_{ST}\left[Z(\mathbf{s}_i, t_k), Z(\mathbf{s}_j, t_l)\right] \\
&= \mathrm{Cov}_S\left(\|\mathbf{s}_i - \mathbf{s}_j\|\right) \times \mathrm{Cov}_T\left(|t_k - t_l|\right) \\
&= \mathrm{Cov}_S\left(h_S\right) \times \mathrm{Cov}_T\left(h_T\right)
\end{aligned}
$$

- Assuming a separable covariance structure,

$$\begin{aligned}\Sigma^{i,j,k,l} &= \text{Cov}_{ST}\left[Z(\mathbf{s}_i, t_k), Z(\mathbf{s}_j, t_l)\right] \\ &= \text{Cov}_S\left(\|\mathbf{s}_i - \mathbf{s}_j\|\right) \times \text{Cov}_T\left(|t_k - t_l|\right) \\ &= \text{Cov}_S\left(h_S\right) \times \text{Cov}_T\left(h_T\right)\end{aligned}$$

- Two interpretations for Σ,
  - $\Sigma(h_{\boldsymbol{S}}, h_{\boldsymbol{T}}) = \sigma^{\mathbf{2}}_{\boldsymbol{total}} R_{\boldsymbol{S}}(h_{\boldsymbol{S}}) \otimes R_{\boldsymbol{T}}(h_{\boldsymbol{T}})$ (Rodriguez-Iturbe & Mejia (1974))

- Assuming a separable covariance structure,

$$\begin{aligned}\Sigma^{i,j,k,l} &= \mathrm{Cov}_{ST}\left[Z(\mathbf{s}_i, t_k), Z(\mathbf{s}_j, t_l)\right] \\ &= \mathrm{Cov}_S\left(\|\mathbf{s}_i - \mathbf{s}_j\|\right) \times \mathrm{Cov}_T\left(|t_k - t_l|\right) \\ &= \mathrm{Cov}_S\left(h_S\right) \times \mathrm{Cov}_T\left(h_T\right)\end{aligned}$$

- Two interpretations for Σ,
  - $\Sigma(h_{\boldsymbol{S}}, h_{\boldsymbol{T}}) = \sigma^{\boldsymbol{2}}_{\boldsymbol{total}} R_{\boldsymbol{S}}(h_{\boldsymbol{S}}) \otimes R_{\boldsymbol{T}}(h_{\boldsymbol{T}})$ (Rodriguez-Iturbe & Mejia (1974))
  - $\Sigma(h_{\boldsymbol{S}}, h_{\boldsymbol{T}}) = \sigma^{\boldsymbol{2}}_{\boldsymbol{S}} R_{\boldsymbol{S}}(h_{\boldsymbol{S}}) \otimes \sigma^{\boldsymbol{2}}_{\boldsymbol{T}} R_{\boldsymbol{T}}(h_{\boldsymbol{T}})$

- Assuming a separable covariance structure,

$$\begin{aligned}\Sigma^{i,j,k,l} &= \mathrm{Cov}_{ST}\left[Z(\mathbf{s}_i, t_k), Z(\mathbf{s}_j, t_l)\right] \\ &= \mathrm{Cov}_S\left(\|\mathbf{s}_i - \mathbf{s}_j\|\right) \times \mathrm{Cov}_T\left(|t_k - t_l|\right) \\ &= \mathrm{Cov}_S\left(h_S\right) \times \mathrm{Cov}_T\left(h_T\right)\end{aligned}$$

- Two interpretations for Σ,
  - $\Sigma(h_{\mathbf{S}}, h_{\mathbf{T}}) = \sigma^2_{total}\, R_{\mathbf{S}}(h_{\mathbf{S}}) \otimes R_{\mathbf{T}}(h_{\mathbf{T}})$ (Rodriguez-Iturbe & Mejia (1974))
  - $\Sigma(h_{\mathbf{S}}, h_{\mathbf{T}}) = \sigma^2_{\mathbf{S}} R_{\mathbf{S}}(h_{\mathbf{S}}) \otimes \sigma^2_{\mathbf{T}} R_{\mathbf{T}}(h_{\mathbf{T}})$

- Number of temporal correlation parameters equals the number of temporal observations: saturated temporal correlation model.

- Covariates included in the model:
  - indexed in space: sampling locations intensity

- Covariates included in the model:
  - indexed in space: sampling locations intensity
  - indexed in time: specific contribution of a given survey

Predicted Mn concentration for the 2002 survey and interpolation error values

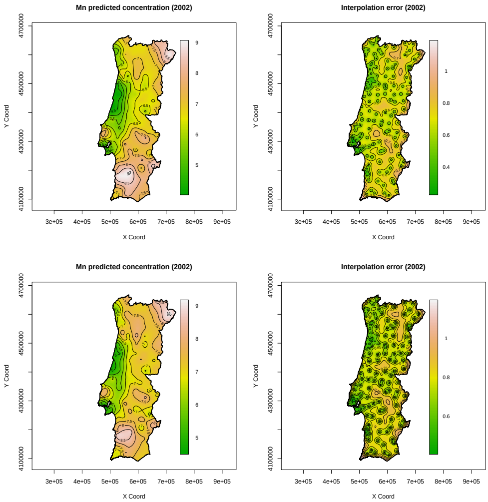
	Covariance as a function of σ^2_{total}		Covariance as a function of σ^2_S and σ^2_T	
	Predicted	Error	Predicted	Error
Min	4.05	0.23	4.53	0.41
Median	7.10	0.72	7.21	0.79
Max	9.05	1.19	9.16	1.19
Mean	6.92	0.78	7.09	0.85
St. dev.	0.92	0.23	0.74	0.19

- Better results in terms of interpolation error comparing with the previous model (Margalho *et al*, 2014)

Predicted Mn concentration for the 2002 survey and interpolation error values

	Covariance as a function of σ^2_{total}		Covariance as a function of σ^2_S and σ^2_T	
	Predicted	Error	Predicted	Error
Min	4.05	0.23	4.53	0.41
Median	7.10	0.72	7.21	0.79
Max	9.05	1.19	9.16	1.19
Mean	6.92	0.78	7.09	0.85
St. dev.	0.92	0.23	0.74	0.19

- Better results in terms of interpolation error comparing with the previous model (Margalho *et al*, 2014)
- Cross-validation study presents equal values of Mean Absolute Prediction Error for either interpretation of Σ



Prediction map for the 2002 survey (left) and the associated interpolation error map (right) for Mn transformed data, considering σ^2_{total} (top row), or σ^2_S and σ^2_T (bottom row)

R Code in the simulation study

- Finding locations

```
gera.locs<-function(num1,num2){
x1<-runif(num1,0,10)
y1<-runif(num1,0,10)
x2<-runif(num2,4.5,5.5)
y2<-runif(num2,4.5,5.5)
x<-c(x1,x2)
y<-c(y1,y2)
locs<-cbind(x,y)
return(locs)}
```

R Code in the simulation study

- Finding locations

```
gera.locs<-function(num1,num2){
x1<-runif(num1,0,10)
y1<-runif(num1,0,10)
x2<-runif(num2,4.5,5.5)
y2<-runif(num2,4.5,5.5)
x<-c(x1,x2)
y<-c(y1,y2)
locs<-cbind(x,y)
return(locs)}
```

- Design matrix

```
design.matrix<-function(intens){
c1<-rep(1,Num.Loc*Num.Temp)
c2<-rep(intens,Num.Temp)
c3<-c(rep(0,Num.Loc),rep(1,Num.Loc),rep(0,Num.Loc))
c4<-c(rep(0,Num.Loc),rep(0,Num.Loc),rep(1,Num.Loc))
M<-cbind(c1,c2,c3,c4)
return(M)}
```

- Spatial correlation matrix

```
mx.corr.spat<-function(locs,phi){
a<-as.matrix(dist(locs))
expon<-exp(-(1/exp(phi))*a)
return(expon) }
```

- Spatial correlation matrix

```
mx.corr.spat<-function(locs,phi){
a<-as.matrix(dist(locs))
expon<-exp(-(1/exp(phi))*a)
return(expon) }
```

- Temporal correlation matrix

```
mx.corr.temp<-function(r12,r13,r23){
qsi<-matrix(c(1,r12,r13,r12,1,r23,r13,r23,1),ncol=3,byrow=T)
return(qsi) }
```

- Covariance matrix

```
matrix.cov<-function(sigma2s,sigma2t,tau2,spat,temp){
mx.sigma<-kronecker(spat,temp)
mx.cov<-exp(sigma2s+sigma2t)*mxsigma+exp(tau2)*diag(Num.Loc*Num.Temp)
return(mx.cov) }
```

- Covariance matrix

```
matrix.cov<-function(sigma2s,sigma2t,tau2,spat,temp){
mx.sigma<-kronecker(spat,temp)
mx.cov<-exp(sigma2s+sigma2t)*mxsigma+exp(tau2)*diag(Num.Loc*Num.Temp)
return(mx.cov) }
```

- Log-likelihood

```
log.lik<-function(ro12,ro13,ro23,sigma2s,sigma2t,tau2,phi,ygeodataST){
ydata<-ygeodataST$data
M<-design.matrix(int50-mean(int50))
corr.spat<-mx.corr.spat(locs.true,phi)
corr.temp<-mx.corr.temp(ro12,ro13,ro23)
V<-matrix.cov(sigma2s,sigma2t,tau2,corr.spat,corr.temp)
beta.hat<-solve(t(M)%*%solve(V)%*%M)%*%t(M)%*%solve(V)%*%ydata
LL<-((-Num.Loc*Num.Temp)/2)*log(2*pi)-(1/2)*log(det(V))-(1/2)*((t(ydata-
M%*%beta.hat))%*%solve(V)%*%(ydata-M%*%beta.hat))
return(LL)}
```

- Function to optimize

```
f.optim<-function(param){
load(file="ygeodataST.RData")
ro12<-param[1]
ro13<-param[2]
ro23<-param[3]
sigma2s<-param[4]
sigma2t<-param[5]
tau2<-param[6]
phi<-param[7]
valor<-log.lik(ro12,ro13,ro23,sigma2s,sigma2t,tau2,phi,ygeodataST)
return(-valor)}
```

- Function to optimize

```
f.optim<-function(param){
load(file="ygeodataST.RData")
ro12<-param[1]
ro13<-param[2]
ro23<-param[3]
sigma2s<-param[4]
sigma2t<-param[5]
tau2<-param[6]
phi<-param[7]
valor<-log.lik(ro12,ro13,ro23,sigma2s,sigma2t,tau2,phi,ygeodataST)
return(-valor)}
```

- Optimization

```
my.likfit<-function(ygeodataST,ini.pars){
save(ygeodataST,file="ygeodataST.RData")
opt<-optim(ini.pars,foptim,method="L-BFGS-
B",lower=lower.pars,upper=upper.pars,control=list(maxit=1000))
return(list(conv=optconvergence, params = optpar)) }
```