Non-gaussian spatiotemporal modeling

Thais C O da Fonseca
Joint work with Prof Mark F J Steel

Department of Statistics
University of Warwick

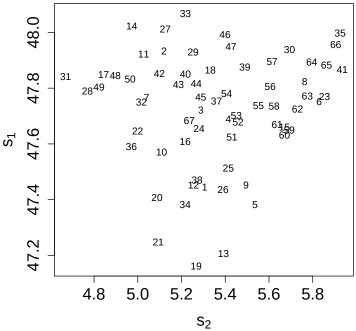
Dec, 2008

Spatiotemporal data

- Due to the proliferation of data sets that are indexed in both space and time, spatiotemporal models have received an increased attention in the literature.
- Maximum temperature data - Spanish Basque Country (67 stations)

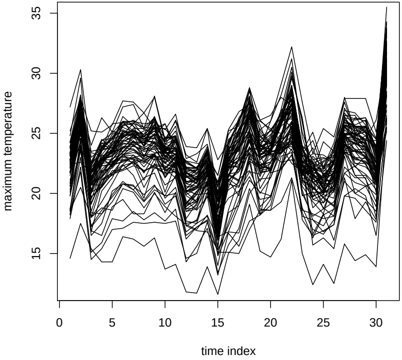
Spatiotemporal data

- Due to the proliferation of data sets that are indexed in both space and time, spatiotemporal models have received an increased attention in the literature.
- Maximum temperature data - Spanish Basque Country (67 stations)



Example

31 time points (july 2006)



Typical problem

- Given: observations $Z(s_i, t_j)$ at a finite number locations $s_i,\ i = 1, \ldots, I$ and time points $t_j,\ j = 1, \ldots, J$.

- Desired: predictive distribution for the unknown value $Z(s_0, t_0)$ at the space-time coordinate (s_0, t_0).

- Focus: continuous space and continuous time which allow for prediction and interpolation at any location and any time.

$$Z(s, t),\ (s, t) \in D \times T,\ \text{where } D \subseteq \Re^d,\ T \subseteq \Re$$

Typical problem

- Given: observations $Z(s_i, t_j)$ at a finite number locations $s_i,\ i = 1, \ldots, I$ and time points $t_j,\ j = 1, \ldots, J$.

- Desired: predictive distribution for the unknown value $Z(s_0, t_0)$ at the space-time coordinate (s_0, t_0).

- Focus: continuous space and continuous time which allow for prediction and interpolation at any location and any time.

$$Z(s,t),\ (s,t) \in D \times T,\ \text{where } D \subseteq \Re^d,\ T \subseteq \Re$$

Typical problem

- Given: observations $Z(s_i, t_j)$ at a finite number locations $s_i,\ i = 1, \ldots, I$ and time points $t_j,\ j = 1, \ldots, J$.
- Desired: predictive distribution for the unknown value $Z(s_0, t_0)$ at the space-time coordinate (s_0, t_0).
- Focus: continuous space and continuous time which allow for prediction and interpolation at any location and any time.

$$Z(s,t),\ (s,t) \in D \times T,\ \text{where } D \subseteq \Re^d,\ T \subseteq \Re$$

General modeling formulation

- The uncertainty of the unobserved parts of the process can be expressed probabilistically by a random function in space and time:

$$\{Z(s,t);(s,t)\in D\times T\}.$$

- We need to specify a valid covariance structure for the process.

$$C(s_1,s_2;t_1,t_2)=\mathrm{Cov}(Z(s_1,t_1),Z(s_2,t_2))$$

General modeling formulation

- The uncertainty of the unobserved parts of the process can be expressed probabilistically by a random function in space and time:

$$\{Z(s,t); (s,t) \in D \times T\}.$$

- We need to specify a valid covariance structure for the process.

$$C(s_1, s_2; t_1, t_2) = \text{Cov}(Z(s_1, t_1), Z(s_2, t_2))$$

Non-gaussian spatiotemporal models

- But building adequate models for these processes is not an easy task.
- One observation of the process ⇒ simplifying assumptions:
  - Stationarity: $Cov(Z(s_1, t_1), Z(s_2, t_2)) = C(s_1 - s_2, t_1 - t_2)$
  - Isotropy: $Cov(Z(s_1, t_1), Z(s_2, t_2)) = C(||s_1 - s_2||, |t_1 - t_2|)$
  - Separability: $Cov(Z(s_1, t_1), Z(s_2, t_2)) = C_s(s_1, s_2)C_t(t_1, t_2)$
  - Gaussianity: The process has finite dimensional Gaussian distribution.
- Models based on Gaussianity will not perform well (poor predictions) if
  - the data are contaminated by outliers;
  - there are regions with larger observational variance;

Non-gaussian spatiotemporal models

- But building adequate models for these processes is not an easy task.
- One observation of the process $\Rightarrow$ simplifying assumptions:
  - Stationarity: $\text{Cov}(Z(s_1, t_1), Z(s_2, t_2)) = C(s_1 - s_2, t_1 - t_2)$
  - Isotropy: $\text{Cov}(Z(s_1, t_1), Z(s_2, t_2)) = C(||s_1 - s_2||, |t_1 - t_2|)$
  - Separability: $\text{Cov}(Z(s_1, t_1), Z(s_2, t_2)) = C_s(s_1, s_2)C_t(t_1, t_2)$
  - Gaussianity: The process has finite dimensional Gaussian distribution.
- Models based on Gaussianity will not perform well (poor predictions) if
  - the data are contaminated by outliers;
  - there are regions with larger observational variance;

Non-gaussian spatiotemporal models

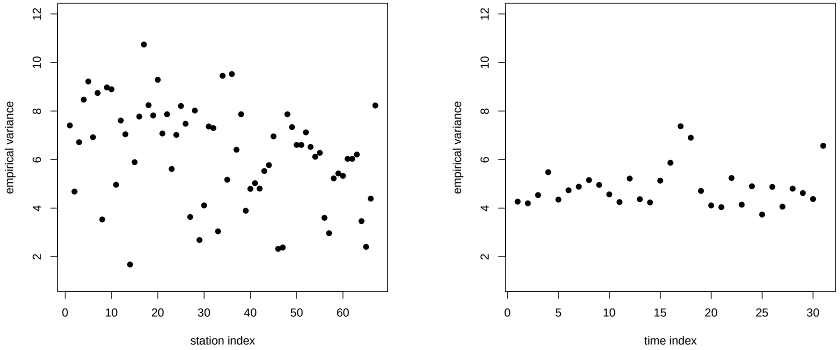
- But building adequate models for these processes is not an easy task.
- One observation of the process $\Rightarrow$ simplifying assumptions:
  - Stationarity: $\text{Cov}(Z(s_1,t_1), Z(s_2,t_2)) = C(s_1 - s_2, t_1 - t_2)$
  - Isotropy: $\text{Cov}(Z(s_1,t_1), Z(s_2,t_2)) = C(||s_1 - s_2||, |t_1 - t_2|)$
  - Separability: $\text{Cov}(Z(s_1,t_1), Z(s_2,t_2)) = C_s(s_1, s_2)C_t(t_1, t_2)$
  - Gaussianity: The process has finite dimensional Gaussian distribution.
- Models based on Gaussianity will not perform well (poor predictions) if
  - the data are contaminated by outliers;
  - there are regions with larger observational variance;

Non-gaussian spatiotemporal models

- But building adequate models for these processes is not an easy task.
- One observation of the process $\Rightarrow$ simplifying assumptions:
  - Stationarity: $\text{Cov}(Z(s_1,t_1),Z(s_2,t_2)) = C(s_1 - s_2, t_1 - t_2)$
  - Isotropy: $\text{Cov}(Z(s_1,t_1),Z(s_2,t_2)) = C(||s_1 - s_2||, |t_1 - t_2|)$
  - Separability: $\text{Cov}(Z(s_1,t_1),Z(s_2,t_2)) = C_s(s_1,s_2)C_t(t_1,t_2)$
  - Gaussianity: The process has finite dimensional Gaussian distribution.
- Models based on Gaussianity will not perform well (poor predictions) if
  - the data are contaminated by outliers;
  - there are regions with larger observational variance;

Example

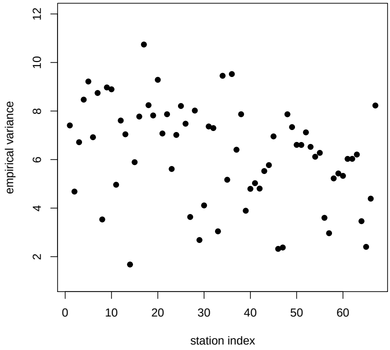
- Maximum temperature data - Spanish Basque Country

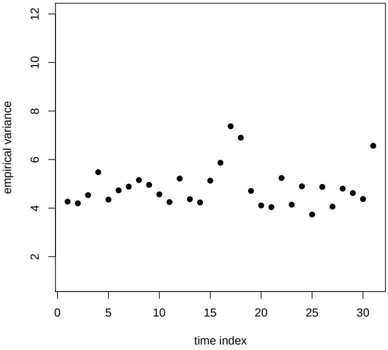


- For this reason, we consider spatiotemporal processes with heavy tailed finite dimensional distributions.

Example

- Maximum temperature data - Spanish Basque Country





- For this reason, we consider spatiotemporal processes with heavy tailed finite dimensional distributions.

We will consider processes that are

- stationary
- isotropic
- nonseparable
- non-Gaussian

Continuous mixture

- Idea: Continuous mixture of separable covariance functions [Ma, 2002].
- It takes advantage of the well known theory developed for purely spatial and purely temporal processes.

Nonseparable model

$$Z(s,t) = Z_1(s;U)Z_2(t;V) \quad (1)$$

(U,V) is a bivariate random vector with correlation c.

Unconditional covariance

$$C(s,t) = \int C_1(s;u)C_2(t;v)dF(u,v) \quad (2)$$

$C(s,t)$ is valid and nonseparable.

Continuous mixture

- Idea: Continuous mixture of separable covariance functions [Ma, 2002].
- It takes advantage of the well known theory developed for purely spatial and purely temporal processes.

Nonseparable model

$$Z(s,t) = Z_1(s;U)Z_2(t;V) \tag{1}$$

(U,V) is a bivariate random vector with correlation c.

Unconditional covariance

$$C(s,t) = \int C_1(s;u)C_2(t;v)dF(u,v) \tag{2}$$

$C(s,t)$ is valid and nonseparable.

Continuous mixture

- Idea: Continuous mixture of separable covariance functions [Ma, 2002].
- It takes advantage of the well known theory developed for purely spatial and purely temporal processes.

Nonseparable model

$$Z(s,t) = Z_1(s;U)Z_2(t;V) \tag{1}$$

(U,V) is a bivariate random vector with correlation c.

Unconditional covariance

$$C(s,t) = \int C_1(s;u)C_2(t;v)dF(u,v) \tag{2}$$

$C(s,t)$ is valid and nonseparable.

In particular, if $C_1(s;u) = \sigma_1 \exp\{-\gamma_1(s)u\}$ and $C_2(t;v) = \sigma_2 \exp\{-\gamma_2(t)v\}$ and $U = X_0 + X_1$ and $V = X_0 + X_2$, where X_i has finite moment generating function M_i, then

Proposition

$$C(s,t) = \sigma^2 M_0(-(\gamma_1(s) + \gamma_2(t)))\, M_1(-\gamma_1(s))\, M_2(-\gamma_2(t)), \;\; (s,t) \in D \times T, \quad (3)$$

where $\gamma_1(s)$ and $\gamma_2(t)$ are spatial and temporal variograms.

For instance, $\gamma_1(s) = ||s/a||^{\alpha}$ and $\gamma_2(t) = |t/b|^{\beta}$.

Notice that $c = corr(U,V)$ measures separability and $c \in [0,1]$.

See [Fonseca and Steel, 2008] for more details.

In particular, if $C_1(s;u) = \sigma_1 \exp\{-\gamma_1(s)u\}$ and $C_2(t;v) = \sigma_2 \exp\{-\gamma_2(t)v\}$ and $U = X_0 + X_1$ and $V = X_0 + X_2$, where X_i has finite moment generating function M_i, then

Proposition

$$C(s,t) = \sigma^2 M_0(-(\gamma_1(s) + \gamma_2(t)))\, M_1(-\gamma_1(s))\, M_2(-\gamma_2(t)),\ \ (s,t) \in D \times T, \qquad (3)$$

where $\gamma_1(s)$ and $\gamma_2(t)$ are spatial and temporal variograms.

For instance, $\gamma_1(s) = ||s/a||^\alpha$ and $\gamma_2(t) = |t/b|^\beta$.

Notice that $c = corr(U,V)$ measures separability and $c \in [0,1]$.

See [Fonseca and Steel, 2008] for more details.

In particular, if $C_1(s;u) = \sigma_1 \exp\{-\gamma_1(s)u\}$ and $C_2(t;v) = \sigma_2 \exp\{-\gamma_2(t)v\}$ and $U = X_0 + X_1$ and $V = X_0 + X_2$, where X_i has finite moment generating function M_i, then

Proposition

$$C(s,t) = \sigma^2 M_0(-(\gamma_1(s) + \gamma_2(t)))\, M_1(-\gamma_1(s))\, M_2(-\gamma_2(t)), \;\; (s,t) \in D \times T, \qquad (3)$$

where $\gamma_1(s)$ and $\gamma_2(t)$ are spatial and temporal variograms.

For instance, $\gamma_1(s) = ||s/a||^\alpha$ and $\gamma_2(t) = |t/b|^\beta$.

Notice that $c = corr(U,V)$ measures separability and $c \in [0,1]$.

See [Fonseca and Steel, 2008] for more details.

In particular, if $C_1(s;u) = \sigma_1 \exp\{-\gamma_1(s)u\}$ and $C_2(t;v) = \sigma_2 \exp\{-\gamma_2(t)v\}$ and $U = X_0 + X_1$ and $V = X_0 + X_2$, where X_i has finite moment generating function M_i, then

Proposition

$$C(s,t) = \sigma^2 M_0(-(\gamma_1(s) + \gamma_2(t)))\, M_1(-\gamma_1(s))\, M_2(-\gamma_2(t)), \;\; (s,t) \in D \times T, \qquad (3)$$

where $\gamma_1(s)$ and $\gamma_2(t)$ are spatial and temporal variograms.

For instance, $\gamma_1(s) = ||s/a||^\alpha$ and $\gamma_2(t) = |t/b|^\beta$.

Notice that $c = corr(U,V)$ measures separability and $c \in [0,1]$.

See [Fonseca and Steel, 2008] for more details.

Mixing in space and time

We consider the process

$$\tilde{Z}(s,t) = \tilde{Z}_1(s;U)\tilde{Z}_2(t;V), \tag{4}$$

Mixing in space

$$\tilde{Z}_1(s;U) = \sqrt{1-\tau^2}\frac{Z_1(s;U)}{\sqrt{\lambda_1(s)}} + \tau\frac{\epsilon(s)}{\sqrt{h(s)}} \tag{5}$$

Mixing in time

$$\tilde{Z}_2(t;V) = \frac{Z_2(t;V)}{\sqrt{\lambda_2(t)}} \tag{6}$$

Mixing in space and time

We consider the process

$$\tilde{Z}(s,t) = \tilde{Z}_1(s;U)\tilde{Z}_2(t;V), \quad (4)$$

Mixing in space

$$\tilde{Z}_1(s;U) = \sqrt{1-\tau^2}\frac{Z_1(s;U)}{\sqrt{\lambda_1(s)}} + \tau\frac{\epsilon(s)}{\sqrt{h(s)}} \quad (5)$$

Mixing in time

$$\tilde{Z}_2(t;V) = \frac{Z_2(t;V)}{\sqrt{\lambda_2(t)}} \quad (6)$$

Mixing in space and time

We consider the process

$$\tilde{Z}(s,t) = \tilde{Z}_1(s;U)\tilde{Z}_2(t;V), \tag{4}$$

Mixing in space

$$\tilde{Z}_1(s;U) = \sqrt{1-\tau^2}\frac{Z_1(s;U)}{\sqrt{\lambda_1(s)}} + \tau\frac{\epsilon(s)}{\sqrt{h(s)}} \tag{5}$$

Mixing in time

$$\tilde{Z}_2(t;V) = \frac{Z_2(t;V)}{\sqrt{\lambda_2(t)}} \tag{6}$$

Process $\lambda_1(s)$

Mixing in space

$$\tilde{Z}_1(s;U) = \sqrt{1-\tau^2}\frac{Z_1(s;U)}{\sqrt{\lambda_1(s)}} + \tau\frac{\epsilon(s)}{\sqrt{h(s)}}$$

- $\lambda_1(s)$ accounts for regions in space with larger observational variance.
- $\lambda_1(s)$ needs to be correlated to induce m.s. continuity of $\tilde{Z}_1(s;U)$, this is equivalent to $E[\lambda_1^{-1/2}(s_i)\lambda_1^{-1/2}(s_{i'})] \rightarrow E[\lambda_1^{-1}(s_i)]$ as $s_i \rightarrow s_{i'}$.
- This is satisfied by $\lambda_1(s) = \lambda, \forall s \Rightarrow$ student-t process. But is does not account for regions with larger variance.
- This is also satisfied by the glg process where $\{ln(\lambda_1(s)); s \in D\}$ is a gaussian process with mean $-\frac{\nu}{2}$ and covariance structure $\nu C_1(.)$. [Palacios and Steel, 2006]

Process $\lambda_1(s)$

Mixing in space

$$\tilde{Z}_1(s;U) = \sqrt{1-\tau^2}\frac{Z_1(s;U)}{\sqrt{\lambda_1(s)}} + \tau\frac{\epsilon(s)}{\sqrt{h(s)}}$$

- $\lambda_1(s)$ accounts for regions in space with larger observational variance.
- $\lambda_1(s)$ needs to be correlated to induce m.s. continuity of $\tilde{Z}_1(s;U)$, this is equivalent to $E[\lambda_1^{-1/2}(s_i)\lambda_1^{-1/2}(s_{i'})] \rightarrow E[\lambda_1^{-1}(s_i)]$ as $s_i \rightarrow s_{i'}$.
- This is satisfied by $\lambda_1(s) = \lambda, \forall s \Rightarrow$ student-t process. But is does not account for regions with larger variance.
- This is also satisfied by the glg process where $\{ln(\lambda_1(s)); s \in D\}$ is a gaussian process with mean $-\frac{\nu}{2}$ and covariance structure $\nu C_1(.)$. [Palacios and Steel, 2006]

Process $\lambda_1(s)$

Mixing in space

$$\tilde{Z}_1(s;U) = \sqrt{1-\tau^2}\frac{Z_1(s;U)}{\sqrt{\lambda_1(s)}} + \tau\frac{\epsilon(s)}{\sqrt{h(s)}}$$

- $\lambda_1(s)$ accounts for regions in space with larger observational variance.
- $\lambda_1(s)$ needs to be correlated to induce m.s. continuity of $\tilde{Z}_1(s;U)$, this is equivalent to $E[\lambda_1^{-1/2}(s_i)\lambda_1^{-1/2}(s_{i'})] \longrightarrow E[\lambda_1^{-1}(s_i)]$ as $s_i \longrightarrow s_{i'}$.
- This is satisfied by $\lambda_1(s) = \lambda, \forall s \Rightarrow$ student-t process. But is does not account for regions with larger variance.
- This is also satisfied by the glg process where $\{ln(\lambda_1(s)); s \in D\}$ is a gaussian process with mean $-\frac{\nu}{2}$ and covariance structure $\nu C_1(.)$. [Palacios and Steel, 2006]

Process $\lambda_1(s)$

Mixing in space

$$\tilde{Z}_1(s;U) = \sqrt{1-\tau^2}\frac{Z_1(s;U)}{\sqrt{\lambda_1(s)}} + \tau\frac{\epsilon(s)}{\sqrt{h(s)}}$$

- $\lambda_1(s)$ accounts for regions in space with larger observational variance.
- $\lambda_1(s)$ needs to be correlated to induce m.s. continuity of $\tilde{Z}_1(s;U)$, this is equivalent to $E[\lambda_1^{-1/2}(s_i)\lambda_1^{-1/2}(s_{i'})] \to E[\lambda_1^{-1}(s_i)]$ as $s_i \to s_{i'}$.
- This is satisfied by $\lambda_1(s) = \lambda, \forall s \Rightarrow$ student-t process. But is does not account for regions with larger variance.
- This is also satisfied by the glg process where $\{ln(\lambda_1(s)); s \in D\}$ is a gaussian process with mean $-\frac{\nu}{2}$ and covariance structure $\nu C_1(.)$. [Palacios and Steel, 2006]

Process $\lambda_1(s)$

Mixing in space

$$\tilde{Z}_1(s;U) = \sqrt{1-\tau^2}\frac{Z_1(s;U)}{\sqrt{\lambda_1(s)}} + \tau\frac{\epsilon(s)}{\sqrt{h(s)}}$$

- $\lambda_1(s)$ accounts for regions in space with larger observational variance.
- $\lambda_1(s)$ needs to be correlated to induce m.s. continuity of $\tilde{Z}_1(s;U)$, this is equivalent to $E[\lambda_1^{-1/2}(s_i)\lambda_1^{-1/2}(s_{i'})] \rightarrow E[\lambda_1^{-1}(s_i)]$ as $s_i \rightarrow s_{i'}$.
- This is satisfied by $\lambda_1(s) = \lambda, \forall s \Rightarrow$ student-t process. But is does not account for regions with larger variance.
- This is also satisfied by the glg process where $\{ln(\lambda_1(s)); s \in D\}$ is a gaussian process with mean $-\frac{\nu}{2}$ and covariance structure $\nu C_1(.)$. [Palacios and Steel, 2006]

Process $h(s)$

Mixing in space

$$\tilde{Z}_1(s;U) = \sqrt{1-\tau^2}\frac{Z_1(s;U)}{\sqrt{\lambda_1(s)}} + \tau\frac{\epsilon(s)}{\sqrt{h(s)}}$$

- $h(s)$ accounts for traditional outliers (different nugget effects).
- We consider the detection of outliers jointly in the estimation procedure and the variable $h_i = h(s_i), i = 1, \ldots, I$ are considered latent variables
- Their posterior distribution indicate outlying observations (h_i close to 0).
- We consider
  - $log(h_i) \sim N(-\nu_h/2, \nu_h)$.
  - $h_i \sim Ga(1/\nu_h, 1/\nu_h)$.

Process $h(s)$

Mixing in space

$$\tilde{Z}_1(s;U) = \sqrt{1-\tau^2}\frac{Z_1(s;U)}{\sqrt{\lambda_1(s)}} + \tau\frac{\epsilon(s)}{\sqrt{h(s)}}$$

- $h(s)$ accounts for traditional outliers (different nugget effects).
- We consider the detection of outliers jointly in the estimation procedure and the variable $h_i = h(s_i), i = 1, \ldots, I$ are considered latent variables
- Their posterior distribution indicate outlying observations (h_i close to 0).
- We consider
  - $log(h_i) \sim N(-\nu_h/2, \nu_h)$.
  - $h_i \sim Ga(1/\nu_h, 1/\nu_h)$.

Process $h(s)$

Mixing in space

$$\tilde{Z}_1(s;U) = \sqrt{1-\tau^2}\frac{Z_1(s;U)}{\sqrt{\lambda_1(s)}} + \tau\frac{\epsilon(s)}{\sqrt{h(s)}}$$

- $h(s)$ accounts for traditional outliers (different nugget effects).
- We consider the detection of outliers jointly in the estimation procedure and the variable $h_i = h(s_i), i = 1, \ldots, I$ are considered latent variables
- Their posterior distribution indicate outlying observations (h_i close to 0).
- We consider
  - $log(h_i) \sim \mathrm{N}(-\nu_h/2, \nu_h)$.
  - $h_i \sim \mathrm{Ga}(1/\nu_h, 1/\nu_h)$.

Process $h(s)$

Mixing in space

$$\tilde{Z}_1(s;U) = \sqrt{1-\tau^2}\frac{Z_1(s;U)}{\sqrt{\lambda_1(s)}} + \tau\frac{\epsilon(s)}{\sqrt{h(s)}}$$

- $h(s)$ accounts for traditional outliers (different nugget effects).
- We consider the detection of outliers jointly in the estimation procedure and the variable $h_i = h(s_i), i = 1, \ldots, I$ are considered latent variables
- Their posterior distribution indicate outlying observations (h_i close to 0).
- We consider
  - $log(h_i) \sim \mathrm{N}(-\nu_h/2, \nu_h)$.
  - $h_i \sim \mathrm{Ga}(1/\nu_h, 1/\nu_h)$.

Process $\lambda_2(t)$

Mixing in time

$$\tilde{Z}_2(t;V) = \frac{Z_2(t;V)}{\sqrt{\lambda_2(t)}}$$

- $\lambda_2(t)$ accounts for sections in time with larger observational variance.
- This can be seen as a way to adress the issue of volatility clustering, which is common in finantial time series data.
- We consider the log gaussian process where $\{ln(\lambda_2(t)); t \in T\}$ is a gaussian process with mean $-\frac{\nu_2}{2}$ and covariance structure $\nu_2 C_2(.)$.

Process $\lambda_2(t)$

Mixing in time

$$\tilde{Z}_2(t;V) = \frac{Z_2(t;V)}{\sqrt{\lambda_2(t)}}$$

- $\lambda_2(t)$ accounts for sections in time with larger observational variance.
- This can be seen as a way to adress the issue of volatility clustering, which is common in finantial time series data.
- We consider the log gaussian process where $\{ln(\lambda_2(t)); t \in T\}$ is a gaussian process with mean $-\frac{\nu_2}{2}$ and covariance structure $\nu_2 C_2(.)$.

Process $\lambda_2(t)$

Mixing in time

$$\tilde{Z}_2(t;V) = \frac{Z_2(t;V)}{\sqrt{\lambda_2(t)}}$$

- $\lambda_2(t)$ accounts for sections in time with larger observational variance.
- This can be seen as a way to adress the issue of volatility clustering, which is common in finantial time series data.
- We consider the log gaussian process where $\{ln(\lambda_2(t)); t \in T\}$ is a gaussian process with mean $-\frac{\nu_2}{2}$ and covariance structure $\nu_2 C_2(.)$.

Predictions

- $(\lambda_{1i}, h_i, \lambda_{2j})$ are considered latent variables and sampled in our MCMC sampler.
- Given $(\lambda_{1i}, h_i, \lambda_{2j})$ the process is gaussian and we can predict at unobserved locations and time points.
- We compare the predictive performance using proper scoring rules [Gneiting and Raftery, 2008]:
  - $LPS(p, x) = -log(p(x))$
  - $IS(q_1, q_2; x) = (q_2 - q_1) + \frac{2}{\xi}(q_1 - x)I(x < q_1) + \frac{2}{\xi}(x - q_2)I(x > q_2)$. We use $\xi = 0.05$ resulting in a 95% credible interval.

Predictions

- $(\lambda_{1i}, h_i, \lambda_{2j})$ are considered latent variables and sampled in our MCMC sampler.
- Given $(\lambda_{1i}, h_i, \lambda_{2j})$ the process is gaussian and we can predict at unobserved locations and time points.
- We compare the predictive performance using proper scoring rules [Gneiting and Raftery, 2008]:
  - $LPS(p, x) = -log(p(x))$
  - $IS(q_l, q_u; x) = (q_u - q_l) + \frac{2}{\xi}(q_l - x)I(x < q_l) + \frac{2}{\xi}(x - q_u)I(x > q_u)$. We use $\xi = 0.05$ resulting in a 95% credible interval.

Predictions

- $(\lambda_{1i}, h_i, \lambda_{2j})$ are considered latent variables and sampled in our MCMC sampler.
- Given $(\lambda_{1i}, h_i, \lambda_{2j})$ the process is gaussian and we can predict at unobserved locations and time points.
- We compare the predictive performance using proper scoring rules [Gneiting and Raftery, 2008]:
  - $LPS(p, x) = -log(p(x))$
  - $IS(q_1, q_2; x) = (q_2 - q_1) + \frac{2}{\xi}(q_1 - x)I(x < q_1) + \frac{2}{\xi}(x - q_2)I(x > q_2)$. We use $\xi = 0.05$ resulting in a 95% credible interval.

Predictions

- $(\lambda_{1i}, h_i, \lambda_{2j})$ are considered latent variables and sampled in our MCMC sampler.
- Given $(\lambda_{1i}, h_i, \lambda_{2j})$ the process is gaussian and we can predict at unobserved locations and time points.
- We compare the predictive performance using proper scoring rules [Gneiting and Raftery, 2008]:
  - $LPS(p,x) = -log(p(x))$
  - $IS(q_1, q_2; x) = (q_2 - q_1) + \frac{2}{\xi}(q_1 - x)I(x < q_1) + \frac{2}{\xi}(x - q_2)I(x > q_2)$. We use $\xi = 0.05$ resulting in a 95% credible interval.

Predictions

- $(\lambda_{1i}, h_i, \lambda_{2j})$ are considered latent variables and sampled in our MCMC sampler.
- Given $(\lambda_{1i}, h_i, \lambda_{2j})$ the process is gaussian and we can predict at unobserved locations and time points.
- We compare the predictive performance using proper scoring rules [Gneiting and Raftery, 2008]:
  - $LPS(p,x) = -log(p(x))$
  - $IS(q_1, q_2; x) = (q_2 - q_1) + \frac{2}{\xi}(q_1 - x)I(x < q_1) + \frac{2}{\xi}(x - q_2)I(x > q_2)$. We use $\xi = 0.05$ resulting in a 95% credible interval.

Data

- This data set has $I = 30$ locations and $J = 30$ time points generated from a Gaussian model with no nugget effect ($\tau^2 = 0$).
- The covariance model is nonseparable Cauchy ($X_i \sim \mathrm{Ga}(\lambda_i, 1)$, $i = 0, 1, 2$) in space and time with $c = 0.5$.
- We contaminated this data set with different kinds of "outliers" in order to see the performance of the proposed models in each situation.

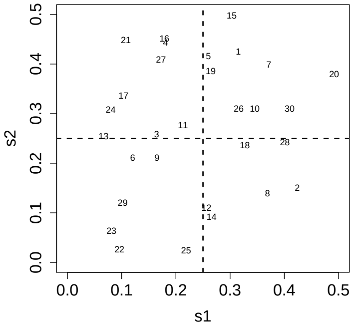
Data

- This data set has $I = 30$ locations and $J = 30$ time points generated from a Gaussian model with no nugget effect ($\tau^2 = 0$).
- The covariance model is nonseparable Cauchy ($X_i \sim \text{Ga}(\lambda_i, 1)$, $i = 0, 1, 2$) in space and time with $c = 0.5$.
- We contaminated this data set with different kinds of "outliers" in order to see the performance of the proposed models in each situation.

Data

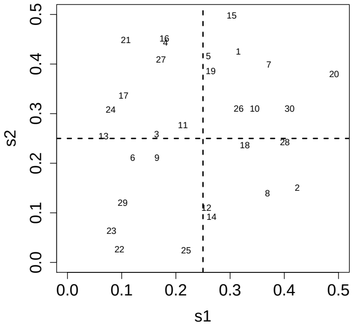
- This data set has $I = 30$ locations and $J = 30$ time points generated from a Gaussian model with no nugget effect ($\tau^2 = 0$).
- The covariance model is nonseparable Cauchy ($X_i \sim \text{Ga}(\lambda_i, 1)$, $i = 0, 1, 2$) in space and time with $c = 0.5$.
- We contaminated this data set with different kinds of "outliers" in order to see the performance of the proposed models in each situation.

Spatial domain



- The proposal for $\lambda_{1i}, h_i,\ i = 1, \ldots, I$ in the MCMC sampler is constructed by dividing the observations in blocks defined by position in the spatial domain.

Spatial domain



- The proposal for $\lambda_{1i}, h_i,\ i = 1, \ldots, I$ in the MCMC sampler is constructed by dividing the observations in blocks defined by position in the spatial domain.

Description and BF

- One location was selected at random (location 7) and a random increment from $\mathrm{Unif}(1.0, 1.5)$ times the standard deviation was added to each observation for this location for the first 20 time points.
- The logarithm of the BF using Shifted-Gamma ($\lambda = 0.98$) estimators:

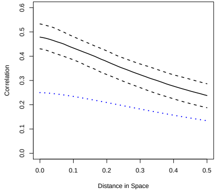
	nug.	h (lognormal)	h (gamma)	λ_1	λ_1 & h (lognormal)
Gaussian	-1	101	98	78	109

Description and BF

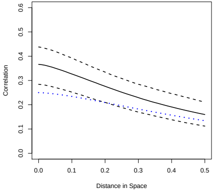
- One location was selected at random (location 7) and a random increment from Unif$(1.0, 1.5)$ times the standard deviation was added to each observation for this location for the first 20 time points.
- The logarithm of the BF using Shifted-Gamma ($\lambda = 0.98$) estimators:

	nug.	h (lognormal)	h (gamma)	λ_1	λ_1 & h (lognormal)
Gaussian	-1	101	98	78	109

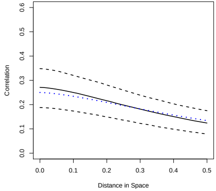
Estimated correlation function - $t_0 = 1$



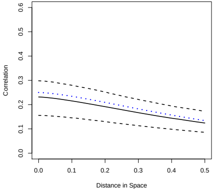
(a) Gaussian



(b) Nongaussian with λ_1

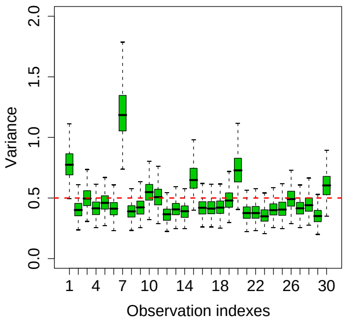


(c) Nongaussian with h and λ_1

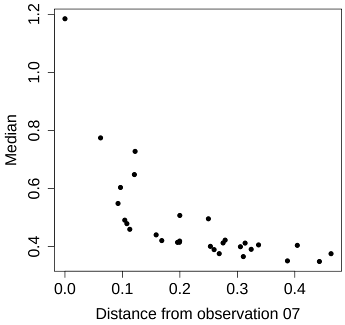


(d) Gaussian (Uncontaminated data)

Nongaussian model with λ_1

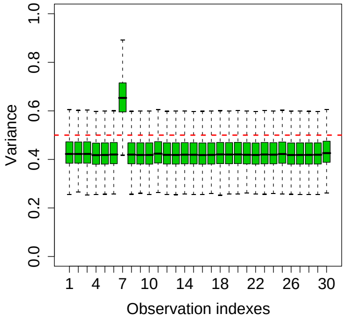


(a) Variance for each location.

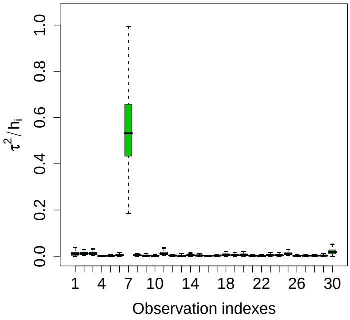


(b) Median of σ_i^2 vs. distance from obs. 7.

Nongaussian model with *h* (lognormal)

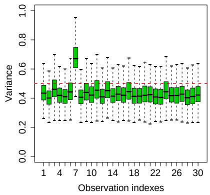


(a) Variance for each location.

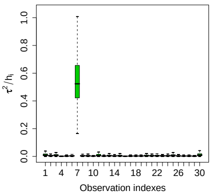


(b) Nugget for each location.

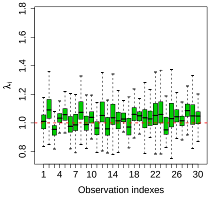
Nongaussian model with λ_1 and h



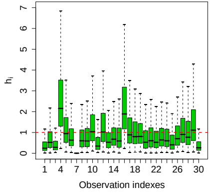
(a) Variance for each location.



(b) Nugget for each location.



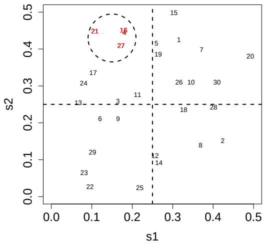
(c) $\lambda_{1i}, i = 1, \ldots, 30.$



(d) $h_i, i = 1, \ldots, 30.$

Description and BF

- A region was selected and an increment from $\mathrm{Unif}(0.5, 1.5)$ times the standard deviation was added to each observation for the first 10 time points.

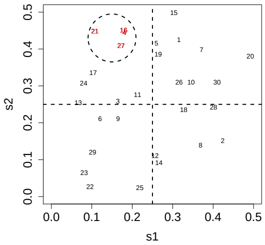


- The logarithm of the BF using Shifted-Gamma ($\lambda = 0.98$) estimators:

	nug.	h (lognormal)	h (gamma)	λ_1	λ_1 & h (lognormal)
Gaussian	44	70	72	75	110

Description and BF

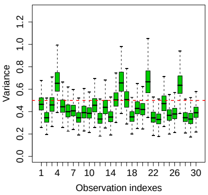
- A region was selected and an increment from Unif$(0.5, 1.5)$ times the standard deviation was added to each observation for the first 10 time points.



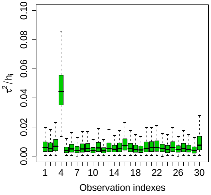
- The logarithm of the BF using Shifted-Gamma $(\lambda = 0.98)$ estimators:

	nug.	h (lognormal)	h (gamma)	λ_1	λ_1 & h (lognormal)
Gaussian	44	70	72	75	110

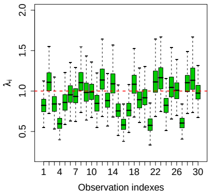
Nongaussian model with λ_1 and h



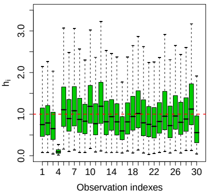
(a) Variance for each location.



(b) Nugget for each location.



(c) $\lambda_{1i}, i = 1, \ldots, 30.$



(d) $h_i, i = 1, \ldots, 30.$

Data 2* - Description and BF

- A region of the spatial domain was selected (locations 4, 16, 21 and 27) and the same random increment from $\text{Unif}(0.5, 1.5)$ times the standard deviation was added to each observation for the first 10 time points.
- The logarithm of the BF using Shifted-Gamma ($\lambda = 0.98$) estimators:

	nug.	h (lognormal)	h (gamma)	λ_1	λ_1 & h (lognormal)
Gaussian	-2	-4	-4	24	20

Data 2* - Description and BF

- A region of the spatial domain was selected (locations 4, 16, 21 and 27) and the same random increment from Unif$(0.5, 1.5)$ times the standard deviation was added to each observation for the first 10 time points.
- The logarithm of the BF using Shifted-Gamma ($\lambda = 0.98$) estimators:

	nug.	h (lognormal)	h (gamma)	λ_1	λ_1 & h (lognormal)
Gaussian	-2	-4	-4	24	20

Description and BF

- The observations at time points 11 to 15 were contaminated by adding a random increment from Unif$(0.5, 1.5)$ times the standard deviation to each observation for all spatial locations.
- The logarithm of the BF using Shifted-Gamma ($\lambda = 0.98$) estimators:

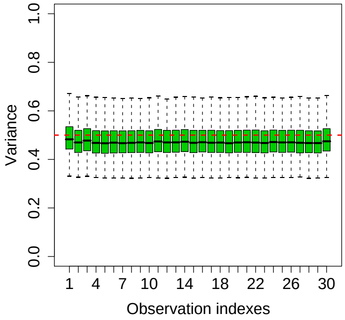
	nug.	h (lognormal)	λ_1	λ_2	$\lambda_1\&\lambda_2$	$\lambda_1\&\lambda_2\&h$
Gaussian	18	44	28	76	112	111

Description and BF

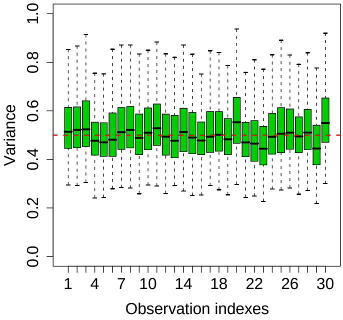
- The observations at time points 11 to 15 were contaminated by adding a random increment from Unif$(0.5, 1.5)$ times the standard deviation to each observation for all spatial locations.
- The logarithm of the BF using Shifted-Gamma ($\lambda = 0.98$) estimators:

	nug.	h (lognormal)	λ_1	λ_2	$\lambda_1 \& \lambda_2$	$\lambda_1 \& \lambda_2 \& h$
Gaussian	18	44	28	76	112	111

Nongaussian models

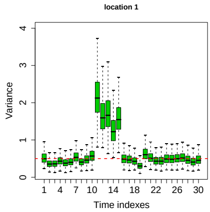


(a) Model with lognormal $h(s)$.

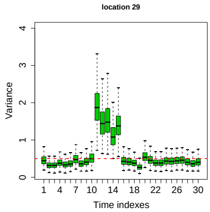


(b) Model with lognormal $h(s)$ and $\lambda_1(s)$.

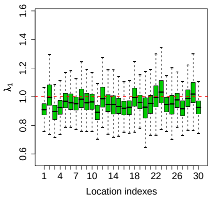
Nongaussian model with λ_1 and λ_2



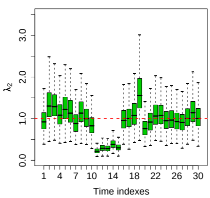
(a) Variance for each time.



(b) Variance for each time.



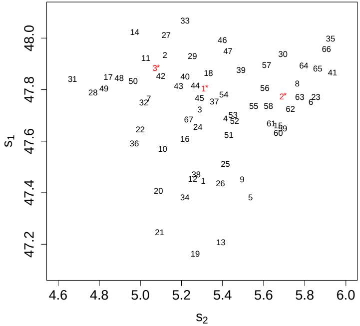
(c) $\lambda_{1i}, i = 1, \ldots, I$.



(d) $\lambda_{2j}, j = 1, \ldots, J$.

Data

(a) Spain and France Map.



(b) Basque Country (Zoom).

Model

- Mean function:

$$\mu(s,t) = \delta_0 + \delta_1 s_1 + \delta_2 s_2 + \delta_3 h + \delta_4 t + \delta_5 t^2$$

- Cauchy covariance function: $X_i \sim \text{Ga}(\lambda_i, 1)$

$$C(s,t) = \left(\frac{1}{1+||s/a||^\alpha}\right)^{\lambda_1}\left(\frac{1}{1+|t/b|^\beta}\right)^{\lambda_2}\left(\frac{1}{1+||s/a||^\alpha+|t/b|^\beta}\right)^{\lambda_0}$$

$\lambda_1 = \lambda_2 = 1$ and $c = \lambda_0/(1+\lambda_0)$.

Model

- Mean function:

$$\mu(s,t) = \delta_0 + \delta_1 s_1 + \delta_2 s_2 + \delta_3 h + \delta_4 t + \delta_5 t^2$$

- Cauchy covariance function: $X_i \sim \mathrm{Ga}(\lambda_i, 1)$

$$C(s,t) = \left(\frac{1}{1+||s/a||^\alpha}\right)^{\lambda_1}\left(\frac{1}{1+|t/b|^\beta}\right)^{\lambda_2}\left(\frac{1}{1+||s/a||^\alpha+|t/b|^\beta}\right)^{\lambda_0}$$

$\lambda_1 = \lambda_2 = 1$ and $c = \lambda_0/(1+\lambda_0)$.

Likelihood

- In order to calculate the likelihood function we need to invert a matrix with dimension 2077×2077.
- We approximate the likelihood by using conditional distributions.
- We consider a partition of Z into subvectors $Z_1, ..., Z_{31}$ where $Z_j = (Z(s_1, t_j), \ldots, Z(s_{67}, t_j))'$ and we define $Z_{(j)} = (Z_{j-L+1}, ..., Z_j)$. Then

$$p(z|\phi) \approx p(z_1|\phi) \prod_{j=2}^{31} p(z_j|z_{(j-1)}, \phi). \tag{7}$$

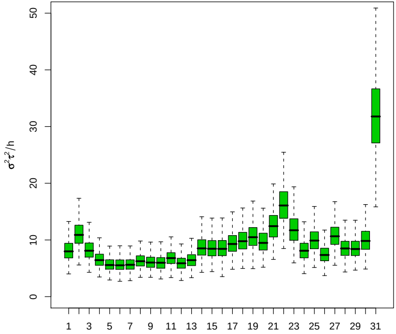
- This means the distribution of Z_j will only depend on the observations in space for the previous L time points.
- In this application we used $L = 5$ to make the MCMC feasible.

Bayes Factor

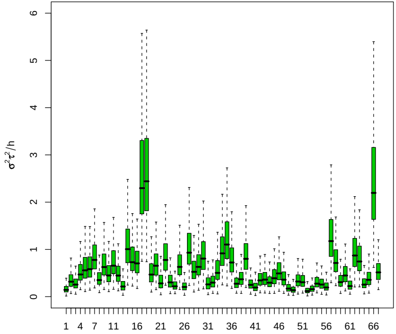
	h	λ_1	λ_1 & h	λ_2	λ_2 & h	λ_1 & λ_2	λ_1, h & λ_2
Shifted gamma	172	148	345	138	279	417	547

Table: The natural logarithm of the Bayes factor in favor of the model in the column versus Gaussian model using Shifted-Gamma ($\lambda = 0.98$) estimator for the predictive density of z.

Model with h and λ_2

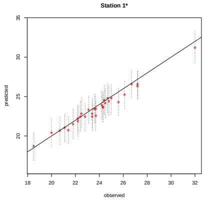


(a) $\sigma^2(1-\tau^2)/\lambda_2$.

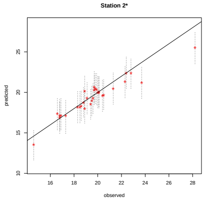


(b) $\sigma^2\tau^2/h$.

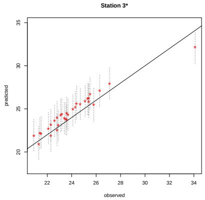
Predicted temperature at the out-of-sample stations



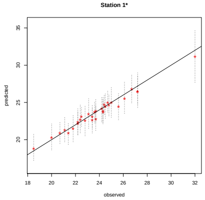
(a) Gaussian Model.



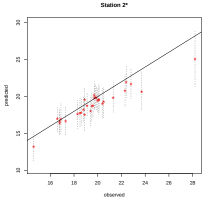
(b) Gaussian Model.



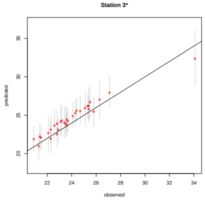
(c) Gaussian Model.



(d) Model with λ_2 & h.



(e) Model with λ_2 & h.



(f) Model with λ_2 & h.

Model comparison

model	Average width	$\bar{IS}$	LPS
Gaussian	3.78	4.35	103.81
h	3.83	4.34	102.04
λ_1	3.74	4.36	105.09
λ_1 & h	3.75	4.48	103.79
λ_2	3.73	3.94	87.33
λ_2 & h	3.73	3.87	86.57
λ_1 & λ_2	4.51	4.65	85.89
λ_1, h & λ_2	3.84	4.02	83.78

References

T Fonseca and M F J Steel
A New Class of Nonseparable Spatiotemporal Models
CRiSM Working Paper 08-13. 2008.

T Gneiting and A E Raftery
Strictly proper scoring rules, prediction and estimation
JASA. (102) 360–378, 2007.

C Ma
Spatio-temporal covariance functions generated by mixtures
Mathematical geology. (34) 965–975, 2002.

M B Palacios and M F J Steel
Non-Gaussian Bayesian Geostatistical Modeling
JASA. (101) 604–618, 2006.