NON-GAUSSIAN SPATIOTEMPORAL MODELING THROUGH SCALE MIXING

Thais C O Fonseca - Universidade Federal do Rio de Janeiro

Joint work with Mark F J Steel - Warwick University

PASI, 2014

1. INTRODUCTION
   - Motivation
2. SPATIOTEMPORAL MODELING
   - Nonseparable models
   - Heavy tailed processes
3. SIMULATION RESULTS
   - Data 1
   - Data 2
4. TEMPERATURE DATA
   - Non-gaussian spatiotemporal modeling
   - Results

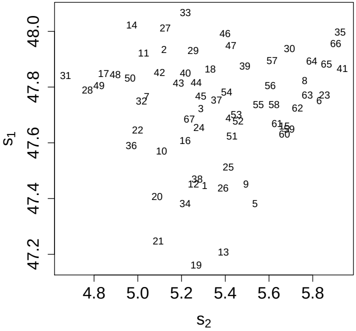
TYPICAL PROBLEM

- 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.

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

SPATIOTEMPORAL DATA

Maximum temperature data - Spanish Basque Country (67 stations)



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.
$$C(s_1,s_2;t_1,t_2)=\mathrm{Cov}(Z(s_1,t_1),Z(s_2,t_2))$$
- Positive definiteness: C has to imply that $\sum_{i=1}^{n} a_i Z(s_i,t_i)$ has positive variance for any $(s_1,t_1),\ldots,(s_n,t_n)$, any real a_1, ..., a_n, and any positive integer n.
- It is quite difficult to check whether a function is positive definite.

NON-GAUSSIAN SPATIOTEMPORAL MODELS

- Simplifying assumptions:
  - Stationarity: $\mathrm{Cov}(Z(s_1,t_1),Z(s_2,t_2)) = C(s_1 - s_2, t_1 - t_2)$
  - Isotropy: $\mathrm{Cov}(Z(s_1,t_1),Z(s_2,t_2)) = C(||s_1 - s_2||, |t_1 - t_2|)$
  - Separability: $\mathrm{Cov}(Z(s_1,t_1),Z(s_2,t_2)) = C_s(s_1,s_2)C_t(t_1,t_2)$
  - Gaussianity: The process has finite dimensional Gaussian distribution.

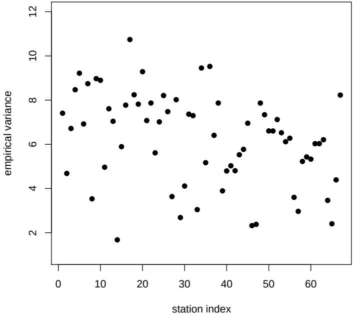
- Models based on Gaussianity will not perform well (poor predictions) if
  - the data are contaminated by outliers;
  - there are regions with larger observational variance;

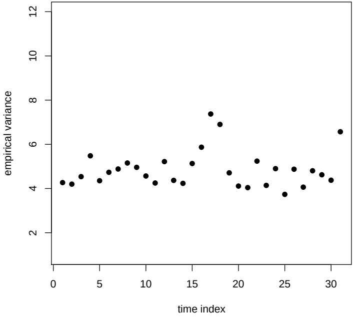
NON-GAUSSIAN SPATIOTEMPORAL MODELS

- 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





We will consider processes that are

- nonseparable in space and time;
- non-Gaussian;

CONTINUOUS MIXTURE

- Some nonseparable models: [Cressie and Huang, 1999], [Gneiting, 2002] and [Ma, 2002].

MIXTURE MODELS [MA, 2002]

$$C(s,t) = \int C_1(s;u)C_2(t;v)dF(u,v) \tag{1}$$

- Idea: convex combinations of valid separable covariance functions are valid and nonseparable functions.
- (U, V) is a bivariate nonnegative random vector with cumulative distribution function F.

MAIN ADVANTAGES

1. One may take advantage of the whole available literature of spatial statistics and time series; C is the unconditional covariance of

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

2. It is natural to make separate modeling decisions regarding the spatial and temporal components, eg. smoothness and long range dependence can be different across space and time.

In particular, if

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

Proposition

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

e.g. $\gamma_1(s) = ||s/a||^\alpha$ and $\gamma_2(t) = |t/b|^\beta$.

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

See [Fonseca and Steel, 2011] for more details.

Mixing in space and time

We consider the process

$$\tilde{Z}(s,t) = \tilde{Z}_1(s;U)\tilde{Z}_2(t;V), \tag{3}$$

Mixing in space

$$\tilde{Z}_1(s;U) = \sqrt{1-\tau^2}\frac{Z_1(s;U)}{\sqrt{\lambda_1(s)}} + \tau\frac{\epsilon(s)}{\sqrt{h(s)}} \tag{4}$$

Mixing in time

$$\tilde{Z}_2(t;V) = \frac{Z_2(t;V)}{\sqrt{\lambda_2(t)}} \tag{5}$$

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.
- If $\lambda_1(s) = \lambda, \forall s \Rightarrow$ student-t process. But is does not account for regions with larger variance.
- We consider 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
  - $log(h_i) \sim \mathrm{N}(-\nu_h/2, \nu_h)$.
  - $h_i \sim \mathrm{Ga}(1/\nu_h, 1/\nu_h)$.

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 $\nu_2 C_2(.)$.

RESULTING CORRELATION

- Conditioning on the latent variables:

$$\begin{aligned} Cov(\tilde{Z}(s_i,t_j),\tilde{Z}(s_{i'},t_{j'})) &= \sigma^2 M_0(-\gamma_1(s)-\gamma_2(t))\frac{M_2(-\gamma_2(t))}{\sqrt{\lambda_2(t_j)\lambda_2(t_j')}} \\ &\times \left[(1-\tau^2)\frac{M_1(-\gamma_1(s))}{\sqrt{\lambda_1(s_i)\lambda_1(s_{i'})}}+\tau^2\frac{I_{s_i=s_{i'}}}{\sqrt{h(s_i)h(s_{i'})}}\right] \end{aligned} \quad (6)$$

$i,i'=1,\ldots,I,\ j,j'=1,\ldots,J,\ s=s_i-s_{i'},\ t=t_j-t_{j'}.$

- Integrating latent variables out:

$$\tilde{\rho}(s,t)=\frac{\exp\left[\frac{\nu}{4}\{C_1^*(s)-1\}+\nu\right]}{\exp(\nu)+\omega^2E(h^{-1})}M_0\{-\gamma_1(s)-\gamma_2(t)\}\ M_1\{-\gamma_1(s)\}\ M_2\{-\gamma_2(t)\} \quad (7)$$

$\omega=\tau^2/(1-\tau^2)$. Throughout, we will use $C_1^*(s)=M_1\{-\gamma_1(s)\}$.

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

- Note that the distribution of λ_1^{pred} is

$$p(\lambda_1^{pred} \mid \lambda_1^{obs}, z^{obs}, \theta) = p(\lambda_1^{pred} \mid \lambda_1^{obs}, \theta)$$

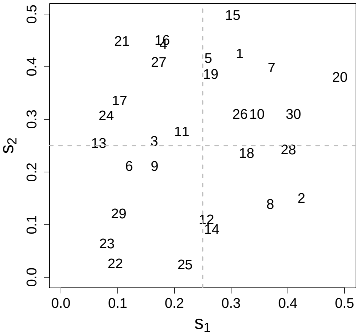
which is a log-Gaussian multivariate distribution.

- Prediction scheme:
  - We sample $log\lambda_1^{pred}$ from a Multivariate Gaussian distribution;
  - And λ_1^{obs} and θ from their posterior distributions;
  - Then z^{pred} is sampled from a Gaussian distribution given $\lambda_1^{obs}, \theta, z^{obs}$.

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.

SPATIAL DOMAIN



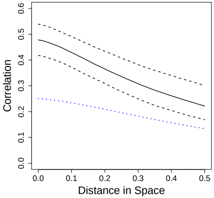
- The proposal for $\lambda_{1i}, h_i,\ i = 1, \ldots, I$ in the MCMC sampler is constructed by dividing the observations in blocks defined by position in the spatial domain.

DESCRIPTION AND BF

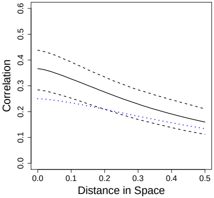
- One location was selected at random (location 7) and a random increment from $\mathrm{Unif}(1.0, 1.5)$ times the standard deviation was added to each observation for this location for the first 20 time points.
- The logarithm of the BF using Shifted-Gamma ($\lambda = 0.98$) estimators:

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

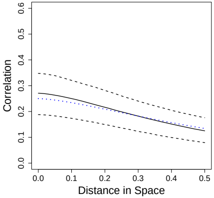
Estimated correlation function - $t_0 = 1$



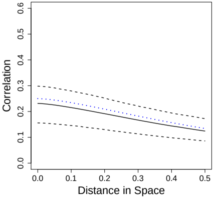
(a) Gaussian



(b) Nongaussian with λ_1

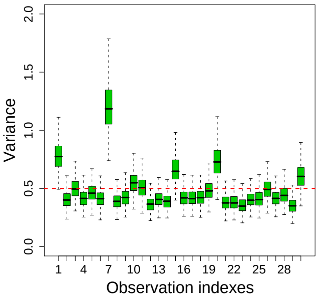


(c) Nongaussian with h and λ_1

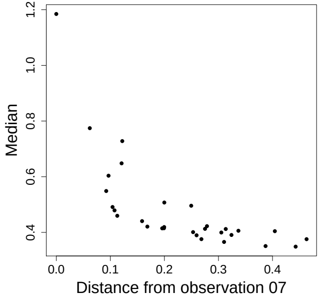


(d) Gaussian (Uncontaminated data)

Nongaussian model with λ_1

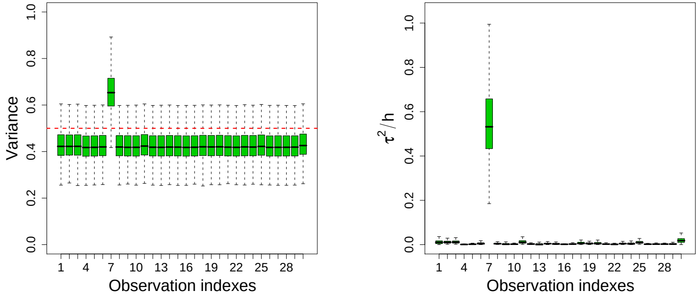


(a) Variance for each location.



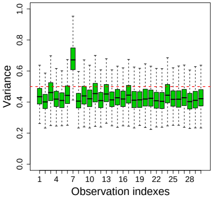
(b) Median of σ_i^2 vs. distance from obs. 7.

NONGAUSSIAN MODEL WITH h (LOGNORMAL)

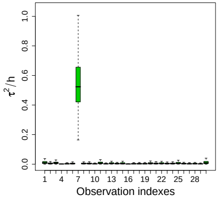


(a) Variance for each location. (b) Nugget for each location.

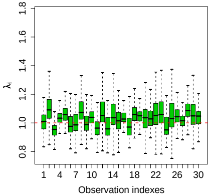
NONGAUSSIAN MODEL WITH λ_1 AND h



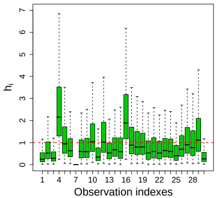
(a) Variance for each location.



(b) Nugget for each location.

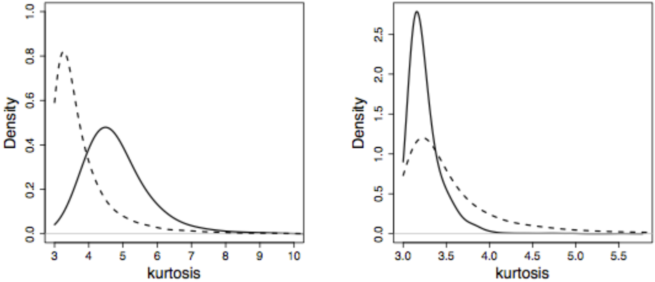


(c) $\lambda_{1i}, i = 1, \ldots, 30.$



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

Kurtosis



(a) Model with λ_1. (b) Model with λ_1 and h.

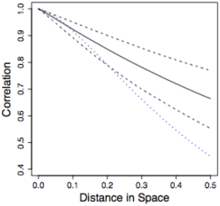
Marginal posterior distribution (full line) and prior distribution (dashed line) of the kurtosis coefficient.

Description and BF

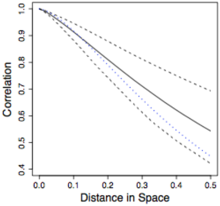
- The observations at time points 11 to 15 were contaminated by adding a random increment from $\mathrm{Unif}(0.5, 1.5)$ times the standard deviation to each observation for all spatial locations.
- The logarithm of the BF using Shifted-Gamma ($\lambda = 0.98$) estimators:

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

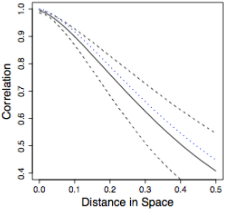
Estimated correlation function - $t_0 = 1$



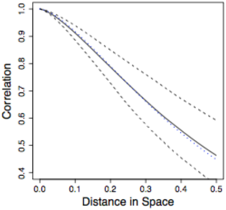
(a) Gaussian.



(b) Model with λ_1 and λ_2.

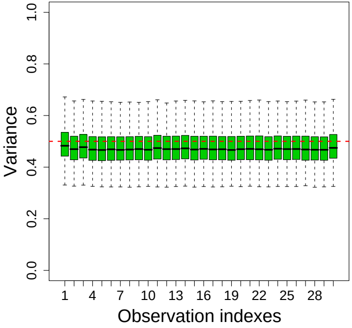


(c) Model with λ_1, λ_2, h.

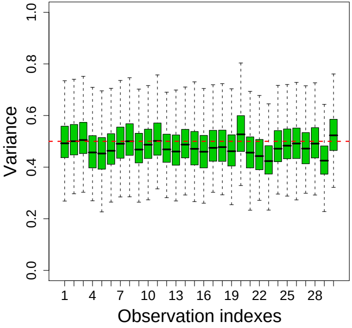


(d) Gaussian (Uncontaminated data).

NONGAUSSIAN MODELS

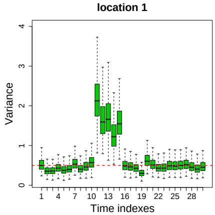


(a) Model with lognormal h.

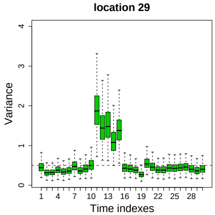


(b) Model with lognormal h and λ_1.

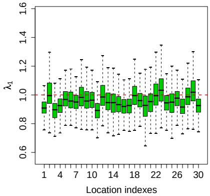
NONGAUSSIAN MODEL WITH λ_1 AND λ_2



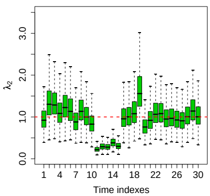
(a) Variance for each time.



(b) Variance for each time.



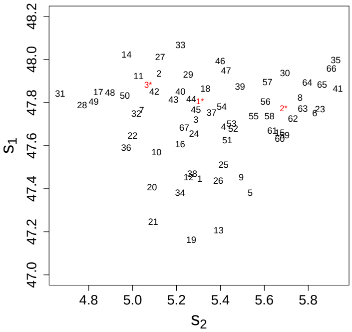
(c) $\lambda_{1i}, i = 1, \ldots, I$.



(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: $X_i \sim \mathrm{Ga}(\eta_i, 1)$

$$C(s_i, s_j, t_i, t_j) = \frac{(1+||s/a||^\alpha)^{-\eta_1}}{\sqrt{\lambda_1(s_i)\lambda_1(s_j)}} \frac{(1+|t/b|^\beta)^{-\eta_2}}{\sqrt{\lambda_2(t_i)\lambda_2(t_j)}} (1+||s/a||^\alpha+|t/b|^\beta)^{-\eta_0},$$

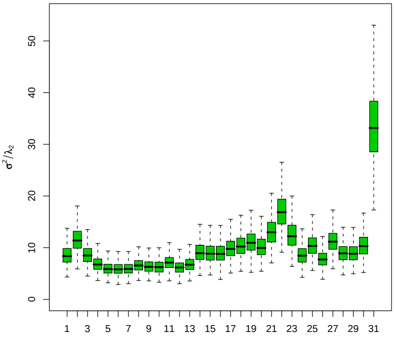
$s = s_i - s_j, t = t_i - t_j$ and $c = \eta_0/(1+\eta_0)$.

BAYES FACTOR

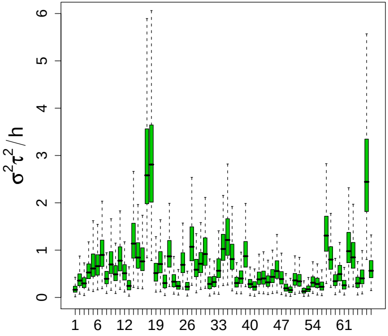
	h	λ_1	λ_1 & h	λ_2	λ_2 & h	λ_1 & λ_2	λ_1, h & λ_2
Shifted gamma	172	148	345	138	279	417	547

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

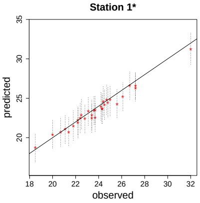


(a) $\sigma^2(1-\tau^2)/\lambda_2$.

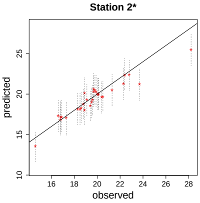


(b) $\sigma^2\tau^2/h$.

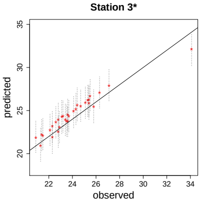
PREDICTED TEMPERATURE AT THE OUT-OF-SAMPLE STATIONS



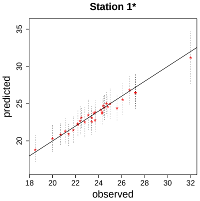
(a) Gaussian Model.



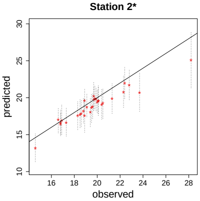
(b) Gaussian Model.



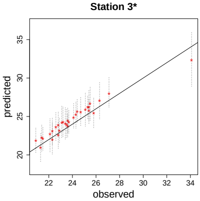
(c) Gaussian Model.



(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	97.25
h	3.83	4.34	112.56
λ_1	3.74	4.36	107.43
λ_1 & h	3.75	4.48	117.20
λ_2	3.73	3.94	76.73
λ_2 & h	3.73	3.87	77.60
λ_1 & λ_2	4.51	4.65	96.35
λ_1, h & λ_2	3.84	4.02	90.30

REFERENCES

N Cressie and H-C Huang
Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions
Journal of the American Statistical Association. (94) 1330–1340, 1999.

T Fonseca and M F J Steel
A General Class of Nonseparable Space-time Covariance Models
Environmetrics. vol 22 issue 2, pages 224-242, 2011.

T Fonseca and M F J Steel
Non-Gaussian Spatiotemporal Modelling through Scale Mixing
Biometrika. vol 98 issue 4, pages 761-774, 2011.

T Gneiting
Nonseparable, Stationary Covariance Functions for Space-Time Data
Journal of the American Statistical Association. (97) 590–600, 2002.

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.