On modelling positive continuous data with spatio-temporal dependence

Moreno Bevilacqua[†], Christian Caamaño [††], Carlo Gaetan[‡]

[†]Instituto de Estadística, Universidad de Valparaíso, Valparaíso, Chile.
[††]Departamento de Estadística, Universidad del Bío-Bío, Concepción, Chile.
[‡]DAIS, Università Ca' Foscari di Venezia, Venice, Italy

April 2, 2018

Abstract

In this paper we aim to propose two models for regression and dependence analysis when dealing with positive spatial or spatio-temporal continuous data. Specifically we propose two (possibly non stationary) random processes with Gamma and Weibull marginals. Both processes stem from the same idea, namely from the transformation of a sum of independent copies of a squared Gaussian random process. We study the associated geometrical and extremal properties. Since maximum likelihood estimation is not feasible, even for relatively small data-set, we suggest to adopt the pairwise likelihood. The effectiveness of our proposal is illustrated through a simulation study that we supplement with a new analysis of well-known dataset, the Irish wind speed data (Haslett and Raftery, 1989). In our example we do not consider a preliminary transformation of the data differently from the previous studies.

Keywords: Linear Prediction; Non-Gaussian data; Pairwise likelihood; Regression; Tail dependence, Wind data.

1 Introduction

Climatology, Environmental Sciences and Engineering, to name some fields, show an increasing interest in statistical analyses of temporal and/or spatial data. In order to model the inherent uncertainty of the data, Gaussian random processes play a fundamental role (see Cressie, 1993; Cressie and Wikle, 2011, for instance). Indeed, they have to offer marginal and dependence modelling in terms of mean and covariance, methods of inference well studied and scalable for large dataset (Heaton et al., 2017) and optimality in the prediction (Stein, 1999).

However a bell shaped distribution of the data could be limiting. In particular, data collected in a range of studies such as wind speeds (Pryor and Barthelmie, 2010), ocean surface currents (Galanis et al., 2012) and rainfalls (Neykov et al., 2014) take continuous positive values and exhibit skewed sampling distributions.

In such situations a large part of the statistical literature have considered non-linear transformations of a Gaussian random process. Examples of these transformed processes are the log-Gaussian random process (Cressie, 1993; De Oliveira, 2006), the trans-Gaussian random processes

(De Oliveira et al., 1997; Allcroft and Glasbey, 2003) and the copula models (Kazianka and Pilz, 2010; Masarotto and Varin, 2012).

A complementary approach specifies consonant distributions, for instance Gamma or Weibull distributions, and uses a hidden Gaussian random process to explain the dependence among the observations. This approach encompasses the spatial generalized linear model (Diggle et al., 1998) and the hierarchical model (Banerjee et al., 2014). Usually the computational burden of the fitting procedure is considerable, but it can be greatly reduced under a Markov assumption about the underlying process (Lindgren et al., 2011).

In this paper we introduce two examples of random processes with Gamma and Weibull marginal distributions for modelling positive spatial or spatio-temporal continuous data. The random processes are obtained by transforming the sum of independent copies of a squared Gaussian random process. Thus our proposal is different from the trans-Gaussian random processes and the copula models since they exploit just one copy. It is also possible to account for non stationary observations by a simple modification of the stationary versions of the processes.

We provides analytical expressions of the bivariate distributions and we illustrates the geometrical properties of the trajectories. Since the Weibull distribution can be used for modelling extreme values (see Opitz, 2016, for instance), we also study the extremal properties of the processes i.e. when the observations overcome a 'high' threshold.

A limitation of our modelling proposal is the difficulty of evaluating the multivariate density that prevents an inference approach based on the likelihood and the derivation of an optimal predictor that minimizes the mean square prediction error. For this reason we investigate the use of a weighted version of the pairwise likelihood (Lindsay, 1988; Varin et al., 2011) for estimating the unknown parameters. Moreover two linear and unbiased predictors are proposed following the approach detailed in De Oliveira (2014).

We have carried out a designed simulation experiment, in the spatio-temporal setting, in order to investigate and compare the sampling properties of the estimators. Then the effectiveness of our models is illustrated through a new analysis of a well-known dataset, the Ireland wind data in Haslett and Raftery (1989). Our study of the wind speed is performed on the raw data without a preliminary transformation differently from the previous literature (Gneiting, 2002b; Stein et al., 2004; Stein, 2005; de Luna and Genton, 2005).

The remainder of the paper is organized as follows. In Section 2, we introduce the random processes and we describe their features. Section 3 starts with a short description of the estimation method and ends with tackling the prediction problem. In Section 4 we report the numerical results of the simulation study and the re-analysis of the Irish Wind speed data. Finally some concluding remarks are consigned to Section 5.

2 Random processes with Gamma and Weibull marginals.

The definition of the random processes starts by considering a 'parent' Gaussian random process $Z := \{Z(s), s \in S\}$, where s represents a location in the domain S. Spatial ($S \subseteq \mathbb{R}^d$) or spatio-temporal examples ($S \subseteq \mathbb{R}^d \times \mathbb{R}_+$) will be considered indifferently. We also assume that Z is stationary with zero mean, unit variance and correlation function $\rho(h) := \mathrm{cor}(Z(s+h), Z(s))$ where $s + h \in S$.

Let $Z_1, \ldots, Z_m$ be m independent copies of Z and define the random process $X_m := \{X_m(s), s \in$

$S\}$ as

$$X_m(s) := \sum_{k=1}^{m} Z_k(s)^2/m. \tag{1}$$

It turns out that X_m is a stationary random process with marginal distribution $\mathrm{Gamma}(m/2, m/2)$. Here the pairs $m/2, m/2$ are the shape and rate parameter of the Gamma distribution i.e. $\mathrm{E}(X_m(s)) = 1$ and $\mathrm{var}(X_m(s)) = 2/m$ for all s.

A possible drawback for the distribution $\mathrm{Gamma}(m/2, m/2)$ is that it is a limited model for continuous positive data due to the restrictions to the half integers for the shape parameter. A closer look to the problem reveals that overcoming this limitation is not easy.

The Laplace transform of the random vector $X_m(s_1), \ldots, X_m(s_n)$ is given by

$$\mathrm{E}\left(\exp\left(-\sum_{i=1}^{n} \alpha_i X_m(s_i)\right)\right) = \frac{1}{|I + (2/m)AR|^{m/2}}, \tag{2}$$

where I is the $n \times n$ identity matrix, $A = \mathrm{diag}(\alpha_1, \ldots, \alpha_n)$ and $R = \{\rho(s_i, s_j)\}$ is the correlation matrix of $Z(s_1), \ldots, Z(s_n)$ (Vere-Jones, 1997, Proposition 4.3). Therefore the finite dimensional distribution of X_m is a particular instance of the finite dimensional distribution of a β-permanental process Y (Eisenbaum and Kaspi, 2009) for which the Laplace transform satisfies

$$\mathrm{E}\left(\exp\left(-\sum_{i=1}^{n} \delta_i Y(s_i)\right)\right) = \frac{1}{|I + DB|^{\beta}}.$$

In the last formula β is any positive real, $B = \{b_{ij}\}$ is a $n \times n$ matrix such that its inverse B^{-1} is a diagonally dominant M-matrix and $D = \mathrm{diag}(\delta_1, \ldots, \delta_n)$. It turns out that the permanental process has marginal distribution $Y(s_i) \sim \mathrm{Gamma}(\beta, 1/b_{ii})$.

Thus one may be tempted to identify the matrix B with a correlation matrix R that comes from the Gaussian process Z and, possibly, define a process with more flexible marginals. However this is identification holds only for particular instances of Gaussian processes since the random process $Z^2 = \{Z^2(s), s \in S\}$ has to be infinitely divisible (Eisenbaum and Kaspi, 2006). The condition is satisfied, for instance, by the stationary Gaussian random process Z defined on $S = \mathbb{R}$ with exponential correlation function.

The multivariate density of X_m can be expressed as an infinite series where each element of the series is a double sum of order n involving the computation of Laguerre polynomials and a determinant of a $n \times n$ square matrix (Krishnamoorthy and Parthasarathy, 1951). However analytical forms can be derived only in some special cases (Royen, 2004) and this fact prevents their use for the maximum likelihood estimation.

On the other hand the evaluation of the bivariate densities of a pair of observations $X_m(s_1)$ and $X_m(s_2)$ is still feasible and could be used for estimating the parameters (see Section 3). The bivariate distribution of X_m is known as the Kibble bivariate Gamma distribution (Kibble, 1941) with density

$$\begin{aligned} f_{X_m}(x_1, x_2) &= \frac{2^{-m} m^m (x_1 x_2)^{m/2-1}}{\Gamma(m/2)(1-\rho^2)^{m/2}} \left(\frac{m|\rho|\sqrt{x_1 x_2}}{2(1-\rho^2)}\right)^{1-m/2} \exp\left\{-\frac{m(x_1+x_2)}{2(1-\rho^2)}\right\} \\ &\quad \times I_{m/2-1}\left(\frac{m|\rho|\sqrt{x_1 x_2}}{(1-\rho^2)}\right), \end{aligned} \tag{3}$$

where $\rho = \rho(s_1 - s_2)$ and $I_\alpha(x)$ denotes the modified Bessel function of the first kind of order α.

When $m = 2$, X_2 has marginal distribution which is a standard exponential distribution. In such case the bivariate distribution reduces to the Downton distribution (Downton, 1970) with density

$$f_{X_2}(x_1, x_2) = \frac{1}{1-\rho^2} \exp\left[-\frac{x_1 + x_2}{1-\rho^2}\right] I_0\left(\frac{2|\rho|\sqrt{x_1 x_2}}{1-\rho^2}\right). \tag{4}$$

Therefore a random process $U = \{U(s), s \in S\}$ with marginal distribution Weibull$(\kappa, \nu(\kappa))$ can be derived by the transformation

$$U(s) := \nu(\kappa) X_2(s)^{1/\kappa}, \tag{5}$$

where $\nu(\kappa) = \Gamma^{-1}(1 + 1/\kappa)$ and $\kappa > 0$ is a shape parameter. Under this parametrization, $\mathrm{E}(U(s)) = 1$ and $\mathrm{var}(U(s)) = (\Gamma(1 + 2/\kappa)\nu^2(\kappa) - 1)$. Moreover the density of a pair of observations $U(s_1)$ and $U(s_2)$ is easily obtained from (4) and (5), namely

$$f_U(u_1, u_2) = \frac{\kappa^2 (u_1 u_2)^{\kappa-1}}{\nu^{2\kappa}(\kappa)(1-\rho^2)} \exp\left[-\frac{u_1^\kappa + u_2^\kappa}{\nu^\kappa(\kappa)(1-\rho^2)}\right] I_0\left(\frac{2|\rho|(u_1 u_2)^{\kappa/2}}{\nu^\kappa(\kappa)(1-\rho^2)}\right). \tag{6}$$

Both processes are stationary by construction. A non stationary version of X_m that we denote as $G_m = \{G_m(s), s \in S\}$ can be obtained by setting

$$G_m(s) := \mu(s) X_m(s), \tag{7}$$

where $\mu(s) > 0$ is a deterministic function. Then the random process G_m has mean function $\mathrm{E}(G_m(s)) = \mu(s)$ that does not depend on the shape parameter and the heteroscedasticity is induced by the mean function, i.e. $\mathrm{var}(G_m(s)) = 2\mu(s)^2/m$. We can apply a similar argument to (5) and define $W = \{W(s), s \in S\}$ with

$$W(s) := \mu(s) U(s). \tag{8}$$

As in the Gamma case the mean function does not depend on the shape parameter, i.e. $\mathrm{E}(W(s)) = \mu(s)$, and $\mathrm{var}(W(s)) = \mu(s)^2(\Gamma(1 + 2/\kappa)\nu^2(\kappa) - 1)$.

Thus a spatio-temporal regression model can be obtained by assuming that $\log(\mu(s)) = x(s)'\gamma$ where $x(s)$ is a k-dimensional vector of covariates and γ is a k-dimensional vector of (unknown) parameters. Finally note that the bivariate densities of G_m and W can be derived as

$$\begin{aligned} f_{G_m}(y_1, y_2) &= (\mu_1\mu_2)^{-1} f_{X_m}(y_1/\mu_1, y_2/\mu_2), \\ f_W(y_1, y_2) &= (\mu_1\mu_2)^{-1} f_U(y_1/\mu_1, y_2/\mu_2). \end{aligned} \tag{9}$$

2.1 Second-order and geometrical properties

It is easy to show that the correlation function of X_m, $\rho_{X_m}(h)$, is equal to $\rho^2(h)$, the squared of the correlation function of the 'parent' Gaussian random process, and X_m and G_m have the same correlation function.

Using Proposition 1 in Appendix we can also obtain the correlation function of U and W, namely

$$\rho_W(h) = \frac{\nu^{-2}(\kappa)}{[\Gamma(1 + 2/\kappa) - \nu^{-2}(\kappa)]} \left[{}_2F_1\left(-1/\kappa, -1/\kappa; 1; \rho^2(h)\right) - 1\right], \tag{10}$$

where the function

$$ {}_pF_q(a_1, a_2, \ldots, a_p; b_1, b_2, \ldots, b_q; x) := \sum_{k=0}^{\infty} \frac{(a_1)_k, (a_2)_k, \ldots, (a_p)_k}{(b_1)_k, (b_2)_k, \ldots, (b_q)_k} \frac{x^k}{k!} \quad \text{for} \quad p, q = 0, 1, 2, \ldots $$

is the generalized hypergeometric function (Gradshteyn and Ryzhik, 2007) and $(a)_k := \Gamma(a + k)/\Gamma(a)$, for $k \in \mathbb{N} \cup \{0\}$, is the Pochhammer symbol.

Unlike the Gamma case, the correlation of W depends on the shape parameter κ and $\rho_{G_m}(h) = \rho_W(h)$ for $\kappa = 1$. We can notice that both processes have positive correlation function with $0 \leq \rho_{G_m}(h) \leq |\rho(h)|$, $0 \leq \rho_W(h) \leq |\rho(h)|$. In fact the continuous function

$$ g(\rho) := \frac{\nu^{-2}(\kappa)}{[\Gamma(1 + 2/\kappa) - \nu^{-2}(\kappa)]} \left[{}_2F_1\left(-1/\kappa, -1/\kappa; 1; \rho^2\right) - 1 \right] $$

is concave for $\rho \in (-1, 1)$ with $g(-1) = g(1) = 1$ and $g(0) = 0$. It turns out that pairwise zero correlation for G_m and W implies independence as for the Gaussian case. Figure 1 allows to better understand the "shrinking" effect on the correlation ρ_W according different values of the shape parameter κ.

Moreover, the function $g(\rho)$ is infinitely differentiable in a neighbourhood of $\rho = 1$. We can conclude the geometrical properties of X_m and U are strictly linked to the geometric properties of the 'parent' Gaussian process Z. In fact, the k-times mean-square differentiability of a stationary random process is equivalent to the $2k$-times differentiability of its correlation function at $h = 0$ (Stein, 1999, Section 2.4). However, $\rho_{X_m}(h)$ and $\rho_U(h)$ are $2k$-times differentiable at $h = 0$ as well as $\rho(h)$ is. In summary, X_m and U are k-times mean-square differentiable if Z is k-times mean-square differentiable. Finally, it is trivial to see that the sample path continuity and differentiability of X_m and U originate from the 'parent' Gaussian process since the processes are smooth transformations of Z.

2.2 Illustrative examples

In this subsection we illustrate some geometric features of the random processes G and W by means of two examples of parametric correlation function $\rho(h)$, namely

1. the Matérn correlation function (Matérn, 1986), defined as

$$ \mathcal{M}(\|h\|; \alpha, \psi) := \frac{2^{1-\psi}}{\Gamma(\psi)} \left(\|h\|/\alpha\right)^{\psi} \mathcal{K}_{\psi}\left(\|h\|/\alpha\right), \qquad \|h\| \geq 0. \tag{11} $$

   Here, $\alpha > 0$ is the range parameter, $\mathcal{K}_{\psi}$ is a modified Bessel function of the second kind of order $\psi > 0$.

2. the Generalized Wendland correlation function (Gneiting, 2002a), defined as

$$ \mathcal{GW}(\|h\|; \alpha, \psi, \delta) := \begin{cases} \mathrm{B}^{-1}(2\psi, \delta + 1) \int_{\|h\|/\alpha}^{1} u \left(u^2 - (\|h\|/\alpha)^2\right)^{\psi - 1} & \\ \qquad\qquad \times \ (1 - u)^{\delta} \, du & \|h\| < \alpha \\ 0 & \text{otherwise} \end{cases}, \tag{12} $$

   where $\mathrm{B}(\cdot, \cdot)$ is the Beta function, $\alpha > 0$, $\psi > 0$ and $\delta \geq (d + 1)/2 + \psi$. For $\psi = 0$ the correlation is defined as

$$\mathcal{GW}(\|h\|;\alpha,0,\delta) := \begin{cases} (1-\|h\|/\alpha)^{\delta} & \|h\|<\alpha \\ 0 & \text{otherwise} \end{cases}.$$

Both models represent flexible parametric models for the spatial correlation and the parameter ψ allows a continuous parametrization for the smoothness of the associated Gaussian random process (Stein, 1999; Bevilacqua et al., 2018). In particular for a positive integer k, the sample paths of a Gaussian process are k times differentiable if and only if $\psi > k$ in the Matérn case and if and only if $\psi > k - 1/2$ in the Generalized Wendland case. The Generalized Wendland correlation is also compactly supported, an interesting feature from computational point of view.

Figures 2 and 3 show six realizations of the stationary random process W on $S = [0,1]^2$ setting $\mu(s) = 1$ and using the same seed for the pseudo-random generation.

We have firstly considered a Matérn correlation function $\rho(h) = \mathcal{M}(\|h\|;\alpha,\psi)$ with three different parametrization for the smoothness parameter $\psi = 0.5, 1.5, 2.5$ and the shape parameter $\kappa = 10, 3, 1$. The values of the range parameters, $\alpha = 0.034, 0.042, 0.067$, have been chosen in order to obtain a practical range equal to 0.2. The corresponding correlation function $\rho_W(h)$ are plotted (from left to right) in the centre panel of Figure 2 and the bottom panel reports the histograms of the observations.

The simulation experiment has been also conducted for the Generalized Wendland correlation function $\rho(h) = \mathcal{GW}(\|h\|;\alpha,\psi,\delta)$ (see Figure 3). Here, we have set three different values of the parameters $\psi = 0, 1, 2$ that correspond to the same degree of the smoothness of the Matérn correlation function and common values for $\alpha = 0.2$ and $\delta = 4.5$. The shape of the marginal distribution does not change with respect the previous experiment, i.e. $\kappa = 10, 3, 1$.

It is apparent that the correlation $\rho_W(h)$ inherits the change of the differentiability at the origin from $\rho(h)$ when increasing ψ. This changes have consequences on the geometrical properties of the associated random processes. In fact the smoothness of the realizations increase with ψ. Note also the flexibility of this kind of random process when modelling positive data since both positive and negative skewness can be achieved with different values of κ, see Figures 3(g-i).

2.3 Extremal properties

In climatology and environmental data analysis, the Weibull distribution is a commonly used marginal model. For instance, the distribution of wind speeds in latitudes extra to the Tropical zone are well known to comply closely the Weibull distribution. Thus methods based on fitting a Weibull distribution to all available data are often applied to extrapolate sizes of future extreme winds, using the tail of the Weibull distribution. When studying exceedances over a high threshold, standard measures of dependence such as correlation, are typically inappropriate due to the need to extrapolate beyond the observation range. In this context different notions of dependence arise, with the main distinction being between asymptotic dependence and asymptotic independence of exceedances as the threshold level increases. Under spatial asymptotic dependence the likelihood of a large event happening in one location is tightly related to high values being recorded at a location nearby; the opposite is true under asymptotic independence, in which large events might be recorded at one location only, and not in neighbouring locations (Wadsworth and Tawn, 2012).

The *tail dependence coefficient* (Sibuya, 1960; Coles et al., 1999) helps to discriminate between the two classes and it is defined as

$$\chi := \lim_{q\to 1^-} \frac{\Pr(Y_1 > F_1^{\leftarrow}(q), Y_2 > F_2^{\leftarrow}(q))}{\Pr(Y_2 > F_2^{\leftarrow}(q))},$$

where $F_i^{\leftarrow}$, $i = 1, 2$ denote the generalized inverse distribution functions of two observations Y_1 and Y_2 . We say that Y_1 and Y_2 are asymptotically dependent if the limit is positive. The case $\chi = 0$ characterizes asymptotic independence. In the Appendix (Proposition 2) we prove that the Weibull random field (5) is asymptotically independent i.e.

$$\lim_{w\to\infty} \frac{\Pr(W(s_1) > w, W(s_2) > w)}{\Pr(W(s_2) > w)} = 0$$

for any pairs $W_1 := W(s_1)$ and $W_2 := W(s_2)$.

3 Estimation and prediction

3.1 Pairwise likelihood inference

Suppose that we have observed $y_1, \ldots, y_n$ at the locations $s_1, \ldots, s_n$ and let θ be the vector of unknown parameters for the random processes G_m or W. The evaluation of the full likelihood for θ is impracticable for moderately large n. A computationally more efficient alternative (Lindsay, 1988; Varin et al., 2011) combines the bivariate distributions of all possible distinct pairs of observations (y_i, y_j). The pairwise likelihood function is given by

$$\mathrm{pl}(\theta) := \sum_{i<j} \log f(y_i, y_j; \theta) c_{ij} \tag{13}$$

where $f(y_i, y_j; \theta)$ is one of the bivariate densities in (9) and c_{ij} is a non-negative weight. The choice of a cut-off weight, namely $c_{ij} = 1$ if $\|s_i - s_j\| \leq \Delta$, and 0 otherwise, for a positive value of Δ, can be motivated by its simplicity and by observing that that dependence between observations which are distant is weak. Therefore, the use of all pairs may skew the information confined in pairs of near observations (Davis and Yau, 2011; Bevilacqua and Gaetan, 2015).

The maximum pairwise likelihood estimator is

$$\widehat{\theta} := \mathrm{argmax}_\theta \ \mathrm{pl}(\theta)$$

and, arguing as in Bevilacqua et al. (2012) and Bevilacqua and Gaetan (2015), it can be shown that, under increasing domain asymptotics, $\widehat{\theta}$ is consistent and asymptotically Gaussian with asymptotic covariance matrix given by the inverse of the Godambe information

$$\mathcal{G}_n(\theta) := \mathcal{H}_n(\theta)\mathcal{J}_n(\theta)^{-1}\mathcal{H}_n(\theta),$$

where $\mathcal{H}_n(\theta) := \mathrm{E}[-\nabla^2 \,\mathrm{pl}(\theta)]$ and $\mathcal{J}_n(\theta) := \mathrm{Var}[\nabla \,\mathrm{pl}(\theta)]$.

Model selection can be performed by minimizing the pairwise likelihood information criterion, defined as

$$\mathrm{PLIC} := -\,\mathrm{pl}(\widehat{\theta}) + \mathrm{tr}(\mathcal{J}_n(\widehat{\theta})^{-1}\mathcal{H}_n(\widehat{\theta}))$$

which is an analogue of the Akaike information Criterion (AIC) in a pairwise likelihood framework (Varin and Vidoni, 2005).

Since the evaluation of the pairwise likelihood can be split into a number of parallel tasks, it has been implemented in the Open Computing Language (OpenCL) for parallel computing. The implementation is part of `GeoModels`, an upcoming R package.

3.2 Linear prediction

The lack of workable multivariate densities forestalls the use of the conditional distributions for the prediction. Therefore we choose a suboptimal solution following Bellier et al. (2010) and De Oliveira (2014) and we suggest to use two linear predictors for the random variable $Y(s_0)$ at some unobserved location s_0 based on the data at locations $s_1, \ldots, s_n$.

If the mean function in known, we define the error process $E = \{E(s), s \in S\}$ with $E(s) := Y(s)/\mu(s)$, where the process Y is indifferently the non stationary Gamma or Weibull process defined in Section 2.

Now two linear predictors can be considered, namely

$$\hat{Y}_{SK}(s_0) := \mu(s_0) + \sum_{i=1}^{n} \lambda_i [\mathrm{E}(s_i) - 1] \tag{14}$$

and

$$\hat{Y}_{OK}(s_0) := \mu(s_0) \sum_{i=1}^{n} \lambda_i \mathrm{E}(s_i), \qquad \text{with} \qquad \sum_{i=1}^{n} \lambda_i = 1. \tag{15}$$

Both predictors are unbiased, $\mathrm{E}(\hat{Y}(s_0)) = \mathrm{E}(Y(s_0))$, and linear predictors of $Y(s_0)$.

The vector of weights $\lambda = (\lambda_1, \ldots, \lambda_n)'$ is derived minimizing the mean square error $\mathrm{MSE}(s_0) = \mathrm{E}[Y(s_0) - \hat{Y}(s_0)]^2$ with respect to λ. The solutions for the predictors are given by the equations of the simple and ordinary kriging, respectively, (Cressie, 1993, ch. 3) i.e.

$$\lambda_{SK} = \Sigma^{-1} c(s_0)$$

and

$$\lambda_{OK} = \Sigma^{-1} \left(c(s_0) + \frac{(1 - 1_n' \Sigma^{-1} c(s_0))}{1_n' \Sigma^{-1} 1_n} 1_n \right).$$

Here 1_n is the unit vector of dimension n, Σ is a $n \times n$ whose (i, j)-th element is $\mathrm{cov}(E(s_i), E(s_j))$ and $c(s_0) = (\mathrm{cov}(E(s_0), E(s_1)), \ldots, \mathrm{cov}(E(s_0), E(s_n)))'$.

In practice the predictors cannot be evaluated since $\mu(s)$ and $\mathrm{cov}(E(s), E(s'))$ are unknown. For this reason we suggest to use a plug-in estimate for $\mu(s)$ and $\mathrm{cov}(E(s), E(s'))$ using the pairwise likelihood estimates.

Finally, we note that the error process is stationary with marginal distributions $\mathrm{Gamma}(m/2, m/2)$ or $\mathrm{Weibull}(\kappa, \nu(\kappa))$ and this feature can be used for checking the adequacy of the marginal specification as will be seen in Section 4.

4 Numerical results

4.1 Simulation study

We report the results of a simulation study for assessing the performance of the estimation method. We have considered examples of spatio-temporal processes tailored to the models we have used in the real data analysis.

We consider two thousands observations $y_{i,j} = y(s_i, t_j)$ over a space time domain where one hundred locations s_i, $i = 1, \ldots, 100$, are uniformly distributed over the unit square and t_j, $j = 1, \ldots, 20$ with $t_j = j$. We choose the non-separable model (Gneiting, 2002b)

$$\rho(h, u) = \frac{1}{(1 + |u|/\alpha_T)^{2.5}} \exp\left\{ -\frac{||h||/\alpha_S}{(1 + |u|/\alpha_T)^{\alpha_{NS}/2}} \right\} \tag{16}$$

m	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\alpha}_S$	$\hat{\alpha}_T$
1	-0.00032	0.00391	-0.00005	-0.00059
	0.07709	0.11467	0.00633	0.33357
6	0.00081	-0.00136	-0.00007	0.00225
	0.03004	0.04558	0.00493	0.27199
9	-0.00013	0.00009	-0.00002	-0.00485
	0.02519	0.03937	0.00495	0.26383

Table 1: Results for the random process G_m, with 'parent' spatio-temporal correlation (16) according different values of the shape parameter $m = 1, 6, 9$. Each cell contains bias (top) and RMSE (bottom) of the estimates evaluated on 1 000 Monte Carlo runs.

κ	$\hat{\kappa}$	$\hat{\gamma}_1$	$\hat{\gamma}_2$	$\hat{\alpha}_S$	$\hat{\alpha}_T$
1	0.00163	0.00171	-0.00296	-0.00020	-0.00981
	0.01977	0.05277	0.07908	0.00549	0.30890
3	0.00489	0.00078	-0.00099	-0.00020	-0.00981
	0.05930	0.01804	0.02636	0.00549	0.30890
10	0.01498	0.00025	-0.00030	-0.00020	-0.00690
	0.19694	0.00552	0.00790	0.00550	0.30713

Table 2: Results for the random process W, with 'parent' spatio-temporal correlation (16) according different values of the shape parameter $\kappa = 1, 3, 10$. Each cell contains bias (top) and RMSE (bottom) of the estimates evaluated on 1 000 Monte Carlo runs.

with $h = s - s'$, $u = t - t'$, $\alpha_S > 0$, $\alpha_T > 0$ and $0 \leq \alpha_{NS} \leq 1$.

In the simulation experiment spatial range, temporal range and non-separability parameter are fixed at $\alpha_S = 0.15/3$, $\alpha_T = 2$ and $\alpha_{NS} = 0.5$, respectively.

Non-stationarity for G_m and W is induced by taking $\mu(s_i, t_j) = \exp\{\gamma_1 + \gamma_2 u(s_i, t_j)\}$ where $u(s_i, t_j)$ is a value from the $(0, 1)$-uniform distribution. The values of the regression parameters are $\gamma_1 = -1$ and $\gamma_2 = 2$. Furthermore we will consider different values for the shape of the marginal distributions, namely $m = 1, 6, 9$ for G_m and $\kappa = 1, 3, 10$ for W.

Finally two vectors of unknown parameters for the pairwise likelihood (13) are examined, namely $\theta_G = (\gamma_1, \gamma_2, \alpha_S, \alpha_T)'$ and $\theta_W = (\kappa, \gamma_1, \gamma_2, \alpha_S, \alpha_T)'$. In shaping the pairwise likelihood, we opt for a cut-off weight equal to one if $\|s - s'\| \leq 0.1$ and $|t - t'| \leq 2$, zero otherwise, which bypasses an explosion of the number of pairwise terms and shifts focus to relatively short-range distances where dependence matters most.

Table 1 and 2 report the bias and the root of the mean square error (RMSE), obtained by 1 000 Monte Carlo runs. Moreover Figure 4 and 5 contain the boxplots of the estimates for two values of the shape parameter ($m = 1$ and $\kappa = 10$) as illustrative examples.

As general comment the estimates are overall approximatively unbiased, see the values on the top of each cell in Table 1-2. The distribution of the estimates are quite symmetric (see Figure 4, centred around the true value and 5) numerically stable with very few outliers. For the random process G_m the RMSE is affected by the magnitude of the shape parameter. More precisely the RMSE seems to decrease when the shape parameter m increases. In the case of the random process W this pattern is less evident, but the precision of estimates of the regression parameters benefits of the increment of the parameter κ. Finally not surprising is the fact that the temporal parameter α_T is worst estimated since we have less observations along the temporal dimension.

4.2 Irish wind speed data

We consider a dataset of daily average wind speeds $y(s_i, t)$, $i = 1, \ldots, 12$, $t = 1, \ldots, 6\,574$, observed for a period $1961 - 1978$, at twelve synoptic meteorological stations in the Republic of Ireland. An in-depth analysis of these data has been carried out for the first time by Haslett and Raftery (1989) and later the data have been considered in several papers (see, for instance, Gneiting, 2002b; Stein et al., 2004; Stein, 2005; de Luna and Genton, 2005). The original time series contain positive values, except for few zeroes and the marginal distributions are noticeably asymmetric. Moreover the standard deviations are correlated with means that vary over the seasons.

For these reasons the previous analysis have suggested to preliminary transform the data and then extract a common seasonal pattern. These transformations yield to approximately stationary time-series, with bell-shaped marginal distributions, that have been modelled by different Gaussian random processes.

Differently from the previous attempts, we adopt the G_m and W random processes as models of the original data and we replace the sixteen zeros with the mean speed of the previous and next day. The seasonal variations in the level and in the variability are modelled by the common trend

$$\log \mu(s, t; \gamma) = \gamma_0 + \sum_{k=1}^{3} \gamma_{1,k} \cos\left(\frac{2\pi kt}{365}\right) + \gamma_{2,k} \sin\left(\frac{2\pi kt}{365}\right),$$

with $\gamma = (\gamma_0, \gamma_{1,1}, \gamma_{1,2}, \gamma_{1,3}, \gamma_{2,1}, \gamma_{2,2}, \gamma_{2,3})$.

Following Gneiting (2002b) we restrict our attention to the short distance and short lag dependence using the the same correlation function (16), with different values of $\alpha_{NS} = 0, 0.5, 1$. However we note that the space-time separability, i.e. $\rho(h, u) = \rho_1(h)\rho_2(u)$ for two correlation functions ρ_1 and ρ_2, is inherited only by the G_m random process.

A natural competitor of our models is the Log-Gaussian random process $L = \{L(s, t), (s, t) \in S\}$ namely $L(s, t) = \mu(s, t; \gamma) \exp\{\sigma Z(s, t)\}$, $\sigma > 0$, where $Z(s, t)$ is a zero mean, unit variance Gaussian process with correlation $\rho(h, u)$ (see, for instance, De Oliveira, 2006). In such case $\mathrm{E}(L(s, t)) = \mu^*(s, t)$ and

$$\log \mu^*(s, t; \gamma) = \gamma_0^* + \sum_{k=1}^{3} \gamma_{1,k} \cos\left(\frac{2\pi kt}{365}\right) + \gamma_{2,k} \sin\left(\frac{2\pi kt}{365}\right),$$

with $\gamma_0^* = \gamma_0 + \sigma^2/2$.

In order to asses the predictive performances of the different models we split the data set into a training period consisting of years $1961 - 1970$, with $n_{tr} = 3\,289$ days, and a test period for the remaining years, $n_{te} = 3\,285$ days. The sample size ($n = 39\,468$) prevents the use of a full likelihood approach even for the Log-Gaussian model. Therefore the estimation of parameters for the three models is based on the weighted pairwise likelihood, with a cut-off weight equal to one if $\|s - s'\| \leq 350$ jointly with $|t - t'| \leq 1$ and zero otherwise.

Our implementation, in the `GeoModels` package, based on C and OpenCL allows to speed-up five times the computation of the weigthed pairwise likelihood estimator with respect to a standard C implementation. In particular one evaluation of the function needs approximatively 1.1 and 4.9 seconds with or without OpenCL implementation where the time of evaluation (in seconds) has been measured in terms of elapsed time on a laptop with 2.4 GHz processor and 16 GB of memory.

In Table 3 we report the estimates with the standard errors and PLIC values computed using the sub-sampling technique in Bevilacqua et al. (2012). Actually, we have estimated several spec-

ification of G_m for $m = 4, 5, \ldots, 9$ but, for simplicity, we report only our best choice, $m = 6$, according to the PLIC value.

The estimated trends are not dissimilar for the three models. Instead it is evident that the estimates of α_S and α_T entail a stronger spatial and temporal range for Gamma and Weibull models with respect to the Log-Gaussian ones. A comparison of the three models in terms of PLIC leads to a clear discrimination among them, in favour of the first ones.

We remind that $E(s,t) = Y(s,t)/\mu(s,t;\gamma)$ is a random variable with distribution Gamma, Weibull or Log-Gaussian. Thus the residuals $\widehat{r}(s_i,t) = y(s_i,t)/\mu(s_i,t;\widehat{\gamma})$ can be used for checking the goodness-of-fit of the models. The marginal fitting of the Weibull model is extremely good, according Figure 6-(a) in which we compare the quantiles of the residuals with the quantiles of the distribution $\mathrm{Weibull}(\widehat{\kappa}, \Gamma^{-1}(1+1/\widehat{\kappa}))$. Moreover, the choice of (16) as a model for $\rho(h,u)$ is supported by the comparison of the empirical semi-variograms of the residuals with the estimated semi-variograms, see Figures 6-(b-c).

Then we compare the prediction performances of the three models. The root-mean-square error (RMSE) and the mean absolute error (MAE) for each station, namely

$$\mathrm{RMSE}(s_i) = \left\{ \frac{1}{n_{te}} \sum_{t=n_{tr}+1}^{n_{tr}+n_{te}} \left[y(s_i,t) - \hat{Y}(s_i,t) \right]^2 \right\}^{1/2}$$

and

$$\mathrm{MAE}(s_i) = \frac{1}{n_{te}} \sum_{t=n_{tr}+1}^{n_{tr}+n_{te}} |y(s_i,t) - \hat{Y}(s_i,t)|,$$

are customary measures of the prediction performances. Here we consider one-day ahead predictions $\widehat{Y}(s_i,t)$ of the wind speed in the test period using the observations $\{y(s_j, t-k), j = 1, \ldots, 12,\ k = 1, 2, 3\}$ for the last three days.

For the Gamma and the Weibull model we have reported the results obtained by using the simple kriging predictor (14) since we have not found remarkable differences with respect to use the alternative ordinary kriging predictor (15). Instead for the Log-Gaussian model we have chosen the optimal predictor in De Oliveira (2006). As benchmark, we have also considered the naïve predictor $\widehat{Y}(s_i,t) = y(s_i, t-1)$, that uses the observation recorded the day before at the station.

Even though the simple kriging predictor is an suboptimal solution, Gamma and Weibull models outperform again the Log-Gaussian model as we can infer from the last column in Tables 4-5. Overall, the three models performs better than the naïve predictor.

5 Concluding remarks

In this paper we proposed two models for regression and dependence analysis when dealing with positive continuous data. Specifically we propose two non stationary random processes with Gamma and Weibull marginals. Both models are obtained transforming a rescaled sum of independent copies of the squared of a 'parent' Gaussian random process. Our proposal share some similarity with the trans-Gaussian random process. However, unlike many other trans-Gaussian random process, we have showed that nice properties such as stationarity, mean-square continuity, and degrees of mean-square differentiability are inherited from the 'parent' Gaussian random process.

In particular discontinuity of the paths can be induced by choosing a discontinuous correlation function. It is worth noting that a hierarchical spatial model like the generalized linear model (Dig-

	$\alpha_{NS} = 0$			$\alpha_{NS} = 0.5$			$\alpha_{NS} = 1$		
	Gamma	Weibull	Log-Gaussian	Gamma	Weibull	Log-Gaussian	Gamma	Weibull	Log-Gaussian
$\hat{\gamma}_0$	2.3430 (0.0006)	2.3383 (0.0008)	2.1797 (0.0004)	2.3429 (0.0006)	2.3382 (0.0008)	2.1795 (0.0003)	2.3429 (0.0005)	2.3383 (0.0008)	2.1794 (0.0004)
$\hat{\gamma}_{11}$	0.0530 (0.0291)	0.0437 (0.0270)	0.0739 (0.0066)	0.0532 (0.0290)	0.0438 (0.0270)	0.0737 (0.0045)	0.0531 (0.0246)	0.0440 (0.0270)	0.0736 (0.0044)
$\hat{\gamma}_{21}$	0.1367 (0.01058)	0.1457 (0.00754)	0.1060 (0.0052)	0.1372 (0.0106)	0.1457 (0.0075)	0.1058 (0.0037)	0.1372 (0.0090)	0.1457 (0.0075)	0.1058 (0.0038)
$\hat{\gamma}_{12}$	-0.0074 (0.1944)	-0.0102 (0.1113)	-0.0097 (0.0503)	-0.0078 (0.1865)	-0.0102 (0.1115)	-0.0099 (0.0340)	-0.0078 (0.1574)	-0.0103 (0.1108)	-0.0099 (0.0332)
$\hat{\gamma}_{22}$	-0.0288 (0.0515)	-0.0294 (0.0399)	-0.0344 (0.0137)	-0.0290 (0.0514)	-0.0294 (0.0400)	-0.0344 (0.0092)	-0.0291 (0.0433)	-0.0296 (0.0400)	-0.0345 (0.0089)
$\hat{\gamma}_{13}$	-0.0153 (0.09667)	-0.0128 (0.0905)	-0.0257 (0.0191)	-0.0155 (0.0958)	-0.0129 (0.0905)	-0.0258 (0.0130)	-0.0155 (0.0807)	-0.0126 (0.0931)	-0.0258 (0.0128)
$\hat{\gamma}_{23}$	-0.0028 (0.5169)	-0.0070 (0.1609)	0.0003 (1.3805)	-0.0028 (0.5141)	-0.0070 (0.1611)	0.0002 (1.3018)	-0.0027 (0.4471)	-0.0069 (0.1634)	0.0002 (1.5912)
$\hat{\kappa}$		1.9625 (0.0016)			1.9629 (0.0017)			1.9625 (0.0016)	
$\hat{\sigma}^2$			0.4048 (0.0004)			0.4046 (0.0002)			0.4044 (0.0002)
$\hat{\alpha}_s$	640.707 (5.7243)	582.796 (7.6777)	244.388 (0.2319)	641.975 (5.3231)	579.545 (8.2661)	243.371 (0.3513)	642.491 (5.6668)	578.782 (8.1882)	242.233 (0.5147)
$\hat{\alpha}_t$	9.5991 (0.1328)	8.2451 (0.2385)	3.6603 (0.0277)	9.4293 (0.1360)	8.1205 (0.2357)	3.5209 (0.0267)	9.2770 (0.1032)	7.9797 (0.2359)	3.3858 (0.0287)
PLIC	8886705	**8151609**	12805228	**8870918**	8152061	**12283246**	8640458	8156138	14300738

Table 3: Pairwise likelihood estimates for G_m ($m = 6$), W and L processes, their associated standard errors and PLIC values. We highlight the best PLIC for each models in bold.

gle et al., 1998) always induces discontinuity in the path. Another relative merits of the parametrization (7) and (8) is that allows a clear expression and interpretation of the mean function.

We also remark that even though we have limited to ourselves to a continuous Euclidean space, our models can be extended to a spherical domain (Gneiting, 2013; Porcu et al., 2016) or to a network space. In this respect the G_m random process should represent a generalization of the Markov model in Warren (1992).

Two limitations are evident in our proposal: the shape parameter of G_m is restricted to the half-integer values and the lack of an amenable expression of the density outside of the bivariate case. In Section 2 we have evoked the connection of G_m with the permanental processes. This is an interesting line of research that we plan to follow in the future mimicking the approach described in Møller and Rubak (2010).

About the second issue, we have showed that an inferential approach based on the pairwise likelihood could be a solution for estimating the unknown parameter. On the other hand probabilities of multivariate events could be evaluated by Monte Carlo method since the random processes can be quickly simulated. However our solution to the conditional prediction, based on a linear predictor, is limited and deserves further consideration even if in our real data application has been performed well.

Acknowledgement

Partial support was provided by FONDECYT grant 1160280, Chile for Moreno Bevilacqua, by grant CONICYT-PCHA, Doctorado Nacional, 2015-21150156 for Christian Caamaño and by the IRIDE program of the Ca' Foscari University of Venice for Carlo Gaetan.

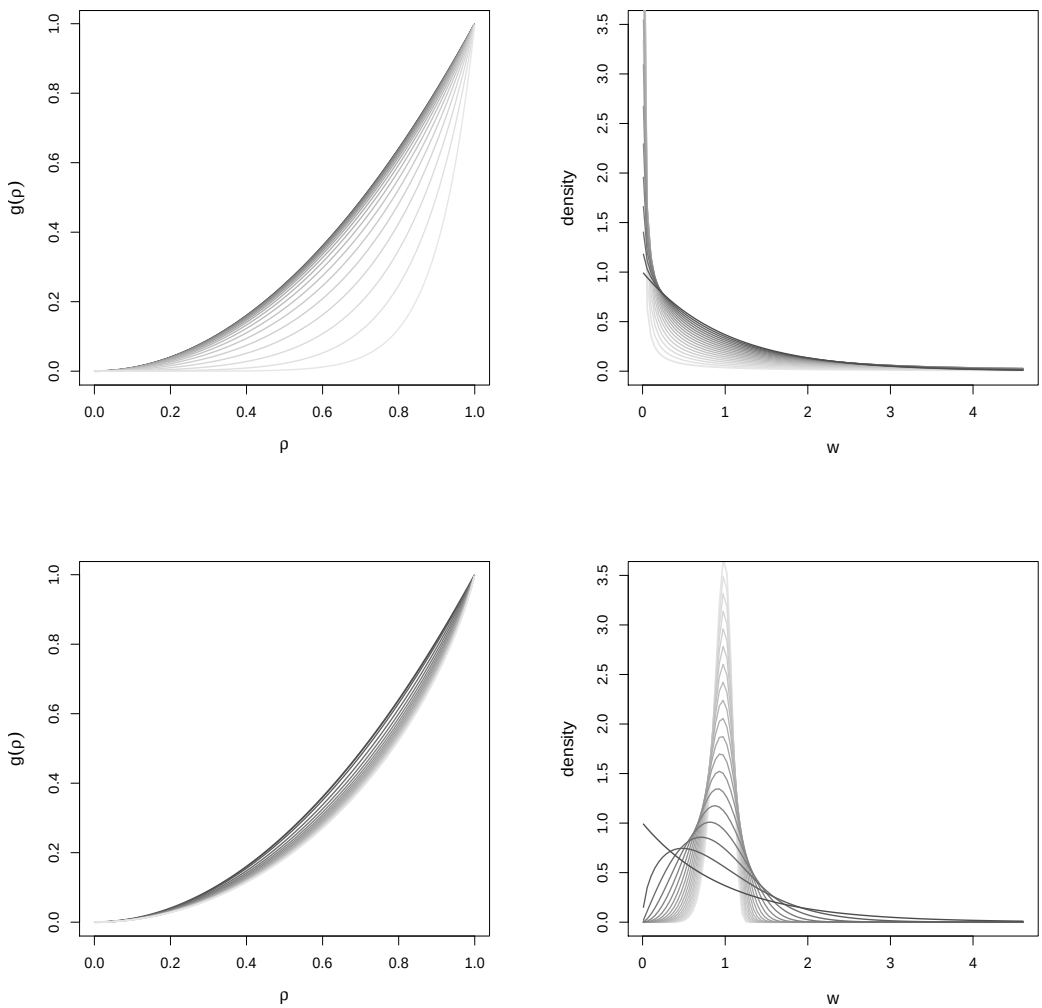


Figure 1: Plot of the function $g(\rho)$ and the associated Weibull marginal density, according different value of κ. Top line: $\kappa = 0.10, 0.15, \ldots, 1$ ($\kappa = 1$ corresponds to the black line). Bottom line $\kappa = 1, 1.5, \ldots, 10$ ($\kappa = 1$ corresponds to the black line).

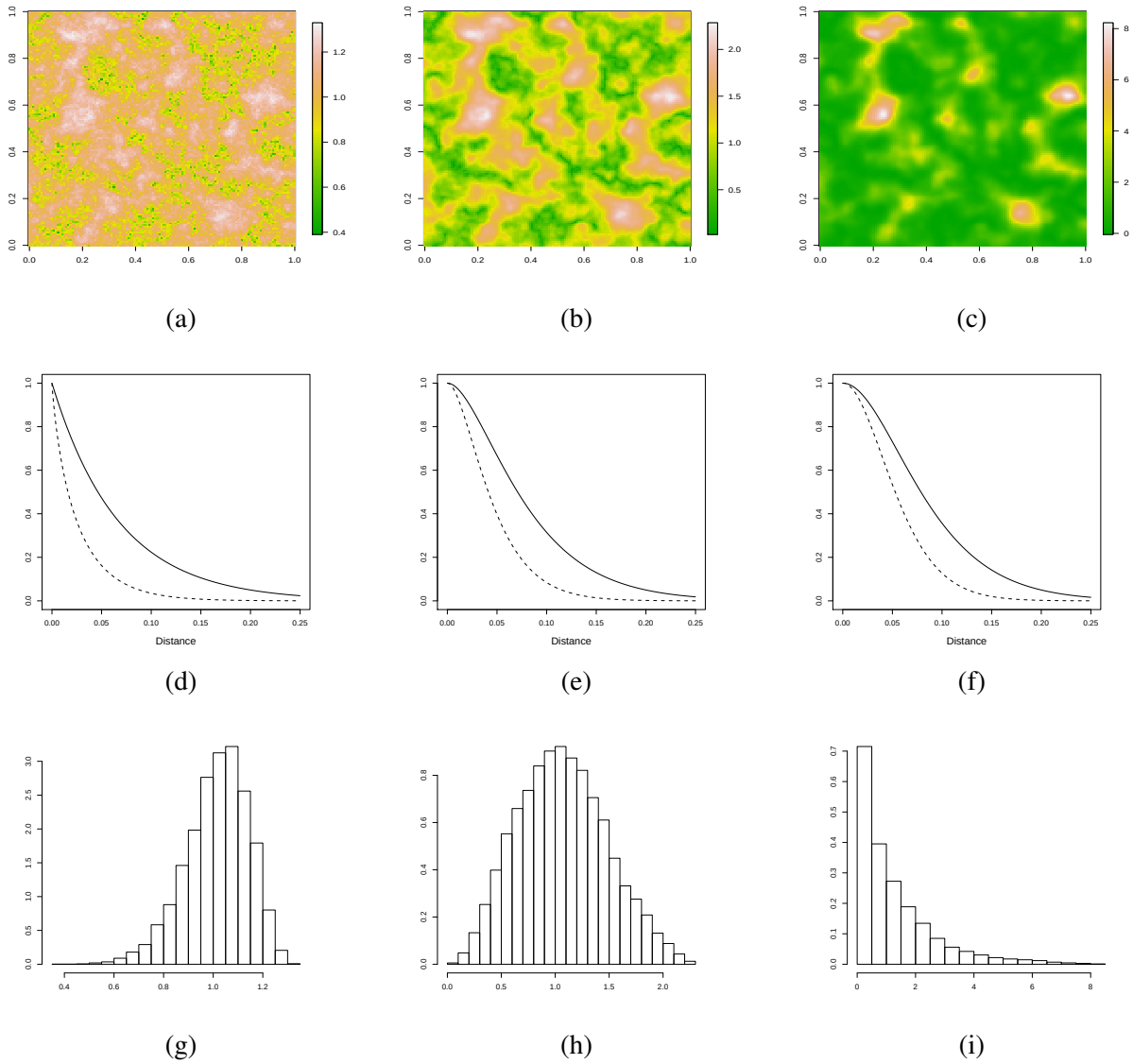


Figure 2: Top: realizations of a stationary random process W with $\mu(s) = 1$ and correlation given under the image. Centre: comparison between $\rho(h) = \mathcal{W}(\|h\|; \alpha, \psi, 4.5)$ (solid line) and the corresponding $\rho_W(h)$ (dashed line) for different values of α, κ ψ and ($\kappa = 10, 3, 1$, $\psi = 0.5, 1.5, 2.5$ from left to right). Bottom: histograms of the observations.

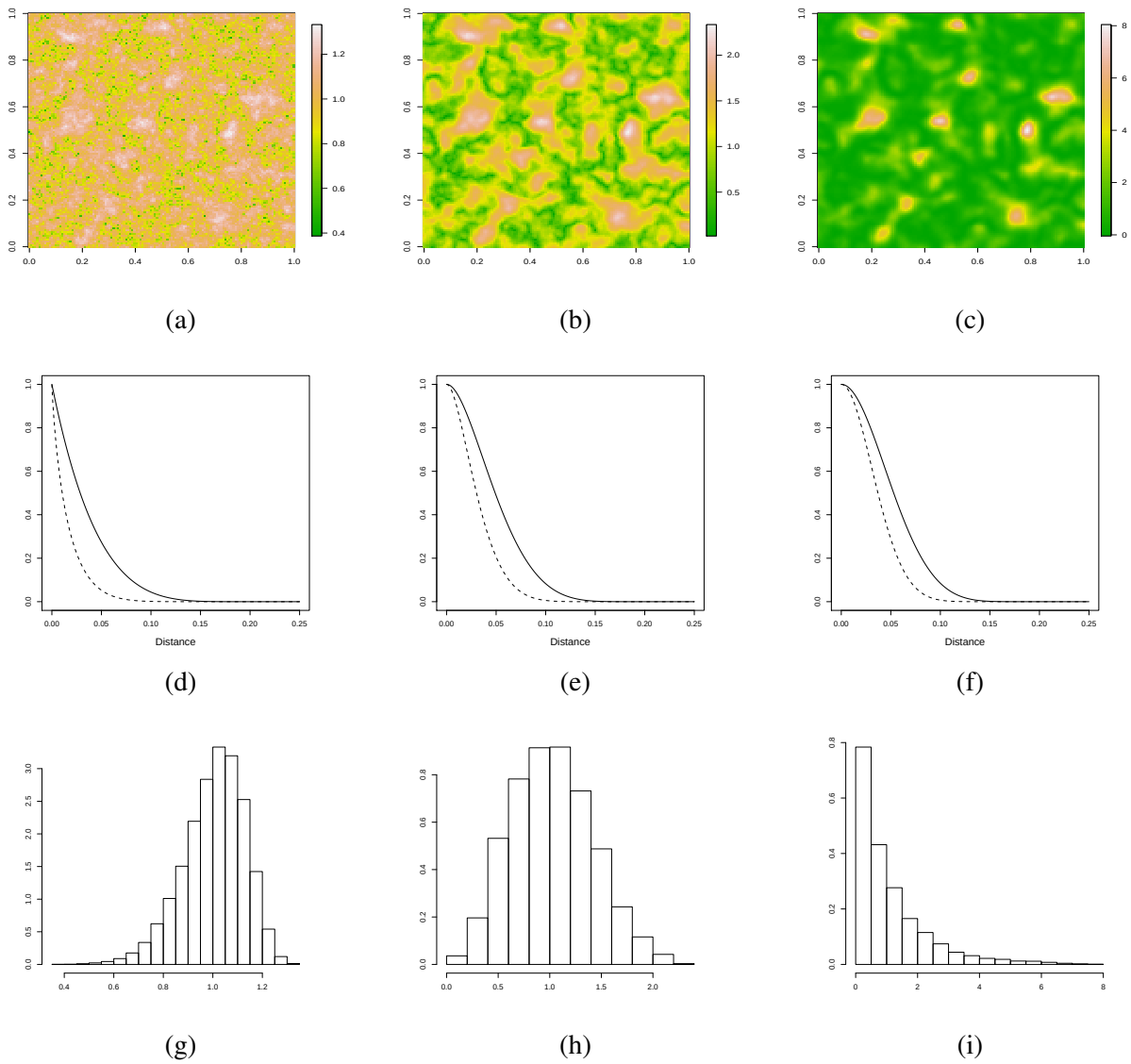


Figure 3: Top: realizations of a stationary random process W with $\mu(s) = 1$ and correlation given under the image. Centre: comparison between $\rho(h) = \mathcal{GW}(\|h\|; 0.2, \psi, 4.5)$ (solid line) and the corresponding $\rho_W(h)$ (dashed line) for different values of ψ and κ ($\psi = 0, 1, 2$, $\kappa = 10, 3, 1$ from left to right). Bottom: histograms of the observations.

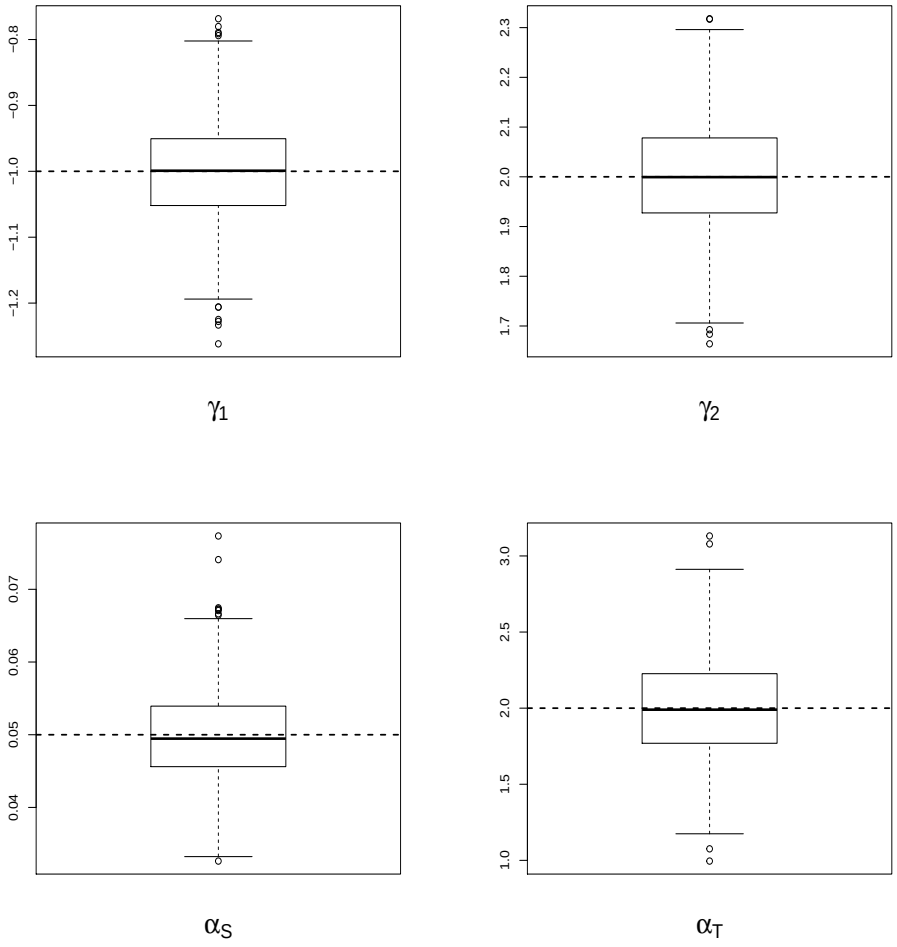


Figure 4: Boxplots of the estimates obtained from 1 000 Monte Carlo runs, for the random process G, with 'parent' spatio-temporal correlation (16) and shape parameter $m = 1$.

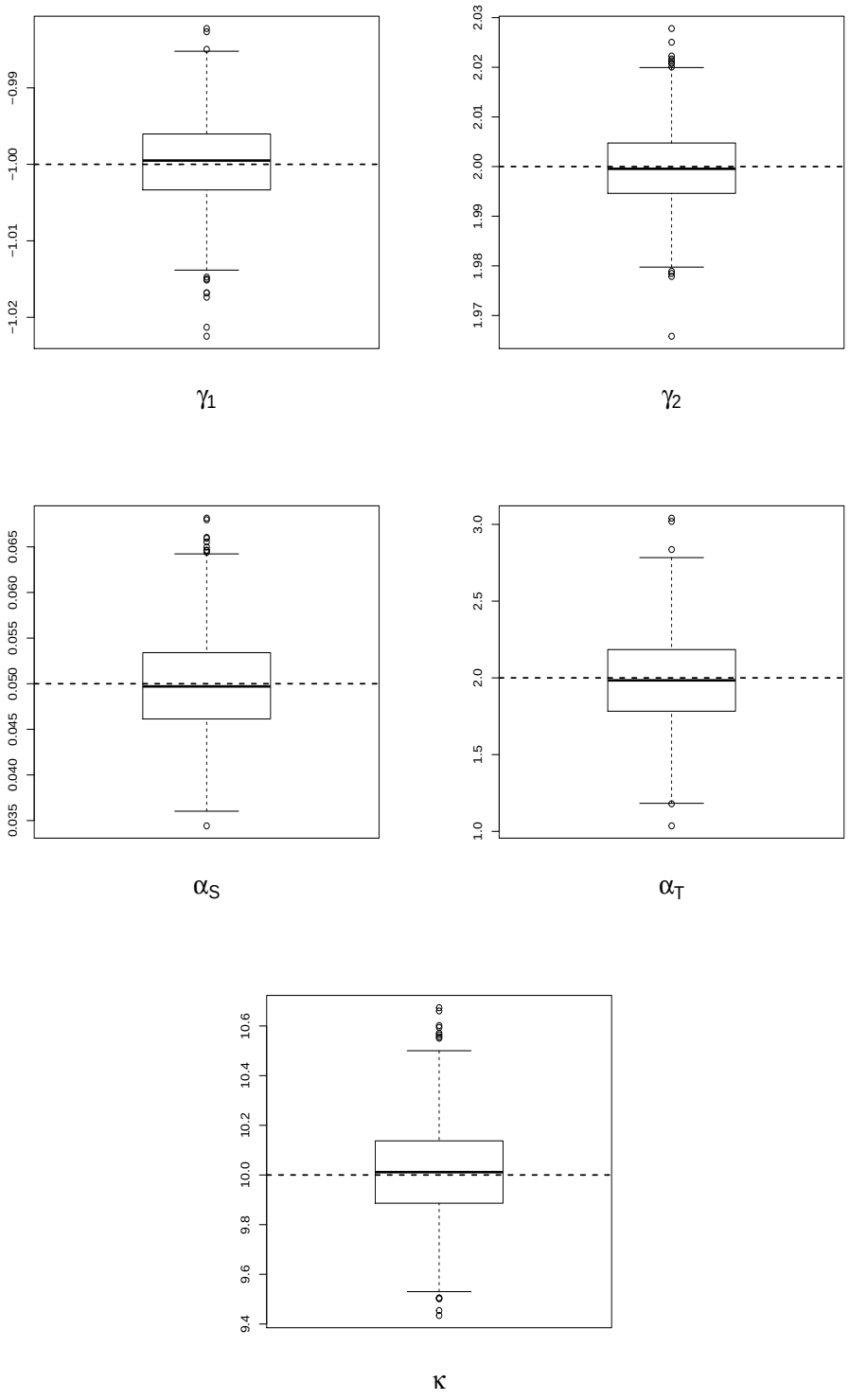
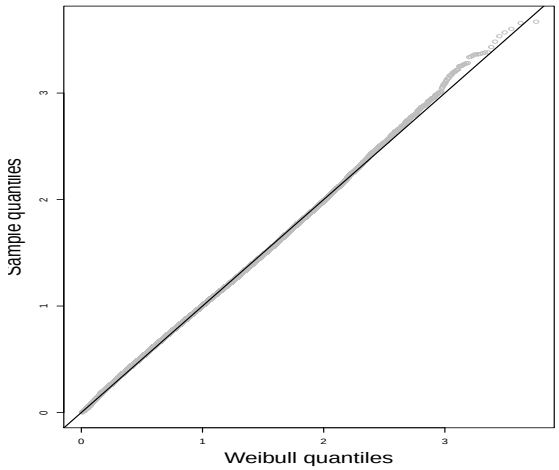
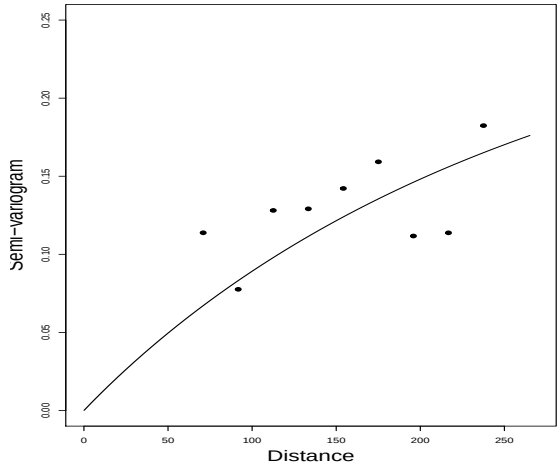


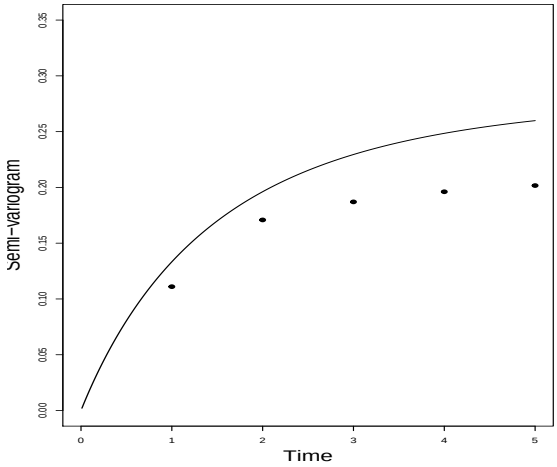
Figure 5: Boxplots of the estimates obtained from 1 000 Monte Carlo runs, for the random process W, with 'parent' spatio-temporal correlation (16) and shape parameter $\kappa = 10$.



(a)



(b)



(c)

Figure 6: (a) Q-Q plot of the residuals against the estimated quantiles of a Weibull distribution. Empirical semi-variogram (dotted points) of the residuals and estimated theoretical semi-variogram (solid line) along the spatial distance (b) and the time-lag (c)

	RPT	VAL	ROS	KIL	SHA	BIR	DUB	CLA	MUL	CLO	BEL	MAL	Total
Naïve	5.551	4.989	5.051	3.404	4.539	3.663	4.289	4.211	3.859	4.246	5.499	6.230	4.711
$\alpha_{NS} = 0$													
Gamma	4.859	4.375	4.366	**3.437**	4.003	**3.489**	3.815	**3.791**	**3.450**	**3.813**	4.900	5.827	**4.234**
Weibull	4.839	**4.364**	**4.319**	3.611	**4.000**	3.599	**3.799**	3.812	3.464	3.824	4.924	5.949	4.265
Log-Gaussian	**4.773**	4.398	4.322	3.553	4.064	3.638	3.885	3.914	3.593	3.939	**4.847**	**5.757**	4.268
$\alpha_{NS} = 0.5$													
Gamma	4.854	4.376	4.353	**3.489**	4.006	**3.526**	3.809	**3.802**	**3.458**	**3.821**	4.905	5.842	**4.243**
Weibull	4.837	**4.367**	**4.313**	3.670	**4.004**	3.639	**3.796**	3.823	3.475	3.833	4.931	5.971	4.278
Log-Gaussian	**4.762**	4.405	4.296	3.675	4.081	3.730	3.879	3.948	3.627	3.970	**4.853**	**5.782**	4.293
$\alpha_{NS} = 1$													
Gamma	4.849	4.379	4.343	**3.544**	4.011	**3.565**	3.805	**3.815**	**3.469**	**3.830**	4.910	5.857	**4.253**
Weibull	4.834	**4.370**	4.307	3.736	**4.009**	3.685	**3.794**	3.835	3.486	3.844	4.938	5.991	4.292
Log-Gaussian	**4.753**	4.415	**4.283**	3.815	4.105	3.836	3.882	3.989	3.671	4.010	**4.866**	**5.818**	4.326

Table 4: Root-mean-square-error (RMSE) for one-day ahead predictors of the wind speed at the meteorological stations during the test period (1971-1978). In bold we report the predictor with the lowest RMSE.

	RPT	VAL	ROS	KIL	SHA	BIR	DUB	CLA	MUL	CLO	BEL	MAL	Total
Naïve	4.301	3.834	3.868	2.543	3.465	2.833	3.311	3.254	3.015	3.266	4.269	4.866	3.647
$\alpha_{NS} = 0$													
Gamma	3.771	**3.448**	3.381	**2.816**	**3.148**	**2.816**	**3.029**	**3.039**	**2.759**	**3.043**	3.841	4.551	**3.304**
Weibull	3.752	3.458	**3.339**	3.009	3.163	2.923	3.030	3.078	2.782	3.084	3.866	4.632	3.343
Log-Gaussian	**3.734**	3.478	3.390	2.876	3.222	2.910	3.097	3.126	2.882	3.154	**3.824**	**4.492**	3.349
$\alpha_{NS} = 0.5$													
Gamma	3.766	**3.450**	3.371	**2.867**	**3.153**	**2.847**	**3.025**	**3.049**	**2.767**	**3.052**	3.845	4.561	**3.313**
Weibull	3.749	3.460	**3.332**	3.064	3.168	2.957	3.029	3.088	2.791	3.094	3.872	4.648	3.354
Log-Gaussian	**3.720**	3.483	3.364	2.990	3.240	2.988	3.096	3.156	2.910	3.182	**3.826**	**4.506**	3.372
$\alpha_{NS} = 1$													
Gamma	3.761	**3.452**	3.362	**2.920**	**3.159**	**2.879**	**3.023**	**3.060**	**2.776**	**3.062**	3.849	4.571	**3.323**
Weibull	3.746	3.462	**3.327**	3.115	3.174	2.989	3.027	3.096	2.801	3.104	3.877	4.663	3.368
Log-Gaussian	**3.709**	3.490	3.349	3.120	3.263	3.078	3.103	3.192	2.944	3.218	**3.834**	**4.532**	3.403

Table 5: Mean absolute error (MAE) for one-day ahead predictors of the wind speed at the meteorological stations during the test period (1971-1978). In bold we report the predictor with the lowest MAE.

Appendix

Proposition 1. *The $(a,b)-th$ product moment of any pairs $W_1 := W(s_1)$ and $W_2 := W(s_2)$ is given by*

$$\mathrm{E}(W_1^a\, W_2^b) \;=\; \frac{\mu_1^a \mu_2^b}{\Gamma\left(1+1/\kappa\right)^{a+b}} \Gamma\left(1+a/\kappa\right) \Gamma\left(1+b/\kappa\right) {}_2F_1\left(-a/\kappa, -b/\kappa; 1; \rho^2\right) \quad (17)$$

where $\rho = \rho(h), \quad h = s_1 - s_2$

Proof. Note that equation (6) can be rewritten in terms of the hypergeometric function using the identity

$${}_0F_1(;b;x) = \Gamma(b) x^{(1-b)/2} I_{b-1}(2\sqrt{x}).$$

Then by definition of product moments and using series expansion of hypergeometric function ${}_0F_1$, we have:

$$\begin{aligned}
\mathrm{E}(W_1^a\, W_2^b) &= \frac{\kappa^2 \Gamma\left(1+1/\kappa\right)^{2\kappa}}{(\mu_1\mu_2)^\kappa (1-\rho^2)} \int_0^\infty \int_0^\infty u^{\kappa+a-1} v^{\kappa+b-1} \exp\left\{-\frac{\Gamma\left(1+1/\kappa\right)^\kappa}{(1-\rho^2)} \left[\left(\frac{u}{\mu_1}\right)^\kappa + \left(\frac{v}{\mu_2}\right)^\kappa\right]\right\} \\
&\quad \times {}_0F_1\left(1; \frac{\rho^2 (uv)^\kappa \Gamma\left(1+1/\kappa\right)^{2\kappa}}{(\mu_1\mu_2)^\kappa (1-\rho^2)^2}\right) dudv \\
&= \frac{\kappa^2 \Gamma\left(1+1/\kappa\right)^{2\kappa}}{(\mu_1\mu_2)^\kappa (1-\rho^2)} \sum_{m=0}^\infty \frac{1}{m!\,(1)_m} \left(\frac{\rho^2 \Gamma\left(1+1/\kappa\right)^{2\kappa}}{(\mu_1\mu_2)^\kappa (1-\rho^2)^2}\right)^m \\
&\quad \times \int_0^\infty \int_0^\infty \exp\left\{-\frac{\Gamma\left(1+1/\kappa\right)^\kappa}{(1-\rho^2)} \left[\left(\frac{u}{\mu_1}\right)^\kappa + \left(\frac{v}{\mu_2}\right)^\kappa\right]\right\} dudv \\
&= \frac{\kappa^2 \Gamma\left(1+1/\kappa\right)^{2\kappa}}{(\mu_1\mu_2)^\kappa (1-\rho^2)} \sum_{m=0}^\infty \frac{I(m)}{m!\,(1)_m} \left(\frac{\rho^2 \Gamma\left(1+1/\kappa\right)^{2\kappa}}{(\mu_1\mu_2)^\kappa (1-\rho^2)^2}\right)^m \qquad (18)
\end{aligned}$$

Using Fubini's Theorem and (3.381.4) in Gradshteyn and Ryzhik (2007), we obtain

$$\begin{aligned}
I(\kappa) &= \int_0^\infty u^{\kappa+a+m\kappa-1} \exp\left\{-\frac{\Gamma\left(1+1/\kappa\right)^\kappa}{(1-\rho^2)} \left(\frac{u}{\mu_1}\right)^\kappa\right\} du \\
&\quad \times \int_0^\infty v^{\kappa+b+m\kappa-1} \exp\left\{-\frac{\Gamma\left(1+1/\kappa\right)^\kappa}{(1-\rho^2)} \left(\frac{v}{\mu_2}\right)^\kappa\right\} dv \\
&= \kappa^{-2} \Gamma\left(1+a/\kappa+m\right) \Gamma\left(1+b/\kappa+m\right) \\
&\quad \times \left(\frac{\mu_1^\kappa (1-\rho^2)}{\Gamma\left(1+1/\kappa\right)^\kappa}\right)^{1+a/\kappa+m} \left(\frac{\mu_2^\kappa (1-\rho^2)}{\Gamma\left(1+1/\kappa\right)^\kappa}\right)^{1+b/\kappa+m} \qquad (19)
\end{aligned}$$

Combining equations (18) and (19), we obtain

$$\begin{aligned}\mathrm{E}(W_1^a\, W_2^b) &= \frac{\mu_1^a \mu_2^b (1-\rho^2)^{1+(a+b)/\kappa}\Gamma\left(1+a/\kappa\right)\Gamma\left(1+b/\kappa\right)}{\Gamma\left(1+1/\kappa\right)^{a+b}} \\ &\quad \times {}_2F_1\left(1+a/\kappa, 1+b/\kappa; 1; \rho^2\right)\end{aligned}$$

Finally, using Euler transformation, we obtain (17). □

Proposition 2. *The Weibull random field (5) is asymptotically independent i.e.*

$$\lim_{w\to\infty} \frac{\Pr(W_1 > w, W_2 > w)}{\Pr(W_2 > w)} = 0$$

for any pairs $W_1 := W(s_1)$ *and* $W_2 := W(s_2)$

Proof. Denote $W_i = c_i X_i^{1/\kappa}$ and $x_i = (w_i/c_i)^\kappa$. We have

$$\begin{aligned}\Pr(W_1 > w_1, W_2 > w_2) &= \Pr(X_1 > x_1, X_2 > x_2) \\ &= \frac{1}{1-\rho^2}\int_{x_1}^{\infty}\int_{x_2}^{\infty} \exp\left(-\frac{u+v}{1-\rho^2}\right) I_o\left(\frac{2\sqrt{uv\rho^2}}{1-\rho^2}\right) dudv\end{aligned}$$

Since $I_0(z) \sim (2\pi)^{-1/2} z^{-1/2}\exp(z)$ for $z \to \infty$ then

$$\begin{aligned}\Pr(X_1 > x_1, X_2 > x_2) &\sim \frac{1}{\sqrt{2\pi}(1-\rho^2)}\int_{x_1}^{\infty}\int_{x_2}^{\infty} \exp\left(-\frac{u+v}{1-\rho^2}\right)\left(\frac{2\sqrt{uv\rho^2}}{1-\rho^2}\right)^{-1/2} \\ &\quad \times \exp\left(\frac{2\sqrt{uv\rho^2}}{1-\rho^2}\right) dudv \\ &= \frac{1}{\sqrt{2\pi}(1-\rho^2)}\int_{x_1}^{\infty}\int_{x_2}^{\infty} \left(\frac{2\sqrt{uv\rho^2}}{1-\rho^2}\right)^{-1/2} \\ &\quad \times \exp\left(\frac{2\sqrt{uv\rho^2}}{1-\rho^2} - \frac{u+v}{1-\rho^2}\right) dudv \\ &= \frac{1}{\sqrt{2\pi}(1-\rho^2)}\left(\frac{2\sqrt{\rho^2}}{1-\rho^2}\right)^{-1/2}\int_{x_1}^{\infty}\int_{x_2}^{\infty} (uv)^{-1/4} \\ &\quad \times \exp\left(\frac{2\sqrt{uv\rho^2}}{1-\rho^2} - \frac{u+v}{1-\rho^2}\right) dudv \\ &= \frac{1}{2\sqrt{\pi}(1-\rho^2)^{3/2}|\rho|^{1/2}} A(x_1, x_2)\end{aligned}$$

with $A(x_1, x_2) := \int_{x_1}^{\infty}\int_{x_2}^{\infty} (uv)^{-1/4}\exp\left(\frac{2\sqrt{uv\rho^2}}{1-\rho^2} - \frac{u+v}{1-\rho^2}\right) dudv.$

But $\sqrt{uv} \le (u+v)/2$, so

$$\frac{2\sqrt{uv\rho^2}}{1-\rho^2} - \frac{u+v}{1-\rho^2} \le \frac{(u+v)|\rho|}{1-\rho^2} - \frac{u+v}{1-\rho^2} = -\frac{(u+v)(1-|\rho|)}{1-\rho^2} = -\frac{(u+v)}{1+|\rho|}$$

and

$$\begin{aligned} A(x_1,x_2) &\leq \int_{x_1}^{\infty}\int_{x_2}^{\infty}(uv)^{-1/4}\exp\left(-\frac{u+v}{1+|\rho|}\right)dudv \\ &= \int_{x_1}^{\infty}u^{-1/4}\exp\left(-\frac{u}{1+|\rho|}\right)du\int_{x_2}^{\infty}v^{-1/4}\exp\left(-\frac{v}{1+|\rho|}\right)dv \end{aligned}$$

Denote $\Gamma(s,z) := \int_z^{\infty} u^{s-1}\exp(-u)du$ the complementary gamma function. We have

$$\int_z^{\infty} u^{s-1}\exp(-au)du = \frac{\Gamma(s,az)}{a^s}, \qquad \text{for } a>0$$

We know that $\Gamma(s,z) = \exp(-z)z^{s-1}(1+O(z^{-1}))$ as $|z|\to\infty$, then

$$\begin{aligned} A(x_1,x_2) &\leq \int_{x_1}^{\infty}u^{-1/4}\exp\left(-\frac{u}{1+|\rho|}\right)du\int_{x_2}^{\infty}v^{-1/4}\exp\left(-\frac{v}{1+|\rho|}\right)dv \\ &= (1+|\rho|)^2\Gamma\left(3/4,\frac{x_1}{1+|\rho|}\right)\Gamma\left(3/4,\frac{x_2}{1+|\rho|}\right) \end{aligned}$$

For $x\to\infty$, we have

$$\begin{aligned} A(x,x) &\leq (1+|\rho|)^2\left\{\Gamma\left(3/4,\frac{x}{1+|\rho|}\right)\right\}^2 \\ &\leq (1+|\rho|)^2\exp\left(-2\frac{x}{1+|\rho|}\right)\left(\frac{x}{1+|\rho|}\right)^{-1/2}(1+O(x^{-1})) \end{aligned}$$

and

$$\begin{aligned} \Pr(X_1>x,X_2>x) &\leq \frac{1}{2\sqrt{\pi}(1-\rho^2)^{3/2}|\rho|^{1/2}}(1+|\rho|)^2\exp\left(-2\frac{x}{1+|\rho|}\right) \\ &\quad \left(\frac{x}{1+|\rho|}\right)^{-1/2}(1+O(x^{-1})) \end{aligned}$$

Since $\Pr(X_i>x)=\exp(-x)$, $i=1,2$ we have

$$\lim_{w\to\infty}\frac{\Pr(X_1>(w/c_1)^{\kappa},X_2>(w/c_2)^{\kappa})}{\Pr(X_1>(w/c_2)^{\kappa})} = \lim_{x\to\infty}\frac{\Pr(W_1>w,W_2>w)}{\Pr(W_2>x)} = \chi = 0.$$

□

References

Allcroft, D. and C. Glasbey (2003). A latent Gaussian Markov random field model for spatio-temporal rainfall disaggregation. *Journal of the Royal Statistical Society: Series C 52*, 487–498.

Banerjee, S., B. P. Carlin, and A. E. Gelfand (2014). *Hierarchical Modeling and Analysis for Spatial Data* (second ed.). Boca Raton: FL: CRC Press.

Bellier, E., P. Monestiez, and C. Guinet (2010). Geostatistical modelling of wildlife populations: a non-stationary hierarchical model for count data. In P. Atkinson and C. Lloyd (Eds.), *geoENV VII Geostatistics for Environmental Applications. Springer*, pp. 1–12.

Bevilacqua, M., T. Faouzi, R. Furrer, and E. Porcu (2018). Estimation and prediction using generalized Wendland functions under fixed domain asymptotics. *The Annals of Statistics*. to appear.

Bevilacqua, M. and C. Gaetan (2015). Comparing composite likelihood methods based on pairs for spatial Gaussian random fields. *Statistics and Computing 25*, 877–892.

Bevilacqua, M., C. Gaetan, J. Mateu, and E. Porcu (2012). Estimating space and space-time covariance functions for large data sets: a weighted composite likelihood approach. *Journal of the American Statistical Association 107*, 268–280.

Coles, S., J. Heffernan, and J. Tawn (1999). Dependence measures for extreme value analyses. *Extremes 2*, 339–365.

Cressie, N. (1993). *Statistics for Spatial Data* (revised ed.). New York: Wiley.

Cressie, N. and C. Wikle (2011). *Statistics for Spatio-Temporal Data.* New York: Wiley.

Davis, R. and C.-Y. Yau (2011). Comments on pairwise likelihood in time series models. *Statistica Sinica 21*, 255–277.

de Luna, X. and M. Genton (2005). Predictive spatio-temporal models for spatially sparse environmental data. *Statistica Sinica 15*, 547–568.

De Oliveira, V. (2006). On optimal point and block prediction in log-Gaussian random fields. *Scandinavian Journal of Statistics 33*, 523–540.

De Oliveira, V. (2014). Poisson kriging: A closer investigation. *Spatial Statistics 7*, 1–20.

De Oliveira, V., B. Kedem, and D. A. Short (1997). Bayesian prediction of transformed Gaussian random fields. *Journal of the American Statistical Association 92*, 1422–1433.

Diggle, P., J. Tawn, and R. Moyeed (1998). Model-based geostatistics. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 47*, 299–350.

Downton, F. (1970). Bivariate exponential distributions in reliability theory. *Journal of the Royal Statistical Society: Series B 32*, 408–417.

Eisenbaum, N. and H. Kaspi (2006). A characterization of the infinitely divisible squared gaussian processes. *The Annals of Probability 34*, 728–742.

Eisenbaum, N. and H. Kaspi (2009). On permanental processes. *Stochastic Processes and their Applications 119*, 1401 – 1415.

Galanis, G., D. Hayes, G. Zodiatis, P. Chu, Y.-H. Kuo, and G. Kallos (2012). Wave height characteristics in the Mediterranean Sea by means of numerical modeling, satellite data, statistical and geometrical techniques. *Marine Geophysical Research 33*, 1–15.

Gneiting, T. (2002a). Compactly supported correlation functions. *Journal of Multivariate Analysis 83*, 493–508.

Gneiting, T. (2002b). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association 97*, 590–600.

Gneiting, T. (2013). Strictly and non-strictly positive definite functions on spheres. *Bernoulli 19*, 1327–1349.

Gradshteyn, I. and I. Ryzhik (2007). *Table of Integrals, Series, and Products* (eight ed.). Cambridge, MA: Academic Press.

Haslett, J. and A. Raftery (1989). Space-time modelling with long-memory dependence: assessing irelands wind-power resource (with discussion). *Applied Statistics 38*, 1–50.

Heaton, M. J., A. Datta, A. Finley, R. Furrer, R. Guhaniyogi, F. Gerber, R. B. Gramacy, D. Hammerling, M. Katzfuss, F. Lindgren, D. W. Nychka, F. Sun, and A. Zammit-Mangion (2017). Methods for analyzing large spatial data: A review and comparison. *ArXiv e-prints*.

Kazianka, H. and J. Pilz (2010). Copula-based geostatistical modeling of continuous and discrete data including covariates. *Stochastic Environmental Research and Risk Assessment 24*, 661–673.

Kibble, W. F. (1941). A two-variate gamma type distribution. *Sankhyā: The Indian Journal of Statistics (1933-1960) 5*, 137–150.

Krishnamoorthy, A. S. and M. Parthasarathy (1951). A multivariate Gamma-type distribution. *The Annals of Mathematical Statistics 22*, 549–557.

Lindgren, F., H. Rue, and J. Lindström (2011). An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach. *Journal of the Royal Statistical Society: Series B 73*, 423–498.

Lindsay, B. (1988). Composite likelihood methods. *Contemporary Mathematics 80*, 221–239.

Masarotto, G. and C. Varin (2012). Gaussian copula marginal regression. *Electronic Journal of Statistics 6*, 1517–1549.

Matérn, B. (1986). *Spatial Variation: Stochastic Models and their Applications to Some Problems in Forest Surveys and Other Sampling Investigations* (2nd ed.). Heidelberg: Springer.

Møller, J. and E. Rubak (2010). A model for positively correlated count variables. *International Statistical Review / Revue Internationale de Statistique 78*, 65–80.

Neykov, N. M., P. N. Neytchev, and W. Zucchini (2014). Stochastic daily precipitation model with a heavy-tailed component. *Natural Hazards and Earth System Sciences 14*, 2321–2335.

Opitz, T. (2016). Modeling asymptotically independent spatial extremes based on Laplace random fields. *Spatial Statistics 16*, 1 – 18.

Porcu, E., M. Bevilacqua, and M. Genton (2016). Spatio-temporal covariance and cross covariance functions of the great circle distance on a sphere. *Journal of the American Statistical Association 111*, 888–898.

Pryor, S. C. and R. J. Barthelmie (2010). Climate change impacts on wind energy: a review. *Renewable and Sustainability Energy Reviews 14*, 430–437.

Royen, T. (2004). Multivariate Gamma distributions II. In *Encyclopedia of Statistical Sciences*, pp. 419–425. New York: John Wiley & Sons.

Sibuya, M. (1960). Bivariate extreme statistics. *Annals of the Institute of Statistical Mathematics 11*, 195–210.

Stein, M. (1999). *Interpolation of Spatial Data. Some Theory of Kriging*. New York: Springer-Verlag.

Stein, M. (2005). Space-time covariance functions. *Journal of the American Statistical Association 100*, 310–321.

Stein, M., Z. Chi, and L. Welty (2004). Approximating likelihoods for large spatial data sets. *Journal of the Royal Statistical Society: Series B 66*, 275–296.

Varin, C., N. Reid, and D. Firth (2011). An overview of composite likelihood methods. *Statistica Sinica 21*, 5–42.

Varin, C. and P. Vidoni (2005). A note on composite likelihood inference and model selection. *Biometrika 52*, 519–528.

Vere-Jones, D. (1997). Alpha-permanents and their applications to multivariate Gamma, negative binomial and ordinary binomial distributions. *New Zealand Journal of Mathematics 26*, 125–149.

Wadsworth, J. and J. Tawn (2012). Dependence modelling for spatial extremes. *Biometrika 99*, 253–272.

Warren, D. (1992). A multivariate gamma distribution arising from a Markov model. *Stochastic Hydrology and Hydraulics 6*, 183–190.