Skew-normality for climatic data and dispersal models for plant epidemiology: when application fields drive spatial statistics

D. Allard and S. Soubeyrand

Biostatistique et Processus Spatiaux (BioSP), INRA, Site Agroparc, 84914 Avignon, France.

Running title: Skew-normality and dispersal models

Corresponding author: D. Allard `allard@avignon.inra.fr`

Summary

Developments in spatial statistics have a long standing tradition of being drawn by specific applications. In this paper we illustrate this point by showing how research driven by two specific needs, namely sensitivity of agriculture to climate change and plant epidemiology has lead to new developments in two branches of spatial statistics, respectively random field theory and spatio-temporal point processes. In a first part, it is shown how skew-normal distributions allow to define skew-normal random fields and how they can be used to build a weather generator. In a second part, models for dispersal of propagules that can account for rough anisotropies and a model for group dispersal are presented. The distribution of the furthest dispersed point, a quantity of great interest for assessing expansion speed, is derived.

Keywords: anisotropy; Gaussian random field; group dispersal; skew-normal distribution spatial point process; weather generator

1 Introduction

In the introductory chapter of the "Handbook of Spatial Statistics" (Gelfand et al., 2010) summarizing the historical development of spatial statistics as a recognizable sub-branch of statistics, P. Diggle points out that spatial statistics has this specificity that it was pioneered

by scientists working in different areas of applications (agriculture with R. Fisher, mining industry with G. Matheron and forestry with B. Matérn) before being brought to the statistical community by two seminal papers by Besag (1974) and Ripley (1977). Since then, spatial statistics has been enriched with many important contributions, in particular in the field of asymptotic theory for geostatistics (Stein, 1999), model based geostatistics (Diggle et al., 1998; Diggle and Ribeiro, 2007) and hierarchical modeling (Banerjee et al., 2004), multivariate modelling (Wackernagel, 2003), spatio-temporal processes (Cressie and Wikle, 2011) and marked point processes (Schlather et al., 2004). In terms of applications, from the traditional geoscience area, spatial statistics has spread towards any science collecting and analyzing spatial data, e.g. ecology, disease mapping, epidemiology, geography economics and astronomy.

Developments in spatial statistics have continued to be drawn by specific applications. It is our aim to illustrate this point from our own research perspective. We are statisticians members of BioSP, a statistical and mathematical group working on developments and applications of spatio-temporal processes in the French National Agricultural Research Institute. Our research must thus bring value to the analysis of the data brought by our collaborators. Yet, at the same time, we are also motivated by the development of new spatio-temporal statistical models and methods. We will thus show how research driven by specific needs, namely sensitivity of agriculture to climate change and plant epidemiology and ecology has lead to new developments in two branches of spatial statistics, respectively random field theory and spatio-temporal point processes. In Section 2 we will present a class of random fields for modelling skewed data such as temperature, humidity or radiation. It will also be shown how skew-normal distributions are used for building a weather generator. In Section 3, it will be shown how dispersal studies in plant epidemiology and ecology lead to build new anisotropic and random intensity functions for spatial point processes. Because we are interested in dispersal of propagules, unusual characteristics of spatial point processes are

investigated, e.g. the distribution of the furthest point of a point process which is related with the expansion speed of a disease or a plant population.

2 Skew-normal processes for environmental and climatic data

The ubiquitous assumption of normality for modeling spatial and spatio-temporal data can be understood for many reasons. A major one is that the multivariate normal distribution is completely characterized by its first two moments. In addition, the stability of multivariate normal distribution under summation and conditioning offers tractability and simplicity. But for a wide range of applications in environment, climate studies and ecology Gaussian spatial or spatio-temporal models cannot reasonably be fitted to the observations.

How to best integrate spatial or temporal skewness (when necessary) is still an open debate and many different approaches can be proposed. The classical strategy, well known in geostatistics (Wackernagel, 2003; Chilès and Delfiner, 1999), consists of transforming a spatial field into a Gaussian one. Besides the problem of choosing the adequate transform, there are a few difficulties associated with this method. Another strategy is to assume that the random field follows a specific process. The family of the complete skew-normal (CSN) distribution is an extension of the Gaussian distribution which admits skewness while at the same time retaining most of its tractability: it is closed under summation, marginalization and conditioning (González-Farías et al., 2004; Genton, 2004). Allard and Naveau (2007) proposed a novel way of modeling skewness for spatial data by working with this family. Karimi and Mohammadzadeh (2010) have used the CSN distribution for Bayesian spatial regression. It is however time consuming and can be difficult to implement.

We shall first make a brief presentation of the multivariate closed skew-normal distribution. We shall then show how this family can be used for modelling spatial data. We will then turn to climatic data in the view of building a weather generator.

2.1 The multivariate closed skew-normal distribution

An n-dimensional random vector $\mathbf{Y}$ is said to follow a multivariate closed skew-normal distribution, denoted by $\text{CSN}_{n,m}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}, \boldsymbol{\nu}, \boldsymbol{\Delta})$, if its density function is of the form

$$f(\mathbf{y}) = c_m \, \phi_n(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) \, \Phi_m(\mathbf{D}'(\mathbf{y} - \boldsymbol{\mu}); \boldsymbol{\nu}, \boldsymbol{\Delta}), \text{ with } c_m^{-1} = \Phi_m(\mathbf{0}; \boldsymbol{\nu}, \boldsymbol{\Delta} + \boldsymbol{D}'\boldsymbol{\Sigma}\boldsymbol{D}), \tag{1}$$

where $\boldsymbol{\mu} \in \mathbb{R}^n$, $\boldsymbol{\nu} \in \mathbb{R}^m$, are vectors of expectation, $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$ and $\boldsymbol{\Delta} \in \mathbb{R}^{m \times m}$ are covariance matrices, $\mathbf{D} \in \mathbb{R}^{n \times m}$ is a transformation matrix and $\phi_n(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ and $\Phi_n(\mathbf{y}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ are the pdf and cdf, respectively, of the n-dimensional normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, and $\mathbf{D}'$ is the transpose of the matrix $\mathbf{D}$. If $\mathbf{D} = \mathbf{0}$, the density (1) reduces to the multivariate normal one. From its moment generating function it is in theory possible to derive the expectation, variance and skewness coefficients of $\mathbf{Y}$, but these expressions are not easy to handle, even for small dimensions of m. In a spatial context, it implies that only simple structures for $\boldsymbol{\Delta}$ and $\mathbf{D}$ can be investigated. We will show below one possible parametrizations for a spatial CSN model.

The $\text{CSN}_{n,m}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}, \boldsymbol{\nu}, \boldsymbol{\Delta})$ distribution defined by (1) can be generated in the following way. Let $\mathbf{U}$ be a Gaussian vector of dimension m and consider the augmented Gaussian vector $(\mathbf{U}', \mathbf{Z}')'$ such that

$$\begin{pmatrix} \mathbf{U} \\ \mathbf{Z} \end{pmatrix} \overset{\text{d}}{=} N_{m+n}\left(\begin{pmatrix} \boldsymbol{\nu} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Delta} + \mathbf{D}'\boldsymbol{\Sigma}\boldsymbol{D} & -\mathbf{D}'\boldsymbol{\Sigma} \\ -\boldsymbol{\Sigma}\boldsymbol{D} & \boldsymbol{\Sigma} \end{pmatrix} \right), \tag{2}$$

where $\overset{\text{d}}{=}$ denotes equality in distribution. Then it is straightforward to show that, conditional on $\mathbf{U} \leq \mathbf{0}$, the random vector

$$\mathbf{Y} = \boldsymbol{\mu} + [\mathbf{Z} | \mathbf{U} \leq \mathbf{0}] \tag{3}$$

is distributed according to the $\text{CSN}_{n,m}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \mathbf{D}, \boldsymbol{\nu}, \boldsymbol{\Delta})$ distribution defined in (1). The notation $\mathbf{U} \leq \mathbf{0}$ corresponds to $U_i \leq 0$, for all $i = 1, \ldots, m$. This construction offers in theory a wide range of possible models depending on the choice of $\boldsymbol{\mu}$, $\boldsymbol{\nu}$, $\boldsymbol{\Delta}$, $\boldsymbol{\Sigma}$ and $\mathbf{D}$, but considerations seen above will limit the range of models that can be used in practice. Since the parameter

$\boldsymbol{\nu}$ in (1) is a redundant quantity in the expression $\Phi_m(\mathbf{D}'(\mathbf{y}-\boldsymbol{\mu});\boldsymbol{\nu},\boldsymbol{\Delta})$, it is set to $\boldsymbol{\nu}=\mathbf{0}$ in the rest of this paper.

2.2 Building a spatial skew-normal process

We now define a CSN random process $\{Y(s)\},\ s\in\mathcal{D}$ as a generalization of (3)

$$Y(s) \stackrel{\mathrm{d}}{:=} \mu(s) + [Z(s) \mid \mathbf{U}\leq\mathbf{0}].$$

Thus, for any n and any vector $\mathbf{Z}=(Z(s_1),\ldots,Z(s_n))'$, Equation (3) holds. The vector $\mathbf{Y}$ can also be expressed as a sum of two independent processes. Let us introduce

$$\begin{pmatrix}\mathbf{U}\\ \mathbf{V}\end{pmatrix} \stackrel{\mathrm{d}}{=} N_{m+n}\left(\begin{pmatrix}\mathbf{0}\\ \mathbf{0}\end{pmatrix}, \begin{pmatrix}\boldsymbol{\Delta}+\mathbf{D}'\boldsymbol{\Sigma}\boldsymbol{D} & \mathbf{0}\\ \mathbf{0} & \mathbf{I}_n\end{pmatrix}\right),$$

where $\mathbf{I}_n$ represents the identity matrix of size n. Then the vector $\mathbf{Z}$ is set equal to $-\mathbf{F}\mathbf{U}+\mathbf{G}^{1/2}\mathbf{V}$ with $\mathbf{F}=\boldsymbol{\Sigma}\boldsymbol{D}(\boldsymbol{\Delta}+\mathbf{D}'\boldsymbol{\Sigma}\boldsymbol{D})^{-1}$ and $\mathbf{G}=\boldsymbol{\Sigma}-\boldsymbol{\Sigma}\boldsymbol{D}(\boldsymbol{\Delta}+\mathbf{D}'\boldsymbol{\Sigma}\boldsymbol{D})^{-1}\mathbf{D}'\boldsymbol{\Sigma}$. This is a multivariate Gaussian vector with zero-mean vector and covariance $\boldsymbol{\Sigma}$ and, more importantly, the bivariate couple $(\mathbf{U}',\mathbf{Z}')'$ satisfies (2). The independence of $\mathbf{U}$ and $\mathbf{V}$ allows us to write

$$\mathbf{Y} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + \mathbf{F}[\mathbf{U} \mid \mathbf{U}\geq\mathbf{0}] + \mathbf{G}^{1/2}\mathbf{V}. \tag{4}$$

In Naveau and Allard (2004) and Allard and Naveau (2007), several parametrizations are explored. In particular, the integer m is assumed to be a known quantity, large enough in order to impose sufficient skewness. Allard and Naveau (2007) proposed that $m=n$, $\boldsymbol{\Delta}=\boldsymbol{\Sigma}$ and that $\boldsymbol{\Sigma}$ corresponds to a covariance function $C(h)$ computed at all pairs of locations s_i, s_j, with $1\leq i,j\leq n$. Such a construction corresponds to equating the hidden points which define implicitly the elements of $\mathbf{U}$ with the observed ones. This model is thus referred to as the *homotopic model*. To simplify the interpretation of the model, it is further assumed that $\mathbf{D}=\delta\mathbf{I}_n$, where $\delta\in\mathbb{R}$ is a single parameter controlling the skewness. When $\delta=0$, $\mathbf{Y}$ is independent of $\mathbf{U}$, i.e. $\mathbf{Y}$ is a Gaussian vector with expectation $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. As δ^2 increases, the vector $\mathbf{Y}$ tends to a proper truncated Gaussian vector as δ^2 increases.

With this parametrization, (4) becomes

$$\mathbf{Y} \stackrel{\mathrm{d}}{=} \boldsymbol{\mu} + \frac{\delta}{1+\delta^2}[\mathbf{U} \mid \mathbf{U} \geq \mathbf{0}] + \frac{1}{\sqrt{1+\delta^2}}\boldsymbol{\Sigma}^{1/2}\mathbf{V}. \tag{5}$$

The vector $\mathbf{Y}$ is thus a weighted sum of a sample from a stationary process with covariance function proportional to $C(h)$ and a vector $\mathbf{U}$ conditional on $\mathbf{U} \geq 0$, whose covariance function is $(1+\delta^2)C(h)$.

For computing the moments of $\mathbf{Y}$, results on moments of the truncated multivariate normal distribution given in Tallis (1961) have been used in the decomposition (5). For ease of exposition we will now assume that $\boldsymbol{\mu} = \mu\ (1, \ldots, 1)'$.

The expectation of the experimental variogram computed on the data, denoted $E[\hat{\gamma}_Y(h)]$, can be related to expressions involving second order moments of $\mathbf{U}$. Based on these moments, it is thus possible to estimate the paramerers μ and $\gamma(h) = C(0) - C(h)$ of the model in (5). The method is described in details in Allard and Naveau (2007). It consists in: (i) estimating the range and smoothness parameters of the variogram of $\mathbf{Y}$ for a given parametric family (e.g. Matérn class) based on the experimental variogram computed on the data; (ii) computing the corresponding second order quantitites involved in $E[\hat{\gamma}_Y(h)]$; (iii) estimating δ and σ^2 by equating $\hat{\gamma}_Y(h)$ to its expectation $E[\hat{\gamma}_Y(h)]$; iv) estimating μ from the theoretical expression of $E[\mathbf{Y}]$. Steps iii) and iv) correspond to a method of moment estimators.

This model and the estimation of its parameters are now illustrated in Figure 1. The top left panel shows one realization at $n = 100$ sample locations within the unit square of the spatial homotopic model with $(\mu, \sigma^2, \delta) = (1, 1, 3)$ and an exponential covariance function $C(h) = \exp\{||h||/0.1\}$. The histogram, strongly positively skewed, shows clearly that the model is capable of generating a fair degree of skewness with $\delta = 3$. The sample variogram (bottom left) indicates a spatial structure up to a distance of around 0.3, i.e. the practical range. However, the variance values corresponding to the sill in the variogram is around 0.7, smaller than the value of 1 taken by σ, in agreement with theoretical results. The estimated parameters are $(\hat{\mu}, \hat{\sigma}^2, \hat{\delta}, \hat{a}) = (1.30, 1.15, 3sh.27, 0.062)$. The theoretical variogram

corresponding to these parameters is also displayed (dashed line). The bottom right panel shows the average of 100 sample variograms with the theoretical variogram corresponding to the input parameters. The fit between sample and theoretical variograms is clear.

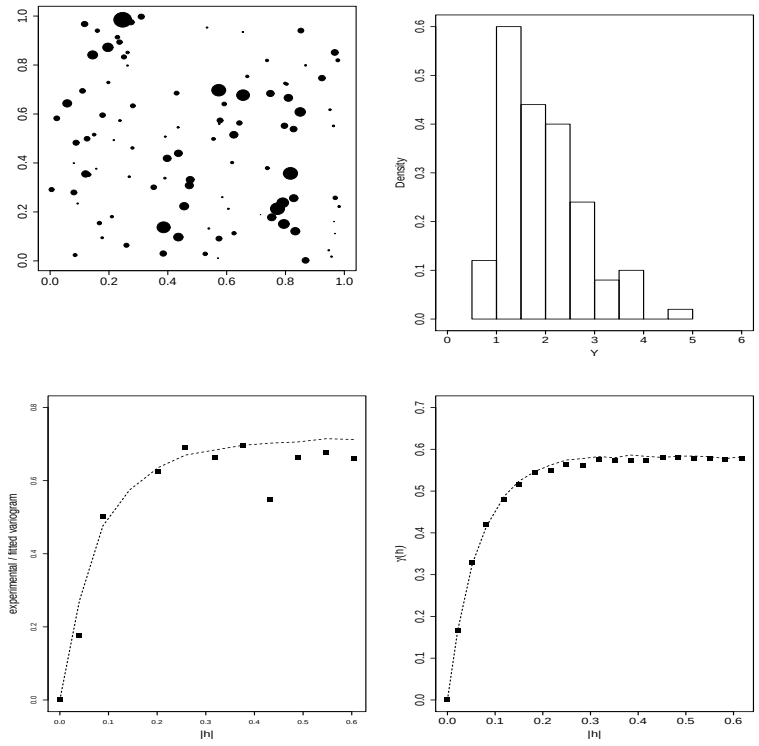

Figure 1: CSN spatial process with $\mu = 1$, $\delta = 3$ and exponential covariance function $C(h) = \exp\{-||h||/0.1\}$. Top left, simulation at $n = 100$ locations; top right, histogram. Bottom left, sample and theoretical variogram with estimated parameters; bottom right averaged sample variogram and theoretical variogram with input parameters.

This family of skewed-random fields offers the advantage of having very few parameters, while integrating the classical spatial structures used in geostatistics and adding the possibility of generating a large degree of skewness throughout a single skewness parameter. This model is easy to simulate, and we have shown that estimation of the variogram was possible and accurate. However, many computational issues remain, in particular for spatial proediction. We hope they will spur more research in the near future.

2.3 Wheater generators

Agronomical models need a large variety of daily weather data as inputs to model past, present and future variability for yields, soil carbon, etc (Brisson et al., 2003). Such daily inputs have to be simulated quickly and easily for long time periods at given locations and for given climatic conditions. According to an old aphorism, "Climate is what you expect; weather is what you get". Middle school pupils would say: "Climate tells you what clothes to buy, but weather tells you what clothes to wear". Stochastic weather generators aim at reproducing the statistical distributional properties of meteorological variables of a given climate. In other words, for given climatic parameters (mean, (co-)variability of temperature, precipitation, etc.), a weather generator will provide stochastically generated time series simulating this climate.

Conceptually, the majority of statistical weather generators can be classified into two categories. The first one consists in pooling out analog days from a database of past observations according to a given criterion and sampling from these analog days (Rajagopalan et al., 1997; Rajagopalan and Lall, 1999; Boé and Terray, 2008). An important drawback resides in the incapability of creating new time series, i.e. unobserved meteorological situations. To alleviate this undesirable feature, the second category of weather generators is based on stochastically drawing random realizations from a statistical model whose parameters have been estimated on a database of past observations (katz, 1996; Semenov and Barrow, 1997).

WACS-Gen is a stochastic parametric weather generator based on weather states and multivariate closed skew-normal distributions. The general structure of WACS-Gen, which involves mixtures of CSN distributions is first briefly described. It will then be shown how parameters are estimated from recorded series and how stochastic simulations of new series are performed. Finally, WACS-Gen will be illustrated on a climatic time series recorded from 1973 to 1992 in Colmar (France) on five daily variables: maximum temperature T_x, minimum temperature T_n, precipitation P, wind speed V and total radiation R. A more

detailed presentation of WACS-Gen can be found in Flecher et al. (2010).

2.4 General structure of WACS-Gen

Seasonal variations

A normalization pre-processing of the data is first performed in order to filter out long term and seasonal trends on each variable. The median and the average absolute deviation (mean of absolute difference between the variable and its median) are smoothed by a spline function. This smoothed median is then subtracted to the studied variable and the difference is rescaled by the smoothed average absolute deviation. Estimation of all other parameters is done on the residuals which are to be stochastically simulated. Residuals will be studied independently within the four following seasons: December-January-February (DJF), March-April-May (MAM), June-July-August (JJA) and September-October-November (SON). Precipation residuals, P, are further transformed independently for all seasons into $\tilde{P} = \Phi_1^{-1}(G(P))$ where G represents the fit by a Gamma cdf.

Weather states

Meteorologically similar days are grouped together in "weather states" with use of the `mclust` package in `R` (Fraley and Raftery, 2003). The transitions between successive weather states are modeled by a first-order homogeneous Markov chain with N_W states, where N_W is estimated from the data along with the clustering. Once the clustering has been achieved, the weather state transition probability from w to w' is simply estimated by counting the transitions on the recorded series.

Multivariate modelling

To account for the skewness of the residuals, a multivariate CSN is fitted for each weather state to the vector of residuals. It was found that mixtures of skew-normal densities could adequately be fitted on $(\tilde{P}, R, V, T_n, T_x)$. Note that for dry weather states, the variable $\tilde{P} = 0$ is not modeled. In the following, the number of modeled variables (either 4 or 5) will

be denoted k. To make the parameter estimation possible, it was assumed, without loosing the skew-normal flexibility, that $\mathbf{D} = \mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{S}$ and $\mathbf{\Delta} = \mathbf{I}_k - \mathbf{S}^2$ where $\mathbf{\Sigma}^{-\frac{1}{2}}\mathbf{\Sigma}^{-\frac{1}{2}} = \mathbf{\Sigma}^{-1}$, $\mathbf{I}_k$ is the k-dimensional identity matrix and $\mathbf{S}$ is a diagonal matrix with elements in $[-1, 1]$. With this parametrization, the multivariate CSN density (1) becomes

$$f(\mathbf{y}) = 2^{-k}\ \phi_k(\mathbf{y}; \boldsymbol{\mu}, \mathbf{\Sigma})\ \Phi_k(\mathbf{S}\mathbf{\Sigma}^{-\frac{1}{2}}(\mathbf{y} - \boldsymbol{\mu}); \mathbf{0}, \mathbf{I}_k - \mathbf{S}^2), \tag{6}$$

which will be denoted $CSN_k^*(\boldsymbol{\mu}, \mathbf{\Sigma}, \mathbf{S})$. The vector $\boldsymbol{\mu}$ parametrizes the shift from $\mathbf{0}$ of the residuals for a given weather state, $\mathbf{\Sigma}$ their covariance matrix and diag($\mathbf{S}$) their skewness.

Temporal correlation

Temporal persistence is then modeled in the following way. Consider two consecutive days t and $t+1$ and their associated weather states, say w_t and w_{t+1}. The transformed vector

$$\tilde{\mathbf{X}}_{w_t} = \mathbf{\Sigma}_{w_t}^{-1/2}\left(\mathbf{X}_{w_t} - \boldsymbol{\mu}_{w_t}\right) \sim \mathrm{CSN}_k^*(\mathbf{0}, \mathbf{I}_k, \mathbf{S}_{w_t}), \tag{7}$$

can be shown to have independent margins distributed as $\mathrm{CSN}_1^*(0, 1, \delta_{w,i})$. The pairwise structure between the i^{th} components of $\tilde{\mathbf{X}}_{w_t}$ and $\tilde{\mathbf{X}}_{w_{t+1}}$ is then modeled by a correlation coefficient depending on the variable, weather states and season. The vector $\tilde{\mathbf{X}}_{w_{t+1}}$ is then back-transformed according to $\mathbf{X}_{w_{t+1}} = \mathbf{\Sigma}_{w_{t+1}}^{-1/2}\tilde{\mathbf{X}}_{w_{t+1}} + \boldsymbol{\mu}_{w_{t+1}}$.

Parameter estimation

Conditionally to the weather state w, a maximum likelihood approach ignoring temporal dependence is implemented to estimate the parameters of the distribution $\mathrm{CSN}_k^*(\boldsymbol{\mu}_w, \mathbf{\Sigma}_w, \boldsymbol{S}_w)$. The estimates are only slightly changed if the temporal dependence is taken into account in the estimation procedure, which will be thus ignored. The temporal correlation coefficients are estimated via the weighted moment approach described in Flecher et al. (2009).

2.5 Simulation algorithm

After estimating the parameters, the following algorithm simulates realizations of the five dimensional vector of interest:

1. A season is chosen and one $\tilde{\mathbf{X}}(0)$ is randomly drawn.

2. The transition probabilities are used to generate a Markov chain sequence of weather states: $w_1, \ldots, w_t, \ldots, w_T$.

3. For $t = 1, \ldots, T$

   (a) Given $\tilde{\mathbf{X}}(t) = \mathbf{x}_t$ and two consecutive weather states, w_t and w_{t+1}, a realization of the vector $\tilde{\mathbf{X}}(t+1)$ taking into account temporal dependence is drawn.

   (b) $\tilde{\mathbf{X}}(t+1)$ is multiplied by $\boldsymbol{\Sigma}_{w_{t+1}}^{1/2}$ and $\boldsymbol{\mu}_{w_{t+1}}$ is added in order to provide a proper residual $\mathbf{X}(t+1)$.

4. Residuals are multiplied by the absolute deviation and the median is added to account for trends and seasonal effects.

2.6 Some results

Colmar is located in the north east part of France at an altitude of 175 meters. A twenty year series is available from 1973 to 1992 for the five daily variables under study. Annual precipitation amounts are about 530mm and the frequency of rainy days is about 1/4. The climate is characterized by warm summers from June to September and cold winters. Both oceanic and continental climate trades can affect this site. This produces an important variability on the daily meteorology.

The mean behaviour of WACS-Gen has been illustrated in Flecher et al. (2010) on this series. In particular, it was shown that univariate and bivariate densities are very well fitted with the use of mixture of CSN densities. Then, on a set of thirty stochastic simulations of 20 years, densities, correlations between variables and temporal correlations were whown to be correctly reproduced.

Here, we illustrate how variability is reproduced. Figure 2.6 illustrates the variability generated by WACS-Gen on T_n for one 20 year long simulated series and compares it with

the recorded series. Each year, the mean value and the standard deviation for each calendar month is computed. These statistics are then summarized with boxplots, one boxplot per month. Comparison of monthly averages shows that the central tendency is quite well reproduced. However, the year-to-year variability of monthly averages is too low during the winter while at the same time the monthly standard deviations are too high. This can be explained by a lack of temporal persistance wich leads to less variable monthly averages and higher monthly variances. Temporal persistance is further illustrated in Figure 2.6 which shows the statistical frequency of freezing spells, i.e. number of consecutive days with $T_n < 0$. Although the total number of freezing days is comparable (1143 on measured data; 1017 on simulated ones), the simulated series show clearly more frequent short dry spells and less frequent long ones. These statistics are very difficult to reproduce, since one single day with positive minimum temperature is sufficient to break down a freezing spell into two shorter ones.

Concerning precipitation intensity, a qq-plot analysis (not shown here) revealed that precipitation distributions are well reproduced by the generator at the exception of the largest value The highest observed precipitation is 70 mm, while the highest simulated precipitations are in the range 40-68 mm. Note however that the mean precipitation is well reproduced and that on other sites than Colmar the opposite situation (simulated highest precipitations larger than measured ones) has be observed (results not reported here).

In conclusion, WACS-Gen does indeed reproduce correctly central tendencies of climate. It reproduces less correctly the variability, the statistics related to temporal persistance and extreme values.

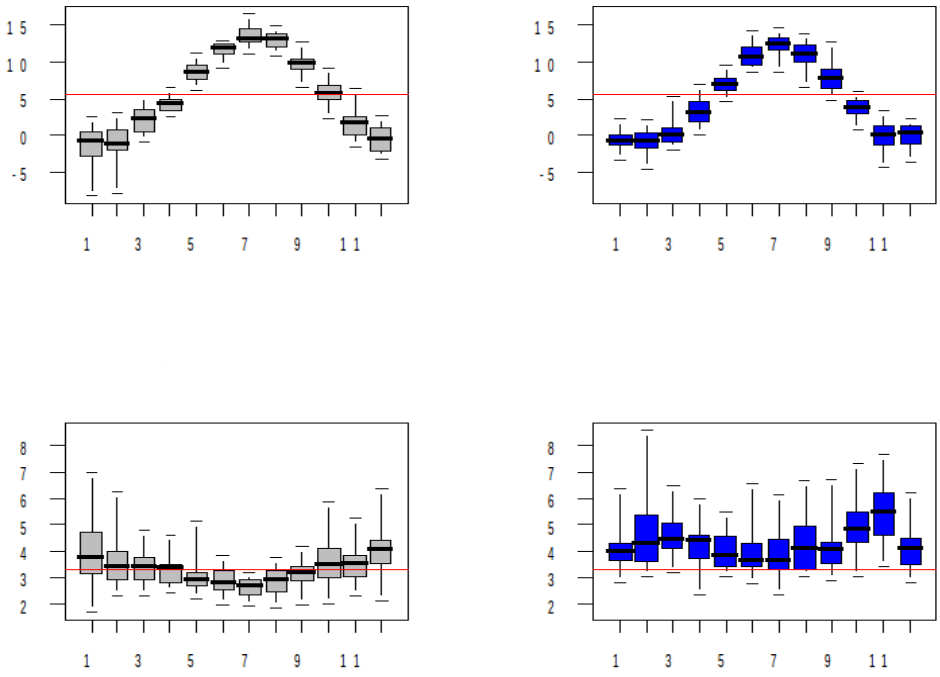

Figure 2: Minimum temperature on the Colmar series. Top row: boxplots of monthly averages. Bottom row: boxplots of monthly standard deviations. Left: measured data. Right: simulated data.

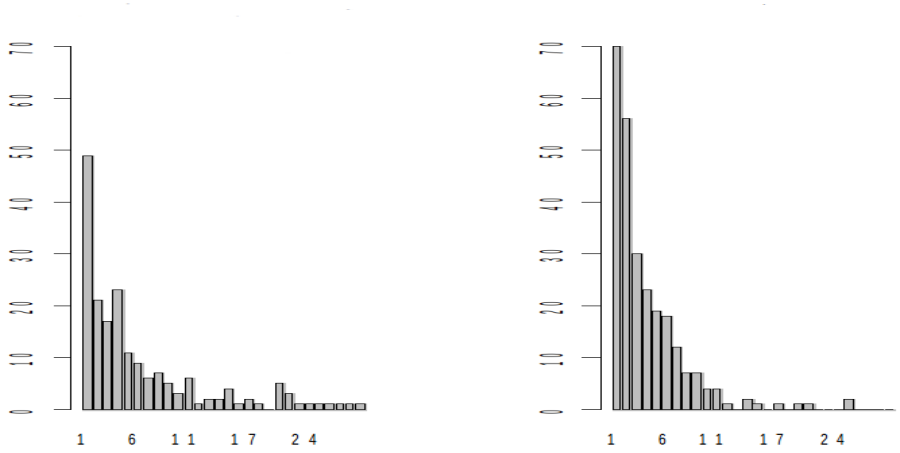

Figure 3: Distribution of freezing spells on the Colmar series. Left: measured data. Right: simulated data.

3 Dispersal kernels in spatial point processes for plant epidemiology and ecology

3.1 Introduction

Estimating the dispersal of pollen grains, seeds or spores (small propagules which allow an airborne plant disease to propagate) is of major concern to better assess and predict the spread of plants and airborne plant diseases. Over a two-dimensional planar space, the dispersal kernel is the probability density function of the random deposit location of a propagule released by a source located at the origin. The shape of the kernel is an important topic in dispersal studies because it strongly influences the spatio-temporal dynamics of organisms, the spatial clustering and the spreading speed (Minogue, 1989; Mollison, 1977). The most investigated feature of dispersal kernels is the decrease rate along radial distances and many models have been proposed (Austerlitz et al., 2004; Klein et al., 2006; Tufto et al., 1997), some of them allowing for long distance dispersal (heavy kernel tails). Non-monotonous dispersal kernels have also been built to represent an increase at short distances and a decrease at large distances of the density of deposited propagules (Colbach and Sache, 2001; Stoyan and Wagner, 2001). In the following, we focus on two other kernel features which have been less studied than the decrease rate but which however strongly influence the spatio-temporal dynamics of organisms: anisotropic dispersal and group dispersal. The next two sections illustrate the usefulness of spatial point patterns, spatial Brownian motions and Gaussian random processes to represent anisotropic dispersal and group dispersal.

3.2 Anisotropic dispersal

In the models under consideration here, deposit locations of propagules emitted by a point source at the origin form a spatial Poisson point pattern (Illian et al., 2008) with inhomogeneous intensity decreasing along radial directions. In addition the decrease along radial directions is anisotropic, i.e. it varies with respect to the direction.

Models based on 3D spatial Brownian motions describing spores transports (Klein et al., 2003; Stockmarr, 2002; Tufto et al., 1997) allow the introduction of anisotropy by adding trends to the horizontal components of the Brownian motions. Models presented in Soubeyrand et al. (2008b) allow for 3D anisotropy by combining anisotropic planar dispersal kernels and markov chains describing vertical moves of propagules. Another set of approaches consists in incorporated von Mises functions commonly used to describe the distributions of circular data (Fisher, 1995) into dispersal kernels to achieve anisotropy. Thus, in the models in Herrmann et al. (2011); Wagner et al. (2004); Wälder et al. (2009), the Euclidean distance from the source is multiplied by a function of the direction, typically a von Mises function. The approach that we followed in Soubeyrand et al. (2007a, 2008a, 2009a) also uses von Mises functions or more generally functions defined on the circle but these functions are used to modify the parameter of the dispersal kernel as well as the source strength. This approach, presented below, accounts for a double anisotropy in the dispersal of propagules.

3.2.1 "Anisotropizing" dispersal kernels

Let us consider a point source located at the origin of the planar space $\mathbb{R}^2$. Suppose that the deposit locations of propagules form a Poisson point process with intensity at location $x \in \mathbb{R}^2$ proportional to the isotropic exponential dispersal kernel:

$$f_{iso}(x) = \frac{1}{2\pi\beta_{iso}^2} \exp\left(-\frac{||x||}{\beta_{iso}}\right),$$

where $||\cdot||$ is the Euclidean distance and $\beta_{iso} > 0$ is the dispersal parameter. This isotropic kernel has been generalized in Soubeyrand et al. (2007a) into a doubly anisotropic exponential dispersal kernel:

$$f(x) = \frac{\alpha(\phi)}{\beta(\phi)^2} \exp\left(-\frac{||x||}{\beta(\phi)}\right), \tag{8}$$

where ϕ is the angle made by x, $\alpha(\cdot)$ is a circular probability density function (pdf) and $\beta(\cdot)$ is a positive circular function, both defined over $[0, 2\pi)$. It can be easily verified that f is a probability density function over $\mathbb{R}^2$. The function $\alpha(\phi)$ is related to the density of deposit

locations of propagules in direction ϕ, while $\beta(\phi)$ is the dispersal parameter in direction ϕ: the larger $\beta(\phi)$, the further in expectation propagules are deposited in direction ϕ. Other isotropic kernels can obviously be "anisotropized" in the same way, such as e.g. Gaussian kernels or geometric kernels (Soubeyrand et al., 2009a).

3.2.2 Specifying the anisotropy functions α and β

It was first proposed in Soubeyrand et al. (2007a) to specify α and β using von Mises functions which are regular and unimodal functions defined over the circle,

$$\alpha(\phi) = \frac{1}{2\pi I_0(\sigma_\alpha)} \exp\{\sigma_\alpha \cos(\phi - \mu_\alpha)\} \quad , \quad \beta(\phi) = \frac{\beta_0}{2\pi I_0(\sigma_\beta)} \exp\{\sigma_\beta \cos(\phi - \mu_\beta)\},$$

where $\mu_\alpha \in [0, 2\pi)$ is the mean dispersal direction and $\sigma_\alpha \geq 0$ is the dispersion around μ_α; $\mu_\beta \in [0, 2\pi)$ is the direction along which propagules are deposited the furthest in expectation; $\sigma_\beta \geq 0$ the dispersion of dispersal distances around μ_β and where $\beta_0 > 0$ is a multiplicative constant measuring how far from the source propagules are deposited; and $I_0(\sigma) = (2\pi)^{-1} \int_0^{2\pi} \exp\{\sigma(\theta - \mu)\} d\theta$. Obviously, other parametric forms than von Mises could be used, e.g. mixtures of von Mises functions, cardioid functions, wrapped Cauchy or normal functions; see Fisher (1995).

These functions are very regular. To take into account rougher anisotropies, Soubeyrand et al. (2008a) used circular Gaussian random processes (GRP) to specify the anisotropy functions with

$$\alpha(\phi) \propto \exp(Z_\alpha(\phi)) \quad \text{and} \quad \beta(\phi) = \exp(Z_\beta(\phi)),$$

where Z_α and Z_β are the realizations of two independent stationary circular GRPs with circular correlation functions (CCF) C_α and C_β. CCFs are generally obtained by using the chordal distance as the argument of a correlation function defined over $\mathbb{R}$. Let C denote a valid correlation function over $\mathbb{R}$, then $C(2\sin(\phi/2))$ is a valid correlation function on the circle with radius one. To model the dispersal of yellow rust spores, Soubeyrand et al. (2008a)

built the following CCF without resorting to this technique:

$$C(\phi) = 1 - \sin^{\delta}(\phi/2), \qquad \phi \in [0, 2\pi), \tag{9}$$

where $\delta \in (0, 2)$. This CCF is continuous but not differentiable at the origin so that a GRP with CCF (9) is mean square continuous but not mean square differentiable. It can therefore be used for rough processes. A model selection criteria showed that CCF (9) achieves the best fit to these data among a large class of function.

3.2.3 Estimation

For propagules whose sizes are measured in micrometers (e.g. spores, pollen grains) and propagules which are not detected easily in fields, orchards or forests, we do not generally observe the point pattern formed by the deposited propagules. Instead we may observe discrete counts: numbers of spores collected in traps, numbers of symptoms on plants, numbers of diseased plants or numbers of seedlings. The observed count data are modeled as random variables whose distributions depend on the dispersal kernel; see Soubeyrand et al. (2007b) who discuss the change of scales between the dispersal processes and the observation of data. Finally, a likelihood may be written and inference based on this likelihood may be performed. For example, Soubeyrand et al. (2007a) considered counts of wheat plants infected by the yellow rust and fitted to data the kernel (8) incorporating von Mises functions by specifying a binomial observation process and using a Newton-Raphson algorithm to maximize the likelihood. Exploiting the same data set, Soubeyrand et al. (2008a) fitted to data the kernel (8) incorporating circular GRPs by specifying a binomial observation process and using a Markov Chain Expectation Maximization algorithm (Wei and Tanner, 1990) to obtain maximum likelihood estimates of parameters and latent variables of their hierarchical model; see Figure 4. Soubeyrand et al. (2009b) studied the spatio-temporal spread of powdery mildew infecting *plantago lanceolata*, the data consisting in presence patterns of powdery mildew in a set of about 4000 host plant patches. They included the kernel (8) incorporating von

Mises functions in their spatio-temporal model and fitted this model to data using a Markov chain Monte Carlo (Robert and Casella, 1999) algorithm to assess posterior distributions of parameters.

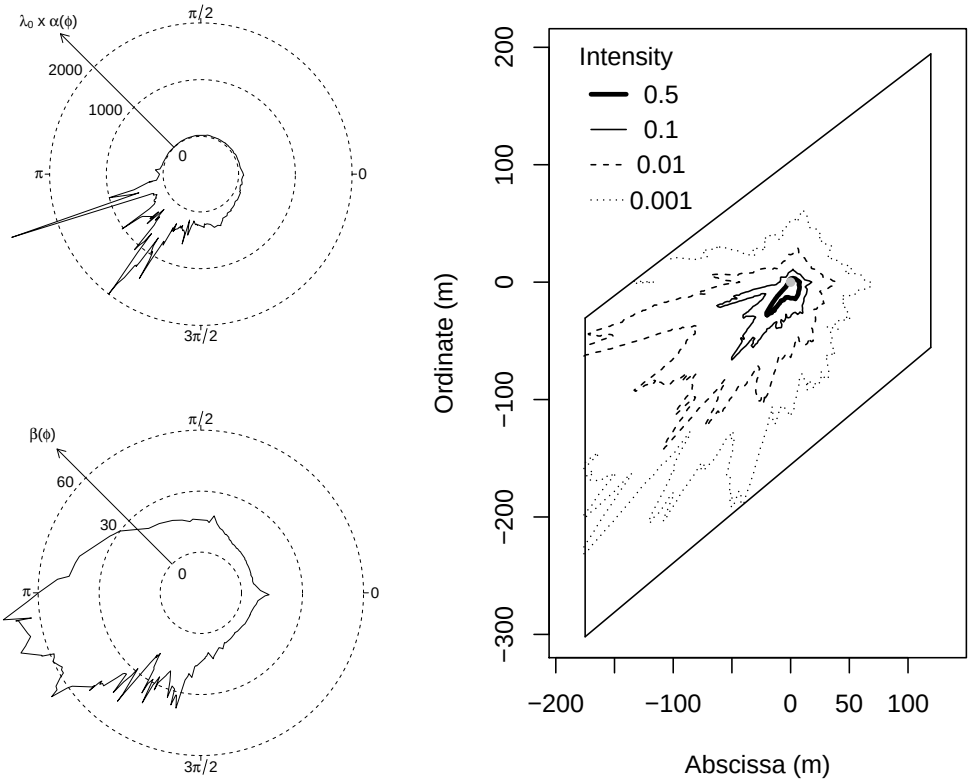

Figure 4: Estimates of the anisotropy functions α (up to a multiplicative constant λ_0) and β based on circular GRPs (left). Estimate of the resulting anisotropic dispersal kernel up to the multiplicative constant λ_0 (right). The product of λ_0 and the kernel is the intensity function of the point process formed by deposited propagules.

3.3 Group dispersal

Group dispersal occurs when several propagules are released because of a wind gust, transported in the air into a more or less limited volume and deposited over a more or less limited area. In propagation models for airborne plant pathogens and plants, deposit locations of

propagules are usually assumed to be independently and identically drawn under the dispersal kernel. When group dispersal occurs the independence assumption is not valid anymore. To represent group dispersal, Soubeyrand et al. (2011) resorted to a hierarchical structure of dependence: at the first stage of the hierarchy, groups are independently dispersed; at the second stage, propagules within each group are dispersed independently but conditionally on the group transport. The resulting model can be viewed as a doubly stochastic point process model also called Cox process (Illian et al., 2008). The model in Soubeyrand et al. (2011) and some of its properties are described in what follows. One of the properties, namely the distribution of the furthest point, is not commonly studied in point process theory but is especially relevant in dispersal studies.

3.3.1 Group dispersal model (GDM)

Consider a single point source of propagules located at the origin of the planar space $\mathbb{R}^2$. The deposit location vector X_{jn} of the n-th propagule of group j ($j \in \{1, \ldots, J\}$ and $n \in \{1, \ldots, N_j\}$) is assumed to satisfy

$$X_{jn} = X_j + B_{jn}(\nu||X_j||),$$

where $X_j = E(X_{jn} \mid X_j)$ is the final location vector of the center of group j, B_{jn} is a centered Brownian motion describing the relative movement of the n-th propagule in group j with respect to the group center, ν is a positive parameter and $||\cdot||$ denotes the Euclidean distance.

The random variables J, N_j, X_j and the random processes $\{B_{jn} : n = 1, \ldots, N_j\}$ are mutually independent. The number of groups J is Poisson distributed with mean value λ. The N_j are independently drawn from the counting distribution p_{μ,σ^2} defined over $\mathbb{N}$ with mean and variance parameters μ and σ^2, respectively. The group center locations X_j are independently and identically drawn from the pdf f_{X_j}. This function can be characterized by features usually associated with classical dispersal kernels, see Sections 3.1 and 3.2: the

decrease of f_{X_j} at the origin is more or less steep, the tail of f_{X_j} is more or less heavy, the shape of f_{X_j} is more or less anisotropic. The Brownian motions B_{jn} defined over $\mathbb{R}^2$ are centered, independent and with independent components. They are stopped at time $t_j = \nu||X_j||$. The distance between the source and the location X_j is used, up to the scaling parameter ν, as a time surrogate. Thus, the further a group is transported, the most the particules forming the group are spread with respect to the group center. The value of ν determines the strength of the relative spread from the group center. It follows that $B_{jn}(\nu||X_j||)$ follow independent and centered normal distributions with variance matrices $\nu||X_j||I$ where I is the 2×2 identity matrix. Figure 5 which shows a simulation of the GDM clearly illustrates the existence of groups whose extent increase with distance from the point source.

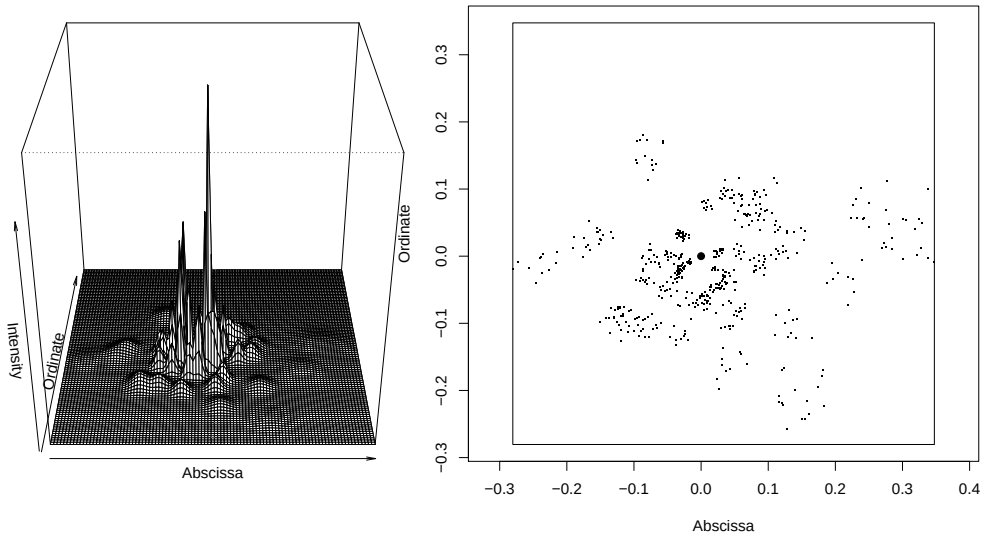

Figure 5: Simulation of the group dispersal model with a single point source. Left: conditional density of the deposit location of a propagule whose group it belongs to is unknown given the locations of the group centers. Right: deposit locations of particles obtained under the density shown on the left; the large dot indicates the location of the point source.

3.3.2 Moments and generation of clusters

A first understanding of the ability of the GDM to generate multiple foci was achieved by studying the moments of the number of propagules $Q(x+dx)$ deposited in the infinitesimal surface $x+dx$ centered around x. The expectation, variance and covariance are derived in Soubeyrand et al. (2011):

$$\begin{aligned} E\{Q(x+dx)\} &= \lambda\mu f_{X_{jn}}(x)dx \\ V\{Q(x+dx)\} &= \lambda[\mu f_{X_{jn}}(x)dx + (\sigma^2+\mu^2-\mu)E\{\phi_{\nu,X_j}(x)^2\}(dx)^2] \\ \operatorname{cov}\{Q(x_1+dx), Q(x_2+dx)\} &= \lambda(\sigma^2+\mu^2-\mu)E\{\phi_{\nu,X_j}(x_1)\phi_{\nu,X_j}(x_2)\}(dx)^2, \end{aligned}$$

where $f_{X_{jn}}$ is the probability density function of X_{jn}. It it thus the dispersal kernel of propagules, satisfying

$$f_{X_{jn}}(x) = \int_{\mathbb{R}^2} f_{X_{jn}|X_j}(x \mid y) f_{X_j}(y)dy = \int_{\mathbb{R}^2} \phi_{\nu,y}(x) f_{X_j}(y)dy.$$

Hence, $\phi_{\nu,y}(x) = \frac{1}{2\pi\nu||y||}\exp\left(-\frac{(x-y)'(x-y)}{2\nu||y||}\right)$, $E\{\phi_{\nu,X_j}(x)^2\} = \int_{\mathbb{R}^2}\phi_{\nu,y}(x)^2 f_{X_j}(y)dy$ and $E\{\phi_{\nu,X_j}(x_1)\phi_{\nu,X_j}(x_2)\} = \int_{\mathbb{R}^2}\phi_{\nu,y}(x_1)\phi_{\nu,y}(x_2)f_{X_j}(y)dy$. It has to be noted that for a counting distribution over $\mathbb{N}$ characterized by mean $\mu > 0$ and variance σ^2, the quantity $\sigma^2+\mu^2-\mu$ is positive; it is zero if and only if $\mu = 0$ (which implies that $\sigma^2 = 0$) or $(\mu,\sigma) = (1,0)$. This implies that the covariance given above for the GDM is non-negative.

These moments were especially used to compare point patterns obtained under the GDM with $\mu = 1$ and under the independent dispersal model (IDM) with $\mu = 1$ and $\sigma = 0$. The extra-variance and the positive spatial covariance characterizing the GDM induce the occurrence of aggregates (or clusters) in space, while such aggregates are not expected under the IDM. These aggregates are at the origin of secondary foci visible in a spatio-temporal context (i.e. when deposited propagules become sources at the next time step) without resorting to heavy-tailed dispersal kernels or to spatial heterogeneity.

3.3.3 Furthest propagule

Of particular interest is the distribution of the furthest deposited propagule, because it determines the spreading speed (or invasion speed) in a spatio-temporal context: the more concentrated the propagules, the lowest the spreading speed. Let R_j^{max} be the distance between the origin and the furthest deposited propagule of group j. Obviously, $R_j^{max} = 0$ if no propagule is dispersed in group j. The distribution of R_j^{max} is thus zero inflated. Analytical expressions of its pdf, $f_{R_j^{max}}$, and cumulative distribution function, $F_{R_j^{max}}$, can be found in Soubeyrand et al. (2011).

Let now R^{max} denote the distance from the source to the furthest deposited propagule: $R^{max} = \max\{R_j^{max} : j \in \mathcal{J}\}$. Under the GDM, the distribution of the distance between the origin and the furthest deposited propagule is zero inflated and satisfies:

$$P(R^{max} = 0) = \exp\left[\lambda\{p_{\mu,\sigma^2}(0) - 1\}\right] \quad \text{and} \quad f_{R^{max}}(r) = \lambda f_{R_j^{max}}(r) \exp\{\lambda(F_{R_j^{max}}(r) - 1)\},$$

for $r > 0$. The material provided allowed the analytic study of the probability that R^{max} is larger than a distance $r > 0$:

$$P(R^{max} \geq r) = \int_r^{+\infty} f_{R^{max}}(s)ds.$$

It especially followed that for every GDM and IDM characterized by the same dispersal kernel for the propagules and the same expected number of dispersed propagules $\lambda\mu$, the furthest propagule under the GDM has less chance to be at a distance greater than any $r > 0$ than the furthest propagule under the IDM. Therefore, the population of propagules is expected to be more concentrated under the GDM than under the IDM. In other words, the average expansion speed under the GDM is expected to be lower than the average expansion speed under the IDM.

4 Discussion for further research

Based on work undertaken in two different fields of applications, the studies presented in this article illustrate how application fields can bring the researcher to investigate new spatial models and their features. This participates to the iterative process of research: in any scientific field, once new results have been stated, one may be interested in refining them or understanding the discrepancies between the results and the reality. This in turn leads to new models and methods. This iterative process lead us to study skewed normal distributions and anisotropic and group dispersal. On these two subjects further research are required.

Weather generators in general and WACS-Gen in particular present currently some limitations: 1) when multivariate they are local instead of being regionalized; 2) being designed to reproduce first and second moments (means, variances and correlations) they do not reproduce correctly extreme values. Further research will be targeted at proposing models and methods for a multivariate, multilocal extension of the weather generator WACS-gen with a view towards the modeling of extreme values. For this we will benefit from recent developments in the spatial modeling of extreme values (Bacro and Gaetan, 2012).

Regarding the dispersal studies presented above, research can be extended in many directions. Some of these directions concern the spatiotemporal versions of the models which have been presented. Building a spatiotemporal model including the anisotropic dispersal with circular GRPs and a spatiotemporal model including group dispersal is easy. More challenging is the study of the properties of these models (e.g. expansion speed) and the implementation of accurate estimation techniques because the spatiotemporal models will include very large sets of latent variables.

The models presented in this paper share the common concern to adequatly model right-skewed data and extreme values, such as extreme temperatures or extreme dispersal distances. In relation to these statistical concepts is the agronomically oriented question of assessing the impact of extreme or rare events: extreme droughts, extreme heat waves, fast or long

propagation of pests or new species.

References

Allard D. and Naveau, P. (2007). A new spatial skew-normal random field model. *Communications in Statistics*, **36**, 1821–1834.

Austerlitz, F., C. W. Dick, C. Dutech, E. K. Klein, S. Oddou-Muratorio, P. E. Smouse, and V. L. Sork (2004). Using genetic markers to estimate the pollen dispersal curve. *Molecular Ecology* **13**, 937–954.

Banerjee, S. Carlin, B.P. and Gelfand, A.E. (2004). *Hierarchical Modelling and Analysis for Spatial Data*. Chapman & Hall/CRC Press, Boca-Raton.

Besag, J.E. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, series B*, **34**, 75–83.

Boé, J. and Terray, L. (2008). A weather type approach to analysing winter precipitation in France: twentieth century trends and influence of anthropogenic forcing. *Journal of Climate*, **21**, 3118–3133.

Brisson, N., Gary, C., Justes, E., Roche, R., Mary, B., Ripoche, D., Zimmer, D., Sierra, J., Bertuzzi, P., Burger, P., Bussiere, F., Cabidoche, Y.M., Cellier, P., Debaeke, P., Gaudillere, J.P., Henault, C., Maraux, F., Seguin, B., Sinoquet, H. (2003). An overview of the crop model STICS. *European Journal of Agronomy*, **18**, 309–332.

Chilès J.P. and Delfiner P. (1999). *Geostatistics: Modeling Spatial Uncertainty*. John Wiley & Sons, New-York.

Colbach, N. and I. Sache (2001). Blackgrass (alopecurus myosuroides huds.) seed dispersal from a single plant and its consequences on weed infestation. *Ecological modelling* **139**(2-3), 201–219.

Cressie, N. and Wikle, C.K. (2011). *Statistics for Spatio-temporal data*. Wiley, Hoboken.

Diggle, P.J., Tawn J.A. and Moyeed, R.A. (1998). Model-based geostatistics (with discussion). *Applied Statistics*, **47**, 299–350.

Diggle, P.J. and Ribeiro P.J. (2007). *Model-Based Geostatistics*. Springer, New-York.

Fisher, N. I. (1995). *Statistical Analysis of Circular Data*. Cambridge University Press.

Flecher, C., Naveau, Ph. and Allard, D. (2009). Estimating the Closed Skew-Normal distributions parameters using weighted moments *Statistics and Probability Letters*, **79**, 1977–1984.

Flecher, C., Naveau P., Allard D. and Brisson, N. (2010). A Stochastic Daily Weather Generator for Skewed Data, *Water Ressource Research*, **46**, W07519.

Fraley, C. and Raftery, A.E. (2003). Enhanced model-based clustering, density estimation, and discriminant analysis software: MCLUST. *Journal of Classification*, **20**, 263–286.

Gelfand, A.E., Diggle, P.J., Fuentes, M. and Guttorp, P. (Eds.) (2010) *Handbook of Spatial Statistics*. Chapman & Hall/CRC Press, Boca-Raton.

Genton, M.G. (Ed.) (2004) *Skew-Elliptical Distributions and Their Applications: A Journey Beyond Normality*. Chapman & Hall/CRC, Boca Raton.

González-Farías, G., Domínguez-Molina, J. and Gupta, A., (2004). Additive properties of skew-normal random vectors. *Journal of Statistical Planning and Inference*, **126**, 521–534.

Herrmann, I., T. Herrmann, and S. Wagner (2011). Improvements in anisotropic models of single tree effects in cartesian coordinates. *Ecological Modelling* **222**(7), 1333 – 1336.

Illian, J., A. Penttinen, H. Stoyan, and D. Stoyan (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. Wiley.

Karimi, O. and Mohammadzadeh, M. (2010). Bayesian spatial regression models with closed skew normal correlated errors and missing observations. *Statistical Papers* DOI 10.1007/s00362-010-0329-2.

Katz, R.W (1996). The use of stochastic models to generate climate scenarios. *Climate Change*, **32**, 237–255.

Klein, E. K., C. Lavigne, X. Foueillassar, P.-H. Gouyon, and C. Larédo (2003). Corn pollen dispersal: quasi-mechanistic models and field experiments. *Ecological Monographs* **73**, 131–150.

Klein, E. K., C. Lavigne, and P.-H. Gouyon (2006). Mixing of propagules from discrete sources at long distance: comparing a dispersal tail to an exponential. *BMC Ecology* **6**, 3.

Minogue, K. P. (1989). Diffusion and spatial probability models for disease spread. In M. J. Jeger (Ed.), *Spatial Components of Plant Disease Epidemics*, pp. 127–143. Prentice Hall.

Mollison, D. (1977). Spatial contact models for ecological and epidemic spread. *J. R. Statist. Soc. B* **39**, 283–326.

Naveau, P. Allard, D. (2004) Modeling skewness in spatial data analysis without data transformation. In: Leuangthong, O., Deutsch, C., (Eds.) *Proceedings of the Seventh International Geostatistics Congress*, 929-938. Springer, Dordrecht.

Rajagopalan, B., Lall, U., Tarboton, D.G. and Bowles, D.S. (1997). Multivariate non parametric resampling sheme for generation of daily weather variables. *Stochastic Hydrology and Hydraulics*, **11**, 523–547.

Rajagopalan, B. and Lall, U. (1999). A k-nearest-neighbor simulator for daily precipitation and other weather variables. *Water Resources Research*, **35-10**, 3089–3101.

Ripley, B.D. (1977) Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society, series B*, **39**, 172–266.

Robert, C. P. and G. Casella (1999). *Monte Carlo Statistical Methods*. New York: Springer.

Schlather, M. Ribeiro, P. and Diggle, P.J. (2004) Detecting dependence between marks and locations of marked point processes. *Journal of the Royal Statistical Society, series B*, **66**, 79–83.

Semenov, M.A. and Barrow, E.M. (1997) Use of stochastic weather generator in the developpement of climate change senarios. *Climate Change*, **35**, 397–414.

Soubeyrand, S., J. Enjalbert, A. Kretzschmar, and I. Sache (2009). Building anisotropic sampling schemes for the estimation of anisotropic dispersal. *Annals of Applied Biology* **154**(3), 399–411.

Soubeyrand, S., J. Enjalbert, and I. Sache (2008). Accounting for roughness of circular processes: Using gaussian random processes to model the anisotropic spread of airborne plant disease. *Theoretical Population Biology* **73**, 92–103.

Soubeyrand, S., J. Enjalbert, A. Sanchez, and I. Sache (2007). Anisotropy, in density and in distance, of the dispersal of yellow rust of wheat: Experiments in large field plots and estimation. *Phytopathology* **97**, 1315–1324.

Soubeyrand, S., L. Held, M. Höhle, and I. Sache (2008b). Modelling the spread in space and time of an airborne plant disease. *Journal of the Royal Statistical Society C* **57**, 253–272.

Soubeyrand, S., A. L. Laine, I. Hanski, and A. Penttinen (2009). Spatio-temporal structure of host-pathogen interactions in a metapopulation. *The American Naturalist* **174**, 308–320.

Soubeyrand, S., L. Roques, J. Coville, and J. Fayard (2011). Patchy patterns due to group dispersal. *Journal of Theoretical Biology* **271**, 87–99.

Soubeyrand, S., G. Thébaud, and J. Chadœuf (2007). Accounting for biological variability and sampling scale: a multi-scale approach to building epidemic models. *Journal of The Royal Society Interface* **4**(16), 985–997.

Stockmarr, A. (2002). The distribution of particles in the plane dispersed by a simple 3-dimensional diffusion process. *Journal of Mathematical Biology* **45**, 461–469.

Stein, M.L. (1999) *Interpolation of Spatial Data*. Springer, New-York.

Stoyan, D. and S. Wagner (2001). Estimating the fruit dispersion of anemochorous forest trees. *Ecological Modelling* **145**(1), 35–47.

Tallis, G.M. (1961). The moment generating function of the truncated multi-normal Distribution. *J. Roy. Stat. Soc. B* **23**, 223–229

Tufto, J., S. Engen, and K. Hindar (1997). Stochastic dispersal processes in plant populations. *Theoretical Population Biology* **52**, 16–26.

Wackernagel, H. (2003) *Multivariate Geostatistics, third edition*. Springer, Berlin.

Wagner, S., K. Walder, E. Ribbens, and A. Zeibig (2004). Directionality in fruit dispersal models for anemochorous forest trees. *Ecological modelling* **179**(4), 487–498.

Wälder, K., W. Näther, and S. Wagner (2009). Improving inverse model fitting in trees–anisotropy, multiplicative effects, and bayes estimation. *Ecological Modelling* **220**(8), 1044–1053.

Wei, G. C. G. and M. A. Tanner (1990). Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *J. of the American Statistical Association* **85**, 699–704.