AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE
10.1002/2014WR016455

Key Points:
- Latent censored Gaussian spatiotemporal field
- Marginals given by gamma and generalized Pareto distribution
- New estimation methodology for covariance of censored bivariate Gaussian

Correspondence to:
A. Baxevani,
baxevani@ucy.ac.cy

Citation:
Baxevani, A., and J. Lennartsson (2015), A spatiotemporal precipitation generator based on a censored latent Gaussian field, *Water Resour. Res.*, *51*, 4338–4358, doi:10.1002/2014WR016455.

Received 24 SEP 2014
Accepted 16 MAY 2015
Accepted article online 20 MAY 2015
Published online 18 JUN 2015

A spatiotemporal precipitation generator based on a censored latent Gaussian field

Anastassia Baxevani[1] and Jan Lennartsson[2]

[1]Department of Mathematics and Statistics, University of Cyprus, Nicosia, Cyprus, [2]Chalmers University of Technology and Gothenburg University, Gothenburg, Sweden

Abstract A daily stochastic spatiotemporal precipitation generator that yields precipitation realizations that are quantitatively consistent is described. The methodology relies on a latent Gaussian field that drives both the occurrence and intensity of the precipitation process. For the precipitation intensity, the marginal distributions, which are space and time dependent, are described by a composite model of a gamma distribution for observations below some threshold with a generalized Pareto distribution modeling the excesses above the threshold. Model parameters are estimated from data and extrapolated to locations and times with no direct observations using linear regression of position covariates. One advantage of such a model is that stochastic generator parameters are readily available at any location and time of the year inside the stationarity regions. The methodology is illustrated for a network of 12 locations in Sweden. Performance of the model is judged through its ability to accurately reproduce a series of spatial dependence measures and weather indices.

1. Introduction

Precipitation modeling has attracted a lot of attention the last 50 years or so, mainly due to its many applications in hydrology, ecology, agriculture, and so many other aspects of everyday life. Realistic sequences of precipitation are key inputs in many models, and synthetic data are used routinely to fill in gaps both in time and/or space in historical records.

Traditionally, most models were individual location models although, recently, there have been many attempts to construct spatiotemporal stochastic models capable of generating temporally and spatially correlated fields of precipitation. These statistical models, also known as weather generators, can be used among others, to generate synthetic precipitation records where none are available, to extrapolate short observed records and to downscale climate change scenarios in both space and time. Additionally, they have the great advantage of providing not only with precipitation estimates but also quantitative measures of uncertainty. Daily rainfall simulators are by far the most common, partly because of the wide availability of data in this time scale, and partly due to an abundance of process-based models, like electricity demand models or crop models [see e.g., *Robertson et al*., 2007], that are driven by daily inputs.

The most prominent characteristic of precipitation is that its distribution is a mixed distribution with a discrete part corresponding to occurrence or not of precipitation, and a continuous part modeling the nonzero amounts. As a consequence, precipitation generators can be built up using the familiar chain-dependent stochastic model, a two-stage approach, modeling separately the discrete and continuous part [see e.g., *Todorovic and Woolhiser*, 1975; *Katz*, 1977; *Wilks*, 1998, and references therein]. The occurrence component is usually modeled by means of Markov chains of different orders [see *Katz*, 1977; *Lennartsson et al*., 2008], including nonhomogeneous Markov chains [see *Katz and Parlange*, 1995; *Furrer and Katz*, 2008], semi-Markov models, and multistate Markov chains [*Flecher et al*., 2010]. On the other hand, the most common models for the precipitation part are parametric statistical non-Gaussian models to which the underlying atmospheric physical processes are related only implicitly [see *Hutchinson*, 1986; *Stern and Coe*, 1984, among others]. Alternatively, *Bell* [1987], *Hutchinson* [1995], *Glasbey and Nevison* [1997], *Durban and Glasbey* [2001], *Allcroft and Glasbey* [2003], and *Sigrist et al*. [2012] have used a monotonic data transformation to achieve marginal normality. Using a different approach, *Hughes et al*. [1999] and *Ailliot et al*. [2009] have suggested instead use of hidden Markov models. Variable is defined so that dry conditions correspond to censored values below a given threshold.

Recently, *Allard and Bourotte* [2015] and *Sigrist et al*. [2012] have additionally assumed that both parts of the precipitation process—occurrence and intensity—can be modeled using the same latent Gaussian process. Various well known transformation functions have been suggested, for example, *Allcroft and Glasbey* [2003] use a quadratic power function, *Sigrist et al*. [2012] use a power function, and *Allard and Bourotte* [2015] use a power-exponential function to transform the Gaussian value to the desired intensities. Recently, *Kleiber et al*. [2012] applied a two-part transformation function, with one part being the standard normal distribution and the second part be given by the inverse of a gamma distribution. We found, in accordance with *Allard and Bourotte* [2015], that although the gamma distribution model is adequate for low and moderate rainfall, it fails to model properly the most extreme intensities.

On the other hand, a variety of methods have been suggested to tackle the problem of multisite rainfall precipitation. In a simplistic approach, the sites are mutually independent in space and time given the weather type [see *Zucchini and Guttorp*, 1991; *Hughes and Guttorp*, 1994], an autologistic model was used in *Bellone et al*. [2000] and *Bardossy and Plate* [1992]. *Ailliot et al*. [2009] used a censored, power-transformed Gaussian distribution for daily rainfall, while *Kleiber et al*. [2012] developed a multiscale extension of the chain-dependent model. Alternatively, *Buishand and Brandsma* [2001] used techniques based on resampling, *Hughes et al*. [1999] and *Charles et al*. [1999] used techniques based on unobserved underlying weather states, and *Yang et al*. [2005] employed generalized linear models. Recently, *Burton et al*. [2008] presented RainSim, a stochastic rainfall field generator where rainfall fields are sampled from a spatial-temporal Neyman-Scott rectangular pulses process. Finally, *Stehlík and Bárdossy* [2002] suggested transforming the rainfall distribution to approximate normality with zeros corresponding to negative transformed values.

The novelty of our work is that we model simultaneously both the occurrences and intensities using a latent Gaussian field. Using the truncated normal with negative observations being registered as dry days, we model the monotonic transformation function, referred to hereafter as anamorphosis function, as a composition of a gamma and a generalized Pareto distribution for the excesses above a high level. We also suggest a new method for estimating the spatiotemporal covariance function.

The paper is organized as follows: the data are introduced in section 2. The proposed model together with the model fitting and parameter estimation methodology is presented in section 3. In section 4, we present the validation results. Finally, in section 5, we discuss the pros and cons of the proposed model and suggest possible future research directions.

2. Data

Daily precipitation records from a small network of 14 weather stations across southern Sweden, covering an area from the West coast to the East coast in the Baltic Sea were used, 12 of them for model fitting and the remaining two for validation. The altitudes of the stations vary from 2 meters above sea level (station 4) to 305 meters (station 7). The data consist of 51 years (2 January 1961 to 31 December 2011) of daily precipitation (days with accumulated precipitation less than 0.1 mm are considered as dry) with less than 10% of observations missing from each station. The observation network is quite dense with distances between the stations ranging between 30 and 290 km. The climate in the area is dominated by the effects of the South Swedish highland, an area situated more than 200 m above the ocean with the prevailing western winds resulting in orographic precipitation in the surrounding areas.

We assume that daily rainfall from the 51 years of each month is 51 realizations of the same stochastic field, i.e., we assume that the stochastic model during each month is stationary and data from each year are an independent realization of that stochastic model. This assumption is not uncommon for meteorological processes. Most of the results will be presented for one location and the months of January and July as representative of the winter and summer season, respectively, but the rest of the months and stations gave similar results.

3. Stochastic Model

In this section, we present the stochastic approach used to model the temporal and spatial dependence for daily precipitation. We suggest using a latent Gaussian field to model the occurrence process and a

transformation of the same latent field to model the nonzero amounts. Use of the same variable for occurrence and intensity is preferred mainly for two reasons. First, the number of parameters is small and second, we avoid the edge effect, in which unrealistically large intensities can be generated near the boundary of dry areas when the occurrence and intensity processes are modeled separately [see *Bardossy and Plate*, 1992; *Wilks*, 1998; *Schleiss et al*., 2014, and references therein].

Hence, we assume the precipitation field $Y(\mathbf{s},t)$ at location $\mathbf{s}$ and time t depends on the latent Gaussian random variable $Z(\mathbf{s},t)$, through the following relation:

$$Y(\mathbf{s},t)=\begin{cases}\psi(Z(\mathbf{s},t)), & \text{if } Z(\mathbf{s},t)>0\\ 0, & \text{if } Z(\mathbf{s},t)\le 0\end{cases}, \tag{1}$$

where ψ is a nondecreasing function referred as anamorphosis. This idea of using meta-Gaussian processes where the marginal distribution of Gaussian process is transformed to non-Gaussian distribution has been suggested by *Guillot* [1999]. The advantage of using a meta-Gaussian random field model is twofold. First, it is compatible with any marginal distribution, even with discontinuous cumulative distribution functions [see *Guillot*, 1999, for more details] and second, the latent Gaussian field is completely described by the first two moments (mean and covariance function) while exhibiting remarkable flexibility.

3.1. Marginal Model

Model (1) suggests that the nonzero amounts of precipitation are modeled as a transformed censored Gaussian random field. Traditionally, at single locations, the precipitation amounts have been modeled using a Box-Cox transformation [see *Box and Cox*, 1964], a quadratic power transformation [see *Glasbey and Nevison*, 1997; *Durban and Glasbey*, 2001] or a power-exponential transformation of the censored Gaussian distribution [see *Allard and Bourotte*, 2015]. In *Lennartsson et al*. [2008], a generalized Pareto (GP) distribution modeled heavy rainfall above a high level. *Kleiber et al*. [2012] used a gamma distribution, while mixtures of exponential distributions were used in *Wilks* [1998].

In this work, our interest is not only in modeling accurately the lower and middle parts of the distribution, where the majority of the data falls, but also the extreme, although rare, amounts of precipitation. The latter have a profound impact on agriculture [see e.g., *Rosenzweig et al*., 2002; *Furrer and Katz*, 2008], infrastructure [see *Larsen et al*., 2008], and the society [see *McMichael et al*., 2006]. Moreover, one of the objectives of this work is to reproduce some of the weather indices [see *Karl et al*., 1999; *Peterson et al*., 2001] and section 4, which are greatly influenced by the tail of the distribution. A gamma distribution alone, although flexible enough, is not quite adequate to model the tail of the distribution since it underestimates large values [see e.g., *Kleiber et al*., 2012]. The fit improves considerably when we use the so-called hybrid gamma and GP distribution [see *Li et al*., 2012; *Furrer and Katz*, 2008]. This hybrid distribution is a result of coupling a gamma distribution with a GP distribution and has its origin in the one introduced by *Carreau and Bengio* [2009], where Gaussian and GP distributions were stitched together.

So we choose the transformation:

$$Y(\mathbf{s},t)=G_{\mathbf{s},t}^{-1}\circ\Phi_{\mu}(Z(\mathbf{s},t)), \tag{2}$$

where $G_{\mathbf{s},t}$ is the cumulative distribution function (cdf) of a hybrid gamma with GP at location $\mathbf{s}$ and time t and Φ_{μ} is the cdf of a censored normal random variable with mean value μ and unit variance. This type of anamorphosis transformation [see *Chilés and Delfiner*, 2012; *Kleiber et al*., 2012] retains the hybrid gamma with GP distribution at the individual locations while it allows different types of spatial correlation between the sites.

For the distribution G, we use the cdf of a hybrid gamma with GP distribution

$$G(x;u)=F_{\gamma}(\min\{x,u\};\alpha,\gamma)+(1-F_{\gamma}(u;\alpha,\gamma))F_u(x;\xi,\sigma,u), \tag{3}$$

where F_{γ} and F_u are the cdfs of a gamma and a GP, respectively. The GP is supported on (u,∞) where u is the threshold to the excesses of which the GP model is fitted. The normalizing factor $1-F_{\gamma}(u)$ ensures that G is a distribution function. It is clear that G has a point of discontinuity at threshold u. *Furrer and Katz* [2008] suggested refining the hybrid model through a compromise by estimating the gamma distribution from all the data, the GP distribution from the data above the threshold and then using only the estimated shape parameter ξ of the GP while adjusting its scale parameter in order to achieve a continuous density.

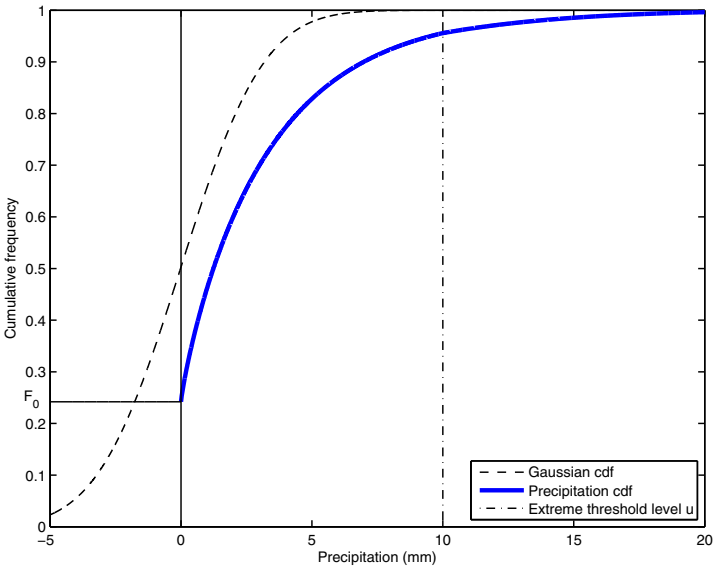

Figure 1. Illustration of the anamorphosis function ψ transforming a Gaussian distribution (dashed line) to the hybrid gamma with generalized Pareto G characterized by an atom at 0.

We have decided not to follow this compromise since the meta-Gaussian model allows for discontinuous distributions [see *Guillot*, 1999] and because the fit of the distribution to data did not improve.

The proposed parametric approach, which can be seen in Figure 1, is a modification of the semiparametric one in *Lennartsson et al.* [2008], where the part below the threshold was modeled using the empirical distribution, and is preferred over that since it simplifies extending the model, in a homogeneous way, in space and therefore making predictions possible at locations where no measurements are available.

Finally, Φ_μ is the cdf of a censored Gaussian random variable with mean value μ and unit variance which is given by

$$\Phi_\mu(z)=\begin{cases}\dfrac{\Phi(z-\mu)-\Phi(-\mu)}{\Phi(\mu)}, & \text{if } z>0\\ 0, & \text{if } z\le 0,\end{cases} \tag{4}$$

where ϕ and Φ are the pdf and cdf of a standard normal random variable, respectively. Therefore, the transformation ψ in (1) is fully parameterized by the parameters of the G distribution, $[\alpha,\gamma,u,\xi,\sigma]$, and the mean value μ of the latent Gaussian field. We next turn to the problem of their estimation.

3.2. Mean Function

There is a clear link between the mean function of the latent field Z and the precipitation data Y through the frequency of wet days:

$$P(\text{wet } t \text{ day at location } \mathbf{s})=\mathrm{P}(\mathrm{Y}(\mathbf{s},t)>0)=P(Z(\mathbf{s},t)>0)=\Phi(\mu(\mathbf{s},t)). \tag{5}$$

Estimation of the mean value is therefore straightforward by simply inverting (5) to obtain

$$\widehat{\mu}(\mathbf{s},t)=\Phi^{-1}(\widehat{p}_{\mathbf{s},t}), \tag{6}$$

where $\widehat{p}_{\mathbf{s},t}$ is the estimate of the proportion of wet days at location $\mathbf{s}$ and for the tth day of the year for $t=1,\ldots,365/366$. Hereafter, we shall denote by $\widehat{x}$ the estimate of any parameter x.

The mean value estimates at each station exhibit seasonality which is accounted for by means of finite Fourier series with spatially varying parameters:

$$\widehat{\mu}(\mathbf{s},t)=\widehat{\mu}_0(\mathbf{s})+\sum_{j=1}^{J}\widehat{\mu}_{j,1}(\mathbf{s})\sin\left(\frac{2\pi j}{365}t+\widehat{\mu}_{j,2}(\mathbf{s})\right),\quad t=1,\ldots,365/366, \tag{7}$$

with J being finite and not exceeding 2. Parameter estimation is done by the Weighted Least Square (WLS) method, while the number of covariates included is determined using the Bayesian Information Criterion (BIC) [see *Schwarz*, 1978]. The BIC gave 1 as optimal value of trigonometric terms in (7) for about half of the stations and 2 for the rest. We have decided to use $J=1$ for all stations since the gain for using the most complex model was not substantial. This resulted to

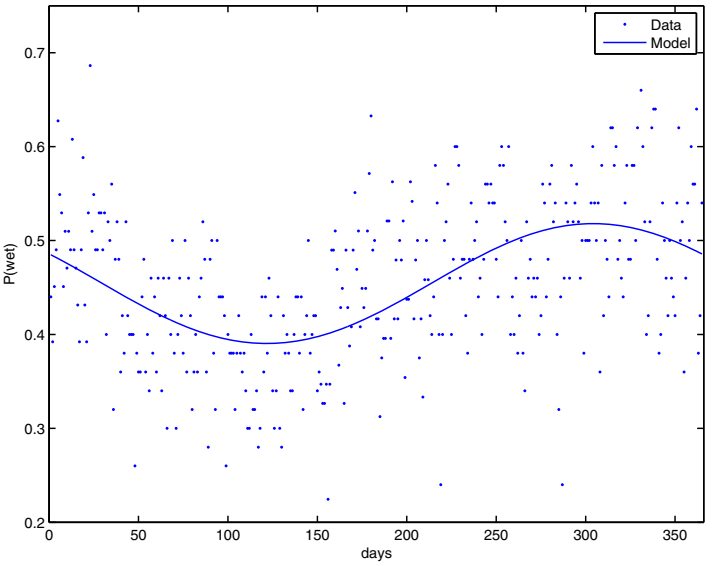

Figure 2. Empirical and model estimates of $\Phi^{-1}(\widehat{p}_{\mathbf{s}_1,t})$.

$$\widehat{\mu}(\mathbf{s},t)=\widehat{\mu}_0(\mathbf{s})+\widehat{\mu}_1(\mathbf{s})\sin\left(\frac{2\pi j}{365}t+\widehat{\mu}_2(\mathbf{s})\right),\quad t=1,\ldots,365/366. \tag{8}$$

Since direct estimates of the mean are available only at the weather stations, as is the case with the rest of the model parameters too, the estimates are interpolated to ungauged locations by regressing the spatial coefficients in (8) on position covariates

$$\widehat{\mu}_i(\mathbf{s})=\delta'_{\mu_i}w(\mathbf{s}),\quad i=0,1,2, \tag{9}$$

where $\delta_{\mu_i}=(\delta_{\mu_i,0},\delta_{\mu_i,1},\delta_{\mu_i,2},\delta_{\mu_i,3})'$ with $\delta_{\mu_i,0}$ being the intercept and $w(\mathbf{s})$ the covariates (1, latitude, longitude, altitude). The two steps in the estimation procedure could probably be combined in one by building a composite likelihood, although this was not explored any further. If available one could additionally include such covariates as climate model output or broad scale atmospheric conditions. Additional covariates could also include slope, vegetation, orientation, or distance from the coast. A different approach was adapted in *Kleiber et al*. [2012], where the authors assumed that the parameter estimates were themselves a realization of some random field.

In Figure 2, we present the proportion of wet days with the model probabilities superimposed for one example location, the city of Halmstad. The cyclic behavior of the series is apparent. In Figure 3, we have gathered the observed and model proportions of wet days for stations 1–12 as estimated for the 15th of each month. Notice that while in January we have precipitation for about half of the time this reduces to a little more than a third of the time during July. Although we consider only linear functions, this basic model exhibits significant flexibility and replicates properly the probability of precipitation at most stations.

3.3. Hybrid Gamma and Generalized Pareto Distribution

The hybrid gamma with the GP distribution is fitted to data in the Peaks over Threshold (POT) framework using a gamma distribution for the values that fall below a given threshold u and the GP distribution for the positive excesses of the u level.

Selection of threshold u is a difficult task in practice since given a data set it is difficult to pinpoint the level at which the extreme value theory is based, and its selection is usually based on the stability properties of the generalized Pareto distribution and the mean residual life plots [see for example, *McNeil*, 1997; *Davison and Smith*, 1990; *Frigessi et al*., 2002]. To circumvent this problem, *Vrac and Naveau* [2007] and *Hundecha et al*. [2009] suggested the use of a dynamic mixture of gamma and generalized Pareto distribution. The drawback with this model is that its application is constrained by functional complexity, numerical instability, and computational demand [see *Li et al*., 2012]. In this work, we have decided to use a fixed level of $u=10$ mm. We recognize that this is a strong assumption but this threshold worked reasonably well for all stations and months and we do not want to delve further into the problem of threshold selection, which is a difficult problem that deserves further attention.

After the threshold level is chosen, a gamma distribution is fitted assuming that data are independent and maximizing a modified version of the likelihood

$$L_{\mathbf{s}}=\Pi_{\{(\mathbf{s},t):Y(\mathbf{s},t)>0\}}p(Y(\mathbf{s},t);\alpha(\mathbf{s},t),\gamma(\mathbf{s},t)), \tag{10}$$

with

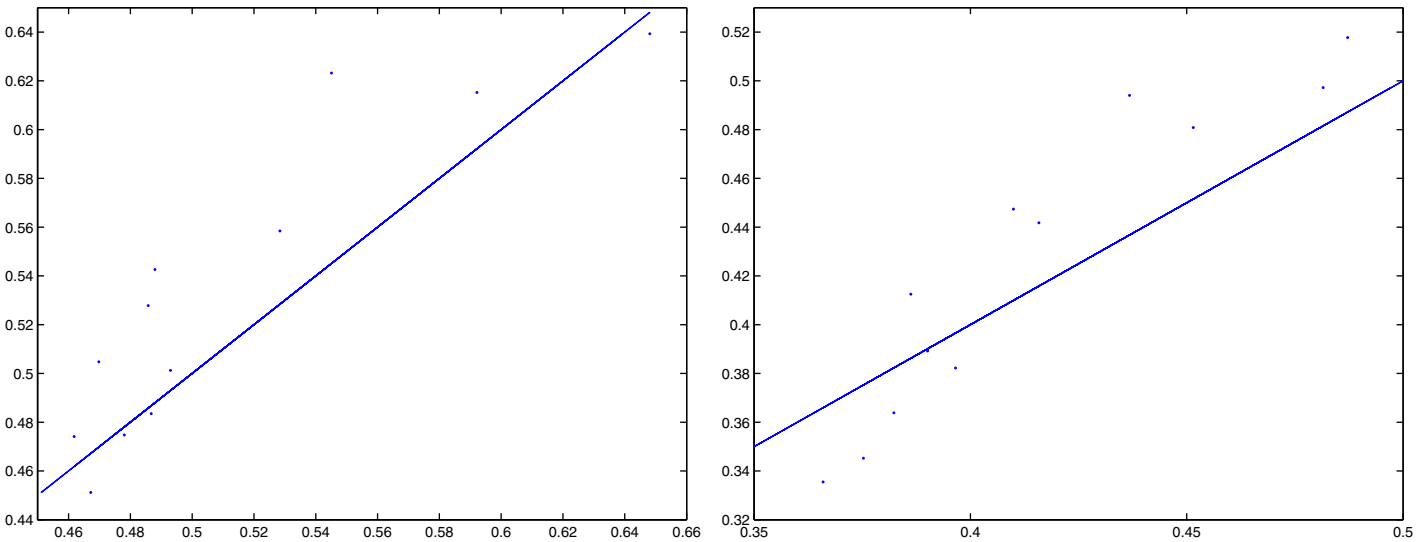

Figure 3. Observed proportion of wet days (*y* axis) and model probability of wet day (*x* axis) on (left) 15 January and (right) 15 July.

$$p(x;\alpha,\gamma)=\begin{cases} f_\gamma(x;\alpha,\gamma), & x\le u \\ \int_u^\infty f_\gamma(x;\alpha,\gamma)dx, & x>u, \end{cases} \tag{11}$$

where $f_\gamma(x;\alpha,\gamma)$ is the pdf of a gamma distribution with parameters α and γ. The modified likelihood with the pdf in (11) is preferred over the usual one, since in this way, the excesses of threshold u are treated as censored and the information they contain is not cast away. *Durban and Glasbey* [2001] used a similar approach to obtain estimates for the trend of censored values. *Ma*ximizing (10) results to estimates of the gamma parameters at the gauge locations. The observed seasonality is accounted for by fitting Fourier harmonics to obtain

$$\widehat{\alpha}(\mathbf{s},t)=\widehat{\alpha}_0(\mathbf{s})+\widehat{\alpha}_1(\mathbf{s})\sin\left(\frac{2\pi}{365}t+\widehat{\alpha}_2(\mathbf{s})\right),\quad t=1,\ldots,365/366 \tag{12}$$

and

$$\widehat{\gamma}(\mathbf{s},t)=\widehat{\gamma}_0(\mathbf{s})+\widehat{\gamma}_1(\mathbf{s})\sin\left(\frac{2\pi}{365}t+\widehat{\gamma}_2(\mathbf{s})\right),\quad t=1,\ldots,365/366. \tag{13}$$

Interpolation to the rest of locations, where no measurements are available, is again by regressing on spatial covariates, i.e.,

$$\widehat{\alpha_i}(\mathbf{s})=\delta'_{\alpha_i}w(\mathbf{s})\quad\text{and}\quad\widehat{\gamma_i}(\mathbf{s})=\delta'_{\gamma_i}w(\mathbf{s}),\quad i=0,1,2, \tag{14}$$

where $\delta_{\alpha_i}, \delta_{\gamma_i}$, and $w(\mathbf{s})$ are as in (9).

Next, the generalized Pareto distribution is fitted to the positive excesses. Dependence in the excesses is reduced by declustering the data and using only one excess per cluster. Clusters are determined using the extremal index, which is defined as the reciprocal of the mean cluster size, and is estimated by log likelihood [see *Suveges and Davison*, 2010, equation (3)]. Running the algorithm on the data using threshold level of 10 mm resulted to cluster lengths ranging from 2 to 8 days with a mean value of 4.7 days. Once data are declustered the shape $\xi(\mathbf{s},t)$ and scale $\sigma(\mathbf{s},t)$ parameters are estimated by maximizing the likelihood

$$\tilde{L}_{\mathbf{s}}=\Pi_{\{(\mathbf{s},t):Y(\mathbf{s},t)\in A\}}p(Y(\mathbf{s}',t);u,\xi(\mathbf{s},t),\sigma(\mathbf{s},t)),$$

where A is a set that contains a single excess per cluster per location and

$$p(x;u,\xi,\sigma)=\frac{1}{\sigma}\left(1+\frac{\xi(x-u)}{\sigma}\right)^{\left(-\frac{1}{\xi}-1\right)}.$$

The generalized Pareto estimates could also be regressed on Fourier harmonics for time dynamics with spatially dependent parameters but as it turns out for the data set in Sweden, this is not necessary since they do not exhibit any cyclic behavior. Spatial interpolation of the generalized Pareto parameters to locations with no observations is by using a regression like the one in (9).

3.4. Covariance Structure

To fully characterize the latent Gaussian random field $Z(\mathbf{s},t)$ what remains is to specify the covariance function $Cov(Z(\mathbf{s}_1,t_1),Z(\mathbf{s}_2,t_2))=C(\mathbf{s}_1,\mathbf{s}_2;t_1,t_2)$.

For the dependence over space and time, we follow the traditional geostatistics paradigm of assuming a parametric covariance function. For simplicity, we have considered only stationary and isotropic covariance functions. While the assumption of stationarity does not seem to be very restrictive, the area of southern Sweden seems to be homogeneous enough, this is not the case with spatial isotropy. The isotropy can be partly corrected by forming an anisotropic covariance function by applying to the isotropic C a non-Euclidean measure formed as Euclidean distance in a linearly transformed spatial coordinate system, as can be seen in section 3.4.1. Additionally as can be clearly seen in the data, precipitation is moving from south/south-west to north/north-east, hence the covariance function should also include some kind of dynamics, a feature that is not accounted for in the present choice of covariance structure but should be further investigated.

Several parametric functions specifying explicitly the joint space-time covariance structure have been fitted. For a collection of valid spatiotemporal covariances, see *Cressie and Huang* [1999], *Ma* [2003], and *Stein* [2005] among others. The covariance structure that gave the best fit in terms of Weighted Least Squares (WLS) criterion is the following:

$$C(\mathbf{h},\tau)=Cov(Z(\mathbf{s},t),Z(\mathbf{s}+\mathbf{h},t+\tau))=\eta\{||\mathbf{h}||=0,\tau=0\}+\frac{(1-\eta)}{a|\tau|+1}e^{-\frac{b||\mathbf{h}||^2}{a|\tau|+1}}, \tag{15}$$

where $\eta>0$ and $\{A\}$ stands for the indicator function that equals one when property A is satisfied and zero otherwise. The function $\eta\{\cdot,\cdot\}$ models the nugget effect which allows for a discontinuity at zero, and is used to account for the undistinguishable microscale variability and measurement error. The nonnegative parameters a and b are the scaling parameters of time and space, respectively, controlling the degree of dependence. For temporal lag equal to 0, the spatial covariance simplifies to Gaussian which allows the latent field to be smooth over space.

3.4.1. Estimation of the Correlation

Estimation of the correlation structure of the latent Gaussian field Z (we assume the random field has unit variance) is challenging for two distinct reasons. First, for dry days, the Gaussian variable is censored and second, there exist no direct observations of the Gaussian variable.

One way of estimating the correlation coefficients at a particular time lag for a censored Gaussian variable is by numerically maximizing the likelihood of the observed bivariate histogram of the censored latent variable [see *Glasbey and Nevison*, 1997] or by maximizing a modified version of the likelihood for censored values [see *Durban and Glasbey*, 2001]. *Guillot* [1999] proposed a method of estimating the covariance function of the latent field which consists of computing the empirical covariance of raw data, then fitting a positive definite function to it, computing the inverse of this function through the Hermite polynomial expansion of the anamorphosis function, fitting a positive definite function to it, and then reversing it again with the use of the same Hermite polynomial expansion of the anamorphosis function to finally obtain an estimate of the desired covariance. We propose an alternative method of moments approach, which we feel is simpler than the method proposed in *Guillot* [1999], by inverting the theoretical expression for the mean of the censored cross product. We turn to this next.

The following relation holds for a bivariate Gaussian random variable $\mathbf{Z}=(Z_1,Z_2)$ with mean $\mu=(\mu_1,\mu_2)$ and correlation ρ (unit variances):

$$EZ_1^+Z_2^+=\int_0^\infty g(x;\mu,\rho)\phi(x)\ dx, \tag{16}$$

where $Z_i^+=max(Z_i,0)$, ϕ denotes the pdf of a standard normal random variable and the function g is given by

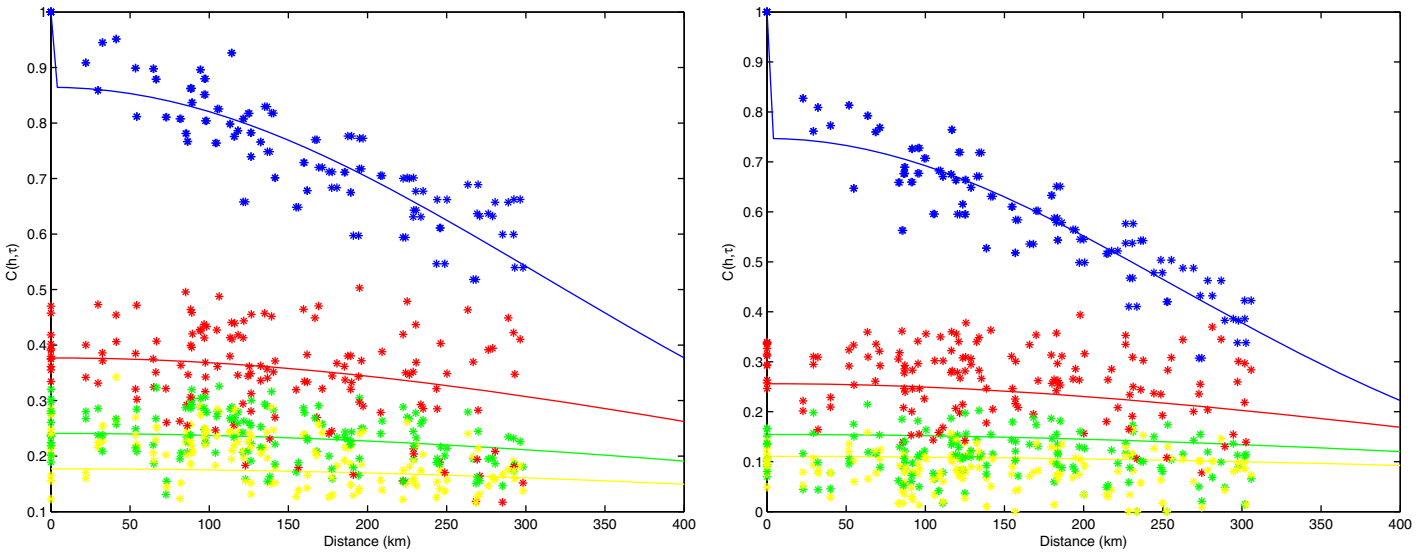

Figure 4. Spatiotemporal covariance function (15) (*lines*) for the latent Gaussian field with empirical covariances (⋆) for (left) January and (right) July for time lags $\tau=0, 1, 2, 3$ from top to bottom.

$$g(x;\mu,\rho)=x\phi(x-\mu_1)\left[(\rho(x-\mu_2)+\mu_1)\Phi\left(\frac{\mu_1+\rho(x-\mu_2)}{\sqrt{1-\rho^2}}\right)+\sqrt{1-\rho^2}\phi\left(\frac{\mu_1+\rho(x-\mu_2)}{\sqrt{1-\rho^2}}\right)\right]. \tag{17}$$

Therefore, using (16), correlation $\rho_{ij}(\tau)$ between locations $\mathbf{s}_i$ and $\mathbf{s}_j$ and time lag τ can be estimated by minimizing

$$min_\rho\left|\overline{z_i^+\cdot z_j^+}-\int_0^\infty g(x;\mu_{ij},\rho_{ij})\phi(x)\ dx\right|, \tag{18}$$

where $\overline{z_i^+\cdot z_j^+}$ denotes the average of the product of the censored values at locations indicated by the subindices for the given τ and μ_{ij} denotes the vector $(\mu(\mathbf{s}_i),\mu(\mathbf{s}_j))$. Note that function g is not a simple function of pairwise correlations which can be analytically inverted. The integral in (18) needs to be computed using numerical integration.

Once estimates $\widehat{\rho}_{ij}(\tau)$ of the pairwise correlations (which, since variances equal one, coincide with $\widehat{Cov}(Z(\mathbf{s}_i,t),Z(\mathbf{s}_j,t+\tau))$ are obtained, the covariance parameters are estimated through the following method of moments minimization:

$$min_{\eta,a,b,\eta,\theta}\sum_t\sum_{i\neq j}n_{obs}(i,j,\tau)\left(\widehat{\rho}_{ij}(\tau)-Cov_M(Z(\mathbf{s}_i,t),Z(\mathbf{s}_j,t+\tau))\right)^2, \tag{19}$$

where Cov_M is the corresponding model covariance and n_{obs} is the number of observations used in the estimation in which varies with location and time of the year.

As model covariance Cov_M, we use the covariance C in equation (15) but with a non-Euclidean distance measure formed as Euclidean distance in a linearly transformed spatial coordinate system. The new coordinate axis is obtained by rotating the old isotropic axis x, y by an angle θ, forming new coordinates

$$\begin{pmatrix}x'\\y'\end{pmatrix}=\begin{pmatrix}\cos\theta & \sin\theta\\-\sin\theta & \cos\theta\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix} \tag{20}$$

and then dilating them by a factor ϵ to form the new coordinates

$$\begin{pmatrix}x''\\y''\end{pmatrix}=\begin{pmatrix}\sqrt{\epsilon} & 0\\0 & 1/\sqrt{\epsilon}\end{pmatrix}\begin{pmatrix}x'\\y'\end{pmatrix}. \tag{21}$$

The distance **h** in the new coordinate system is transformed to

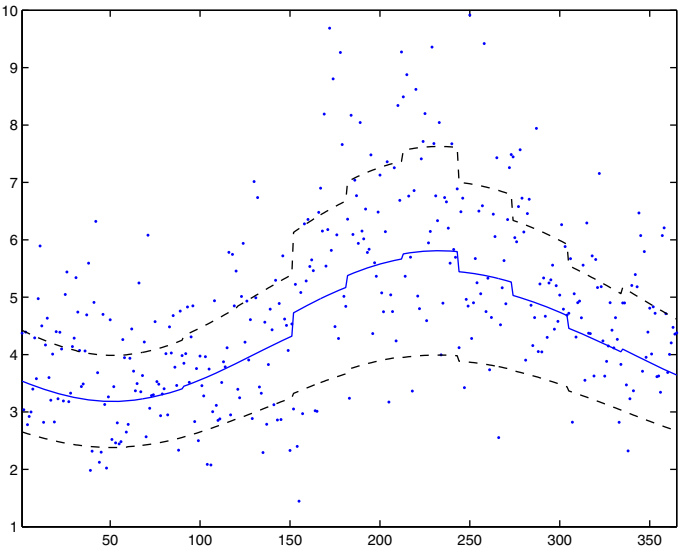

Figure 5. The observed and model daily mean intensity of rainfall during wet days with corresponding approximate 90% confidence interval at Halmstad.

$$||A_{\theta,\epsilon}\mathbf{h}|| = \sqrt{\epsilon(x\cos\theta + y\sin\theta)^2 + \frac{1}{\epsilon}(-x\sin\theta + y\cos\theta)^2}, \tag{22}$$

with isotropy attained for $\epsilon = 1$. There is some nonuniqueness in this formulation since adding π to the anisotropy angle θ gives an identical correlation, and correlation with anisotropy angle $\theta + \frac{\pi}{2}$ and factor $\frac{1}{\epsilon}$ is identical to correlation with angle θ and factor ϵ. This nonuniqueness can be removed by adding some constraints on the ranges of the parameters [see *Haskard*, 2007].

To obtain estimates of the censored values, z^+ we transform the observed precipitation amounts y, as follows:

$$z^+ = \Phi_{\widehat{\mu}}^{-1} \circ F_{emp}(y), \tag{23}$$

with $F_{emp}(\cdot)$ denoting the empirical cdf of the precipitation intensity during each month and Φ_μ as in (4). Then, having obtained estimates $\widehat{\rho}_{ij}$ for different time lags using (18), we fit the model covariance function C in (15) to obtain estimates of the parameters by minimizing (19). The constant $n_{obs}(i,j,\tau)$ equals the number of observations used in the estimation of the empirical covariance between station $\mathbf{s}_i$ on some day and station $\mathbf{s}_j$ after τ days. The minimization is performed for each time lag τ separately and results in a set of parameters $[a, b, \eta, \theta, \epsilon]$. This procedure resulted in a set of 12 (months)$\times$5.

To illustrate the fitting of the spatiotemporal covariance function, Figure 4 shows the fitted spatial covariance function for different time lags together with the empirical covariances for the two example months of January and July. Our method of moments approach to estimating the parameters of the covariance shows reasonable performance. A simulation study showed that the method of moments approach performed considerably better than the modified maximum likelihood (MLE) used in *Durban and Glasbey* [2001] in terms of bias of the estimated correlation. The modified MLE severely overestimated the correlation when the dependence between censored bivariate Gaussian data was strong but performed reasonably well when data were independent. On the other hand, the method of moments approach performed equally well for any type of dependence between data.

4. Model Validation

In this section, we validate the performance of the spatiotemporal stochastic model introduced in section 3. For this, we generate 100 trajectories of the 51 years of data using the proposed stochastic generator and then we examine the results.

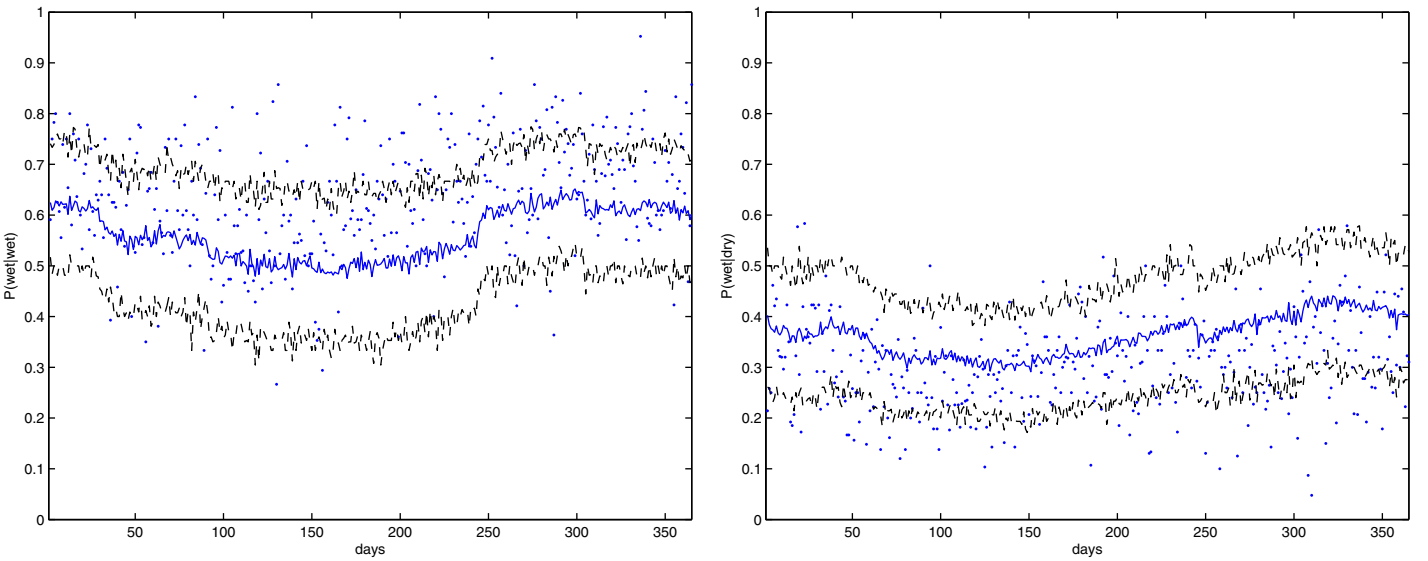

Figure 6. The observed (*dots*) and simulated (*line*) proportions of wet day given previous (left) wet and (right) dry days at Halmstad, with a pointwise 90% confidence interval based on 100 simulations.

4.1. Simulations

To simulate precipitation data at site **s** on day t, we perform the following steps:

1. A realization of a normal random field $Z(\mathbf{s},t)$ with mean function $\widehat{\mu}$ and covariance $\widehat{C}$ is generated.
2. For every location and day, there is zero precipitation if $Z(\mathbf{s},t) \leq 0$.
3. For location and day with positive precipitation, the simulated intensity is set to $Y(\mathbf{s},t)=\widehat{\psi}(Z(\mathbf{s},t))$, where $\widehat{\psi}$ is estimated according to the method described in section 3.3.

4.2. Temporal Model

Mean intensity of precipitation, i.e., the average amounts of precipitation over days with positive precipitation, is a fundamental measure the stochastic weather generators attempt to replicate. Figure 5 presents the observed and model mean intensities of precipitation together with an approximate 90% confidence

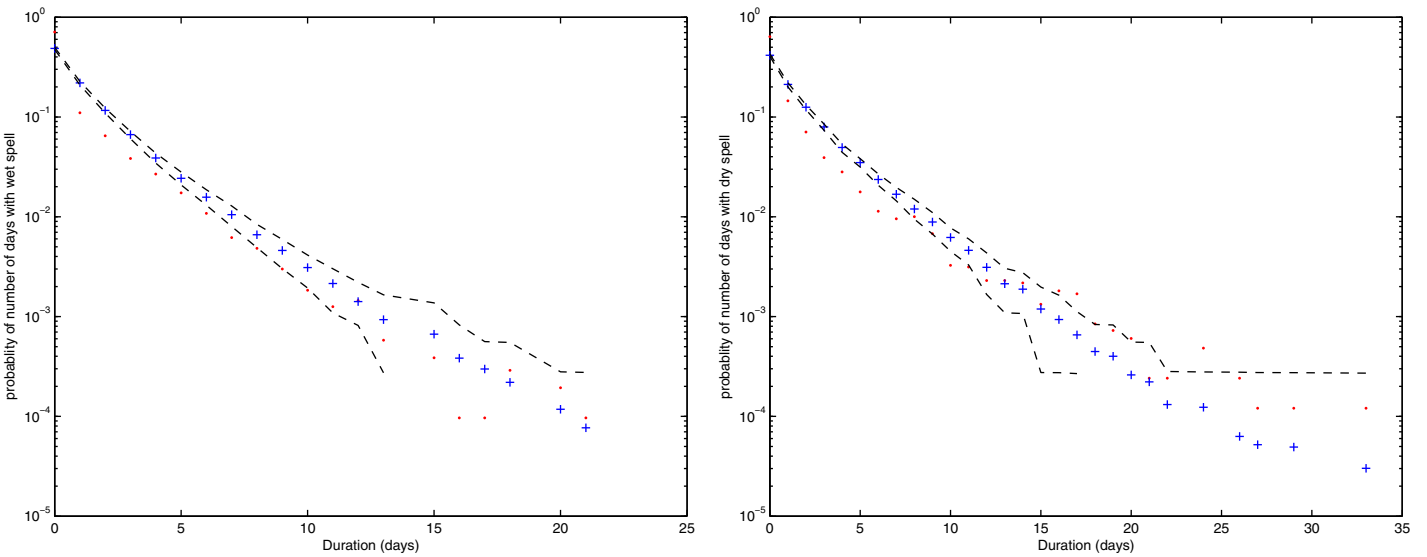

Figure 7. The empirical distribution (dots) of number consecutive (left) wet and (right) dry days, and the ones based on simulations (+) with a 90% confidence interval superimposed at Halmstad.

Table 1. The Weather Indices Concerning Precipitation Defined by ETCCDI

Index	Description
$RX1day_j$	Highest 1 day precipitation amount in month *j*
$RX5day_j$	Highest 5 day precipitation amount in month *j*
R20mm	Number of annual days with precipitation above 20 mm
CDD	Max annual number of consecutive dry days
CWD	Max annual number of consecutive wet days
PRCPTOT	Annual total precipitation

interval computed using the estimated $\widehat{G}$ distribution in (2), for a specific location, the city of Halmstad. The marginal distribution at individual locations reproduces the mean intensity very well. It can be noticed that the variability of the intensity of precipitation is higher during days 150–300, which correspond to the Swedish summer and fall. Similar plots were obtained for the rest of months and weather stations.

We have already seen in Figure 2 that the stochastic generator accurately models the unconditional probability of precipitation over the year. As is well known, the previous day's property of dry or wet greatly influences next day's weather [see e.g., *Lennartsson et al.*, 2008]. For this reason, transition probabilities with previous dry and wet days are estimated. Figure 6 illustrates these observed and simulated transition probabilities at Halmstad together with a pointwise 90% confidence interval based on 100

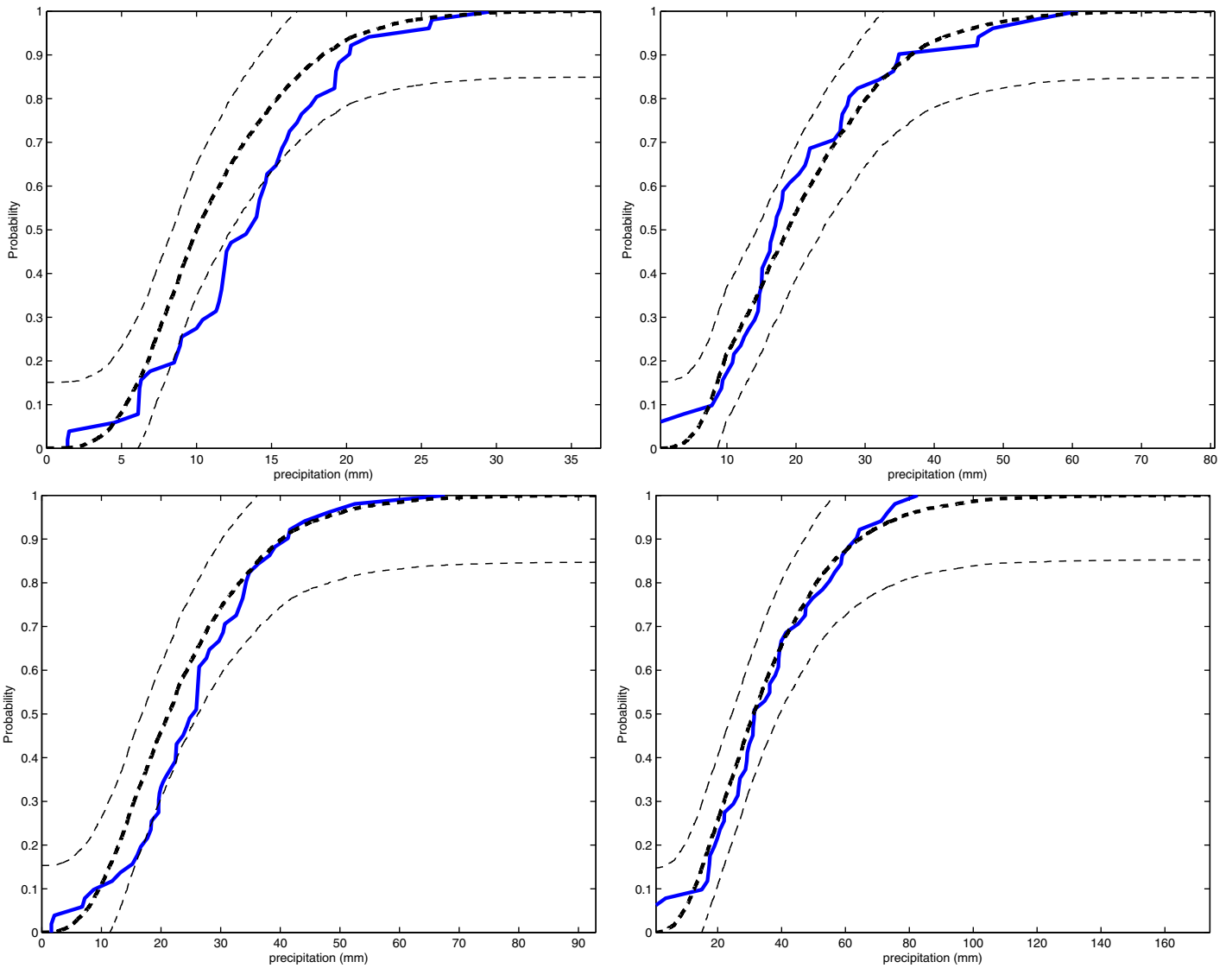

Figure 8. The empirical distribution (*line*) of (top) RX1day and (bottom) RX5day in (left) January and (right) July, and the one based on simulations (*dashed*) with a 90% confidence interval superimposed at Halmstad. (Years with more than 10 days missing were removed. Same holds for Figures 9 and 10.)

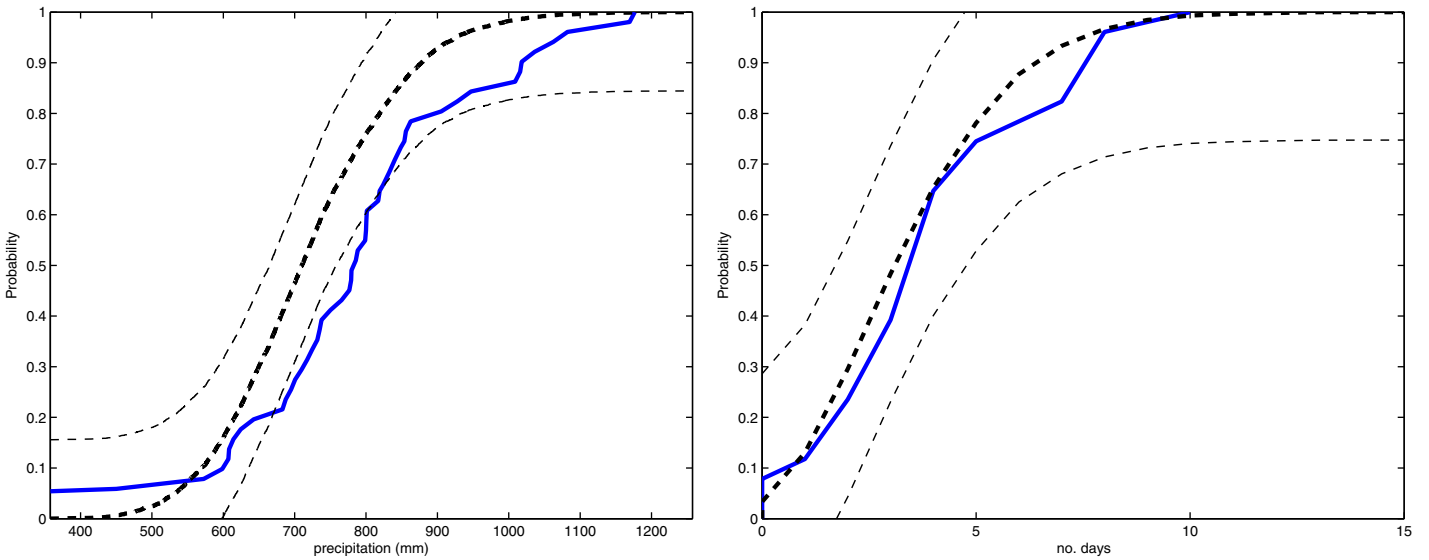

Figure 9. (left) The empirical distribution (*line*) of PRCPTOT, and the one based on simulations (*dashed*) with a 90% confidence interval superimposed at Halmstad. (right) The empirical distribution (*line*) of R20mm, and the one based on simulations (*dashed*) with a 90% confidence interval superimposed at Halmstad.

simulations. As it can be seen the probability of next day to be wet is about 0.2 higher if present day is wet than if it is dry. The weather generator underestimates the transition probability of a wet day given the previous day is wet and it overestimates the corresponding probability given previous day is dry. Since the simulated proportions depend on the choice of level u, the threshold of the GP distribution, the estimated parameters of the marginal distribution as well as the mean function and temporal correlation, there are a lot of possible sources of error which could contribute to the differences from the observed proportions.

We turn now to the dry/wet behavior of the stochastic generator. We remind that the distribution of the length of the wet and dry spells, i.e., the time the precipitation is above zero or equal to zero, respectively, is an essential feature of any weather generator. Notice that these characteristics depend solely on the latent process since they coincide with the time the process spends above (below) the zero level. The

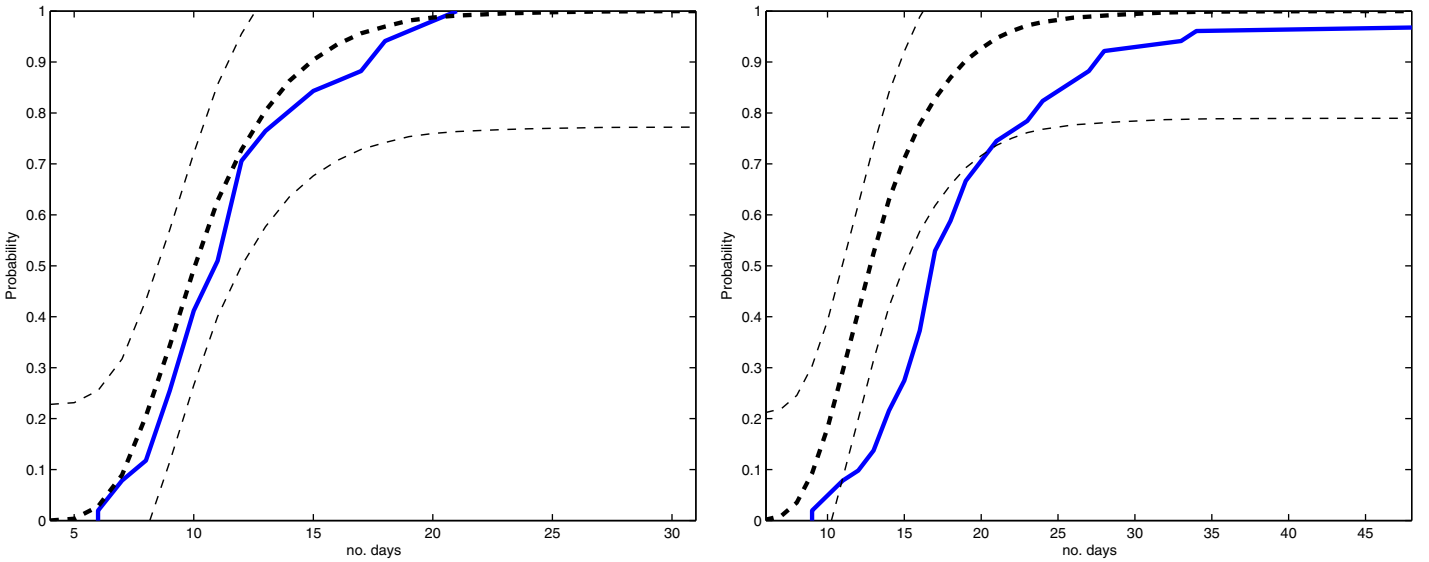

Figure 10. The empirical distribution (*line*) of (left) CWD and (right) CDD and the ones based on simulations (*dashed*) with a 90% confidence interval superimposed at Halmstad.

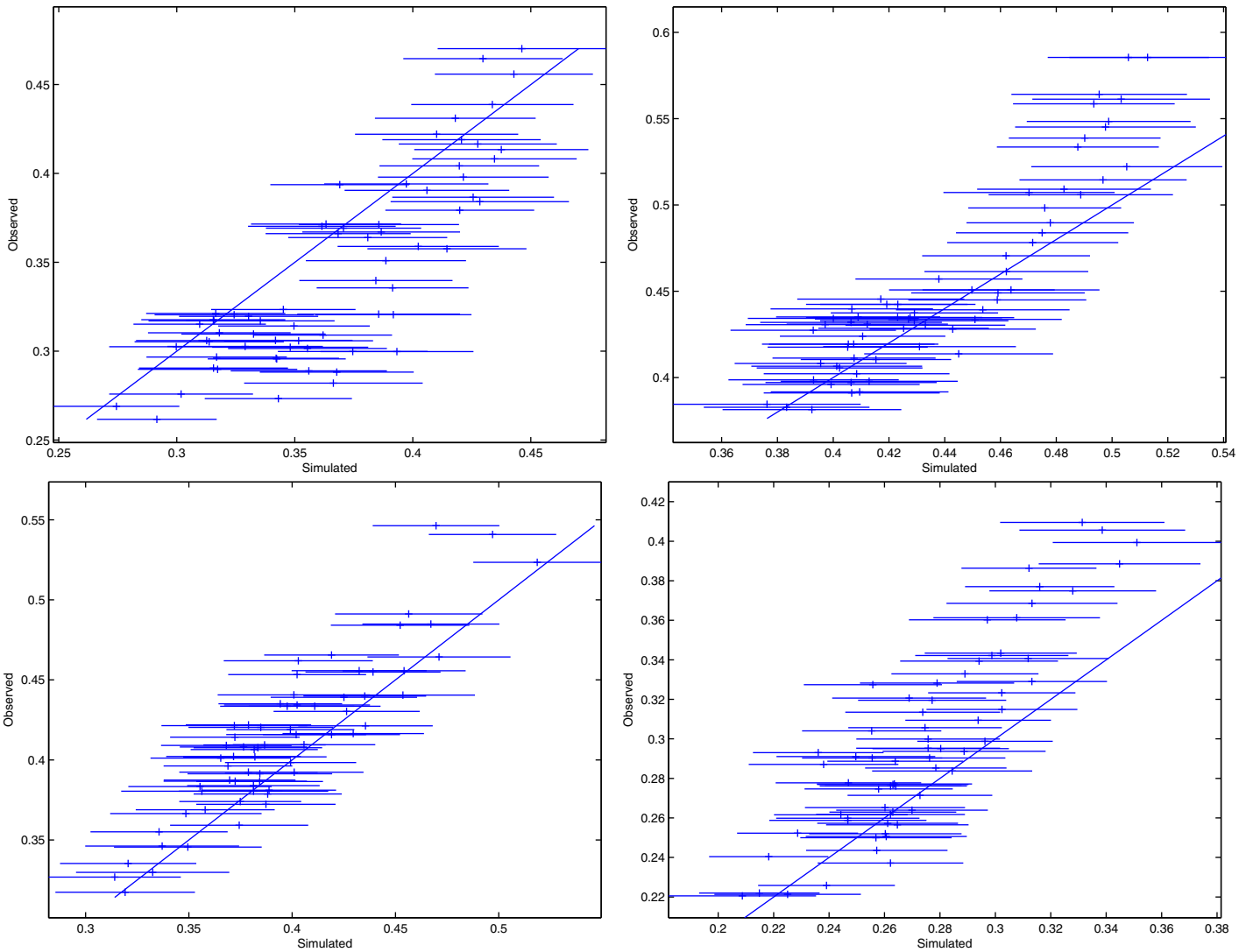

Figure 11. The proportions of (top) simultaneously dry and (bottom) simultaneously wet days in (left) January and (right) July. The proportions of observations are given on *y* axis with the corresponding simulated proportions on the *x* axis with a 90% confidence interval based on 100 simulations given by lines.

empirical distribution of wet (*left*) and dry (*right*) spells can be seen in Figure 7 together with the model distribution and a 90% confidence interval superimposed. The length of the wet spell is very well replicated for any spell length, and the same is true for most of the dry spell lengths.

Finally, we check if the proposed stochastic generator replicates properly extreme weather. In order to quantify extreme properties of weather, the joint work-group CCI/CLIVAR/JCOMM Expert Team on Climate Change Detection and Indices (ETCCDI) (Web page http://etccdi.pacificclimate.org/list_27_indices.shtml) [see also *Karl et al*., 1999; *Peterson et al*., 2001] have suggested a number of weather indices monitoring of which can be used for weather change detection. Six of these indices (see Table 1) are of relevance to this study. Two indices (CDD and CWD) consider only the dry/wet behavior of the precipitation process. The remaining indices (RX1day, RX5day, R20mm, and PRCPTOT) consider both the dry/wet behavior and the marginal distribution. Figures 8–10 illustrate the empirical distribution function of each index with the simulated distribution and 90% confidence interval superimposed.

Figure 8 shows the 1 day maximum (RX1day) and the 5 day accumulated maximum (RX5day) of precipitation for the months of January and July at Halmstad. The model seems to be working better for July than January, which maybe suggests that using the same threshold *u* for all months is not optimal. Note that the results for the RX5day index additionally indicate that the temporal dependence is properly modeled.

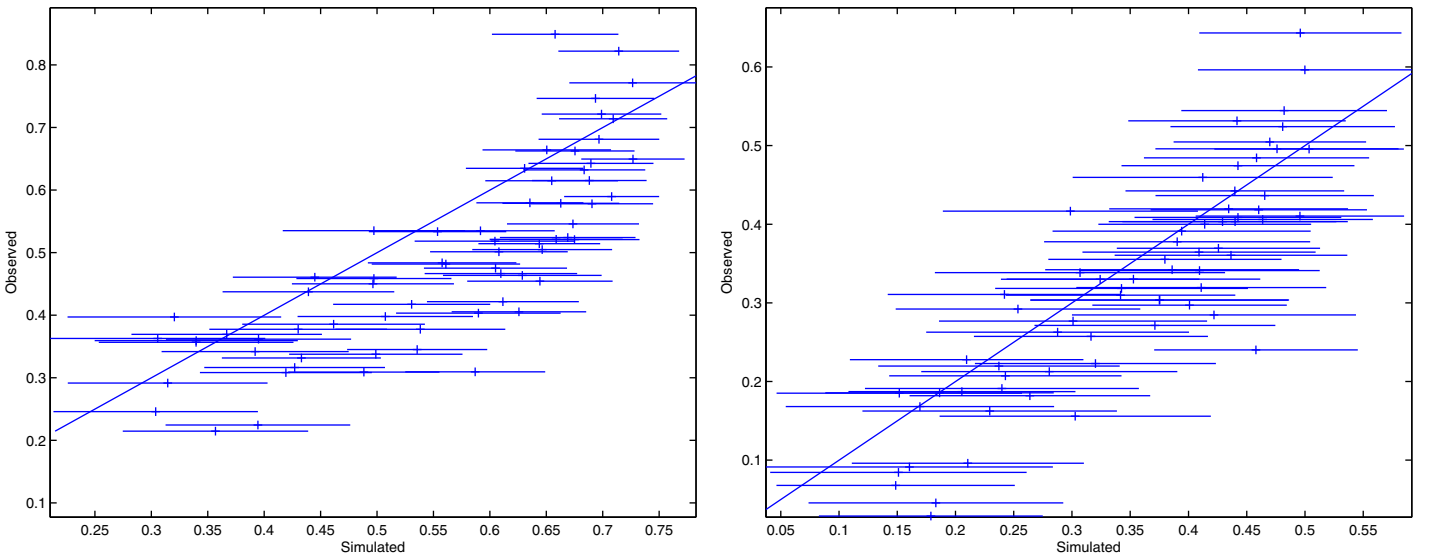

Figure 12. The correlation of intensity in (left) January and (right) July, where the observed correlations are given on *y* axis with the corresponding quantities for the simulations on the *x* axis with a 90% confidence interval given by lines.

Figure 9(left) illustrates the total amount of precipitation per year (PRCPTOT) at Halmstad. The stochastic generator seems to underestimate these quantities. Another essential feature that weather generators should be able to replicate accurately is the annual number of days with extreme weather, which depends on both the latent process and the transformation ψ. Figure 9(right) shows the number of days per year with precipitation above 20 mm at Halmstad. The fit of index R20mm is really good which in another indication that the latent process and the transformation ψ are properly modeled.

Figure 10 shows the maximum number of consecutive dry and wet days per year. The annual maximum number of consecutive dry days (CWD) is very well replicated by the stochastic generator (see Figure 10(left)). On the other hand, the model consistently overestimates the probability of the maximum number of consecutive dry days (see Figure 10(right)). Note that the model has not been tuned to reproduce these statistics.

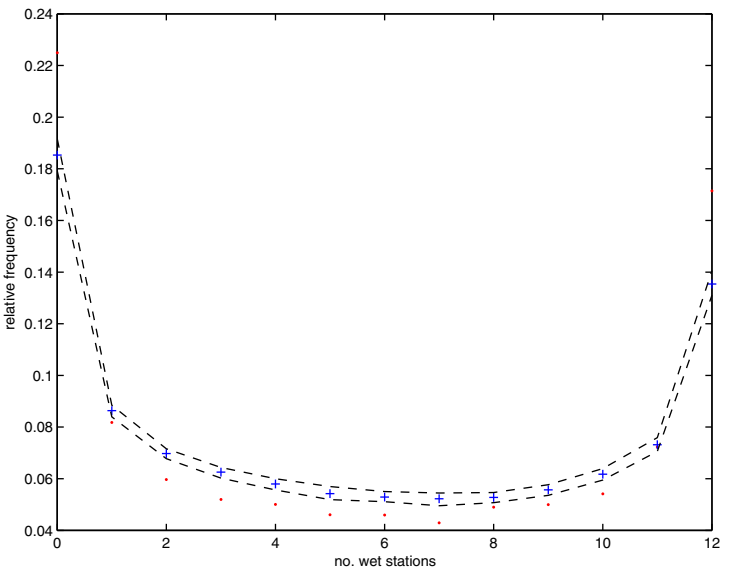

Figure 13. The proportions of observed number of stations that are simultaneously wet and the corresponding quantity for the stochastic model with a 90% confidence interval.

However, both CWD and CDD are very complex measures that only capture the extreme occurrences of wet and dry spells on an annual basis. Overall, the model seems to simulate the temporal behavior of precipitation at individual locations properly.

4.3. Spatial Model

A positive feature of the defined stochastic generator is in correlating occurrences and amounts of precipitation across space. In order to illustrate the fit of spatial dependence, we have considered all 66 pairs of stations and then computed the observed and model proportions of days that were simultaneously dry and simultaneously wet at each station pair in January and July. The

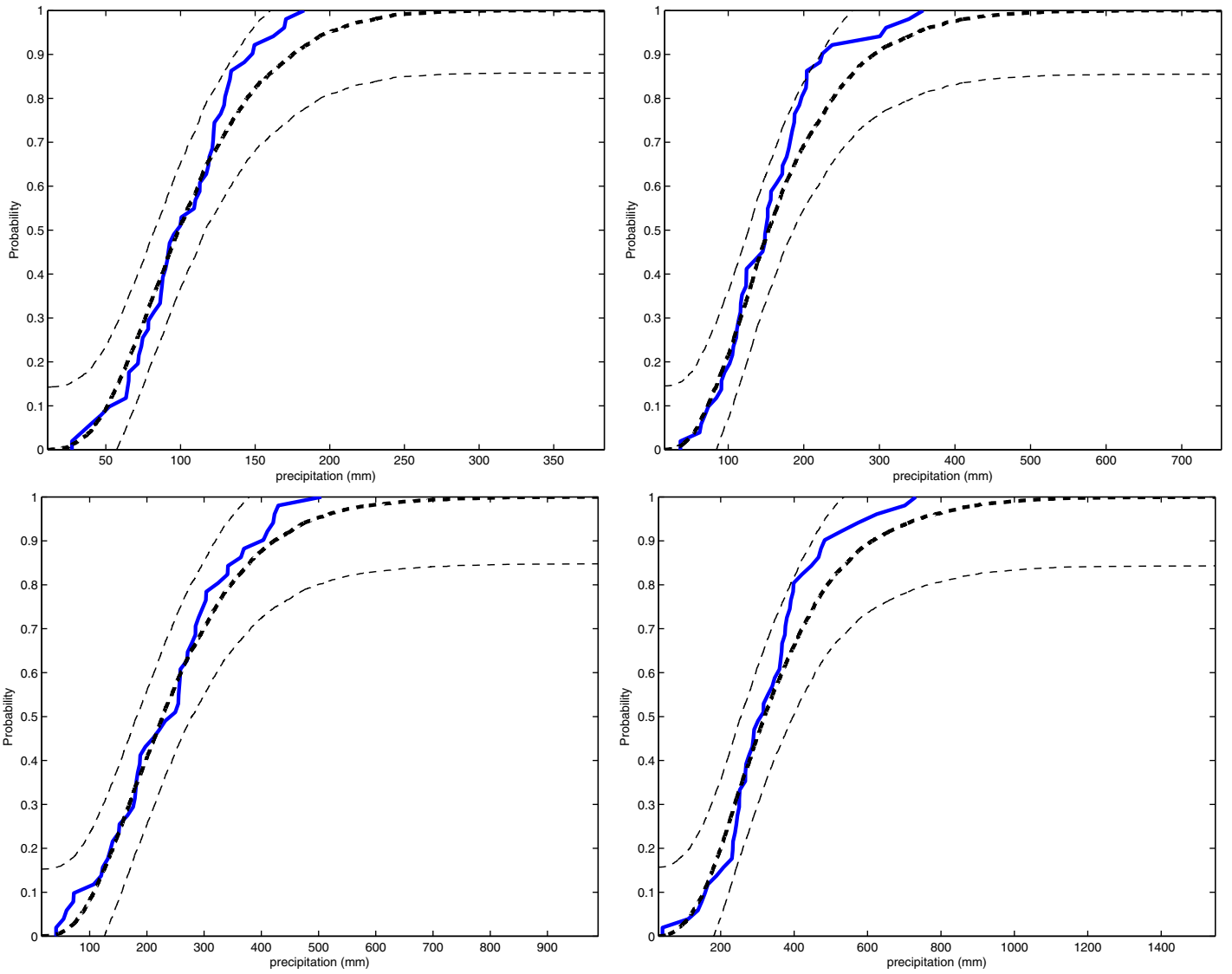

Figure 14. The empirical distribution (line) of (top) RX1day and (bottom) RX5day in (left) January and (right) July, for the domain aggregated data, together with the simulated (*dashed*) and a 90% confidence interval superimposed.

results are gathered in Figure 11. The latent Gaussian field seems to replicate well the simultaneous wet behavior and the simultaneous dry for July but not so well for January.

The stochastic generator's ability to adapt to seasons, due to monthly dependent covariance parameters, allows for a substantial increase in spatial correlation between two stations during winter as compared to summer season. It should not be surprising that the pairwise correlations in January are more variable than in July, since the nugget effect is substantially larger, and hence suppresses the spatial (and temporal) dependence for the summer months. Figure 12 displays the observed and model correlations of precipitation intensity for all station pairs. Note here that only days with simultaneous positive precipitation have been used to estimate the respective correlations. The stochastic generator seems to slightly overestimate the correlations for moderate values (0.4–0.7) during winter months, but overall the fit is satisfactory.

Another interesting feature of the spatial dependence structure is whether it properly replicates the number of wet stations each day. Figure 13 displays the observed proportion of number of stations with positive precipitation. Days where data are missing for at least one of the stations were removed. The stochastic generator seems to replicate quite well the observed frequency of total number of stations with

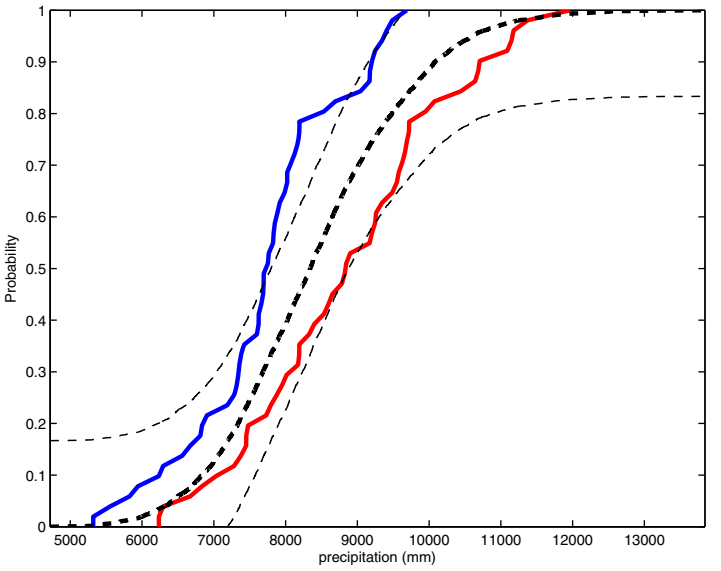

Figure 15. The empirical distribution (*line*) of PRCPTOT for domain aggregated precipitation and the corresponding distributions for the stochastic generator (*dashed*) with a 90% confidence interval based on the 100 simulations, where missing observations are replaced with only dry days (*blue*) or all days corresponding to the mean intensity (*red*).

simultaneously positive precipitation. It is interesting to notice that the probability to have no observed precipitation at any of the stations is about 0.22, while the probability to observe precipitation at all stations is about 0.17 and higher than the probability to observe precipitation in any subset of locations.

In general, we feel that the model replicates spatial aspects of the precipitation process well.

4.4. Spatiotemporal Model

The most difficult feature of multisite precipitation data is the spatiotemporal dependence structure. We have already seen in Figure 4 that the parametric covariance function of the latent Gaussian field fits well the empirical covariance for at least a few time lags. Following *Kleiber et al.* [2012], we validate the fit of the spatiotemporal structure by checking the domain aggregated behavior. The domain aggregated precipitation is the total precipitation from all stations at each given day. Further, a domain aggregated dry spell is defined as the number of days where all stations are dry and similarly, a domain aggregated wet spell is the number of days where at least one station is wet.

So we have computed the relevant weather indices for the domain aggregated data and the results are presented in Figures 14 and 15. The model, as was the case when data from each gauge were treated separately, replicates the weather indices of maximum 1 and maximum 5 day precipitation amount very well.

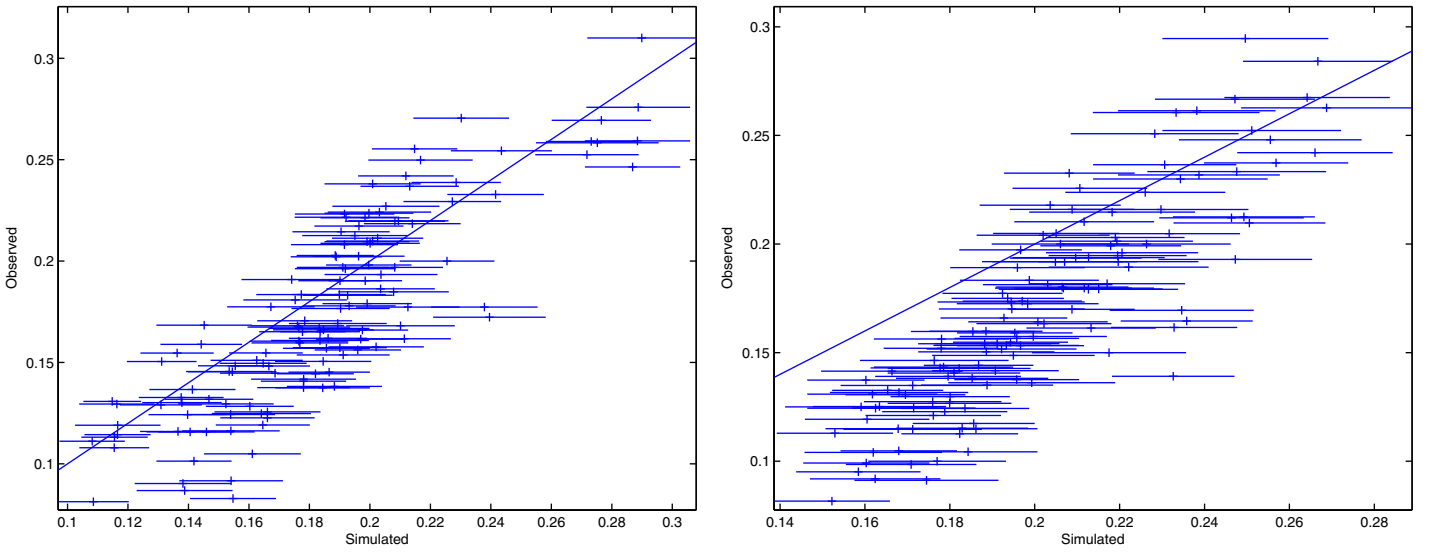

Figure 16. The pairwise lagged occurrence proportions in (left) January and (right) July, where the observed proportions are given on the *y* axis and the corresponding quantities for the stochastic generator are represented on the *x* axis with a 90% confidence interval for the model values given by lines.

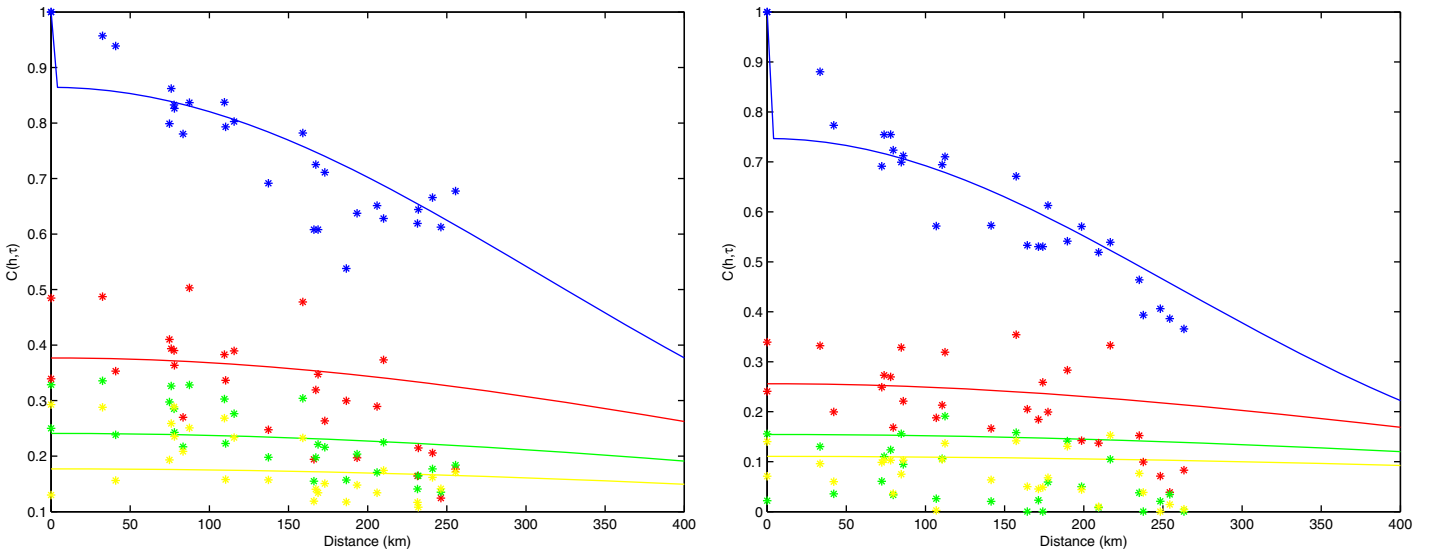

Figure 17. Spatiotemporal covariance function for the latent Gaussian field with empirical covariances for the validation set locations for (left) January and (right) July for time lags $\tau=0$, 1, 2, 3, 4 from top to bottom.

Turning now to the analysis of domain aggregated precipitation, we need to note that missing data may have a substantial influence on the different indices. For this reason, we have computed the distribution of PRCPTOT, having replaced all missing observations with dry days and then with wet days where the amount of precipitation equals the mean intensity. The results are gathered in Figure 15. As we can notice, the annual domain aggregated data are well replicated by the stochastic generator.

Due to missing data, the domain aggregated precipitation becomes too noisy which makes such indices as CWD and CDD as well as the domain aggregated dry and wet spells almost irrelevant.

An alternative way to the spatiotemporal aspects of the generated dry/wet behavior is by examining pairwise lagged occurrence probabilities defined as follows:

$$P(Y(\mathbf{s}_i,t-1)=0, Y(\mathbf{s}_j,t)>0) \text{ and } P(Y(\mathbf{s}_i,t-1)>0, Y(\mathbf{s}_j,t)=0). \quad (24)$$

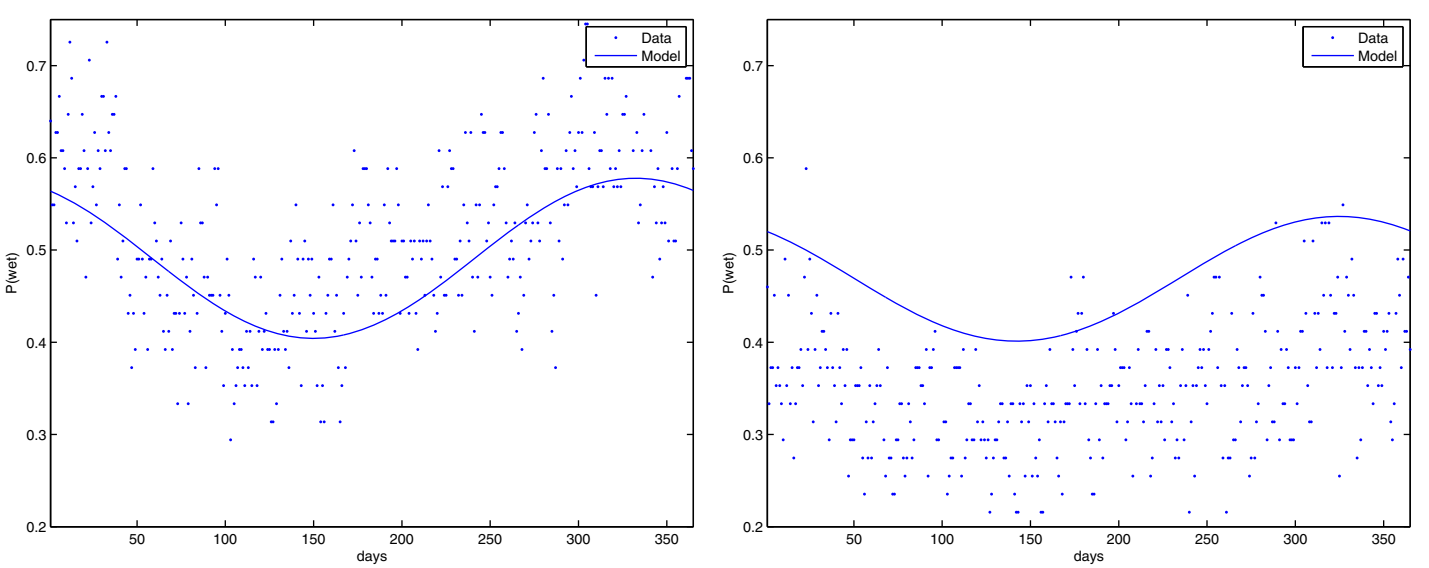

Figure 18. The interpolation model of daily probability of precipitation and the observed proportions of wet day for Borås and Malmslätt.

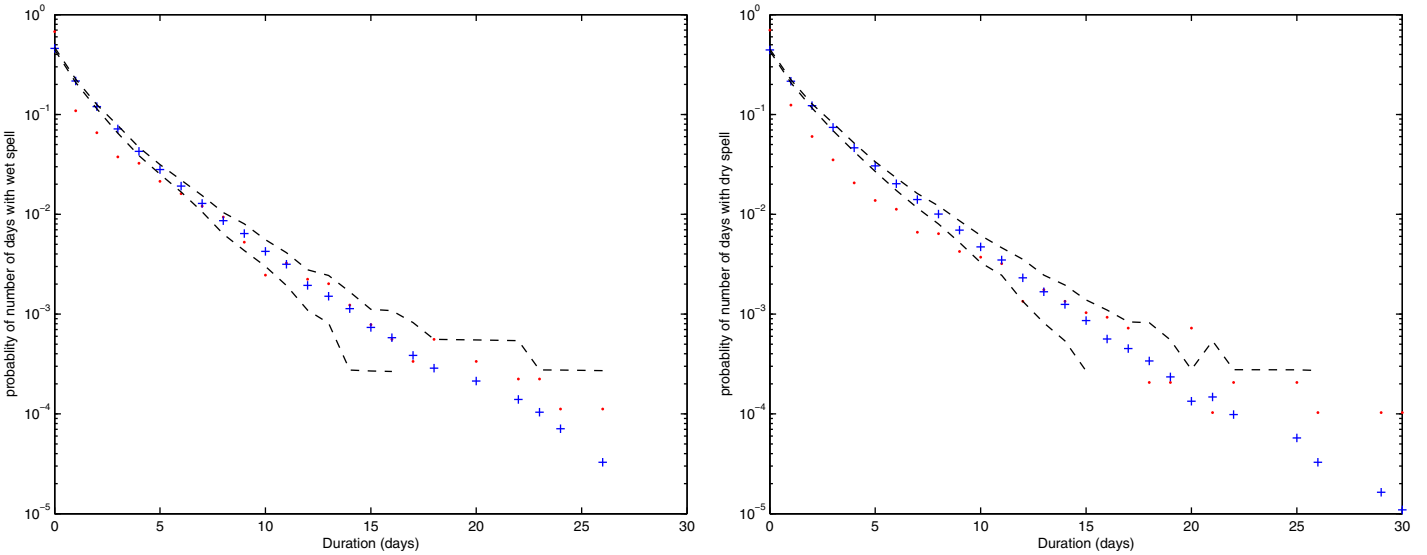

Figure 19. The empirical distribution of number of consecutive (left) wet and (right) dry days, and corresponding distributions for the stochastic generator with a 90% confidence interval at Borås.

Each point in Figure 16 represents observed and model proportions of either one of the pairwise lagged occurrences for January and July. The fit of the pairwise lagged occurrence probabilities in January is good, while the stochastic generator overestimates the quantity for small proportions in July. The deviation is only of order 0.02 and unlikely to have a significant impact in practice.

4.5. Cross-Validation

Finally, what remains to be validated is the ability of the stochastic generator to reproduce the precipitation behavior at locations with no available data. For this reason we have excluded two stations at Borås and Malmslätt, from the data that were used to tune the model parameters and now we check how the model performs at these two stations.

To illustrate the fitting of the spatiotemporal covariance function, Figure 17 shows the model and empirical spatial covariance function for a few time lags for these two stations. Having in mind that these data have

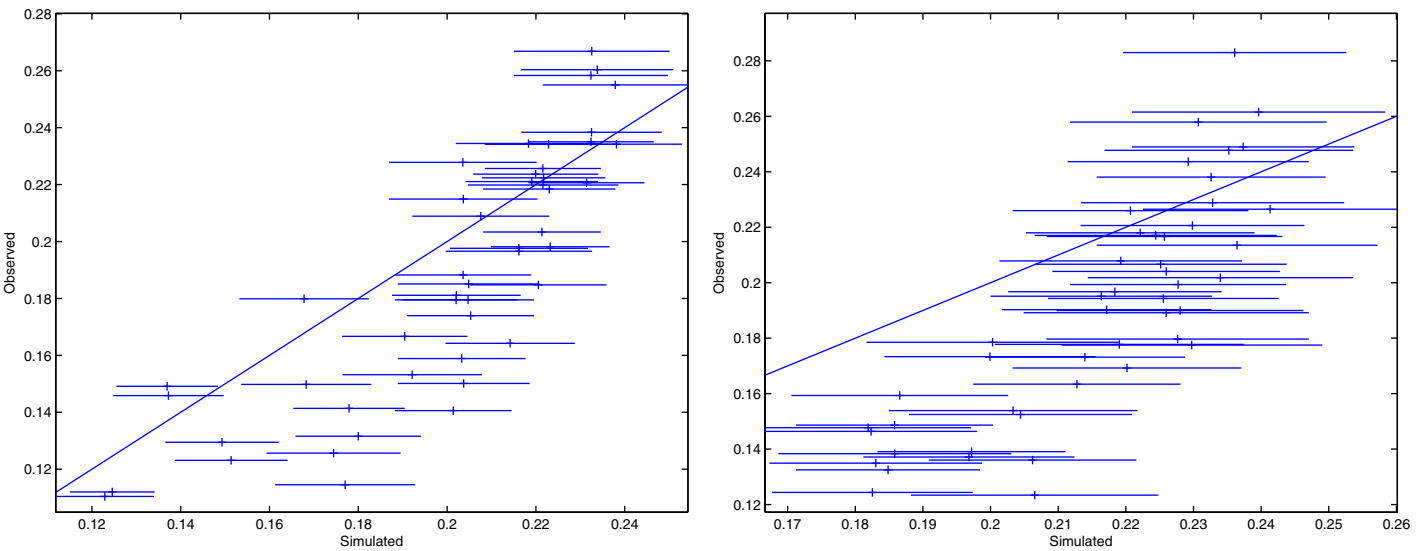

Figure 20. The observed and model pairwise lagged occurrence probabilities as defined in equation (24) for (left) January and (right) July for one station among stations 1–12 and one of the validation stations 13 and 14.

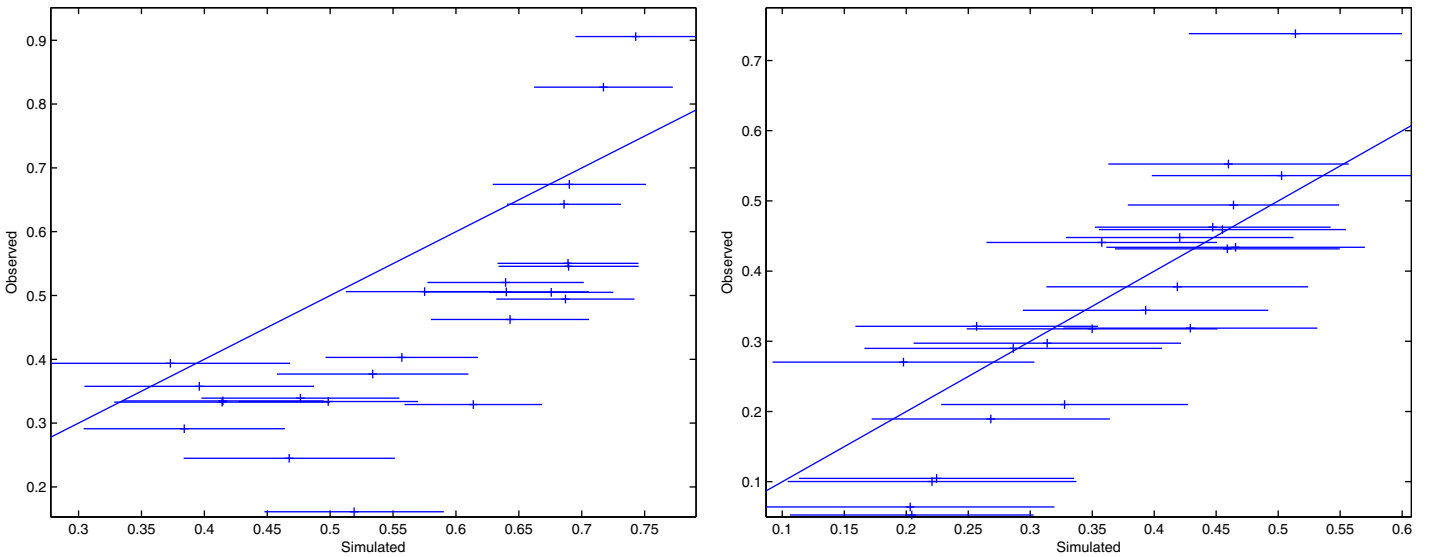

Figure 21. Estimates of observed (*y* axis) and model (*x* axis) correlation of the intensity in (left) January and (right) July, for station used in the fitting and one station not used.

not been used in the calibration of the model, the parametric covariance *C* seems to fit the empirical covariance rather well.

Next, we check the performance of the regression used to interpolate the mean function to ungauged locations. Figure 18 presents the observed daily proportions of wet days with the corresponding probability given by the stochastic generator superimposed. While the results for Borås are good, there is a constant bias of the order of 0.1 in Malmslätt. This could also be due to the fact there is only one station close to this one with data used in the model fitting.

As we have seen a very important characteristic that the weather generator should possess is the ability to reproduce the correct dry/wet behavior. Figure 19 illustrates the observed and model distributions of wet and dry spell lengths, which seem to fit well.

Next, we examine the pairwise lagged occurrence probabilities as defined in (24) with one station being used in the tuning of the model and the second station being one of the two stations left out for validation purposes. These probabilities are gathered in Figure 20. It seems like the model slightly overestimates the probabilities of pairwise lagged occurrence for small proportions ($p < 0.23$) in July. For moderate and high proportions, the bias in the probability is too small and unlikely to have any impact in practice.

Finally, Figure 21 displays observed and model pairwise station correlations of precipitation intensity, where one station was used in the model fitting and the other was not. Note that only days with simultaneous positive precipitation have been used. The fit is rather good for July but for January the model seems to overestimate the pairwise correlations.

Overall, we feel that the stochastic model suggested captures the important aspects of the precipitation data when interpolated to locations with no available information.

5. Conclusions

This paper has presented a spatiotemporal stochastic model for daily precipitation observations based on a common latent Gaussian field driving both the occurrence and intensity process. A composite transformation consisting of a gamma and a generalized Pareto distribution was fitted to the marginal distribution of the nonzero amounts of precipitation. The mean function was estimated using daily observations from weather stations and then interpolated in time and space using regression with Fourier components and location covariates, respectively. A parametric covariance function was used to model the observed

spatiotemporal correlations with the fitting performed using a method of moments approach and certain relations that hold for censored Gaussian moments.

Realistic spatiotemporal artificial sequences of precipitation have been generated and used to validate different aspects of the proposed model. We have shown that while the stochastic generator seems to be able to reproduce the dependence between different days and stations, it does not capture all the essential properties of the mean function and the marginal distribution. The possibility to include other covariates such as climate model output or broad scale atmospheric conditions in order to model the corresponding spatial extrapolation of the parameters should be investigated. The method of combining the anamorphosis transformation with the one stage approach of modeling dry and wet behavior simultaneously seems to be very flexible and to produce the desired results.

Acknowledgments

The data were in open access from the Swedish meteorological and hydrological institute (SMHI), see URL: http://www.smhi.se/en/services/professional-services/data-and-statistics/international-weather-data-and-statistical-analyses-1.33875. The interested parties need to contact the agency to obtain the data. We gratefully acknowledge the very helpful comments/suggestions of the three anonymous reviewers and the Associate Editor, which we feel have improved this work.

References

Ailliot, P., C. Thompson, and P. Thomson (2009), Space time modelling of precipitation by using a hidden Markov model and censored Gaussian distributions, *J. R. Stat. Soc., Ser. B*, *58*(3), 405–426.

Allard, D., and M. Bourotte (2015), Disaggregating daily precipitations into hourly values with a transformed censored latent Gaussian process, *Stochastic Environ. Res. Risk Assess*, *29*(2), pp. 453–462.

Allcroft, D. J., and C. A. Glasbey (2003), A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation, *Biomath. Stat.*, *52*, 487–498.

Bardossy, A., and E. J. Plate (1992), Space-time model for daily rainfall using atmospheric circulation patterns, *Water Resour. Res.*, *28*, 1247–1259.

Bell, T. L. (1987), A space-time stochastic model of rainfall for satellite remote-sensing studies, *J. Geophys. Res.*, *92*, 9631–9643.

Bellone, J., E. Hughes, and P. Guttorp (2000), A hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts, *Clim. Res.*, *15*, 1–12.

Box, G. E. P., and D. R. Cox (1964), An analysis of transformations, *J. R. Stat. Soc., Ser. B*, *26*(2), 211–252.

Buishand, T., and T. Brandsma (2001), Multisite simulation of daily precipitation and temperature in the Rhine basin by nearest-neighbor resampling, *Water Resour. Res.*, *37*, 2761–2776.

Burton, A., C. G. Kilsby, H. J. Fowler, P. S. P. Cowpertwait, and P. E. O'Connell (2008), Rainsim: A spatiotemporal stochastic rainfall modelling system, *Environ. Modell. Software*, *23*(12), 1356–1369.

Carreau, J., and Y. Bengio (2009), A hybrid Pareto model for asymmetric fat-tailed data: The univariate case, *Extremes*, *12*, 53–76.

Charles, S., B. Bates, and J. Hughes (1999), A spatiotemporal model for downscaling precipitation occurrence and amounts, *J. Geophys. Res.*, *104*, 31,657–31,669.

Chilés, J.-P., and P. Delfiner (2012), *Geostatistics: Modeling Spatial Uncertainty*, 2nd ed., Wiley, N. J.

Cressie, N., and H.-C. Huang (1999), Classes of nonseparable, spatio-temporal stationary covariance functions, *J. Am. Stat. Assoc.*, *94*(448), 1330–1340.

Davison, A., and R. Smith (1990), Models for exceedances over high thresholds (with discussion), *J. R. Stat. Soc., Ser. B*, *52*, 237–254.

Durban, M., and C. A. Glasbey (2001), Weather modelling using a multivariate latent Gaussian model, *Agric. For. Meteorol.*, *109*, 187–201.

Flecher, C., P. Naveau, D. Allard, and N. Brisson (2010), A stochastic daily weather generator for skewed data, *Water Resour. Res.*, *46*, W07519, doi:10.1029/2009WR008098.

Frigessi, A., O. Haug, and H. Rue (2002), A dynamic mixture model for unsupervised tail estimation without threshold selection, *Extremes*, *5*(3), 219–235.

Furrer, M. E., and R. Katz (2008), Improving the simulation of extreme precipitation events by stochastic weather generators, *Water Resour. Res.*, *44*, W12439, doi:10.1029/2008WR007316.

Glasbey, C. A., and I. M. Nevison (1997), Rainfall modelling using a latent Gaussian variable, in *Modell Longitudal and Spatially Correlated Data*, Lecture Notes in Statistics 122, pp. 233–242, Springer, N. Y.

Guillot, G. (1999), Approximation of Sahelian rainfall fields with meta-Gaussian random functions. Part 1: Model definition and methodology, *Stochastic Environ. Res. Risk Assess.*, *13*, 100–112.

Haskard, K. A. (2007), An anisotropic Matern spatial covariance model: Reml estimation and properties, PhD thesis, Univ. of Adelaide, Sch. of Agric., Food and Wine, BiometricsSA.

Hughes, J., and P. Guttorp (1994), A class of stochastic models for relating synoptic atmospheric patterns to local hydrologic phenomenon, *Water Resour. Res.*, *30*, 1535–1546.

Hughes, J., P. Guttorp, and S. Charles (1999), A non-homogeneous hidden Markov model for precipitation occurrence, *Appl. Stat.*, *48*(1), 15–30.

Hundecha, Y., M. Pahlow, and A. Schumann (2009), Modeling of daily precipitation at multiple locations using a mixture of distributions to characterize the extremes, *Water Resour. Res.*, *45*, W12412, doi:10.1029/2008WR007543.

Hutchinson, M. F. (1986), Methods of generation of weather sequences, in Agricultural Environments, edited by A. H. Bunting, *C.A.B. International*, vol. 73, pp. 149–157, Wallingford.

Hutchinson, M. F. (1995), Stochastic space-time weather models from ground-based data, *Agric. For. Meteorol.*, *73*, 237–264.

Karl, N., T. R. Nicholls, and A. Ghazi (1999), Clivar/acos/wmo workshop on indices and indicators for climate extremes: Workshop summary, *Clim. Change*, *42*, 3–7.

Katz, R. (1977), Precipitation as a chain-dependent process, *J. Appl. Meteorol.*, *16*, 671–676.

Katz, R., and M. Parlange (1995), Generalizations of chain-dependent processes: Application to hourly precipitation, *Water Resour. Res.*, *31*, 1331–1341.

Kleiber, W., R. W. Katz, and B. Rajagopalan (2012), Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes, *Water Resour. Res.*, *48*, W01523, doi:10.1029/2011WR011105.

Larsen, P., S. Goldsmith, O. Smith, M. Wilson, K. Strzepek, P. Chinowsky, and B. Saylor (2008), Estimating future costs for Alaska public infrastructure at risk from climate change, *Global Environ. Change*, *18*, 442–457.

Lennartsson, J., A. Baxevani, and D. Chen (2008), Modelling precipitation in Sweden using multiple step Markov chains and a composite model, *J. Hydrol.*, *363*(1–4), 42–59.

Li, C., V. Singh, and A. Mishra (2012), Simulation of the entire range of daily precipitation using a hybrid probability distribution, *Water Resour. Res.*, *48*, W03521, doi:10.1029/2011WR011446.
Ma, C. (2003), Families of spatio-temporal stationary covariance models, *J. Stat. Plann. Inference*, *116*, 489–501.
McMichael, A., R. Woodruff, and S. Hales (2006), Climate change and human health: Present and future risks, *Lancet*, *367*, 859–869.
McNeil, A. (1997), Estimating the tails of loss severity distributions using extreme value theory, ASTIN Bulletin, *27*, pp. 117–137.
Peterson, T. C., C. Folland, G. Gruza, W. Hogg, A. Mokssit, and N. Plummer (2001), Report on the activities of the working group on climate change detection and related rapporteurs 1998–2001, *ICPO Publ. Ser. 48*, DCDMP-47, WMO-TD 1071, WMO, Geneva, Switzerland.
Robertson, A. W., A. V. M. Ines, and J. W. Hansen (2007), Downscaling of seasonal precipitation for crop simulation, *J. Appl. Meteorol. Climatol.*, *46*, 677–693.
Rosenzweig, C., F. Tubiello, R. Goldberg, E. Mills, and J. Bloomfield (2002), Increased crop damage in the US from excess precipitation under climate change, *Global Environ. Change*, *12*, 197–202.
Schleiss, M. S., S. Chamoun, and A. Berne (2014), Non-stationarity in intermittent rainfall: "The dry drift," *J. Hydrometeorol.*, *15*(3), 1189–1204.
Schwarz, G. (1978), Estimating the dimension of a model, *Ann. Stat.*, *6*(2), 461–464.
Sigrist, F., H. R. Kunsch, and W. A. Stahel (2012), A dynamic non-stationary spatio-temporal model for short term prediction of precipitation, *Ann. Appl. Stat.*, *6*(4), 1452–1477.
Stehlík, J., and A. Bárdossy (2002), Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation, *J. Hydrol.*, *256*, 120–141.
Stein, M. (2005), Space time covariance functions, *J. Acoust. Soc. Am.*, *100*, 310–321.
Stern, R. D., and R. Coe (1984), A model fitting analysis of daily rainfall data, *J. R. Stat. Soc., Ser. A*, *147*, 1–34.
Suveges, M., and A. C. Davison (2010), Model misspecification in peaks over threshold analysis, *Ann. Appl. Stat.*, *4*(1), 203–221.
Todorovic, P., and D. A. Woolhiser (1975), A stochastic model of n-day precipitation, *J. Appl. Meteorol.*, *14*, 17–24.
Vrac, M., and P. Naveau (2007), Stochastic downscaling of precipitation: From dry events to heavy rainfalls, *Water Resour. Res.*, *43*, W07402, doi:10.1029/2006WR005308.
Wilks, D. S. (1998), Multi-site generalization of a daily stochastic precipitation generation model, *J. Hydrol.*, *210*, 178–191.
Yang, C., R. E. Chandler, V. S. Isham, and H. S. Wheater (2005), Spatio-temporal rainfall simulation using generalized linear models, *Water Resour. Res.*, *41*, W11415, doi:10.1029/2004WR003739.
Zucchini, W., and P. Guttorp (1991), A hidden Markov model for space-time precipitation, *Water Resour. Res.*, *27*, 1917–1923.