A meta-Gaussian distribution for sub-hourly rainfall

Marie Boutigny[1,2], Pierre Ailliot[1*], Aurore Chaubet[2], Philippe Naveau[3] and Benoît Saussol[1]

[1*]Laboratoire de Mathématiques de Bretagne Atlantique, UMR 6205, Université de Brest, France.
[2]SPL Eau Du Ponant, Brest, France.
[3]Laboratoire des Sciences du Climat et de l'Environnement, EstimR, LSCE/IPSL, CNRS-CEA-UVSQ, Université Paris-Saclay, Gif-sur-Yvette, France.

Abstract

Meta-Gaussian models are ubiquitous in the statistical literature. They provide a flexible building block to represent non-Gaussian distributions which inherit modeling and inference methods available in the Gaussian framework. In particular they have been widely used for modeling rainfall distributions. The first step when working with meta-Gaussian models consists in choosing an appropriate transformation which allows to map the Gaussian distribution to the target rainfall distribution. Many transfer functions have been proposed in the literature but most of them are not appropriate to describe heavy-tailed distributions, which is known to be a usual feature for rainfall at sub-daily scales. In this context, we propose and study a new meta-Gaussian model that can handle heavy-tailed observations. It leads to a four parameter model for which each parameter is linked to a different part of the distribution: a first one describes the probability of rainfall occurrence, two of them are related to the lower and upper tailed features of the distribution, and the last one is just a scaling parameter. Theoretical arguments are given to justify the proposed model. A statistical analysis of seven French rain gauges indicates the flexibility of our approach under different climatological regions and different aggregation times, here from six minutes to twenty four hours. Our distribution outperforms other meta-Gaussian models that have been proposed in the literature and, in particular, it captures well heavier tail behaviours below the hourly scale.

Keywords: Rainfall distribution, meta-Gaussian model, heavy-tailed distributions

1 Introduction

Precipitation intensities and frequencies are key variables for many environmental studies, not only in hydrology but also agronomy, meteorology and impact studies (see e.g. [1], [2]). As a consequence, there is an abundant literature on the modeling of daily and monthly rainfall intensity distributions at a single location. The most popular distribution for positive daily precipitation is probably the gamma distribution [3], which also generally provides an adequate fit for precipitation at the monthly scale. Other distributions have been proposed such as the mixed exponential [4], Weibull [5] and log-normal [6]. Comparisons were made for specific data sets. For example, [7] ranked the mixed exponential first and the gamma second whereas [8] ranked the log-normal first, then mixed exponential, gamma and finally exponential. The performance strongly depends on the location of interest as local climate strongly impacts rainfall distribution.

The meta-Gaussian framework is the most usual strategy to construct multivariate models for rainfall distributions. The building block of this approach is to link the phenomena of interest to a Gaussian distribution. A common justification behind all these Gaussian based transformation techniques is that the normal distribution represents a solid, simple and flexible building block to construct multivariate models and handle for example space-time data [9]. Furthermore it allows to model simultaneously the occurrence (dry/wet measurement) and amount of precipitation by censoring the Gaussian variable. As a consequence, meta-Gaussian models have been widely used in the literature on rainfall disaggregation [10–12], downscaling and model correction [13–15], short term or spatial prediction [16, 17], building stochastic weather generators [18–20], data assimilation [21], post-processing precipitation forecasts [22] or merging different data sources [23].

One key ingredient of meta-Gaussian model is the choice of an appropriate transformation to map a Gaussian distribution into a a distribution relevant to describe the phenomenon of interest. The most usual transformation in the statistical literature is probably the Box-Cox transformation [24] which has also been applied for rainfall [9, 23]. Other transformations have been proposed for modeling rainfall specifically, such as the square root transformation in [25], power transformations [18] or power exponential [12]. Another strategy is to use a transformation based on cumulative distributions and quantile functions, which maps the Gaussian distribution to a target distribution such as the gamma distribution ([20]) or the mixed exponential distribution [26], but this generally leads to transformations with relatively complicated analytical expressions.

A particular impetus for this work was the need to describe rainfall distributions at fine time scales (a few minutes) in order to test the sensitivity and robustness of an urban hydrological model in Brest (France). At sub-hourly scales, rainfall measurements are frequently null (when no rain is measured) and discrete due to rain gauge precision. Still, very few events with high intensities strongly skew the density to the right, creating a heavy tailed distribution which needs to be treated with care. The models described above do not produce heavy-tailed distributions and thus may not be appropriate for describing sub-hourly precipitations. More generally, the statistical modeling of the distribution of rainfall accumulated over short time periods is little discussed in the literature.

Extreme value theory promotes the use of the Generalized Pareto distribution (GPD) for modeling exceedances over a high threshold [27, 28]. One difficulty with this approach is that only exceedances above a high threshold are modelled, not the full distribution. Hybrid distributions were proposed in this context [29]. For example, [30] propose a meta-Gaussian model where a Gamma distribution is used to describe low and moderate rainfall, i.e. rainfall below a fixed threshold, and the GPD is used to describe rainfall above the threshold. However, finding the optimal threshold that allows the use of a GPD for the exceedances remains a delicate task for the practitioner. Different methods have been proposed to bypass the threshold selection step. In particular, [31] proposed the Extended GPD model. This leads to a flexible model which can describe the full rainfall distribution range and handle heavy tails. In contrast to a meta-Gaussian model, this approach is based on the uniform distribution, not the Gaussian one. So, this may bring some complexity in some types of extensions, see [32]. In addition, it does not handle dry events.

The main contribution of this paper is to propose a new meta-Gaussian model which is simple but flexible enough to model the full rainfall distributions accumulated over a wide range of time scales including dry, low and heavy precipitations. The paper is organised as follows. The data sets that will be used throughout the paper are introduced in Section 2. The proposed model is introduced in Section 3, together with some theoretical justifications. Section 4 discusses results obtained when fitting the model to the data. Finally conclusions and perspectives are given in Section 5.

All the models discussed in this paper can be fitted with the R package available on Github at https://github.com/mbtgy/tcG.

2 Data

In this study, we consider rainfall data recorded at 7 meteorological stations in France represented on Figure 1. These data were provided by Météo-France. They are available at a 6 minutes time step from 2010 until 2021 (12 years). In order to remove seasonal components, a focus is made on the three months of summer, i.e. June, July and August where the more intense convective

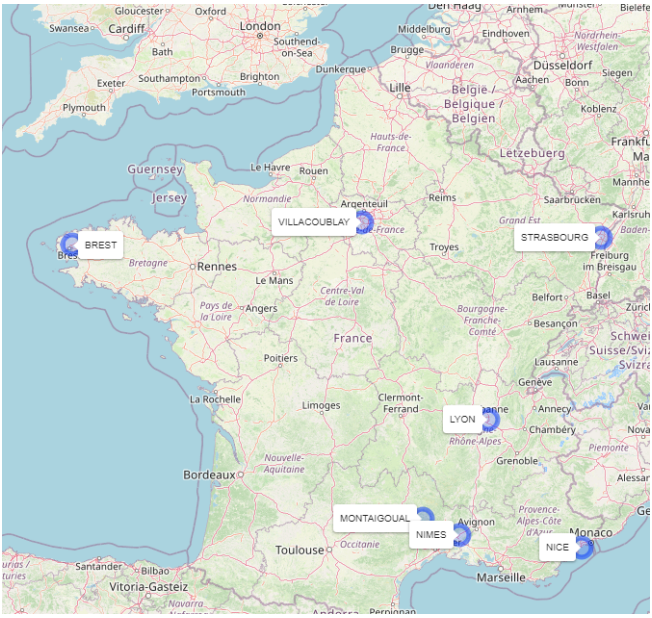

Fig. 1 Locations of our seven rain gauge Metep France stations in France with a six minute time scale recording from 2010 to 2021 (summer months). Different sites correspond to different climates. For example, the westernly climate in the Brittany peninsula strongly differs from the drier southern Mediterranean one.

precipitation events generally occur. All the data were measured using tipping bucket gauges with a 0.2 mm precision.

This work was triggered by the need of high resolution space-time data which can be used as input of an urban hydrological model which describes the sewage system of the city of Brest. One difficulty when trying to merge the different sources of rainfall data available over a particular geographical area such as the watershed of the city of Brest is that they are generally available at different time steps. For example, recent rain gauge typically provide measures of the precipitation accumulated over 3 or 6 minutes, whereas historical measurements are only available at the hourly or daily scale. Other sources of data such as radar, satellite or model outputs may also be available at other time resolutions. In such situation, it is useful to have a simple parametric model which can describe the distribution of precipitation at different time resolutions ranging from a few minutes to daily data. We also consider 6 other stations, located in different climate zones (continental, mountainous and Mediterranean), to check the flexibility of the proposed model.

Figure 2 shows that the temporal resolution has a strong influence on the shape of the rainfall distribution in Brest. Obviously, the percentage of dry records decreases with the duration over which rainfall is accumulated: for 6-minutes data, the percentage of dry measurements is equal to 96.7% whereas it drops to 31.6% for daily data. The empirical distribution of the 6-minutes data (left panel) is highly skewed, with a majority of positive records being

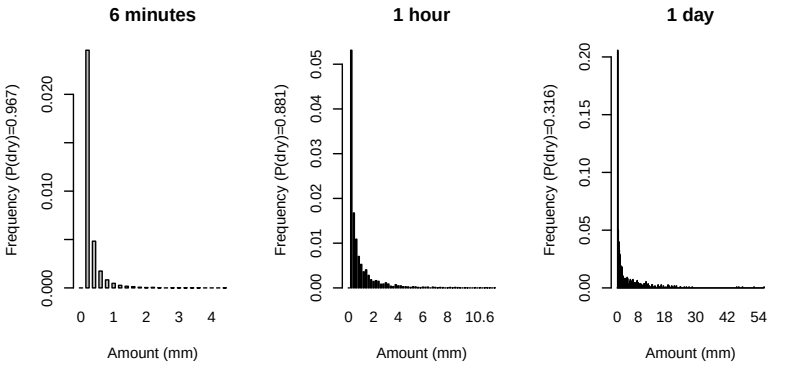

Fig. 2 Empirical distribution of rainfall accumulated over 6 minutes (left), 1 hour (middle) and 1 day (right) in Brest. Dry measurements are not represented but the percentage of dry measurements is indicated in the y-label.

equal to the 0.2 mm precision of the tipping bucket, but also high values corresponding to intense convective events, which creates a heavy tail distribution. The empirical distributions of hourly (middle panel) and daily (right panel) rainfall are also skewed, but it is less pronounced than for 6-minutes data.

Also note that the 0.2 mm precision of the tipping bucket leads to a discretization of the (continuous) distribution of the rainfall. This discretization is clearly visible when looking at the empirical distribution of the 6-minutes rainfall data, whereas it becomes less prominent when looking at hourly or daily data. This has to be taken into account when fitting the continuous model introduced in the next section to the data (see Section 4.1).

To sum up, the goal is to develop a simple parametric model which can describe the distribution of precipitation at different time scales, hence to have a strongly skewed distribution with a discrete component in zero and the ability to produce heavy tails. The next section discusses the choice of such model.

3 Model

3.1 Meta-Gaussian models

A classical approach for modeling rainfall, sometimes called meta-Gaussian model, is to assume that rainfall amounts Y can be linked to a Gaussian variable X with mean μ and variance 1 according to

$$Y = \psi(X)\mathbb{1}_{X\geq 0}, \quad \text{with } X \sim \mathcal{N}(\mu, 1), \tag{1}$$

where $\psi : [0, +\infty[\to]0, +\infty[$ is an increasing function which is generally refereed to as the anamorphosis in the literature and $\mathbb{1}_{X\geq 0}$ is the indicator function equal to 1 if condition $X \geq 0$ is true and 0 otherwise. The operation of such model is schematised in Figure 3. The censorship in 0 produces dry conditions

(step 1 in Figure 3) with a proportion linked to the mean of the Gaussian according to

$$P(Y = 0) = P(X \leq 0) = \Phi(-\mu)$$

where Φ is the cumulative distribution function (cdf) of the standard normal distribution. The transformation ψ acts on the positive part of the distribution which corresponds to wet conditions (step 2 in Figure 3).

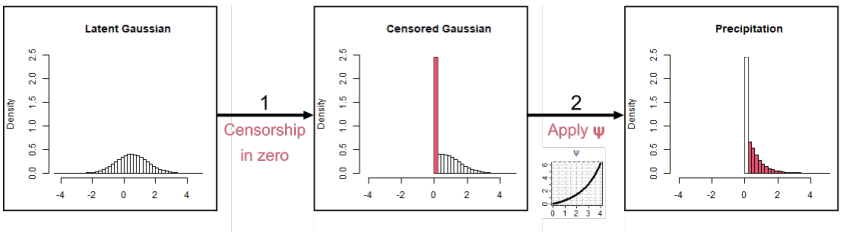

Fig. 3 Schematic functioning of a meta-Gaussian model. The coloured areas in the histograms represent the part of the distribution modified at each step.

The cdf of the random variable Y defined by (1) can be written as

$$F(y) = \begin{cases} \Phi(\psi^{-1}(y) - \mu) & \text{if } y > 0 \\ \Phi(-\mu) & \text{if } y = 0 \end{cases}. \tag{2}$$

Remark that this meta-Gaussian model is general since any positive random variable with a discrete component at the origin like precipitation can be written as (1) using

$$\psi(x) = F^{-1}(\Phi(x - \mu)) \tag{3}$$

where $\mu = -\Phi^{-1}(P(Y = 0))$ and F^{-1} denotes the quantile function (generalized inverse function of the cdf F) of Y. Plugging a non-parametric estimate of the quantile function F^{-1} in (3) allows building non-parametric estimates of ψ, [**?**]see, e.g.,¿lien2013effective, cecinati2017comparing. The dots in Figure 4 show the estimate obtained on the 6-minutes rainfall data in Brest introduced in Section 2. The shape of ψ near zero is linked to the small precipitations. A horizontal tangent at the origin means that they are more low rainfall than expected low values in the censored Gaussian distribution and the density becomes more concentrated in 0 if ψ is flatter at the origin. The growth speed is linked to the upper tail: the convex-exponential shape indicates that the tail is heavier than a Gaussian one. However, parametric approaches are generally favoured in the applications and many models have been proposed for ψ in the literature. The most classical one is probably the power transformation, see [18] and [16],

$$\psi(x) = \sigma x^{1/\alpha}, \tag{4}$$

but other transformations have been proposed. [11] studied a quadratic power function, [14] worked with a simple exponential transform and [12] focused on

$$\psi(x) = \sigma_2(\exp(\sigma_1 x^{1/\alpha}) - 1). \tag{5}$$

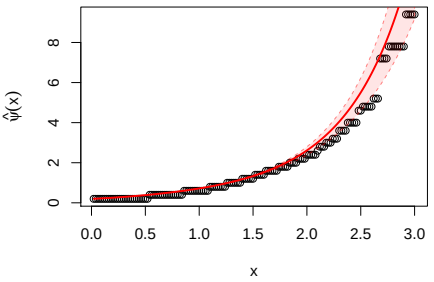

Fig. 4 Non-parametric estimate of the anamorphosis function based on (3) (dots) for the 6-minutes data in Brest-Guipavas. The plain curve corresponds to the parametric model (13) fitted to the data (see Section 4.1), and the dotted line is its 95% confidence interval computed using 500 bootstrap samples.

To force the resulting distribution to match a specific distribution the inverse of a cdf can also be used, as it is the case with the Gamma distribution in [20].

Transformation (4) being the most commonly used, it will be a point of comparison and will be referred to as the classical meta-Gaussian model. Note that it is closely related to the popular Box-Cox transformation [24].

3.2 Lower and upper tails of meta-Gaussian models

The choice of an appropriate anamorphosis function for a particular application is typically a trade-off between model complexity, versatility, tractability and interpretability. In this section, it is advocated that the properties of lower and upper tails of the positive part of the rainfall distribution may also provide interesting insights.

Different studies have shown that rainfall at daily or sub-daily scales are generally heavy tailed [27]. In this situation, ψ should be chosen such that the transformed Gaussian variable defined by (1) is tail equivalent with a Pareto distribution with positive shape parameter ξ. According to B, this holds true if and only if

$$\lim_{x\to\infty} \frac{x\psi(x)}{\psi'(x)} = \frac{1}{\xi}. \tag{6}$$

Solving the differential equation $\frac{x\psi(x)}{\psi'(x)} = \frac{1}{\xi}$ leads to a first function $x \mapsto \exp \frac{\xi x^2}{2}$ which satisfies (6). Then, by re-writting ψ as

$$\psi(x) = \exp \frac{\xi x^2}{2} \exp u(x) \tag{7}$$

- which is always possible - condition (6) becomes

$$\lim_{x\to\infty} \frac{u'(x)}{x} = 0.$$

This condition seems easier to work with as it allows understanding that loosely speaking, the anamorphosis function ψ should increase "like" the function $x \mapsto \exp \frac{\xi x^2}{2}$ when $x \to +\infty$ to get heavy tailed distributions. In particular, one can check that most of the anamorphosis found in the literature - including the classical meta-Gaussian model (4) introduced previously - do not satisfy condition (6). Hence, these transforms are not particularly well suited to model heavy rainfall at small time scales. Interestingly, the model of [12] will be tail equivalent to a Pareto distribution if and only if $\alpha = \frac{1}{2}$ in (5). The Tukey g-and-h distribution, [33–35] [], defined as the random variable

$$\frac{\exp(gX) - 1}{g} \exp \frac{\xi X^2}{2}, \quad \text{with } X \sim \mathcal{N}(0, 1)$$

with h and g real parameters, may not always satisfy condition (6). In addition, it is not directly suited to include the dry event component at 0 and it lacks flexibility to describe the shape of the lower part of the rainfall distribution according to the discussion below.

Concerning non zero but low rainfall, [31] advocated, using arguments of the extreme value theory, that the lower part of the distribution of the positive amount should approximately follow a power-law, i.e. satisfy

$$\lim_{y \downarrow 0} \frac{F(y) - F(0)}{y^\alpha} = C$$

for some positive constant C and shape parameter $\alpha > 0$. In particular, the often used Gamma distribution with shape parameter α satisfies this constraint. These authors proposed and studied different families that meet this criterion. One can wonder if this constraint could also be enforced to the meta-Gaussian model defined by (2). A first order Taylor expansion of F around zero indicates that condition (6) holds true if and only if

$$\psi(x) = x^{\frac{1}{\alpha}} K(x) \tag{8}$$

with K such that $\lim_{x \downarrow 0} K(x)$ exists and is strictly positive. Most of the anamorphosis functions in the hydrological literature, including the classical meta-Gaussian model (4) correspond to the particular case where $K(x)$ is a constant function, and consequently they obey this constraint on low rainfall amount. Remark that the same parameter α in model (5) both controls the shape of the distribution for low and heavy rainfall. This is an undesirable property as is not impossible to create a heavy-tailed distribution with a power shape parameter different from $\alpha = \frac{1}{2}$ for low rainfall.

3.3 Proposed model

According to the previous section, the anamorphosis ψ should be chosen such that conditions (6) and (8) are satisfied in order to obtain a Meta-Gaussian

distribution with a Pareto upper tail and a lower tail that follows a power law. Based on these results, this paper advocates the use of the anamorphosis function

$$\psi(x) = \sigma x^{\frac{1}{\alpha}} \exp \frac{\xi x^2}{2} \tag{9}$$

with $\mu \in \mathbb{R}$, $\sigma \in \mathbb{R}^{+*}$, $\alpha \in \mathbb{R}^{+*}$ and $\xi \in \mathbb{R}$. The distribution of the random variable Y defined through (1) with $X \sim \mathcal{N}(\mu, 1)$ and ψ given by (9) will be referred to as the GP meta-Gaussian distribution with parameter $(\mu, \sigma, \alpha, \xi)$. μ is directly related to the dry probability through (2) and σ is a multiplicative scale parameter. ψ defined by (9) satisfies condition (8) and thus the lower tail of the positive part of the distribution has a power shape with shape parameter α. It also satisfy (6) and thus the upper tail distribution is controlled by the shape parameter ξ. More precisely, if $\xi > 0$, the distribution is tail equivalent with a Pareto distribution with shape parameter ξ. It implies in particular that $E[Y^p] = +\infty$ if $p > \frac{1}{\xi}$. The case $\xi = 0$ corresponds to the classical meta-Gaussian model (4). Negative values for ξ creates an upper bound to the distribution as for the GPD distribution. Indeed, when $\xi < 0$, ψ is strictly monotonic increasing only on the interval $(0, x_{sup})$ with

$$x_{sup} = \sqrt{\left| \frac{-1}{\min(\alpha\xi, 0)} \right|}. \tag{10}$$

The GP meta-Gaussian distribution is thus defined by applying (1) with ψ given by (9) to the Gaussian variable $X \sim \mathcal{N}(\mu, 1)$ truncated at x_{sup}. Remind that truncation means that values above x_{sup} are not observed - unlike the censorship that is used to create the dry component, where the observations above the bound take the value of the bound. The support of the distribution is $[0, y_{sup}]$ with

$$y_{sup} = \sigma \left(\frac{e^{-1}}{\max(-\alpha\xi, 0)} \right)^{\frac{1}{2\alpha}}$$

the upper bound in the precipitation domain. Note that when $\xi \geq 0$ the bounds become $x_{sup} = y_{sup} = +\infty$, so those notations can be used for $\xi \in \mathbb{R}$. When the Gaussian is truncated above x_{sup}, the cdf (2) must be corrected by the probability of truncation (see A).

An advantage of the GP meta-Gaussian transformation over other transformations which satisfy (6) and (8) is the possibility to derive an analytical expression for the inverse of ψ

$$\psi^{-1}(y) = \sqrt{\frac{1}{\alpha\xi} W\left(\alpha\xi \left(\frac{y}{\sigma} \right)^{2\alpha} \right)} \tag{11}$$

where W denotes the Lambert W function [34] defined as the inverse of the function $x \mapsto x \log x$. Note that the Lambert W function is available in usual statistical software which simplifies practical implementation of the model.

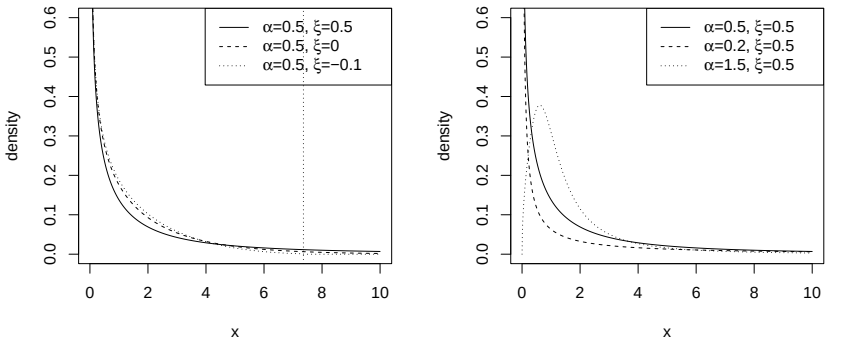

Fig. 5 Probability density function of the GP meta-Gaussian distribution for different parameter values. All the plots were obtained with $\mu = 0$ and $\sigma = 1$. The left panel shows the influence of ξ and the right panel the influence of α. The vertical dotted line on the left plot corresponds to the upper bound y_{sup} of the distribution with $\xi = -0.1$. Only the positive part of the distribution is shown.

In particular, analytical expressions for the cdf and the probability density function (pdf) of the GP meta-Gaussian model can be derived from (2). This simplifies the computation of the likelihood function and the fitting the model to data (see Section 4.1 for more details). Analytical expressions for the finite moments can also be derived, which is not the case for many meta-Gaussian models that can be found in the literature (to the best of our knowledge the classical transform (4) is the only other meta-Gaussian model with analytical moments). Expressions for the pdf, cdf, quantile function and moments of the GP meta-Gaussian model can be found in A.

Figure 5 illustrates the flexibility of the GP meta-Gaussian distribution with different parameter values. The distribution with $\mu = 0$, $\sigma = 1$, $\alpha = 0.5$ and $\xi = 0.5$ is used as a reference since it corresponds to typical parameters values for the rainfall data considered in this study. The left panel shows that the parameter ξ influences the upper tail of the distribution whereas α modifies the shape of the distribution for low rainfall (see right panel) as expected from the theory.

4 Numerical results

4.1 Parameter estimation

This section describes the method used to fit the model introduced in the previous section to the data. As discussed in Section 2 (see also Figure 2) the functioning of a tipping bucket induces a discretizaion of the data which needs to be taken into account.

For that, we assume that

$$P(N = n) = P(n \leq Y < n + \delta) \tag{12}$$

where Y is the (not observed) continuous rainfall, N represents the (observed) discrete measurement with values in $\{0, \delta, 2\delta, ...\}$ and δ denotes the precision of tipping bucket, i.e. 0.2 mm for our data sets.

Applying (12) with $n = 0$ gives

$$P(N = 0) = P(Y < \delta)$$

and thus with the anamorphosis (9) the probability of having a dry measurement depends not only on μ but also on the parameters $\{\sigma, \alpha, \xi\}$ of the anamorphosis function. This creates additional dependence between the different parameters and complicates the interpretation of the parameters. In order to get round this difficulty, the definition of ψ is modified as follows

$$\psi(x) = \delta + \sigma x^{\frac{1}{\alpha}} \exp \frac{\xi x^2}{2}. \tag{13}$$

Note that ψ^{-1} and y_{sup} are consequently modified. We found that the introduction of δ in the anamorphosis function greatly improves the results obtained when fitting the model to the data sets considered in this study.

In order to fit the GP meta-Gaussian model, the maximum likelihood approach is used in this study. When fitting a continuous distribution model to rain gauge data, the likelihood is usually computed directly from the continuous density. However it has been noticed that taking into account the discretization significantly improves the results. More precisely, the discrete log likelihood which is maximised is based on (12)

$$\log \mathcal{L}(\theta) = n_0 \log(\Phi(-\mu)) + \sum_{i:\ n_i > 0} \log \{F(n_i + \delta) - F(n_i)\} \tag{14}$$

where n_0 is the number of dry measurements, $(n_1, ..., n_n)$ the rainfall data, $\theta = (\mu, \alpha, \sigma, \xi)$ the unknown parameter and F the cdf of the distribution of Y (which depends on θ).

We found that the maximum likelihood estimates (MLE) obtained by maximising (14) over θ leads to estimates of α , σ and ξ which are strongly correlated together. The dependence between the estimates of σ and ξ is not surprising since it is well known that a similar behaviour occurs for the GPD distribution: the MLE of the scale and shape parameters are strongly dependent (see e.g. [36] and references therein). We also found a strong dependence between the MLE of α and σ, even when fitting the classical model (4) where $\xi = 0$. This is especially true when fitting the model to rainfall data at a sub-hourly time-step: the discretization has a strong impact (see Figure 2) and this may complicate the estimation of the lower-tail shape parameter α. The

existence of this dependence between the different estimates complicates the interpretation of the results. A usual method to overcome this problem (see e.g. [37]) is to add a penalty term in the likelihood. After some experimentation, we chose to maximise the following penalised log-likelihood function

$$\log \tilde{\mathcal{L}}(\theta) = log\mathcal{L}(\theta) - \frac{(\alpha - 0.5)^2}{\tau^2}.$$

It favours estimates with a shape parameter α close to 0.5 and the hyperparameter τ controls the strength of the penalty. In the numerical results reported below, the value $\tau^2 = 0.001$ is used.

4.2 Results

In order to assess the flexibility and interpretability of the proposed GP Meta-Gaussian model, it was fitted to the rainfall data at the 7 meteorological stations represented on the map of Figure 1. For each of these stations, different temporal resolutions varying from 6-minutes to one day were considered, where lower resolution data are obtained by accumulating 6-minutes data over a longer time interval. In order to reduce the short-term temporal dependence in the data sets, one observation every k observations was retained for the analysis. The value of k depends on the time step between successive observations, ranging form $k = 10$ for 6-minutes data, to $k = 5$ for 12-minutes data, $k = 2$ for 30-minutes data, and $k = 1$ if the time step is larger than 1 hour.

Figure 6 shows quantile-quantile plots for the fitted GP meta-Gaussian model at various time resolutions (6 minutes, 30 minutes, 1 hour and 1 day) for the 7 stations. The global fit of the model is very satisfying at all stations and time resolutions, except maybe for the smaller time steps in Lyon where the fitted model has difficulties in explaining the largest observation. It shows that the model is flexible enough to reproduce rainfall distributions for a variety of climates and temporal resolutions.

The classical meta-Gaussian model with power transformation (4) was also fitted to the data for comparison. Remind that this model is a particular case of the GP meta-Gaussian model when the upper-tail parameter ξ is assumed to be equal to 0. The resulting quantile-quantile plots are superimposed on Figure 6. The fits are generally less satisfactory than with the GP meta-Gaussian model, in particular regarding the upper-tail of the distribution where the power transformation leads to an underestimation of the largest quantiles. This is not surprising since we know from the results given in Section 3.2 that it is not possible to produce heavy-tailed distributions with this model whereas rainfall data generally have heavy tails.

Figure 7 shows the evolution of the model parameters with time resolution (note that the time axis is non linear). The evolution of the parameters is generally smooth. μ is increasing with aggregation, which is expected as there are less and less dry measurements when the period over which rainfall is accumulated increases. The estimate of the scale parameter σ is also generally

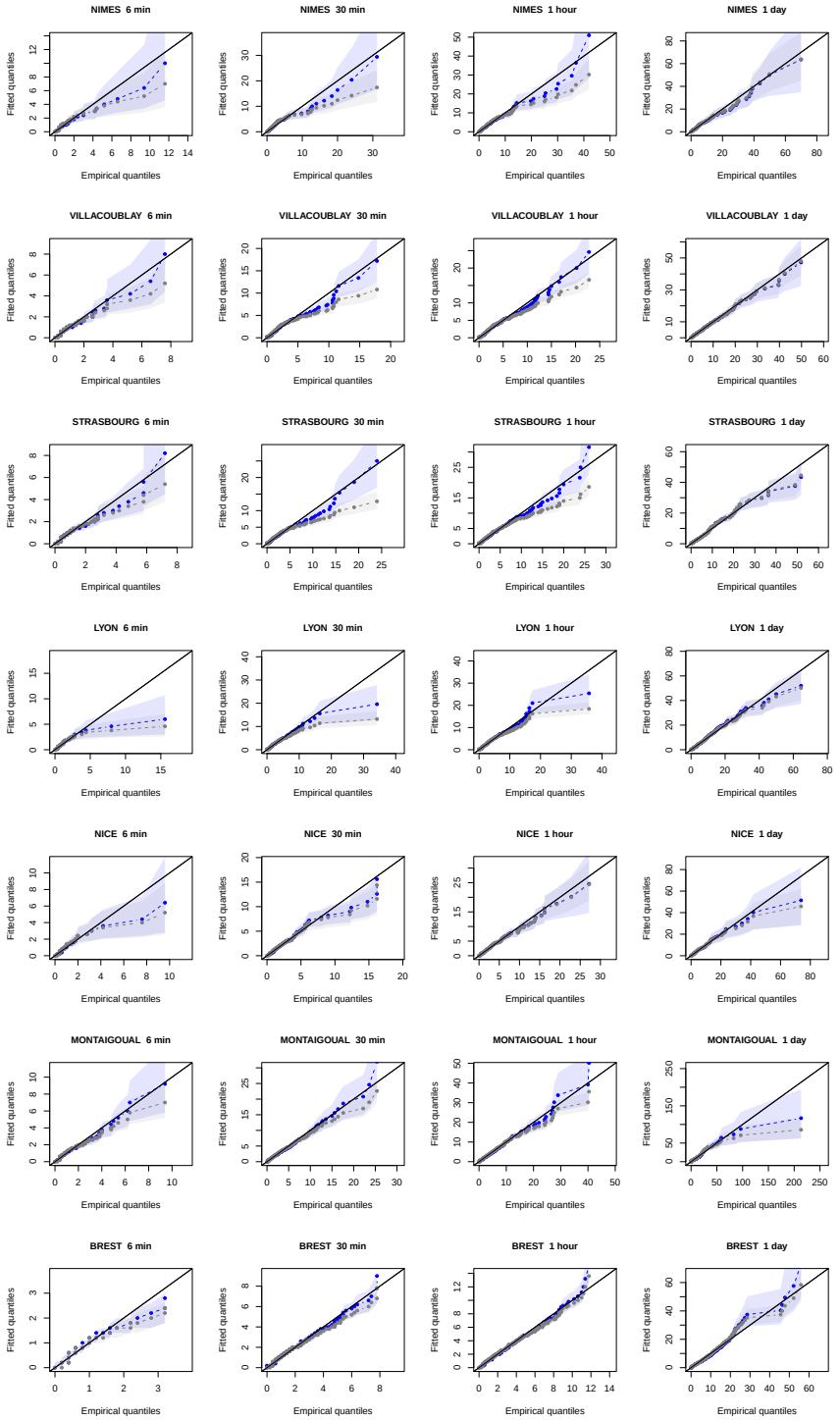

Fig. 6 Quantile-quantile plots for the GP (9) (blue) and power (4) (grey) meta-Gaussian models fitted to the rainfall data. Each column corresponds to a different time resolution and each row to a meteorological station. The light area gives (pointwise) 95% intervals based on 500 non parametric bootstrap replicates.

increasing, which is expected with the sum of (positively correlated) random variables. Remark however that there are some exceptions (e.g. Brest for larger time resolution) which may be due to the correlation between the estimates of σ and the other parameters. As concerns the upper tail parameter ξ, we generally obtain positive values (i.e. heavy-tailed distributions) but the evolution with time resolution seems to be site-dependent. For Nimes, Villacoublay, Strasbourg and Lyon (top panels of Figure 7), the estimated tail parameter ξ is relatively high for 6-minutes data and then decreases when rainfall is accumulated over longer time intervals. This is coherent with the intuition that summing random variable will tend to "gaussianize" them and produce distributions with lighter tails. However, for Nice, Montaigoual and Brest, a different behaviour is observed: the estimated tail parameter ξ is smaller for 6-minutes data and does not clearly decrease with the time resolution. The evolution of the lower tail parameter α is also site dependent. At some stations (Lyon, Strasbourg, Nimes, Villacoublay and Brest) it tends to increase and then decrease with a maximum value reached for hourly data, whereas it is the opposite at Montaigoual. Remind that lower values of α lead to a distribution which is more more "peaky" at the origin. The explanation behind this temporal evolution is not straightforward. The rainfall accumulated over a given time period is the sum of a random number (because of the dry measurements) of correlated (because of the temporal dependence) random variables and hence it may have a complicated behaviour. It thus depends on the climate of the different stations, the characteristics of the rainy events in terms of intensity and duration impacting the distribution of the accumulated rainfalls.

5 Conclusions and perspectives

In this work, we propose a new meta-Gaussian distribution with four parameters that can handle heavy-tailed data with a discrete component at the origin. The proposed GP meta-Gaussian model is tractable and analytical expressions exist for the pdf, cdf, quantile function and for the moments.

It was found that the model is flexible enough to describe the distribution of rainfall over a variety of climates and time resolutions. Comparison with a classical meta-Gaussian model shows what the proposed transform brings to this class of models: a better fit at small time scales due to its capacity to produce heavy tails. The GP meta-Gaussian model is quite similar to the extended GP model [31] in terms of construction but also in terms of performance. The advantage of the meta-Gaussian model is its direct link with a Gaussian variable that allows the use of methods developed for Gaussian data (multivariate, spatiotemporal models, Kalman-like algorithm, etc.) and also the possibility to easily include the discrete component associated to dry conditions.

It was also found that the parameters of the model evolves smoothly with temporal resolution. This could be useful for example to extrapolate the parameters and thus "estimate" the distribution of the rainfall at unobserved time resolution. However, much works remain to be done to understand how

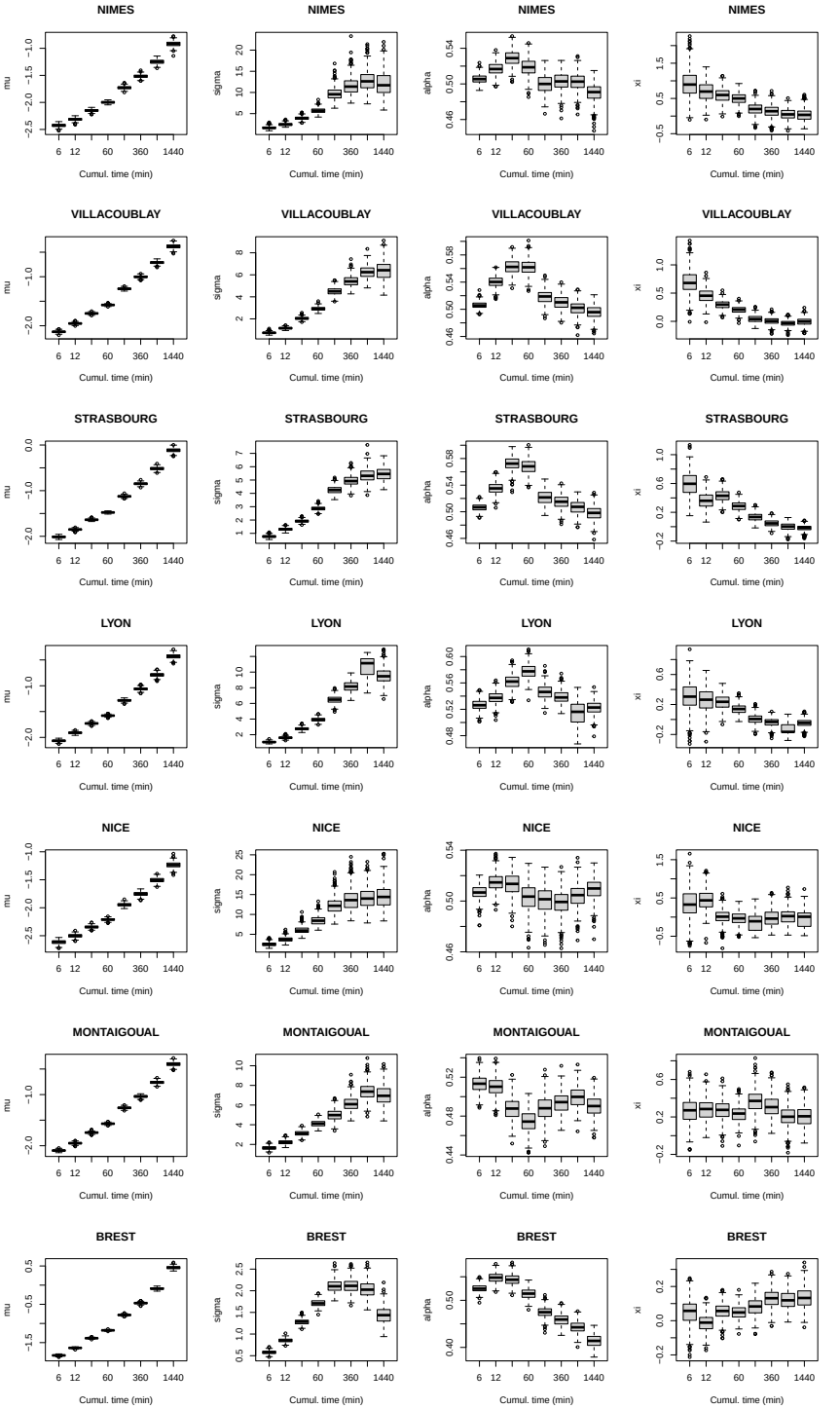

Fig. 7 Parameters estimates of the GP meta-Gaussian model fitted to the data at various time resolution. Each column corresponds to a different parameter of the GP meta-Gaussian and each row to a meteorological station. The boxplots are based on 500 non parametric bootstrap replicates.

the temporal dynamics of the dry and rainy events impacts the distribution of the rainfall accumulated at different temporal resolutions. This will be the topic of future work, with the aim to derive a temporal model compliant with the properties of rainfall at different time scales. Another possible extension is to adapt this meta-Gaussian distribution to spatial rainfall structures, in particular to improve regional frequency analysis, see e.g. [38] and [39].

Appendix A Some Theoretical Properties of the GP Meta-Gaussian Distribution

The density, cdf and quantile function of a meta-Gaussian model as defined in (1) are

$$f(y) = c \times \begin{cases} \phi_\mu\left(\psi^{-1}(y)\right)/\psi'\left(\psi^{-1}(y)\right) & \text{if } y > 0 \\ \Phi_\mu(0) & \text{if } y = 0 \end{cases},$$

$$F(y) = c \times \begin{cases} \Phi_\mu(\psi^{-1}(y)) & \text{if } y > 0 \\ \Phi_\mu(0) & \text{if } y = 0 \end{cases},$$

$$F^{-1}(u) = \begin{cases} \psi(\Phi_\mu^{-1}(u/c)) & \text{if } u > \Phi_\mu(0) \\ 0 & \text{if } u = \Phi_\mu(0) \end{cases},$$

with ϕ_μ and Φ_μ denoting respectively the pdf and cdf of a normal distribution with mean μ. c is the normalisation constant that deals with the probability of truncation when $\xi < 0$ with the GP meta-Gaussian transform. Hence $c = 1$ for the classical transform (4), and for the GP meta-Gaussian transform (9) $c = 1/\Phi_\mu(x_{sup})$, with x_{sup} the upper bound in the Gaussian domain as defined in (10).

An explicit expression of the moments was found for the GP meta-Gaussian distribution when $\xi \geq 0$. Let us write Y_+ the wet measurements.

$$\begin{aligned} E(Y_+^p) &= \frac{1}{\sqrt{2\pi}(1-\Phi(-\mu))} \int_0^{+\infty} \psi(x)^p \exp\left\{-\frac{1}{2}(x-\mu)^2\right\} dx \\ &= \frac{\sigma^p}{\sqrt{2\pi}(1-\Phi(-\mu))} \exp\left(-\frac{\mu^2}{2}\right) \int_0^{+\infty} x^{p/\alpha} \exp\left\{-\frac{1-\xi p}{2}x^2 + \mu x\right\} dx \end{aligned}$$

By identification in [40] (eq. 3.462.1, page 365), with $\gamma = -\mu$, $\nu - 1 = p/\alpha$ and $\beta = (1-\xi p)/2$,

$$E(Y_+^p) = \frac{\sigma^p (1-\xi p)^{-\frac{1}{2}(\frac{p}{\alpha}+1)}}{\sqrt{2\pi}(1-\Phi(-\mu))} \exp\left\{\frac{\mu^2}{2}\left(\frac{1}{2(1-\xi p)} - 1\right)\right\} \Gamma\left(\frac{p}{\alpha}+1\right) D_{-(\frac{p}{\alpha}+1)}\left(-\frac{\mu}{\sqrt{1-\xi p}}\right)$$

Γ is the Gamma function and D_ν can be expressed with Kummer's confluent hypergeometric function of first kind [40], eq. 9.240, page 1028. This expression is valid if $-\alpha < p < 1/\xi$.

Appendix B Pareto Tail for Meta-Gaussian Models

Proposition 1. *Let Z be any positive absolutely continuous random variable with pdf f_Z and with a Pareto survival function $\overline{F}_Z$. Let X be any standardized normal distributed random variable, and let us define the positive random variable*

$$Y \overset{d}{=} \psi(X),$$

where $\overset{d}{=}$ means equality in distribution and $\psi(.)$ represents a continuous and increasing function from the real line to $[0, \infty)$. The two random variables Z and Y are tail-equivalent if and only if

$$\lim_{x \to \infty} \frac{x\psi(x)}{\psi'(x)} = \frac{1}{\xi}, \tag{B1}$$

where ξ corresponds the common positive GP shape parameter of Z.

Proof of Proposition 1: Let ϕ and $\overline{\Phi}$ denote respectively the pdf and survival function of a standard normal distribution X.

Recall that Z and Y are tail-equivalent, if and only

$$\lim_{y \to \infty} \frac{\overline{F}_Z(y)}{\mathbb{P}[Y > y]} = c \in (0, \infty),$$

This condition is satisfied if they have the same tail index. Assuming a Pareto tail with positive shape parameter ξ for Z implies that Z is regularly varying with index $1/\xi$. Proposition A.3.8(b) from [41] recalled that this regular variation type is equivalent to

$$\lim_{z \to \infty} \frac{z \times f_Z(z)}{\overline{F}_Z(z)} = \frac{1}{\xi}.$$

Hence, to show that Y and Z are tail equivalent, one needs to determine under which condition it can be written that

$$\lim_{z \to \infty} \frac{z \times F(z)}{\overline{F}_Y(z)} = \frac{1}{\xi}.$$

where f and $\overline{F}$ denote the pdf and survival function of Y, respectively.

By construction, the survival function of Y equals to

$$\overline{F}_Y(z) = \mathbb{P}[X > \psi^{-1}(z)] = \overline{\Phi}\left[\psi^{-1}(z)\right],$$

The density of Y is

$$f(z) = \left(\psi^{-1}(z)\right)' \phi\left[\psi^{-1}(z)\right].$$

Then one can write

$$\frac{z \times f(z)}{\overline{F}_Y(z)} = \left(z \times \psi^{-1}(z) \times \left(\psi^{-1}(z)\right)'\right) \times \left(\frac{\phi\left[\psi^{-1}(z)\right]}{\psi^{-1}(z)\overline{\Phi}\left[\psi^{-1}(z)\right]}\right).$$

Mill's ratio [41] tells us that the ratio in the last bracket goes to one as $\psi^{-1}(z)$ goes to ∞ (i.e. as z grows). Hence, the condition

$$\lim_{z\to\infty}\left(z \times \psi^{-1}(z) \times \left(\psi^{-1}(z)\right)'\right) = \frac{1}{\xi}, \tag{B2}$$

is equivalent to

$$\lim_{z\to\infty}\frac{z \times f(z)}{\overline{F}_Y(z)} = \frac{1}{\xi}.$$

This is equivalent to have tail equivalence between Z and Y.

Changing variables with $z = \psi(x)$, $x = \psi^{-1}(z)$ and $\left(\psi^{-1}(z)\right)' = dx/dz$, condition (B2) is equivalent to condition (B1).

This is the necessary and sufficient condition on $\psi(.)$ to build a Pareto random variable of tail index ξ from a standardized normal random variable X.

Acknowledgments. The authors acknowledge Météo France for the data sets.

This work was supported by Eau du Ponant SPL, and took place in the context of the MEDISA (Méthodologie de Dimensionnement des Systèmes d'Assainissements) project.

Part of P. Naveau's work work was supported by the French national program (80 PRIME CNRS-INSU), and the European H2020 XAIDA (Grant agreement ID: 101003469). The author also acknowledges the support of the French Agence Nationale de la Recherche (ANR) under reference ANR-20-CE40-0025-01 (T-REX project), and the ANR-Melody (ANR-19-CE46-0011) .

References

[1] Bauer, P., Thorpe, A., Brunet, G.: The quiet revolution of numerical weather prediction. Nature **525**(7567), 47–55 (2015)

[2] Caseri, A., Javelle, P., Ramos, M.-H., Leblois, E.: Generating precipitation ensembles for flood alert and risk management. Journal of Flood Risk Management **9**(4), 402–415 (2016)

[3] Katz, R.W.: Extreme value theory for precipitation: sensitivity analysis for climate change. Advances in water resources **23**(2), 133–139 (1999)

[4] Wilks, D.S.: Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. Agricultural and forest meteorology **93**(3), 153–169 (1999)

[5] Castellvı, F., Mormeneo, I., Perez, P.: Generation of daily amounts of precipitation from standard climatic data: a case study for argentina. Journal of hydrology **289**(1-4), 286–302 (2004)

[6] Shoji, T., Kitaura, H.: Statistical and geostatistical analysis of rainfall in central japan. Computers & Geosciences **32**(8), 1007–1024 (2006)

[7] Woolhiser, D.A., Roldan, J.: Stochastic daily precipitation models: 2. a comparison of distributions of amounts. Water resources research **18**(5), 1461–1468 (1982)

[8] Liu, Y., Zhang, W., Shao, Y., Zhang, K.: A comparison of four precipitation distribution models used in daily stochastic models. Advances in Atmospheric Sciences **28**(4), 809–820 (2011)

[9] Hussain, I., Spöck, G., Pilz, J., Yu, H.-L.: Spatio-temporal interpolation of precipitation during monsoon periods in pakistan. Advances in water resources **33**(8), 880–886 (2010)

[10] Guillot, G., Lebel, T.: Disaggregation of sahelian mesoscale convective system rain fields: Further developments and validation. Journal of Geophysical Research: Atmospheres **104**(D24), 31533–31551 (1999)

[11] Allcroft, D.J., Glasbey, C.A.: A latent gaussian markov random-field model for spatiotemporal rainfall disaggregation. Journal of the Royal Statistical Society: Series C (Applied Statistics) **52**(4), 487–498 (2003)

[12] Allard, D., Bourotte, M.: Disaggregating daily precipitations into hourly values with a transformed censored latent gaussian process. Stochastic environmental research and risk assessment **29**(2), 453–462 (2015)

[13] Maraun, D., Wetterhall, F., Ireson, A., Chandler, R., Kendon, E., Widmann, M., Brienen, S., Rust, H., Sauter, T., Themeßl, M., et al.: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. Reviews of Geophysics **48**(3) (2010)

[14] Rebora, N., Ferraris, L., von Hardenberg, J., Provenzale, A.: Rainfarm: Rainfall downscaling by a filtered autoregressive model. Journal of Hydrometeorology **7**(4), 724–738 (2006)

[15] Zhao, T., Bennett, J.C., Wang, Q., Schepen, A., Wood, A.W., Robertson, D.E., Ramos, M.-H.: How suitable is quantile mapping for postprocessing

gcm precipitation forecasts? Journal of Climate **30**(9), 3185–3196 (2017)

[16] Sigrist, F., Künsch, H.R., Stahel, W.A., *et al.*: A dynamic nonstationary spatio-temporal model for short term prediction of precipitation. The Annals of Applied Statistics **6**(4), 1452–1477 (2012)

[17] Benoit, L., Allard, D., Mariethoz, G.: Stochastic rainfall modeling at sub-kilometer scale. Water Resources Research **54**(6), 4108–4130 (2018)

[18] Bardossy, A., Plate, E.J.: Space-time model for daily rainfall using atmospheric circulation patterns. Water resources research **28**(5), 1247–1259 (1992)

[19] Ailliot, P., Thompson, C., Thomson, P.: Space–time modelling of precipitation by using a hidden markov model and censored gaussian distributions. Journal of the Royal Statistical Society: Series C (Applied Statistics) **58**(3), 405–426 (2009)

[20] Kleiber, W., Katz, R.W., Rajagopalan, B.: Daily spatiotemporal precipitation simulation using latent and transformed gaussian processes. Water Resources Research **48**(1) (2012)

[21] Lien, G.-Y., Kalnay, E., Miyoshi, T.: Effective assimilation of global precipitation: Simulation experiments. Tellus A: Dynamic Meteorology and Oceanography **65**(1), 19915 (2013)

[22] Li, W., Duan, Q., Ye, A., Miao, C.: An improved meta-gaussian distribution model for post-processing of precipitation forecasts by censored maximum likelihood estimation. Journal of Hydrology **574**, 801–810 (2019)

[23] Cecinati, F., Wani, O., Rico-Ramirez, M.A.: Comparing approaches to deal with non-gaussianity of rainfall data in kriging-based radar-gauge rainfall merging. Water Resources Research **53**(11), 8999–9018 (2017)

[24] Box, G.E., Cox, D.R.: An analysis of transformations. Journal of the Royal Statistical Society: Series B (Methodological) **26**(2), 211–243 (1964)

[25] Panofsky, H.A., Brier, G.W., Best, W.H.: Some Application of Statistics to Meteorology. Earth and Mineral Sciences Continuing Education, College of Earth and . . . , ??? (1958)

[26] Wilks, D.: Multisite generalization of a daily stochastic precipitation generation model. journal of Hydrology **210**(1-4), 178–191 (1998)

[27] Papalexiou, S., Koutsoyiannis, D., Makropoulos, C.: How extreme is extreme? an assessment of daily rainfall distribution tails. Hydrology &

Earth System Sciences **17**(1) (2013)

[28] Katz, R.W., Parlange, M.B., Naveau, P.: Statistics of extremes in hydrology. Advances in water resources **25**(8-12), 1287–1304 (2002)

[29] Furrer, E.M., Katz, R.W.: Improving the simulation of extreme precipitation events by stochastic weather generators. Water Resources Research **44**(12) (2008)

[30] Baxevani, A., Lennartsson, J.: A spatiotemporal precipitation generator based on a censored latent g aussian field. Water Resources Research **51**(6), 4338–4358 (2015)

[31] Naveau, P., Huser, R., Ribereau, P., Hannart, A.: Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. Water Resources Research **52**(4), 2753–2769 (2016)

[32] Tencaliec, P., Favre, A.-C., Naveau, P., Prieur, C., Nicolet, G.: Flexible semiparametric generalized pareto modeling of the entire range of rainfall amount. Environmetrics **31**(2), 2582 (2020)

[33] Tukey, J.W.: Modern techniques in data analysis. In: Proceedings of the NSF-Sponsored Regional Research Conference, vol. 7 (1977). Southern Massachusetts University

[34] Goerg, G.M.: The lambert way to gaussianize heavy-tailed data with the inverse of tukey'sh transformation as a special case. The Scientific World Journal **2015** (2015)

[35] Xu, G., Genton, M.G.: Tukey g-and-h random fields. Journal of the American Statistical Association **112**(519), 1236–1249 (2017)

[36] Ribereau, P., Naveau, P., Guillou, A.: A note of caution when interpreting parameters of the distribution of excesses. Advances in Water Resources **34**(10), 1215–1221 (2011)

[37] Coles, S.G., Dixon, M.J.: Likelihood-based inference for extreme value models. Extremes **2**(1), 5–23 (1999)

[38] Carreau, J., Naveau, P., Neppel, L.: Partitioning into hazard subregions for regional peaks-over-threshold modeling of heavy precipitation. Water Resources Research **53**(5), 4407–4426 (2017)

[39] Vannitsem, S., Naveau, P.: Spatial dependences among precipitation maxima over belgium. Nonlinear Processes in geophysics **14**(5), 621–630 (2007)

[40] Gradshteyn, I.S., Ryzhik, I.M.: Table of Integrals, Series, and Products,

7th edn. Academic press, ??? (2007)

[41] Embrechts, P., Klüppelberg, C., Mikosch, T.: Modelling Extremal Events: for Insurance and Finance vol. 33. Springer, ??? (2013)