Stochastic downscaling of precipitation with neural network conditional mixture models

Julie Carreau, Mathieu Vrac

Stochastic downscaling of precipitation with neural network conditional mixture models

Julie Carreau[1] and Mathieu Vrac[2]

Received 15 October 2010; revised 13 July 2011; accepted 4 August 2011; published 4 October 2011.

[1] We present a new class of stochastic downscaling models, the conditional mixture models (CMMs), which builds on neural network models. CMMs are mixture models whose parameters are functions of predictor variables. These functions are implemented with a one-layer feed-forward neural network. By combining the approximation capabilities of mixtures and neural networks, CMMs can, in principle, represent arbitrary conditional distributions. We evaluate the CMMs at downscaling precipitation data at three stations in the French Mediterranean region. A discrete (Dirac) component is included in the mixture to handle the "no-rain" events. Positive rainfall is modeled with a mixture of continuous densities, which can be either Gaussian, log-normal, or hybrid Pareto (an extension of the generalized Pareto). CMMs are stochastic weather generators in the sense that they provide a model for the conditional density of local variables given large-scale information. In this study, we did not look for the most appropriate set of predictors, and we settled for a decent set as the basis to compare the downscaling models. The set of predictors includes the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalyses sea level pressure fields on a 6×6 grid cell region surrounding the stations plus three date variables. We compare the three distribution families of CMMs with a simpler benchmark model, which is more common in the downscaling community. The difference between the benchmark model and CMMs is that positive rainfall is modeled with a single Gamma distribution. The results show that CMM with hybrid Pareto components outperforms both the CMM with Gaussian components and the benchmark model in terms of log-likelihood. However, there is no significant difference with the log-normal CMM. In general, the additional flexibility of mixture models, as opposed to using a single distribution, allows us to better represent the distribution of rainfall, both in the central part and in the upper tail.

Citation: Carreau, J., and M. Vrac (2011), Stochastic downscaling of precipitation with neural network conditional mixture models, *Water Resour. Res., 47*, W10502, doi:10.1029/2010WR010128.

1. Introduction

[2] General circulation models (GCMs) solve the principal physics equations of the dynamics of the atmosphere and of the oceans together with their interactions on a 3-D grid over the globe. GCMs allow us to simulate climate variables and to study the mechanisms of the present, past, and future climate of the Earth [e.g., *Gladstone et al.*, 2005; *Intergovernmental Panel on Climate Change* (*IPCC*), 2007a]. In the last two decades, the IPCC has compared and studied the outputs of different GCMs. These analyses attempt to understand the many processes involved in the current and upcoming climate changes resulting from different greenhouse gas emission scenarios [*IPCC*, 2007a]. In addition, the IPCC seeks to evaluate the potential impacts of climate changes on economy, agriculture, and ecology in the next decades [*IPCC*, 2007b]. Such impact studies require climate simulations at high spatial resolution (small scale), ranging from a few kilometers down to station locations. In particular, precipitation, which is of major importance in agriculture, vegetation, and flood risk assessment, has a strong spatial variability. However, the spatial resolution at which GCMs operate (about 200×200 km) is typically too low to capture such spatial variability. Other reasons why GCMs struggle to reproduce precipitation are related to the features of the distribution of precipitation, namely, boundedness at zero, nonnormality, and the presence of extreme values at local scale with a potential destructive power.

[3] In this context, downscaling techniques have been developed to bridge the gap between large- and small-scale variables [*Hewitson and Crane*, 1996]. There are two different approaches to downscaling. The dynamical approach consists in refining GCMs over a higher-resolution grid. These refined GCMs, called regional climate models (RCMs), operate at a resolution down to about 10 km. RCMs have a high computational cost and thus are often limited in their uses to restricted regions and periods of time. On the other hand, the statistical approach to downscaling proposes statistical models that relate large-scale GCM outputs to local-scale climate

[1]HydroSciences Montpellier, UMR 5569, CNRS/IRD/UM1/UM2, Université de Montpellier 2, Montpellier, France.
[2]Laboratoire des Sciences du Climat et de l'Environnement, IPSL, CNRS/CEA/UVSQ, Gif-sur-Yvette, France.

0043-1397/11/2010WR010128

variables. These statistical downscaling models (SDMs) are, in general, computationally more tractable and can be easily applied to many GCM runs covering large regions and long time periods. Moreover, SDMs offer a great modeling flexibility that has proved useful, for example, in extreme event modeling [e.g., *Vrac and Naveau*, 2007] and for uncertainty assessment [e.g., *Semenov*, 2007]. SDMs generally borrow from one or more of the following sets of methods: transfer functions, stochastic weather generators, and weather typing [e.g., *Maraun et al.*, 2010]. Transfer functions aim at directly translating large-scale data into local-scale values by performing linear or nonlinear regressions. Given large-scale information x, the downscaled local response y is a function $\hat{y}(x)$, which usually estimates $E[Y|\mathbf{X} = x]$ ($\hat{y}(x)$ could estimate another quantile in the case of quantile regression [e.g., see *Cannon*, 2011; *Friederichs and Hense*, 2007]). Among the regression-based transfer functions, we find linear regression [e.g., *Wigley et al.*, 1990; *Huth*, 2002; *Wilby et al.*, 2002; *Busuioc et al.*, 2008; *Goubanova et al.*, 2010], nonlinear parametric models such as polynomial regression [*Hewitson*, 1994; *Sailor and Li*, 1999], nonparametric regression based on splines, generalized additive models [*Vrac et al.*, 2007b; *Salameh et al.*, 2009], and neural networks [e.g., *Snell et al.*, 2000; *Cannon and Whitfield*, 2002; *Haylock et al.*, 2006; *Huth et al.*, 2008; *Ghosh and Mujumdar*, 2008].

[4] Stochastic weather generators (WGs) provide a way to simulate meteorological variables such as precipitation or temperature on the basis of probability density function (pdf) models [e.g., *Semenov and Barrow*, 1997; *Semenov et al.*, 1998; *Wilks*, 1999]. WGs are calibrated so that the simulated observations reproduce the statistical properties of the corresponding local observations. In a downscaling framework, WGs simulate a local variable Y given large-scale information x by building a model for the conditional distribution $Y|\mathbf{X} = x$. To achieve this goal, the parameters of the density model can be seen as functions of some appropriate large-scale information such as American weather regimes [e.g., *Vrac et al.*, 2007a], the North Atlantic Oscillation index [e.g., *Yang et al.*, 2005], or other large-scale climate variables (see *Wilks and Wilby* [1999] for a review). Thus, changes in large-scale variables are transferred into the local-scale density parameters so that the simulated observations evolve accordingly [*Vrac and Naveau*, 2007].

[5] The last set of methods, the weather typing methods, seeks to cluster and classify large-scale atmospheric circulation situations into recurrent weather patterns and assumes that each weather pattern gives rise to similar local-scale meteorological conditions or distributions [e.g., *Huth*, 2001; *Vrac et al.*, 2007c]. Weather typing can serve as a preprocessing step before building transfer functions [e.g., *Huth et al.*, 2008] or weather generators [e.g., *Schnur and Lettenmaier*, 1998; *Vrac et al.*, 2007a; *Vrac and Naveau*, 2007].

[6] Transfer function methods based on neural networks were applied successfully to downscaling [e.g., *Snell et al.*, 2000; *Cannon and Whitfield*, 2002]. They are able to reproduce the nonlinear relationship between large-scale atmospheric data and precipitation [*Hewitson and Crane*, 1996]. However, as regression algorithms, neural networks are limited in the following respects: they underestimate extreme events [*Haylock et al.*, 2006], they provide only pointwise prediction (i.e., no confidence interval or other measure of uncertainty is provided since twice the same input will produce twice the same output), and they cannot account for trends in the variability. In order to enable neural networks to provide probabilistic information, *Williams* [1998] has proposed the following model for rainfall data. The occurrence and the intensity processes are modeled jointly by means of a mixture with a discrete and a continuous component. The discrete component relates to the occurrence process, and the continuous component, which is taken to be a Gamma density, relates to the rainfall intensity process. The parameters of this two-component mixture depend on predictor variables by the functions computed by a neural network. The error function of regression neural networks is the mean-square error, which leads to the estimation of the conditional expectation. In Williams' model, the error function is the conditional log-likelihood of the two-component mixture. In this way, the neural network provides information on the whole distribution of rainfall, not only the conditional expectation. This kind of conditional density model based on a neural network can be seen as a continuous extension of quantile regressions [e.g., *Cannon*, 2011; *Friederichs and Hense*, 2007]. *Williams* [1998] used lagged observations of precipitation as predictor variables to model trend and seasonality of rainfall. Recently, *Haylock et al.* [2006] and *Cawley et al.* [2007] proposed Williams' model as a way for neural networks to model predictive uncertainty in a downscaling application. In the latter, large-scale atmospheric variables are taken as predictor variables for the two-component mixture parameters. Such a model belongs to the class of stochastic WGs since it provides a conditional density model of the local variable given large-scale information.

[7] In this paper, we extend Williams' model by considering a mixture of distributions rather than a single Gamma distribution to model rainfall intensity. Mixtures are flexible nonparametric density estimators that can account for asymmetric and multimodal distributions [e.g., *Priebe*, 1994; *McLachlan and Peel*, 2000]. Our proposed downscaling model is thus a conditional mixture model (CMM) in which one of the components is discrete to model rainfall occurrence. Mixture parameters are estimated by a single layer feed-forward neural network given predictor variables. We evaluate the performances at modeling rainfall intensity of three CMMs that differ in the type of continuous densities (Gaussian, log-normal, or hybrid Pareto) they use as mixture components. We compare CMMs with the two-component conditional mixture from *Williams* [1998] that we use as a benchmark model. We compare these four stochastic models at downscaling precipitation at three rain gauge stations in the French Mediterranean area. The distribution of precipitation is allowed to evolve according to the large-scale atmospheric information in all four stochastic downscaling models. The difference resides in the density model chosen for rainfall intensity. In CMMs, given the state of the atmosphere, described by the predictor variables, precipitation can be in one of several states or regimes, which can be thought of as representing the smaller-scale processes. These regimes are represented by the mixture components.

[8] This paper is structured as follows. Section 2 describes in detail the conditional mixture models. Section 3 presents the precipitation and large-scale data of our downscaling

application together with the preprocessing, training, and model selection steps. Results in terms of log-likelihood and analyses of conditional quantiles and climatological characteristics of the downscaling application are given in section 4. Section 5 provides discussions and conclusions.

2. Statistical Downscaling Models

[9] We adopt the following notation. Let Y be a random variable representing the precipitation process at a given station and let $\mathbf{X}$ be a vector of random variables representing the predictors that include large-scale atmospheric information. Lowercase letters y and x refer to values taken by the corresponding random variables.

2.1. Conditional Mixture Models

[10] Following *Williams* [1998], we consider a mixture with a discrete component to model jointly the occurrence and intensity processes of precipitation:

$$\phi(y;\psi) = (1-\alpha)\delta(y) + \alpha\phi_0(y;\psi_0), \tag{1}$$

where α is the probability of rain occurrence, δ is the Dirac function, which is such that $\int_{-\infty}^{\infty} f(z)\delta(z-a)\mathrm{d}z = f(a)$ and $\delta(z-a) = 0$ for $z \neq a$, and $\phi_0(\cdot;\psi_0)$ is the density model with parameter ψ_0 for rainfall intensity. Therefore, $(1-\alpha)\delta(y)$ handles the "no rain" events while $\alpha\phi_0(\cdot;\psi_0)$ handles the case of positive rainfall. The parameter vector of the mixture in equation (1) is thus $\psi = (\alpha, \psi_0)$. In the work by *Williams* [1998], $\phi_0(\cdot;\psi_0)$ is the Gamma density. We propose to use mixture models instead:

$$\phi_0(y;\psi_0) = \sum_{j=1}^{m} \pi_j f(y;\theta_j), \tag{2}$$

where $f(\cdot;\theta_j)$ is a density with parameter vector θ_j, π_j is the weight of component j, and ψ_0 is the vector that concatenates all the mixture parameters $(\pi_1,\ldots,\pi_m,\theta_1,\ldots,\theta_m)$.

[11] We can take into account the dependence of the distribution of precipitation on predictor variables $\mathbf{x}$ by considering the parameters of the mixture of equation (1) as functions of $\mathbf{x}$: $\psi(\mathbf{x}) = [\alpha(\mathbf{x}), \pi_1(\mathbf{x}),\ldots,\pi_m(\mathbf{x}),\theta_1(\mathbf{x}),\ldots,\theta_m(\mathbf{x})]$. In the downscaling application, the predictors are large-scale atmospheric variables. In the conditional mixture model, precipitation can be thought of as being in one of several states or regimes given the state of the atmosphere, as represented by the predictor variables. These states are not directly observed. The hidden states could be seen as resulting from subscale processes that are not accounted for by the large-scale atmospheric variables. Each of the hidden states is modeled by a component of the mixture. The mixture weight $\pi_j(\mathbf{x})$ gives the probability of occurrence of state j, and $f(y;\theta_j(\mathbf{x}))$ is the density of intensity given state j. This view of CMMs bears similarities with nonhomogeneous hidden Markov models (NHMMs) [*Bellone et al.*, 2000]. For NHMMs, the hidden state represents a weather pattern and is assumed to follow a first-order Markov chain whose transition probability depends on the predictors. This is one main difference with CMMs, which capture the serial dependence through the predictor variables exclusively. In NHMMs, the distribution of rainfall in a given hidden state (the so-called emission density) is similar to the component density in the CMMs.

[12] Some authors [e.g., *Hewitson and Crane*, 1996] have shown that the relationship between large-scale atmospheric variables and precipitation is nonlinear. To take this into account, a convenient way to implement the functions in $\psi(\mathbf{x})$ in the conditional mixture is by means of a one-layer feed-forward neural network [*Bishop*, 1995]. Such neural networks are flexible nonparametric models that can, in principle, approximate any continuous function [see *Hornik*, 1991]. We implemented the feed-forward neural network in a standard way but added an extra linear connection between the input variables and the neural network outputs so that the linear model is a special case of the neural network corresponding to zero hidden units. Let H be the number of hidden units. Each hidden unit h, with $h = 1,\ldots,H$, computes a linear combination of the predictors x_i, which is then nonlinearly transformed by means of the hyperbolic tangent (tanh):

$$z_h = \tanh\left(\sum_{i=1}^{d} \upsilon_{h,i}x_i + \upsilon_{h,0}\right), \tag{3}$$

where $\upsilon_{h,i}$ are the input layer weights linking the predictors to the hidden units. Then, a linear combination of the hidden unit activations z_h plus a linear combination of the predictors (the extra linear connection mentioned above) is transformed by a function g in order to ensure range constraint such as positivity:

$$\psi_j = g\left(\sum_{h=1}^{H} w_{j,h}z_h + \sum_{i=1}^{d} \tilde{\upsilon}_{j,i}x_i + w_{j,0}\right), \tag{4}$$

where $w_{j,h}$ are the hidden unit weights which compute the non-linear part, $\tilde{\upsilon}_{j,i}$ are the linear weights of the extra linear connection and g is chosen according to the mixture parameter ψ_j (see *Carreau and Bengio* [2009b] for more details on the function g). Let ω represent the neural network weights, i.e., $\upsilon_{h,i}$ in equation (3) and $w_{j,h}$ and $\tilde{\upsilon}_{h,i}$ in equation (4). Then the conditional mixture can be written as

$$\phi_\omega(y|x) = \phi(y;\psi_w(\mathbf{x})), \tag{5}$$

where $\phi(\cdot;\psi)$ is defined in equation (1) and $\psi_\omega(\mathbf{x})$ emphasizes that the mixture parameters depend on the neural network weights ω. Those weights are calibrated by minimizing the negative log-likelihood of the conditional mixture over the training set. The optimization is done with a conjugate gradient descent algorithm, and the gradient is computed with the back-propagation algorithm [*Rumelhart et al.*, 1986]. To avoid local minima, the optimization is restarted several times from different initial values of the neural network weights, and the weights leading to the lowest training error are kept. Depending on the type of CMMs and on the data set, three to five restarts seem to be enough. This procedure helps to stabilize the optimization and hence the performance of each model. The complexity level of the conditional mixture, that is, its degree of adaptiveness, is controlled by both the number of hidden units of the neural network and the number of components of the

mixture. Those are called hyperparameters. The hyperparameters have to be carefully selected in order to trade off bias (the model misfit with respect to the data) and variance (also called overfitting or learning by heart). This can be done by selecting the number of hidden units and components via the so-called cross-validation method, which will be described in section 3.2.

2.2. Three Families of CMMs

[13] We evaluate three conditional mixture models, all with a discrete component, which differ in the type of mixture components (i.e., $f(\cdot;\theta_j)$ in equation (2)) and compare them with the two-component conditional mixture from *Williams* [1998]. We took Gaussian, log-normal, and hybrid Pareto as mixture components. Since the intensity of rain is strictly positive, we ensure that the Gaussian and the hybrid Pareto mixtures have only positive density on the positive axis by truncation:

$$\phi_0(y;\psi_0) = \begin{cases} \tilde{\phi}_0(y;\psi_0)/\left[1-\tilde{\Phi}_0(0;\psi_0)\right] & \text{if } y<0 \\ 0 & \text{otherwise} \end{cases}, \quad (6)$$

where $\tilde{\phi}_0(\cdot;\psi_0)$ is either the Gaussian or the hybrid Pareto mixture and $\tilde{\Phi}_0(\cdot;\psi_0)$ is the corresponding cumulative distribution function. The conditional Gaussian mixture was initially presented by *Bishop* [1994]. It combines the approximation capabilities of neural networks and mixture models and can, in principle, represent arbitrary conditional distributions. This motivates the use of conditional Gaussian mixtures to model rainfall intensity. However, when the data set is small and the distribution is heavy tailed, Gaussian mixtures might underestimate the upper tail of the distribution [*Carreau and Bengio*, 2009a]. Since precipitation in the French Mediterranean area where the rain gauges are located is typically heavy tailed [*Delrieu et al.*, 2005], we also considered conditional mixtures with log-normal and hybrid Pareto [*Carreau and Bengio*, 2009a] components. The log-normal density is often employed to model positive intensities of precipitation [see, e.g., *Cho et al.*, 2004]. Indeed, the shape of the log-normal distribution, i.e., its asymmetry and its support on the positive axis, is more suited to model precipitation data. Also, log-normal mixtures are often considered to model moderately heavy tailed data [*McNeil*, 1997; *Frigessi et al.*, 2002]. However, the log-normal might suffer from the same difficulty to model heavy-tailed distribution as the Gaussian. This is because, like the Gaussian, the log-normal tail eventually decreases exponentially fast, whereas the heavy-tailed distribution decreases polynomially fast [*Embrechts et al.*, 1997]. So we propose the hybrid Pareto as a mixture component to explicitly take extreme observations into account. The generalized Pareto distribution (GPD) has been put forward as a model that can approximate all kinds of tails (exponential, polynomial, and finite) [*Pickands*, 1975]. The GPD is designed to model only large observations. The hybrid Pareto provides a smooth extension of the GPD to the whole real axis and makes possible seamless inclusion of the GPD in a mixture while inheriting its tail approximation properties. Another way to include the GPD into a mixture model in a downscaling application was proposed by *Vrac and Naveau* [2007].

[14] The hybrid Pareto is made of a Gaussian stitched together with a generalized Pareto so as to ensure continuity of the density and of its derivative [*Carreau and Bengio*, 2009a]. The hybrid Pareto density is given by

$$h(y;\mu,\sigma,\xi) = \begin{cases} f(y;\mu,\sigma)/\gamma & \text{if } y \le u \\ g(y-u;\xi,\beta)/\gamma & \text{otherwise} \end{cases}, \quad (7)$$

where $f(y;\mu,\sigma)$ is the Gaussian density with parameters $\mu \in \mathbb{R}$ and $\sigma > 0$, $u \in \mathbb{R}$ is the junction point or threshold, $\gamma = 1 + \int_{-\infty}^{\alpha} f(y;\mu,\sigma)\mathrm{d}y$ is the normalization factor, and $g(y-u;\xi,\beta)$ is the generalized Pareto density with scale parameter $\beta > 0$ and tail index parameter $\xi \in \mathbb{R}$:

$$g(y-u;\xi,\beta) = \begin{cases} \frac{1}{\beta}\left(1+\frac{\xi}{\beta}(y-u)\right) & \text{if } \xi \neq 0 \\ \frac{1}{\beta}\exp\left(-\frac{y-u}{\beta}\right) & \text{if } \xi = 0 \end{cases}. \quad (8)$$

In addition to the location and scale parameters μ and σ of the Gaussian, the hybrid Pareto has a third parameter, the tail index ξ, which characterizes the heaviness of the tail of the distribution. Positive ξ indicates a polynomially decreasing tail, i.e., a heavy tail; the tail is called exponential or light when $\xi = 0$ and is called finite for negative ξ. Since precipitation in the French Mediterranean area where the rain gauges are located is typically heavy tailed [*Delrieu et al.*, 2005], we will focus on $\xi > 0$ for the hybrid Pareto tail index parameter. Because of the continuity constraints, the junction point u and the scale parameter of the GPD β are functions of ξ, μ, and σ. Note that, by construction, the hybrid Pareto inherits the tail approximation property of the generalized Pareto. The tail heaviness of a mixture of hybrid Pareto is driven by the component with the heaviest tail. The tail index that characterizes the mixture is then given by $\xi^* = \max_j \xi_j$, where ξ_j denotes the tail index parameter of component j. In the conditional case, the tail index depends on the predictor variables $\xi^*(x) = \max_j \xi_j(x)$. In our downscaling application, it means that the behavior of the tail of the distribution of the precipitation process is allowed to vary with regard to the large-scale atmospheric conditions. *Carreau and Bengio* [2009a, 2009b] have shown that in most cases, the hybrid Pareto mixture, conditional or not, is able to provide a decent estimator of the tail index of the data. It is more challenging in the case of conditional density estimation because of the introduction of large uncertainties in the tail index estimation [*Friederichs*, 2010]. For this reason, a penalty term to control the tail index parameter estimation within a hybrid Pareto mixture was proposed by *Carreau et al.* [2009].

2.3. Penalty Term for Tail Index Estimation

[15] Introducing a penalty term is similar to the approach taken by *Coles and Dixon* [1999]. They argued, by analogy with the probability weighted moment (PWM) estimator [*Hosking et al.*, 1985], that the performance of the maximum likelihood estimator (MLE) in small samples could be improved by imposing a restriction similar to $\xi < 1$ (which implies that the expectation is finite). This assumption leads to a reduced variance of the PWM estimator at the cost of a negative bias [*Coles and Dixon*, 1999]. *Coles*

and Dixon [1999] show that adding a penalty that enforces a similar prior assumption into the MLE leads to similar improvement of the MLE estimator.

[16] The penalty designed for the hybrid Pareto mixture is based on the following assumptions. One or a few components should have tail index parameters high enough to model the upper tail of precipitation data correctly. On the other hand, most components should have their tail index parameters close to zero and be dedicated to modeling the central part of the distribution. Previous studies [e.g., *Gardes and Girard*, 2010] lead us to assume that the conditional tail index $\xi^*(x)$ in the region where the studied rain gauges are located varies between 0.16 and 0.26. We thus suggest penalizing the log-likelihood of the hybrid Pareto mixture according to a density, examples of which are shown in Figure 1. This density has one mode at zero (for the majority of components with light tail indexes that are expected to model the central part of the distribution) and a second mode at 0.2 that covers the interval [0.16,0.26] (for the few components that are expected to model the upper tail of rainfall distribution.) The bimodal densities of Figure 1 are given by a two-component mixture made of an exponential and a Gaussian:

$$p(\xi;\eta,\rho) = \eta\exp\{-\eta x\}/2 + \exp\left\{-(\xi-0.2)^2/(2\rho^2)\right\}/\left(2\sqrt{2\pi\rho}\right), \quad (9)$$

where η and ρ are the parameters of the exponential and the standard deviation of the Gaussian, respectively. The exponential puts one mode at zero while the Gaussian is centered at 0.2. The estimation then consists in maximizing a penalized log-likelihood obtained by adding the logarithm of the bimodal density to the usual log-likelihood. In practice, for the conditional mixture, the neural network weights ω are found by minimizing

$$\mathcal{L}(\omega) = -\sum_{i=1}^{n}\log[\phi_\omega(y_i|\mathbf{x}_i)] - \frac{\lambda}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\log p\left(\xi_{i,j};\eta,\rho\right),$$

where n is the number of observations, m is the number of mixture components, $\phi_\omega(y_i|\mathbf{x}_i)$ is the conditional mixture defined in equation (5), evaluated at point i, $p\left(\xi_{i,j};\eta,\rho\right)$ is the bimodal density defined in equation (9) with $\xi_{i,j} = \xi_j(\mathbf{x}_i,\omega)$, and λ controls the trade-off between minimizing the negative log-likelihood and the penalty. The penalty term introduces extra hyperparameters, namely, λ, η, and ρ. These hyperparameters will be chosen by cross-validation; see section 3.2. As η and ρ influence the range of values taken by the tail index parameters of the mixture, we restrict these hyperparameters to vary as follows: $\eta \in \{60,20\}$ and $\rho \in \{0.02,0.03,0.05\}$. This gives the possibility for the penalty to adapt to the rain gauge–specific tail behavior without letting the tail index parameters take unrealistic values.

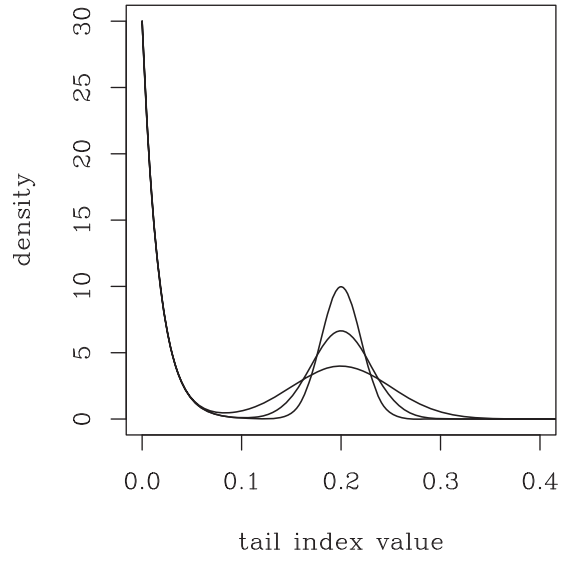

Figure 1. Three examples of bimodal density that reflect our prior assumptions regarding the tail index parameters of the hybrid Pareto mixture in the downscaling application. Most components are assumed to have tail index parameters values near zero (the first mode of the density) to model the central part of the distribution, and some tail index parameters take values near 0.2 (the second mode) to model the upper tail of the distribution.

3. Data Sets and Calibration of the Models

3.1. Large- and Local-Scale Data

[17] The local-scale data are precipitation from three rain gauge stations: Orange, Sète, and Le Massegros. These stations are located in the Cévennes-Vivarais region, which is part of the French Mediterranean area. Because of the Mediterranean influence and of the mountainous back country, the Cévennes-Vivarais region is well known for intense rain events, especially in the fall [*Delrieu et al.*, 2005]. For each rain gauge, we have daily rainfall measurements over 46 years (from 1 January 1959 to 31 December 2004) extracted from the European Climate Assessment & Dataset (ECA&D [*Klein Tank et al.*, 2002]). Precipitation values smaller than 1 mm are set to zero in order to discard values possibly resulting from measurement error.

[18] In this work, we did not seek the best set of predictors to drive the downscaling models. Our goal is to illustrate and compare the performances and advantages of the four downscaling models. For this, we selected a set of predictors with a decent predictive power to drive all four downscaling models, including the benchmark two-component model, so that the comparison among models is fair. Although many large-scale atmospheric variables are relevant to downscale precipitation, we only include the sea level pressure (SLP) field because it has been shown to be a good predictor of precipitation [e.g., *González-Rouco et al.*, 2000]. We selected a subset of the SLP field from the National Centers for Environmental Prediction/National Center for Atmospheric Research (NCEP/NCAR) reanalysis data [*Kalnay et al.*, 1996]. We decided to use a simple geometry (a rectangle of reanalysis grid cells centered on the rain gauges) and then proceeded by trial and error to select the size of the rectangle. We found a 6×6 grid to be a reasonable choice to provide large enough regional information in order to capture the relevant large-scale synoptic information.

[19] In addition to the SLP reanalysis data, we include as predictors three date variables representing the year, the

month, and the week of an observation. The year variable encodes the year as the difference with a reference year (here 1970). This allows us to model a potential trend. The month variable is built from the circular difference with the month of January: this difference goes from zero (January) to six (July) and decreases again to 1 (December). The same kind of computation is done to produce the week variable, but the circular difference operation is taken over 52 weeks. The circular differences are then smoothed in order to have continuous variables in the range [0,1]. These circular difference variables (see *Carreau et al.* [2009] for another application of the date variables) are an alternative to using sine and cosine [see *Williams*, 1998] in order to provide a way for the conditional mixtures to characterize seasonality in the distribution of precipitation.

[20] Seasonal signals might be present both in the SLP and the date variables. Therefore, although we are aware that a more classical application of principal component analysis (PCA) [*Jolliffe*, 1986] would separate the treatment of these two kind of variables, we apply PCA on all 39 initial variables (36 SLP plus 3 date variables). Our use of PCA serves merely to reduce dimensionality and remove redundancy among the predictors. The observed predictors are centered and scaled before the application of PCA in order to ensure that their values are in the same range. We extract the four principal components in order to keep 90% of the variance of the data. The four projected predictors are further standardized to have zero mean and unit standard deviation (this preprocessing is required to facilitate the training of neural networks).

3.2. Training and Hyperparameter Selection

[21] Training refers to the optimization of the parameters of the downscaling models (in all four cases, the neural network weights) with respect to an error function that measures the misfit between the model and a data set, called the training set. The error function is chosen according to the task at hand. Since we deal here with conditional density estimation, the error function is the conditional negative log-likelihood (eventually with a penalty term). Hyperparameters, such as the number of hidden units, control the complexity level of an algorithm, which often is proportional to the number of parameters. The higher the complexity level is, the better the fit of the model to the training set becomes. Therefore, hyperparameters cannot be selected by minimizing the error function on the training set. Theoretically, the optimal hyperparameters should minimize simultaneously bias (the model misfit) and variance (also called overfitting). In practice, since the underlying process is unknown, bias and variance have to be estimated. This can be achieved with the cross-validation procedure [*Bishop*, 1995], which computes an estimation of the validation error (the value of the error function on data that are not used for training). The validation error can be decomposed into a sum of bias plus variance and some noise (the residual part of the error that cannot be fit). Let's define a set $\mathcal{H}$ that contains possible values from which to choose the hyperparameters. We implemented a fivefold cross-validation as follows.

[22] 1. Split the training set into five subsets, called folds: $L_1, \ldots, L_5$.

[23] 2. Leave one of the folds, L_k, aside.

[24] 3. For each $h \in \mathcal{H}$, train the model (i.e., determine its parameters given $h \in \mathcal{H}$) on the four remaining folds.

[25] 4. Evaluate the error function $E_k(h)$ (the conditional negative log-likelihood) on the left-aside fold L_k.

[26] 5. Return to step 2 until all fives are left aside in turn.

[27] 6. Select the best hyperparameter: $h^* = \arg\min_h \sum_{k=1}^{5} E_k(h)$.

[28] For the benchmark Gamma conditional model, the only hyperparameter is the number of hidden units. For the conditional mixtures described in sections 2.1 and 2.2, there is a second hyperparameter, which is the number of mixture components. For the hybrid Pareto conditional mixture, there are three additional hyperparameters for the penalty term. The number of hidden units are chosen in the set $\{0,2,4,8\}$, where zero hidden units means that only a linear function is computed by the hidden layer of the neural network, and the number of mixture components is chosen among $\{1,2,4,8\}$. The penalty hyperparameters of the hybrid Pareto CMMs vary as follows: $\lambda \in \{0,1,10,20,60,100\}$, $\eta \in \{60,120\}$, and $\rho \in \{0.02,0.03,0.05\}$. Therefore, in the cross-validation procedure, we choose among four hyperparameters for the benchmark model, 16 combinations for the Gaussian and log-normal CMMs, and 576 combinations for the hybrid Pareto CMM. For this last model, hyperparameter selection takes about 3 days of computation time on a single CPU.

[29] The 46 year data set is split into a training set of 25 years (from 1 January 1959 to 31 December 1983) and a test set of 21 years (from 1 January 1984 to 31 December 2004). The training set is first used to select the hyperparameters of each SDM with the fivefold cross-validation method. Once the hyperparameters are selected, each model is trained anew on the whole training set. The test set serves exclusively for comparison and evaluation of the SDMs.

4. Results and Analyses

4.1. Global Comparisons

[30] The hybrid Pareto conditional mixture is the most complex model and requires careful and longer training to address properly the issue of conditional tail index estimation. In order to determine if using such a model is justified by the data, we first compare the other three downscaling models in terms of average of differences in conditional log-likelihood with the hybrid Pareto CMM on the test set:

$$\frac{1}{n}\sum_{i=1}^{n}\left[\log \phi_{\omega}^{h}(y_i|x_i) - \log \phi_{\omega}(y_i|x_i)\right], \tag{10}$$

where $\phi_{\omega}^{h}(y_i|x_i)$ is the hybrid Pareto CMM density and $\phi_{\omega}(y_i|x_i)$ is the density of one of the other three models. Positive values indicate that the hybrid Pareto CMM performs better. The evaluation of the performances (or differences in performance) on the test set provides a fair way to compare models even if the number of parameters might vary across models. The risk of overfitting resulting from too many parameters is implicitly taken into account because the performances are evaluated on new data, which did not serve for training or hyperparameter selection. Therefore, we do not need to introduce a penalty term for the number of parameters in the model, as is the case in popular fit criteria

Table 1. Hyperparameters Selected via Cross-Validation for the Three Rain Gauge Stations for Conditional Mixture Models by Type of Components[a]

	Hybrid Pareto	Gaussian	Log-normal	Benchmark
		Orange		
(h, m)	(2, 2)	(0, 4)	(2, 2)	$h = 2$
(λ, η, ρ)	(20, 120, 0.03)	–	–	–
		Sète		
(h, m)	(2, 2)	(0, 8)	(2, 2)	$h = 2$
(λ, η, ρ)	(20, 120, 0.05)	–	–	–
		Le Massegros		
(h, m)	(2, 4)	(4, 4)	(2, 2)	$h = 4$
(λ, η, ρ)	(10, 60, 0.03)	–	–	–

[a]The hyperparameters are the number of hidden units and of components selected (h, m) and penalty parameters (λ, η, ρ) for hybrid Pareto conditional mixtures. The only hyperparameter for the benchmark model is the number of hidden units h.

such as the Bayesian information criterion [*Schwarz*, 1978], which are measured on the training set.

[31] Table 1 presents the hyperparameters selected by fivefold cross-validation on the training set for the four downscaling models. Table 2 shows the average of differences in conditional log-likelihood on the test set along with standard errors for the three competing models (Gaussian CMM, log-normal CMM, and Gamma benchmark) on the three rain gauge stations. The cases where the hybrid Pareto CMM performed significantly better are in bold. We see that the hybrid Pareto outperforms the Gaussian CMM and the Gamma benchmark on all three stations. However, we cannot really distinguish the hybrid Pareto CMM from the log-normal CMM on the basis of this criterion. It is not so surprising that log-normal CMMs have a good performance because the data set is fairly large (which helps training) and the asymmetry and positive support of the log-normal are well suited to model rainfall data.

[32] We illustrate the forthcoming analyses on the Orange station only since the two other stations provide similar insights into the differences between SDMs. We computed the mixture parameters corresponding to the predictors on the test set and then randomly generated precipitation data according to the conditional mixture parameters. This gives us realizations of the precipitation process according to each SDM over the test set. We repeated this a thousand times. To visually assess how realistic a model is at reproducing the precipitation process, we looked at QQ-plots (on logarithmic scale) of the simulated values against the observations in the test set for values greater than 1 mm. This is illustrated in Figure 2 for the hybrid Pareto CMM (the other two CMMs are not shown because they are similar) and the benchmark model. Models that are in accordance with the data should be close to the diagonal line. We see that the benchmark model is less apt at modeling both the central part (overestimation) and the upper part (underestimation) of the distribution.

[33] For each of the 1000 generated time series, we computed the frequency of wet and dry spells of at least k days. These frequency counts were then normalized to obtain proportions. We took the 5% and 95% empirical quantiles over the 1000 replications for each spell length. We also computed the proportion of wet and dry spells of at least k days from the test data. The resulting wet and dry spell confidence intervals together with the spells computed from the observations are compared on the logarithmic scale in Figure 3 for the hybrid Pareto CMM and the benchmark model, with the other two CMMs giving almost identical results. From these wet and dry spell plots, we can conclude that the separate modeling of the occurrence process (for both CMMs and the benchmark model) by means of the discrete mixture component allows us to capture most of the serial dependence, although the shortest wet spell (2 and 3 days) probabilities are slightly underestimated. This dependence is taken into account implicitly by the neural network, which estimates rainfall probability given the predictor variables. Serial dependence could also be explicitly taken into account by including lagged observations in the predictor variables.

[34] As shown in this section, there is a significant gain (in terms of log-likelihood and through the analyses of the QQ-plots) from using a mixture model instead of a single Gamma distribution to model rainfall intensity. In the following analyses, we thus focus on the hybrid Pareto and log-normal CMMs, which are the most likely models for our downscaling application.

4.2. Seasonal Cycles

[35] In this section, we analyze climatologies on the Orange test set. For each day of the 21 years in the test set, we computed the 99% quantile $y_{0.99}(\mathbf{x})$ and the rain probability $\alpha(x)$ according to the SDMs. Then, for each of the 365 days of the generic year (or 366 considering leap years), we estimated empirical quantiles of levels 5%, 50%, and 95% from the 21 values of $y_{0.99}(\mathbf{x})$ and $\alpha(x)$ computed from the models. In order to compare the SDMs with the observations, we also computed the empirical 99% quantile from the test data for each day of the 365/366 days of the year and the proportion of rainy days. For CMMs, conditional quantiles are obtained by solving numerically the following equation for y_p: $\Phi_\omega(y_p|\mathbf{x}) = p$, where $p \in [0,1]$ is the quantile level (i.e., a probability) and $\Phi_\omega(y_p|\mathbf{x})$ is the distribution function associated with the conditional mixture in equation (5). Note that y_p is equal to zero for all $p < 1 - \alpha(x)$, the

Table 2. Average Differences in Log-Likelihood on the Test Set for the Three Rain Gauge Stations Between the Hybrid Pareto CMM and the Other Models[a]

	Gaussian	Log-normal	Benchmark
Orange	**0.02146 (0.003139)**	0.0022512 (0.001910)	**0.02275 (0.002866)**
Sète	**0.01595 (0.003034)**	−0.003530 (0.001647)	**0.01847 (0.002690)**
Le Massegros	**0.01948 (0.006671)**	−0.004606 (0.002121)	**0.02068 (0.003005)**

[a]Standard errors are given in parentheses. Positive numbers indicate that the hybrid Pareto CMM performed better. Significant differences are in bold.

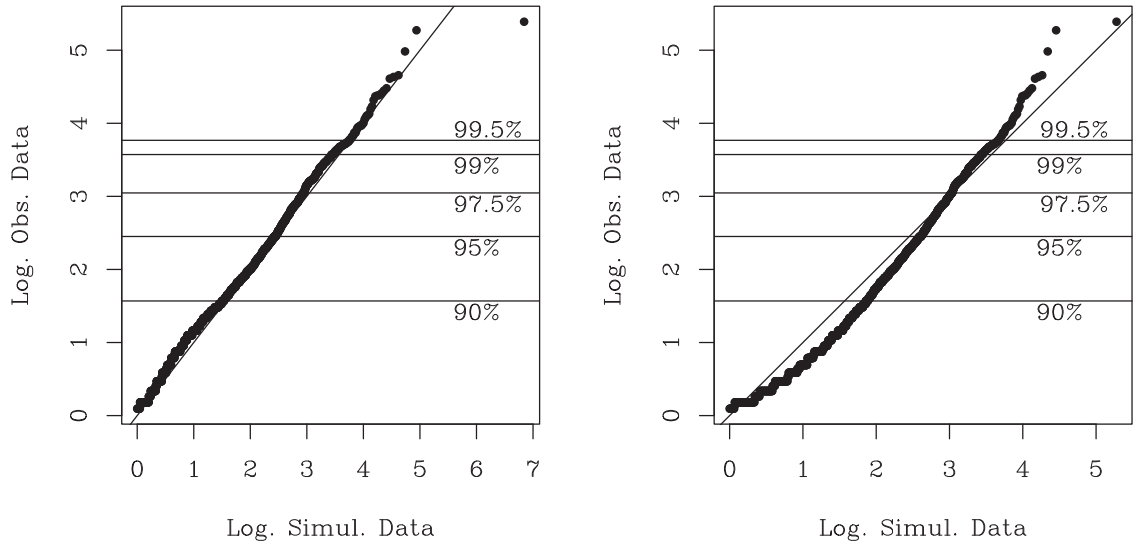

Figure 2. QQ-plots on a logarithmic scale of the simulated precipitation versus observations >1 mm on the Orange test set for (left) the hybrid Pareto conditional mixture and (right) the benchmark model. The horizontal lines are the empirical unconditional quantiles from observations of the test set.

no-rain probability, since in that case y_p falls in the discrete part of the distribution.

[36] Figure 4 shows the daily seasonal cycle of the rain probability and the daily seasonal cycle of the 99% conditional quantile for the Orange test data. The black line is the climatology computed from the observations in the test set: in Figure 4 (left), this is the proportion of rainy days, whereas in Figure 4 (right) it is the empirical 99% quantile. The gray band is the 90% confidence interval based on the empirical quantiles of 5% and 95% levels computed from the hybrid Pareto CMM, and the white line corresponds to the empirical median. In Figure 4 (left), the empirical quantiles are estimated from the 21 modeled conditional probabilities of rain $\alpha(x)$ per day, and in Figure 4 (right) they are estimated on the 21 modeled conditional quantiles of the 99% level, $y_{0.99}(\mathbf{x})$, per day. It is not meaningful to look at more central conditional quantiles because of the presence of the discrete component, which makes these conditional quantiles hard to interpret. From Figure 4 (left), we can identify two seasonal modes, around March (month 3) and October (month 10), which translates into higher probabilities of rain around these two months, while summer (i.e., around July) presents smaller probabilities of rain. This is in agreement with the observations over the test set, showing the same features. Regarding the 99% quantile in Figure 4 (right), the picture is a bit less clear, but we can nevertheless identify two modes in the median (white line) just after March and around fall (months 9 and 10). Larger rainfall amounts are thus expected in spring and fall. The climatologies are based on empirical quantiles computed on 21 values (because of the 21 years in the test set). Few positive observations occurred for a given day out of the 21, and empirical quantiles are thus not very stable (especially at the 99% level). This explains the spiky features in Figure 4. The seasonal cycles for the other three SDMs are very similar and are thus not shown.

[37] In order to get some understanding of the way conditional mixtures approximate the conditional distribution of precipitation intensity, we now look at the daily seasonal cycle of the mixture parameters (the continuous part of the mixture; see equation (1)) on the test set. The mixture parameters are deterministic functions of the predictor variables and therefore vary with the values taken by these. Just as before, for each of the 365/366 days of the year, empirical quantiles of the 5%, 50%, and 95% levels are estimated from the 21 values of the mixture parameters on the test set. Figure 5 depicts the climatology of the hybrid Pareto CMM mixture weight, tail index, location, and scale parameters over the Orange test set. For this station, the cross-validation procedure selected two components in the mixture. The white and the black lines represent the empirical median of the mixture parameter values for each component. The gray bands are the 90% confidence intervals based on the empirical quantiles of the 5% and 95% levels. The seasonal cycle of the mixture weights (or priors) in Figure 5 (top left) shows the predominance of one or the other component in the density. For the hybrid Pareto CMM, we see that in July (month 7), the component associated with the black line tends to dominate, although not significantly, as we see that the confidence intervals overlap. The seasonal cycle of the tail index parameters is shown in Figure 5 (top right). We observe a slight decrease of the tail index parameter from both components in the summer, but globally, they are around the value 0.2, which is in agreement with the penalty of the MLE (see equation (9)). For the Orange data, the penalty hyperparameters η and ρ selected via cross-validation imply that the tail index parameters have just one mode centered on 0.2. For the other two data sets, Sète and Le Massegros, the penalty hyperparameters selected entail that the tail index parameters have two modes centered on zero and on 0.2 with different heights. These results are not shown. Figure 5 (bottom) depicts the cycles of the location and scale parameters of the hybrid Pareto components. There is a strong seasonal signal with a clear bump of the white line component for both location and scale parameters around summertime,

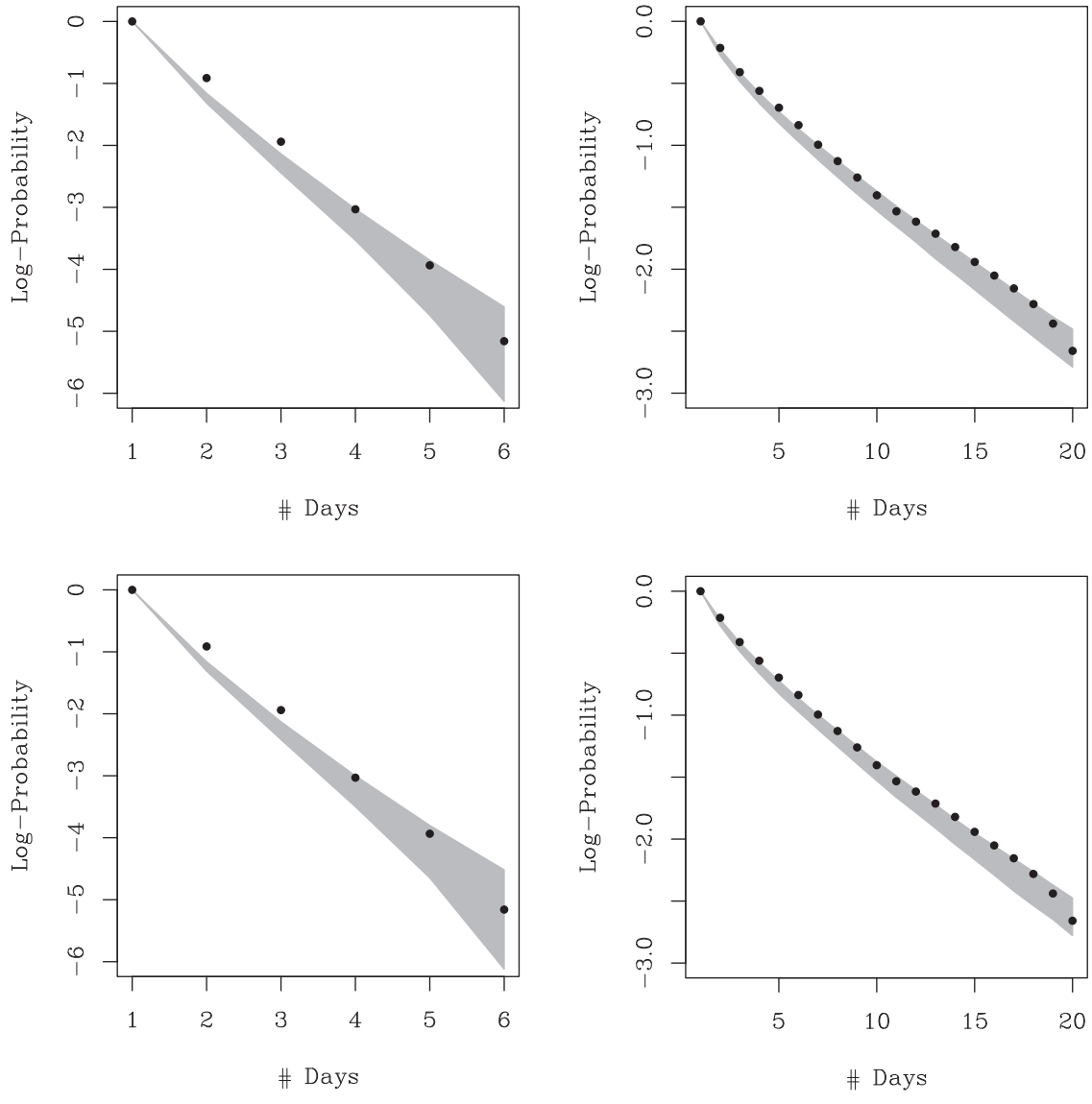

Figure 3. Logarithm of the proportion of (left) wet spells and (right) dry spells of at least k days for the Orange test data (dots) together with 90% empirical confidence interval (gray band) from (top) the hybrid Pareto CMM and (bottom) the benchmark model. The predictor variables allow implicit modeling of serial dependence of rainfall occurrence and provide similar spell probabilities in all four downscaling models.

thereby providing a larger probability of high-intensity rainfall. Keeping in mind that this continuous mixture density is conditional on the fact that it rains ($Y > 0$), it means that when it rains, the rainfall intensity is potentially high. However, we saw in Figure 4 (left) that the probability of rain reaches a low in the middle of summer. This reflects common knowledge about rainfall in this area of France: summer rain is rare but intense.

[38] As a means of comparing the log-normal versus the hybrid Pareto CMMs, we also include the climatologies of the log-normal conditional mixture parameters on the Orange test data; see Figure 6. In this case as well, two components were selected via cross-validation. From the climatology of the mixture weights in Figure 6 (top), we observe a pattern similar to the one of the hybrid Pareto CMM: one of the components dominates in the summer. In this case, the difference is more accentuated. The log-normal distribution has two parameters, μ and σ^2, which are directly linked to the Gaussian parameters. However, unlike the Gaussian, they do not represent the expectation and the variance of the log-normal. So instead of the μ and σ^2, we have computed the climatologies of the expectation and the variance of the log-normal components, which are given by $e^{\mu+\sigma^2/2}$ and $e^{2\mu+\sigma^2}\left(e^{\sigma^2}-1\right)$, respectively. These are plotted in Figure 6 (middle and bottom). In summer, we see the same bump as in Figure 5 (bottom), albeit less pronounced for the expectations (Figure 6, middle). However, the values taken by the variance (Figure 6, bottom) are much higher: the full 90% confidence interval takes values up to about 600, which makes for a standard deviation up to about 25. This means that the tail is not heavy enough to

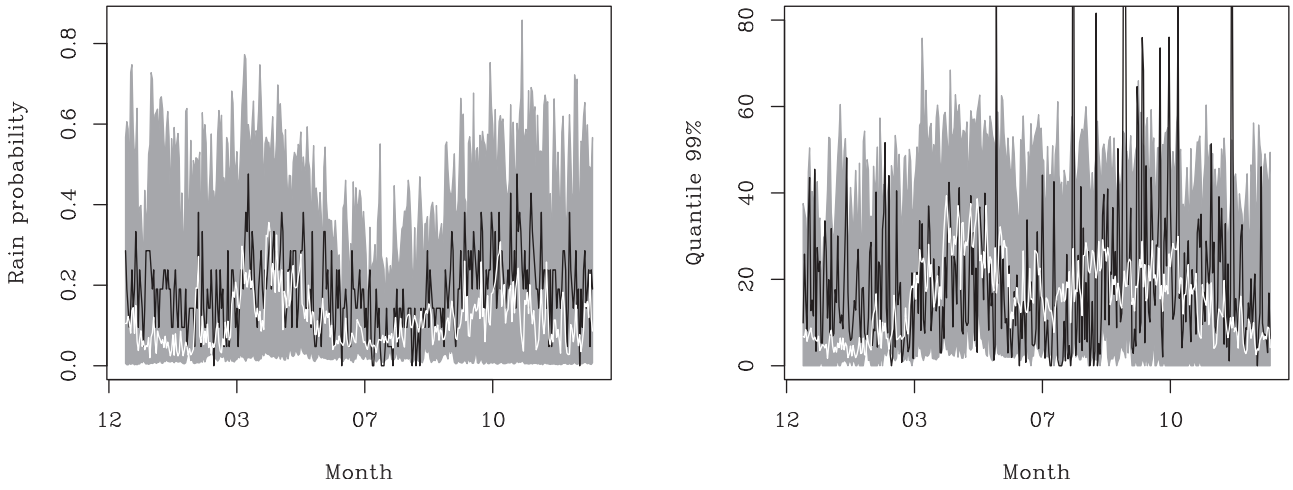

Figure 4. Daily seasonal cycles of (left) the rain occurrence probability and (right) the 99% quantile from the observations (black line) together with an empirical 90% confidence interval (gray band) and median (white line) from the hybrid Pareto conditional mixture for the Orange station test data. Peaks in the seasonal cycle of the occurrence process are visible in spring and fall.

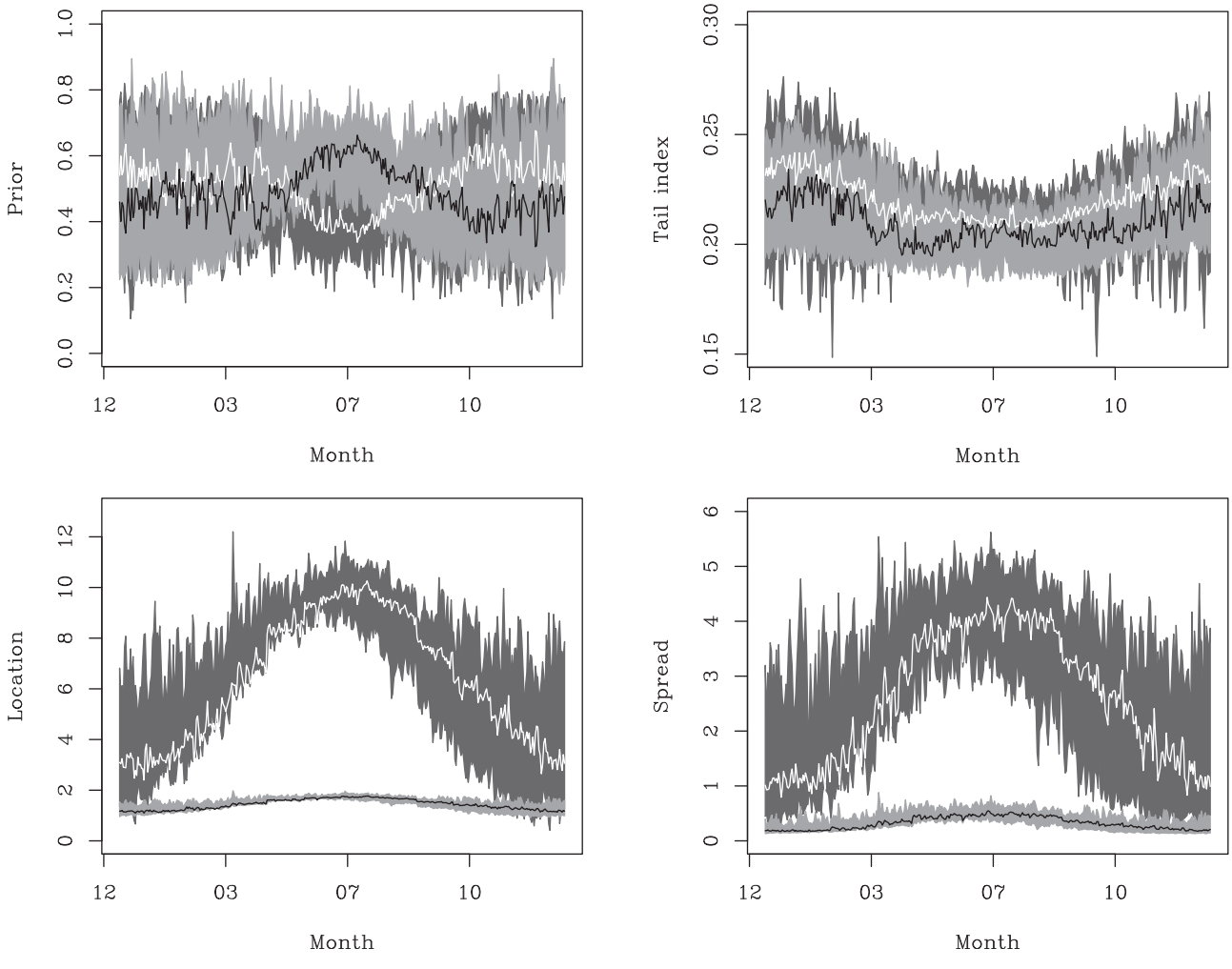

Figure 5. Daily seasonal cycles of the hybrid Pareto conditional mixture parameters (top left to bottom right: mixture weights π_j, tail index parameters ξ_j, location parameters μ_j, and scale parameters σ_j) together with an empirical 90% confidence interval. The mixture has two components whose parameters are represented by the black and white lines.

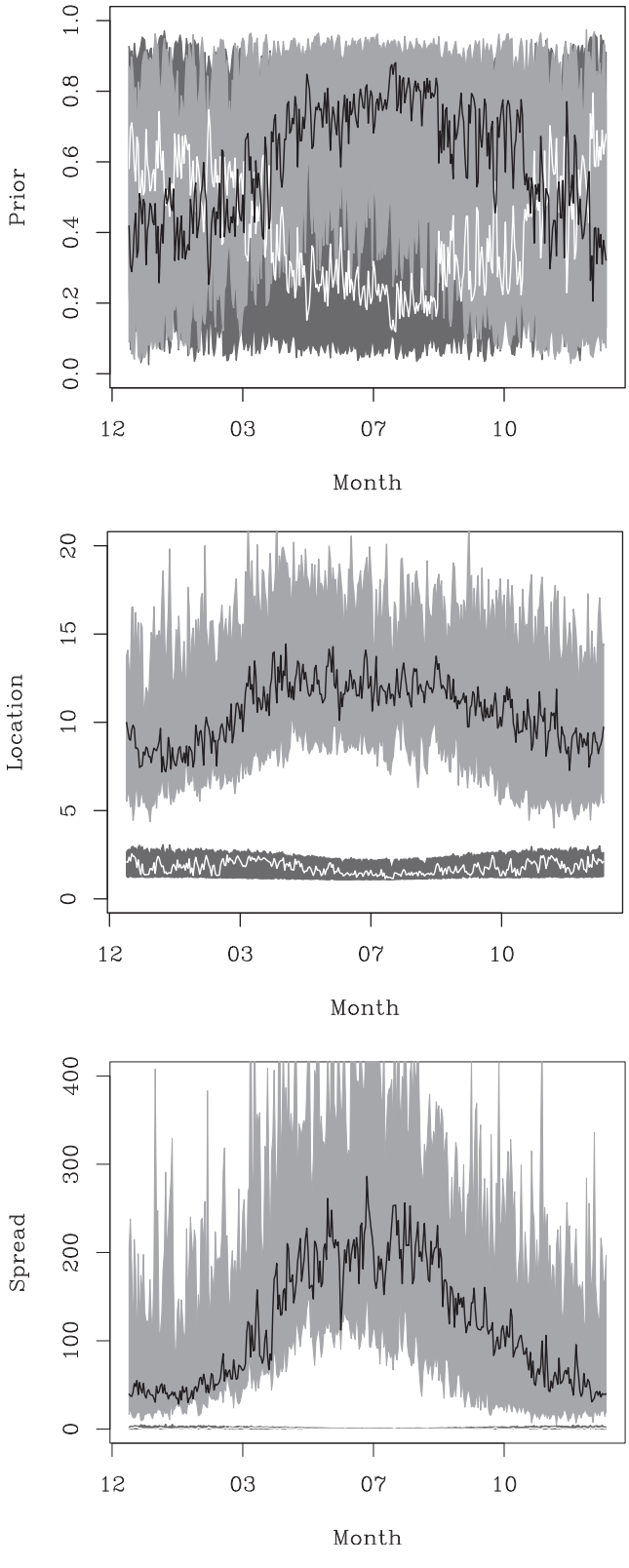

Figure 6. Daily seasonal cycles of the log-normal conditional (top) mixture weights π_j, (middle) component expectations $e^{\mu_j+\sigma_j^2/2}$, and (bottom) component variances $e^{2\mu_j+\sigma_j^2}\left(e^{\sigma_j^2}-1\right)$ together with an empirical 90% confidence interval. The mixture has two components whose parameters are represented by the black and white lines.

correctly represent rainfall. To compensate, the log-normal conditional mixture needs to inflate the variance of one of the components in order to take properly extreme rainfalls into account. We also note that the component with large location and scale in the summer has the largest prior. Therefore, the log-normal CMM is not very realistic. The hybrid Pareto CMM, because of the tail index parameters, does not need to increase the variance. However, the fine-tuning of the penalty term of the MLE (see equation (9)) requires some extra care and greatly influences the final model.

4.3. Analyses of Conditional Events

[39] This section concerns analyses of two specific rain events at the Orange station from the test set: the wet spell with the highest volume of rain (322 mm on 8–9 September 2002) and the longest wet spell (9 days, from 24 April 1993 to 2 May 1993). Our goal is to check whether the SDMs' continuous part can capture rainfall intensity during particularly severe events that might be challenging to reproduce. Results are shown for the hybrid Pareto CMM only. In Figure 7, modeled conditional quantiles of the 95%, 99%, and 99.9% levels, i.e., $y_{0.95}(\mathbf{x})$, $y_{0.99}(\mathbf{x})$, and $y_{0.999}(\mathbf{x})$, are shown in light, medium, and dark gray, respectively, for the hybrid Pareto CMM for these two rain events. The black vertical lines represent the observed precipitation on each day. We can see how the hybrid Pareto CMM adapts to the atmospheric conditions that yield a sequence of dry days or little rain, then a high or long volume of precipitation, and then back to a drier period. If the model is right, we expect the observed precipitation to be below the quantile level $p \times 100\%$ of the time. In other words, we expect the observed precipitation to exceed the modeled conditional 99.9% quantile on average 0.01%, which means, given the test set size, about one observation. The largest observation in the training set for the Orange station is 137 mm, whereas the largest observation in the test set is 220 mm. The latter appears at the date 09/08 in Figure 7 (left) and corresponds to a quantile level of 99.99% according to the hybrid Pareto CMM. On average, the tested SDMs are pretty accurate with regard to conditional quantile modeling; see Table 3.

[40] Another way to look at the evolution of the model through a rain event is by looking directly at the conditional densities, which are associated with different atmospheric conditions, that is, for different predictors. More precisely, we depicted the conditional densities of Y given the predictors x and given that $Y > 0$. For the conditional mixtures, it corresponds to the continuous part of the mixture in equation (1): $\alpha(\mathbf{x})\phi_0(y;\psi_\omega(\mathbf{x}))$. These conditional densities are illustrated in Figure 8 for the hybrid Pareto CMM on the wet spell with the highest volume of rain in the Orange test data. This is the same rain event as in Figure 7 for the conditional confidence intervals. Figure 8 (left) shows the central part of the conditional densities, while Figure 8 (right) represents the upper tails in logarithmic scale. Each curve corresponds to a different day, which is connected in the legend with the amount of rain observed on that day. The days in the legend are presented in chronological order (from top to bottom). We see from Figure 8 (left) that the density of the rain intensity is very low, almost flat, for days where no precipitation occurred

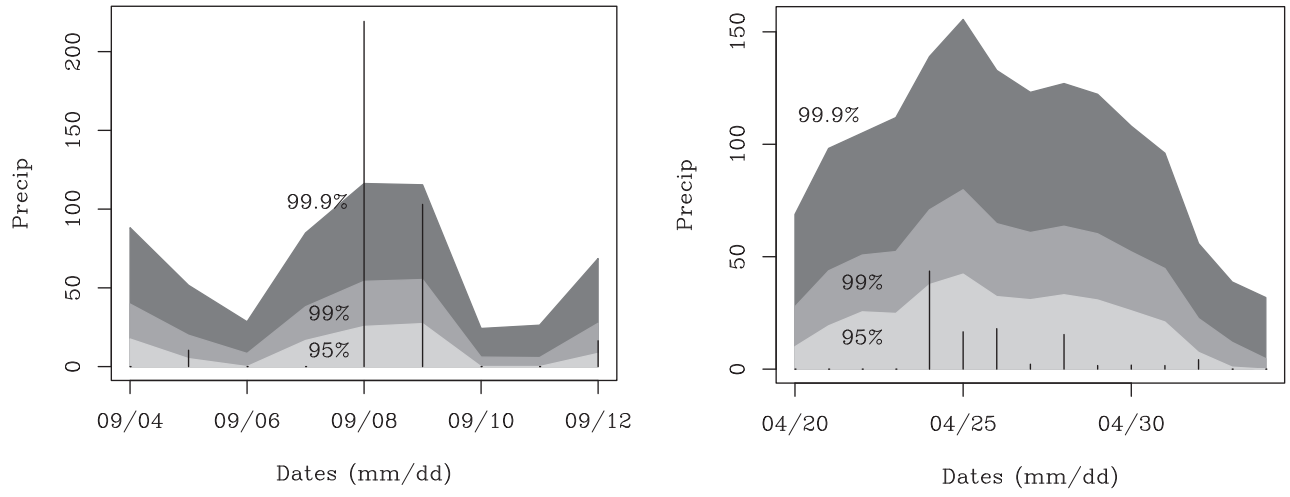

Figure 7. Hybrid Pareto conditional mixture: conditional quantiles $y_{0.95}(\mathbf{x})$ (light gray), $y_{0.99}(\mathbf{x})$ (medium gray), and $y_{0.999}(\mathbf{x})$ (dark gray) on (left) the wet spell with the highest volume of rain and (right) the longest wet spell from the Orange test data. Vertical lines represent the observed precipitation. The largest observed rainfall, 220 mm on 8 September (Figure 7, left), corresponds to a quantile level of 99.99% for the hybrid Pareto CMM.

(curves in warm hues). The corresponding tails in Figure 8 (right) decrease rapidly. The yellow curve, which is the last day with zero precipitation preceding the wet spell, is an exception; this density curve is similar to the density curves of the intense wet days (in green hues). For the days of heavy rains, the upper tail of the conditional density is heavier. We performed similar analyses on the other data sets and with the other three SDMs. There are differences (not shown) caused by the choice of the statistical model for rainfall intensity and reflecting the underlying assumptions. For instance, during an intense rain event, the upper tails of the Gaussian mixture density curves decrease much more rapidly than the upper tails of the hybrid Pareto CMM, indicating a lesser risk of extreme rainfalls for the Gaussian CMM. In general, from these analyses, we see that the conditional mixture model is very responsive to a change in atmospheric conditions. The mixture parameters change to adjust the shape of the distribution of precipitation in a consistent way (heavier upper tail is associated with extreme rainfalls).

5. Conclusions and Discussion

[41] The focus of this paper is on conditional mixture models (CMMs), which to our knowledge, are used for the first time in a downscaling context and open interesting ways to study the interactions between large- and small-scale climate variables. CMMs are flexible stochastic downscaling models: a discrete component in the mixture serves to simulate the occurrence process, whereas the continuous mixture part simulates the intensity process. The continuous mixture is made of either one of the following types of component: Gaussian, log-normal, or hybrid Pareto. The mixture parameters, i.e., mixture weights and both discrete and continuous components parameters, are functions of predictor variables (including large-scale information) and are computed by means of neural networks. CMMs extend the two-component mixture proposed initially by *Williams* [1998], which has a discrete component like CMMs to model the occurrence process but relies on a single density, the Gamma, for the rainfall intensity process. This simpler model, which was used in a downscaling context [*Haylock et al.*, 2006; *Cawley et al.*, 2007], acted as a benchmark model.

Table 3. Percentage of Observations Below Estimated Quantiles of Various Probability Levels ($p = 0.05$, 0.5, 0.9, 0.0975, 0.99, 0.999) for the Hybrid Pareto, Gaussian, and Log-normal CMMs and the Benchmark Model for the Three Rain Gauge Stations on the Test Set[a]

	Hybrid Pareto	Gaussian	Log-normal	Benchmark
		Orange		
$p = 0.05$	–	0	–	–
$p = 0.5$	44.1	50.9	45	46.3
$p = 0.9$	88.6	88.8	89.6	89.9
$p = 0.95$	94.2	94.7	94.4	94.9
$p = 0.975$	96.4	97.2	96.7	96.8
$p = 0.99$	98.5	98.9	98.5	98.5
$p = 0.999$	99.7	99.8	99.8	99.6
		Sète		
$p = 0.05$	–	–	–	–
$p = 0.5$	47.8	51.4	48.9	47.4
$p = 0.9$	89.6	89.7	90.6	91.7
$p = 0.95$	94.3	95.1	94.6	95.1
$p = 0.975$	97.4	97.7	97.5	97.5
$p = 0.99$	98.8	99.1	98.9	98.9
$p = 0.999$	99.7	100	99.9	99.7
		Le Massegros		
$p = 0.05$	8.3	16.7	15.6	9.7
$p = 0.5$	47.6	48.7	48.4	50.4
$p = 0.9$	89.1	89.2	89.0	89.8
$p = 0.95$	95.0	95.0	95.2	95.1
$p = 0.975$	97.4	97.5	97.6	97.2
$p = 0.99$	99.0	98.9	99.0	98.5
$p = 0.999$	99.9	99.9	99.9	99.8

[a]Missing values are because the estimated quantiles fall in the discrete part of the distribution. If the model is correct, the percentage of observations should be close to $p \times 100$.

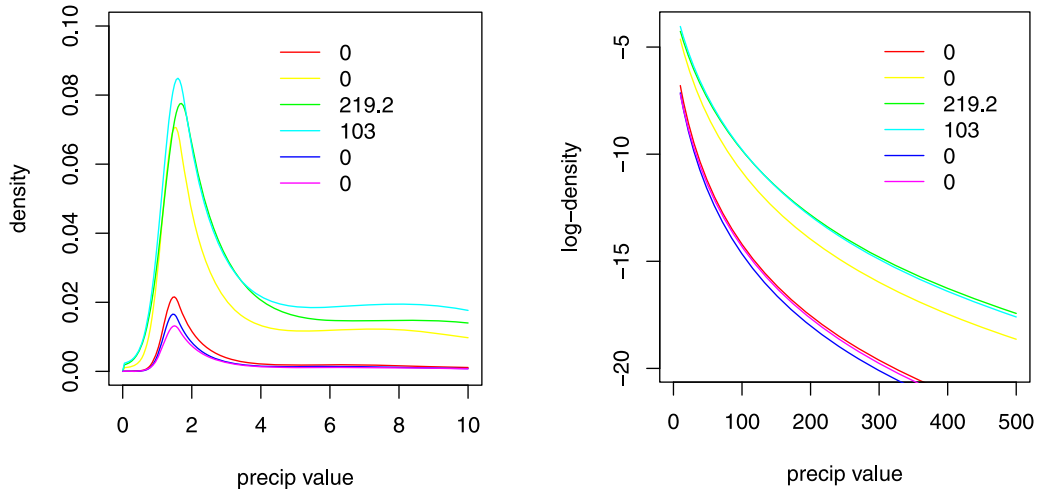

Figure 8. (left) Central part and (right) upper tail in logarithmic scale of the conditional densities $\alpha(\mathbf{x})\phi_0(y;\psi_\omega(\mathbf{x}))$ for the hybrid Pareto CMM day by day for a period comprising the wet spell with the highest volume of rain in the test Orange data (same period as in Figure 7, left). Each daily density is represented with a different color, which is represented in the legend in chronological order, from top to bottom, with the amount of rainfall observed on that day.

[42] We draw the following conclusions from our analyses of the three stations in the French Mediterranean area. First, conditional mixture models have a clear advantage over the benchmark model in terms of flexibility to represent both the central and the extremal part of rainfall intensity distribution. Second, modeling the occurrence process separately with a discrete component allows the implicit introduction of serial dependence through the predictor variables and hence the reproduction of wet and dry spell sequences. Finally, the choice of component in CMMs depends on the data. In our case, Gaussian components are not well suited. Log-normal CMMs offer a good performance and are more straightforward to implement than hybrid Pareto CMMs. However, the assumption of heavy tails of the hybrid Pareto CMM seems more realistic for the precipitation data considered in this work.

[43] The downscaling models considered in this work are all weather generators: they downscale the whole conditional distribution of precipitation from which we can answer all kind of questions. (1) Conditional quantiles can be computed, and from these, climatologies can be examined and confidence intervals can be constructed. (2) Conditional densities give insight into the influence of the atmospheric information on the distribution of precipitation. (3) It is easy to simulate rainfall and check whether the features of observed precipitation, such as wet and dry spells, are well captured by the models.

[44] Although conditional mixtures are rather complex models, some understanding of the modeling mechanisms can be gained by looking at the climatologies of the mixture parameters as functions of the covariates. We believe that the multiple benefits from the conditional mixtures compensate for the extra work of implementation. Note that a package in the R language [*R Development Core Team*, 2010] named CondMixt has been developed for this study and should be made available in the near future.

[45] This study has multiple perspectives and future works. For example, the choice of appropriate predictors should be made with care and requires further analyses. In this paper, our goal was to illustrate and compare the performances and advantages of the proposed downscaling models. We provided the downscaling models with a decent set of predictors without looking for the best set of predictors. The results from the hyperparameter selection for the mixtures show that single-component models are inadequate. This could indicate that more than one component is required, but it could also be due to inappropriate selection of predictors. Hence, although the comparisons between models are fair, a more complete set of predictors describing better the physical processes at play could yield more accurate statistical properties and simulations for all models.

[46] Another interesting perspective would be to evaluate climate change in precipitation according to these downscaling models. For this, we need to validate the use of GCM outputs as predictors in CMMs trained on reanalysis data. Indeed, this is essential to assess the reliability of the "couple" (GCM and CMM) and to give confidence in their present projections before applying CMMs to downscale distributions of rainfall under various potential future greenhouse gas emission scenarios. This would help us to evaluate the impact of climate change on very important features of rainfall, such as (interannual or statistical) variability, seasonality, or extremes.

[47] Finally, a challenging extension of this downscaling approach would be to take into account spatial dependencies between different rain gauges. This would allow joint modeling of precipitation at multiple sites with multivariate CMMs. Possible approaches to model the spatial dependence structure of precipitation include the use of copulas [*Nelsen*, 2006] and the approach suggested by *Cannon* [2008] to change the error function in order to encourage the model to match the observed covariance matrix. The resulting coherent spatial simulations should preserve observed dependencies and would provide tools to understand their temporal evolutions during a control time period,

past or future climate, completing the information brought by CMMs to study the many facets of climate changes.

[48] **Acknowledgments.** The authors thank the AssimileX and ACQWA project teams. M. Vrac has been partly funded by the GIS-REGYNA project. J. Carreau has been partly funded by the Fonds québecois de la recherche sur la nature et les technologies (FQRNT).

References

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

Bishop, C. M. (1994), Mixture density networks, technical report, Aston Univ., Birmingham, U. K.

Bishop, C. M. (1995), *Neural Networks for Pattern Recognition*, Clarendon, Oxford, U. K.

Busuioc, A., R. Tomozeiu, and C. Cacciamani (2008), Statistical downscaling model based on canonical correlation analysis for winter extreme precipitation events in the Emilia-Romania region, *Int. J. Climatol.*, *28*, 449–464.

Cannon, A. J. (2008), Probabilistic multisite precipitation downscaling by an expanded Bernouilli-Gamma density network, *J. Hydrometeorol.*, *9*, 1284–1300.

Cannon, A. J. (2011), Quantile regression neural networks: Implementation in R and application to precipitation downscaling, *Comput. Geosci.*, *37*(9), 1277–1284.

Cannon, A. J., and P. H. Whitfield (2002), Downscaling recent streamflow conditions in British Columbia, Canada using ensemble neural network models, *J. Hydrol.*, *259*, 136–151.

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

Carreau, J., and Y. Bengio (2009b), A hybrid Pareto mixture for conditional asymmetric fat-tailed distributions, *IEEE Trans. Neural Networks*, *20*(7), 1087–1101.

Carreau, J., P. Naveau, and E. Sauquet (2009), A statistical rainfall-runoff mixture model with heavy-tailed components, *Water Resour. Res.*, *45*, W10437, doi:10.1029/2009WR007880.

Cawley, G. C., G. J. Janacek, M. R. Haylock, and S. R. Dorling (2007), Predictive uncertainty in environmental modelling, *Neural Networks*, *20*, 537–549.

Cho, H.-K., K. P. Bowman, and G. R. North (2004), A comparison of Gamma and lognormal distributions for characterizing satellite rain rates from the tropical rainfall measuring mission, *J. Appl. Meteorol.*, *43*, 1586–1597.

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

Delrieu, G., et al. (2005), The catastrophic flash-flood event of 8–9 September 2002 in the Gard region, France: A first case study for the Cévennes Vivarais Mediterranean hydrometeorological observatory, *J. Hydrometeorol.*, *6*(1), 34–52.

Embrechts, P., C. Kluppelberg, and T. Mikosch (1997), *Modelling Extremal Events*, Springer, New York.

Friederichs, P. (2010), Statistical downscaling of extreme precipitation events using extreme value theory, *Extremes*, *13*, 109–132.

Friederichs, P., and A. Hense (2007), Statistical downscaling of extreme precipitation events using censored quantile regression, *Mon. Weather Rev.*, *135*, 2365–2378.

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

Gardes, L., and S. Girard (2010), Conditional extremes from heavy-tailed distributions: An application to the estimation of extreme rainfall return levels, *Extremes*, *13*, 177–204.

Ghosh, S., and P. P. Mujumdar (2008), Statistical downscaling of GCM simulations to streamflow using relevance vector machine, *Adv. Water Resour.*, *31*(1), 132–146.

Gladstone, R., et al. (2005), Mid-Holocene NAO: A PMIP2 model intercomparison, *Geophys. Res. Lett.*, *32*, L16707, doi:10.1029/2005GL023596.

González-Rouco, J. F., H. Heyen, E. Zorita, and F. Valero (2000), Agreement between observed rainfall trends and climate change simulations in the southwest of Europe, *J. Clim.*, *13*, 3057–3065.

Goubanova, K., V. Echevin, B. Dewitte, F. Codron, K. Takahashi, P. Terray, and M. Vrac (2010), Statistical downscaling of sea-surface wind over the Peru-Chile upwelling region: Diagnosing the impact of climate change from the IPSL-CM4 model, *Clim. Dyn.*, *36*(7–8), 1365–1378.

Haylock, M. R., G. C. Cawley, C. Harpham, R. L. Wilby, and C. M. Goodess (2006), Downscaling heavy precipitation over the United Kingdom: A comparison of dynamical and statistical methods and their future scenarios, *Int. J. Climatol.*, *26*, 1397–1415.

Hewitson, B. (1994), Regional climates in the GISS general circulation model: Surface air temperature, *J. Clim.*, *7*(2), 283–303.

Hewitson, B. C., and R. G. Crane (1996), Climate downscaling: Techniques and application, *Clim. Res.*, *7*, 85–95.

Hornik, K. M. (1991), Approximation capabilities of multilayer feedforward networks, *Neural Networks*, *4*, 251–257.

Hosking, J. R. M., J. R. Wallis, and E. F. Wood (1985), Estimation of the generalized extreme-value distribution by the method of probability-weighted moments, *Technometrics*, *27*, 251–261.

Huth, R. (2001), Disaggregating climatic trends by classification of circulation patterns, *J. Climatol.*, *21*, 135–153.

Huth, R. (2002), Statistical downscaling of daily temperature in central Europe, *J. Clim.*, *15*, 1731–1742.

Huth, R., S. Kliegrová, and L. Metelka (2008), Non-linearity in statistical downscaling: Does it bring an improvement for daily temperature in Europe?, *J. Climatol.*, *28*, 465–477.

Intergovernmental Panel on Climate Change (2007a), *Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by S. Solomon et al., Cambridge Univ. Press, Cambridge, U. K.

Intergovernmental Panel on Climate Change (2007b), *Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by M. L. Parry et al., Cambridge Univ. Press, Cambridge, U. K.

Jolliffe, I. T. (1986), *Principal Component Analysis*, Springer, New York.

Kalnay, E., et al. (1996), The NCEP/NCAR 40-year reanalysis project, *Bull. Am. Meteorol. Soc.*, *77*(3), 370–471.

Klein Tank, A. M. G., et al. (2002), Daily dataset of 20th-century surface air temperature and precipitation series for the European climate assessment, *Int. J. Climatol.*, *22*, 1441–1453.

Maraun, D., et al. (2010), Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user, *Rev. Geophys.*, *48*, RG3003, doi:10.1029/2009RG000314.

McLachlan, G., and D. Peel (2000), *Finite Mixture Model*, John Wiley, New York.

McNeil, A. J. (1997), Estimating the tails of loss severity distributions using extreme value theory, *Astin Bull.*, *27*, 117–137.

Nelsen, R. B. (2006), *An Introduction to Copulas*, Springer, New York.

Pickands, J. (1975), Statistical inference using extreme order statistics, *Ann. Stat.*, *3*, 119–131.

Priebe, C. E. (1994), Adaptive mixtures, *J. Am. Stat. Assoc.*, *89*, 796–806.

R Development Core Team (2010), *R: A Language and Environment for Statistical Computing*, R Found. for Stat. Comput., Vienna.

Rumelhart, D. E., G. E. Hinton, and R. J. Williams (1986), Learning internal representations by error propagation, in *Parallel Distributed Processing: Explorations in the Macrostructure of Cognition*, vol. 1, pp. 318–362, MIT Press, Cambridge, Mass.

Sailor, D. J., and X. Li (1999), A semi-empirical downscaling approach for predicting regional temperature impacts associated with climatic change, *J. Clim.*, *12*, 103–114.

Salameh, T., P. Drobinski, M. Vrac, and P. Naveau (2009), Statistical downscaling of near surface wind field over complex terrain in southern France, *Meteorol. Atmos. Phys.*, *103*, 253–265.

Schnur, R., and D. Lettenmaier (1998), A case study of statistical downscaling in Australia using weather classification by recursive partitioning, *J. Hydrol.*, *212–213*, 362–379.

Schwarz, G. (1978), Estimating the dimension of a model, *Ann. Stat.*, *6*, 461–464.

Semenov, M. A. (2007), Development of high-resolution UKCIP02-based climate change scenarios in the UK, *Agric. For. Meteorol.*, *144*, 127–138.

Semenov, M. A., and E. M. Barrow (1997), Use of a stochastic weather generator in the development of climate change scenarios, *Clim. Res.*, *35*, 397–414.

Semenov, M. A., R. J. Brooks, E. M. Barrow, and C. W. Richardson (1998), Comparison of the WGEN and the LARS-WG stochastic weather generators in diverse climates, *Clim. Res.*, *10*, 95–107.

Snell, S. E., S. Gopal, and R. K. Kaufmann (2000), Spatial interpolation of surface air temperatures using artificial neural networks: Evaluating their use for downscaling GCMs, *J. Clim.*, *13*, 886–895.

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.

Vrac, M., M. Stein, and K. Hayhoe (2007a), Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing, *Clim. Res.*, *34*, 169–184.

Vrac, M., P. Marbaix, D. Paillard, and P. Naveau (2007b), Non-linear statistical downscaling of present and LGM precipitation and temperatures over Europe, *Clim. Past*, *3*, 669–682.

Vrac, M., K. Hayhoe, and M. Stein (2007c), Identification and inter-model comparison of seasonal circulation patterns over North America, *J. Climatol.*, *27*, 603–620.

Wigley, T. M. L., P. D. Jones, K. R. Briffa, and G. Smith (1990), Obtaining sub-grid scale information from coarse resolution general circulation model output, *J. Geophys. Res.*, *95*, 1943–1953, doi:10.1029/JD095iD02p01943.

Wilby, R. L., C. W. Dawson, and E. M Barrow (2002), SDSM—A decision support tool for the assessment of regional climate change impacts, *Environ. Modell. Software*, *17*(2), 145–157.

Wilks, D. S. (1999), Multisite downscaling of daily precipitation with a stochastic weather generator, *Clim. Res.*, *11*, 125–136.

Wilks, D. S., and R. L. Wilby (1999), The weather generation game: A review of stochastic weather models, *Prog. Phys. Geogr.*, *23*(3), 329–357.

Williams, M. P. (1998), Modelling seasonality and trends in daily rainfall data, *Adv. Neural Inf. Process. Syst.*, *10*, 985–991.

Yang, C., R. E. Chandler, and V. S. Isham (2005), Spatial-temporal rainfall simulation using generalized linear models, *Water Resour. Res.*, *41*, W11415, doi:10.1029/2004WR003739.

J. Carreau, HydroSciences Montpellier, UMR 5569, CNRS/IRD/UM1/UM2, Université de Montpellier 2, Case MSE, place Eugène Bataillon, F-34095 Montpellier CEDEX 5, France. (julie.carreau@univ-montp2.fr)

M. Vrac, Laboratoire des Sciences du Climat et de l'Environnement, IPSL, CNRS/CEA/UVSQ, Orme des Merisiers, F-91191 Gif-sur-Yvette, France. (mathieu.vrac@lsce.ipsl.fr)