



HAL
open science

A spatial hybrid approach for downscaling of extreme precipitation fields

Aurélien Bechler, Mathieu Vrac, Liliane Bel

► **To cite this version:**

Aurélien Bechler, Mathieu Vrac, Liliane Bel. A spatial hybrid approach for downscaling of extreme precipitation fields. *Journal of Geophysical Research: Atmospheres*, 2015, 120 (10), pp.4534-4550. 10.1002/2014JD022558 . hal-01184830

HAL Id: hal-01184830

<https://hal.science/hal-01184830>

Submitted on 6 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

RESEARCH ARTICLE

10.1002/2014JD022558

Key Points:

- A hybrid approach combines statistical downscaling and conditional simulations
- Spatial modeling and downscaling of extremes
- An application on real precipitation data shows encouraging results

Correspondence to:

A. Bechler,
abechler@hotmail.com

Citation:

Bechler, A., M. Vrac, and L. Bel (2015), A spatial hybrid approach for downscaling of extreme precipitation fields, *J. Geophys. Res. Atmos.*, 120, 4534–4550, doi:10.1002/2014JD022558.

Received 10 SEP 2014

Accepted 25 MAR 2015

Accepted article online 10 APR 2015

Published online 21 MAY 2015

A spatial hybrid approach for downscaling of extreme precipitation fields

Aurélien Bechler^{1,2,3}, Mathieu Vrac³, and Liliane Bel^{1,2}

¹INRA UMR 518, Paris, France, ²AgroParisTech UMR 518, Paris, France, ³LSCE-IPSL, Centre d'Études de Saclay, Gif-sur-Yvette, France

Abstract For a few decades, climate models are used to provide future scenarios of precipitation with increasingly higher spatial resolution. However, this resolution is not yet sufficient to describe efficiently what happens at local scale. Dynamical and statistical methods of downscaling have been developed and allow us to make the link between two levels of resolution and enable us to get values at a local scale based on large-scale information from global or regional climate models. Nevertheless, both the extreme behavior and the spatial structures are not well described by these downscaling methods. We propose a two-step methodology, called spatial hybrid downscaling (SHD), to solve this problem. The first step consists in applying a univariate (i.e., one-dimensional) statistical downscaling to link the high- and low-resolution variables at some given locations. Once this 1d-link is performed, a conditional simulation algorithm of max-stable processes adapted to the extremal t process enables us to get conditional distributions of extreme precipitation at any point of the region. An application is performed on precipitation data in the south of France where extreme (Cevenol) events have major impacts (e.g., floods). Different versions of the SHD approach are tested. Most of them show particularly good results regarding univariate and multivariate criteria and overcome classical downscaling techniques tested in comparison. Furthermore, these conclusions are robust to the choice of the 1d-link functions tested and to the choice of the conditioning points to drive the conditional local-scale simulations performed by the SHD approach.

1. Introduction

Future climate projections from global climate models (GCMs) offer a way to investigate future trends of climatic events. One main issue comes from the spatial resolution of these model outputs as the GCMs give values on a grid of about 200 km of resolution [Intergovernmental Panel on Climate Change (IPCC), 2012]. Hence, when the process of interest is very localized, as, for example, for extreme precipitation, the GCM outputs are not relevant. Furthermore, most of the GCMs are built to characterize the mean behavior of some variables and thus are not able to produce realistic rare and very intense (i.e., extreme) local events. For these reasons, the GCM outputs cannot be used directly as inputs in impact models (e.g., agriculture, water resources, and pollution) that need local-scale data. A commonly used solution consists in simulating more local values by performing downscaling techniques.

There are two main downscaling families. The first one is a dynamical approach based on regional climate models (RCMs) [e.g., Laprise, 2008; Rummukainen, 2010] and may be seen as a regional version of the global models, aiming to solve the physical equations of the atmosphere dynamics, but at a higher spatial resolution than the GCM (from 50 km to 5 km). Thus, these RCMs are able to solve explicitly a part of the physical processes related to meteorological and hydrological information involved at a more local scale. Nevertheless, even with the increasing computational means, the complexification due to the improvement of these regional equations leads to the possibility of using RCMs only over limited regions and time periods. Moreover, these RCMs may still be at a resolution too rough to be used in a local framework by impact models and need to be themselves downscaled.

More recently, another approach called statistical downscaling has emerged as an efficient alternative. This approach does not explicitly solve the equations of the dynamics of the atmosphere but is based on the statistical modeling of existing relations between large-scale variables (coming from GCMs or RCMs) and variables observed at a more local scale. Hence, due to its statistical formulation, this approach enables fast simulations of local variables and the low computational resources requested makes possible the modeling of the uncertainty associated to the local projections. In practice, statistical downscaling gathers a high

diversity of very different methods based on transfer functions [e.g., *Friederichs and Hense, 2007; Vrac et al., 2007; Levavasseur et al., 2011*], weather generators [e.g., *Richardson et al., 1998; Wilks, 2012*], clustering-based techniques [*Huth, 2001; Boé et al., 2006*], and bias correction approaches [e.g., *Déqué, 2007; Piani et al., 2010; Vrac et al., 2012*].

Some of these statistical downscaling techniques are spatialized; i.e., the generated values are spatially correlated and respect the spatial structure of the observations. This spatial characteristic is of huge importance to generate fields of spatially coherent values. This spatial coherence can be achieved through different ways, as, for example, clustering or multisite downscaling [see, e.g., *Wilks, 2012*, for a recent review], and is of huge importance for various impact models [e.g., *Fowler and Ekström, 2009*]. This is also true for extreme events for which downscaling requires to adapt the existing tools [e.g., *Fowler et al., 2010; Kallache et al., 2011; Vrac and Naveau, 2007; Carreau and Vrac, 2011*].

More generally, the understanding and projection of extreme events are major challenges. Indeed, as they are inherently rare and often unexpected, the society is not ready to face them. One of the particularities of the extreme events is that their probability decreases rapidly when the intensity increases and so does the damages and the cost of protection against such events. Both analysis of the last decade observations and future projections from climate models tend to suggest that the occurrence and the intensity of extreme events may increase in the following years around the world [*IPCC, 2013*] and in Europe [*Vautard et al., 2014*]. In that case, an increasing number of catastrophes such as storms, flooding, or crop failures may be expected. In addition to their direct effects in terms of human lives or building destruction, extreme events will also have indirect impacts such as increasing costs for building specific infrastructures or higher insurance premiums. Additionally, these extreme events may also have consequences on global population displacements [*Kirtman et al., 2013; IPCC, 2013, section 11.3.2.5 (pp. 990–992) of Chapter 11*]. Then, a better understanding and modeling of extreme events is a crucial need for both economical and societal points of view. From a mathematical point of view, the statistical theory of extreme values appears to be the suitable probabilistic framework for modeling their frequency and intensity. The extreme value theory (EVT) is well established for univariate stationary series [*Beirlant et al., 2004*], and there is now a growing effort to develop EVT models able to reproduce spatial variability and consistency and to deal with nonstationary problems [e.g., *Panagoulia et al., 2014; Jonathan et al., 2014*]. Multivariate extreme distributions appear in climate and environmental applications in order to take into account the dependence between extremes [e.g., *Coles and Tawn, 1994; De Haan and De Ronde, 1998; Schlather and Tawn, 2003; Fawcett and Walshaw, 2012*].

In the present study, however, such meteorological extreme events are not considered as we will focus on seasonal maximum daily values, also called “block-maxima values” in the literature [e.g., *Beirlant et al., 2004*], hereafter referred to as seasonal extremes. For the study of such extremes in a spatial context (i.e., not only multivariate but over a continuous field), max-stable processes are suitable tools for modeling and inference. Initially introduced by *De Haan [1984]*, different parametric models have later been proposed by *R. L. Smith (Max-stable processes and spatial extreme, unpublished manuscript, 1990)*, *Schlather [2002]*, *Kabluchko et al. [2009]*, or *Padoan [2011]*. Max-stable processes have been used in many climate studies such as for extreme rainfall modeling [e.g., *Buishand et al., 2008; Davison et al., 2012; Bechler et al., 2014*], temperature [e.g., *Davison and Gholamrezaee, 2012; Fuentes et al., 2013*], snowfall [e.g., *Gaume et al., 2013*], or snow depth [e.g., *Blanchet and Davison, 2011*].

In this work, we propose a new spatial downscaling approach by explicitly modeling the spatial structure of the data, focusing on seasonal extreme values. This approach is called “spatial hybrid downscaling” (SHD) since it takes advantage of both the physical consistency of the climate model outputs and the statistical spatial properties of a conditional simulation procedure to provide distribution of local-scale variables at any location in a given region. Indeed, conditional simulations allow us to statistically generate values at locations without any recording, in agreement with the surrounding data. Furthermore, while this SHD approach is developed in a general way, it is adapted in the present application to the downscaling of seasonal extreme values by performing conditional simulations of max-stable processes. The application and tests of this SHD model focus on the French Mediterranean basin, which is a key spot in terms of occurrences of extreme events such as heavy precipitation [e.g., *Ducrocq et al., 2008; Lebeaupin et al., 2006; Beaulant et al., 2011; Vrac and Yiou, 2010; Vrac et al., 2012*]. Indeed, major extreme rainfalls are regularly observed in this region close to the Cévennes mountains. Modeling and providing projections, for example, of maps of return levels, as well as

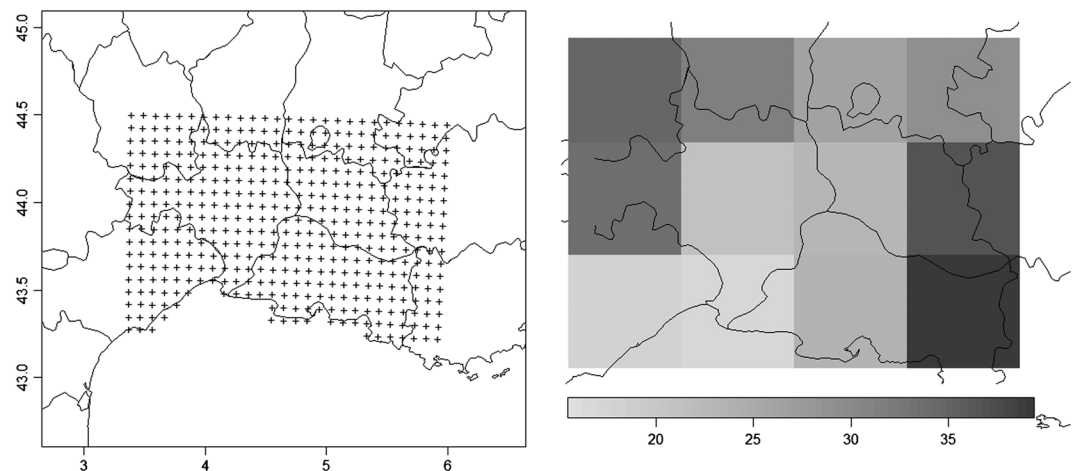


Figure 1. (left) Area of study with a SAFRAN data subset. (right) WRF-MED maximum daily precipitation data for the fall 1989 (in mm).

their evolution in time, for such extreme precipitation called “Cevenol events,” are major challenges in this area as well as important applications of the suggested SHD method since those very intense precipitation events often lead to severe floods in this region [e.g., *Ducrocq et al., 2008*].

The structure of this article is the following. After the presentation of the two data sets (local observations and larger-scale RCM outputs) in section 2, the general methodology of the SHD approach, its adaptation to seasonal extreme values, and all the theoretical considerations are described in section 3. In section 4, different versions of the SHD method are presented, and we propose some scores to assess the quality of the generated values. Then, section 5 displays a comparison of the results from the SHD approach with those from other methods. Different versions of the SHD approach are also discussed in this section. Some conclusions and discussions are finally given in section 6.

2. Large- and Local-Scale Data

As for any statistical downscaling technique, the calibration of our SHD approach requires a reference local-scale data set of the variable of interest, as well as a larger-scale data set to be downscaled.

2.1. Local-Scale Data: The SAFRAN Precipitation

In this paper, the maximum values of autumnal daily precipitation in the southeast of France are of interest. The choice of this season is driven by the fact that the Cevenol events occur in this season when humidity fluxes arise from a warm Mediterranean sea. Furthermore, this choice allows us to assume stationarity, as it is commonly accepted that rainfall data can be considered as stationary within a given season.

Daily precipitation data have been extracted from the SAFRAN data set [*Quintana-Seguí et al., 2008; Vidal et al., 2010*] in the southeast of France covering the region $[3.4^{\circ}\text{E}, 6.0^{\circ}\text{E}] \times [43.2^{\circ}\text{N}, 44.5^{\circ}\text{N}]$ —as represented by the square in Figure 1—with a spatial resolution of $8 \text{ km} \times 8 \text{ km}$, which corresponds to 457 grid cells. SAFRAN is a data set built from an optimal interpolation between ground stations and vertical profiles from a meteorological model to provide a dense gridded network. The basic assumption is that there exists climatically homogeneous zones [*Quintana-Seguí et al., 2008; Vidal et al., 2010*]. Nevertheless, thanks to the quality of this data set, we consider it in the following as observations. Then, for each cell, we computed the autumnal maximum daily precipitation (from 91 days between 1 September and 30 November) for 1960–2007: hence, for each grid cell, we have 48 autumnal maximum daily precipitation values.

2.2. Large-Scale Data: RCM Outputs

The large-scale data to be downscaled in this study are the outputs of the “Weather Research and Forecasting” (WRF) RCM version 3.1.1. The WRF model is a limited area model developed by the National Center for Atmospheric Research (NCAR) [*Skamarock et al., 2008*]. More precisely, the fall (September–October–November) daily precipitation outputs of the model IPSL-WRF311 for 1989–2007 have been downloaded. This corresponds to WRF applied at a 0.44° resolution over the Mediterranean region of the Coordinated Regional

Climate Downscaling Experiment (CORDEX) (<http://wcrp-cordex.ipsl.jussieu.fr/>) exercise to dynamically downscale the ERA-Interim reanalyses [Dee *et al.*, 2011] within the hindcast evaluation context of the Med-CORDEX (<http://www.medcordex.eu/>) initiative. The IPSL-WRF311 model—hereafter referred to as WRF-MED—is one of the few nudged RCMs involved in Med-CORDEX. The nudging technique provides one way to constrain the RCM simulations and keep them from diverging too far from the coarse-scale driving fields [e.g., von Storch *et al.*, 2000]. Twelve WRF-MED grid cells are needed to cover our region of interest (see Figure 1). On each of these grid cells and for each year, we computed the autumnal maximum of the daily precipitation. Figure 1 shows the autumnal maximum precipitation values extracted from the WRF-MED data for the year 1989.

Hereafter, the RCM outputs are considered as large-scale data while SAFRAN data are referred to as local-scale data. For both data sets, the years before 2000 serve as the learning period and the years 2000–2007 are used for the evaluation. This choice appears judicious as it keeps a large learning period and as the test period contains very heterogeneous years, which is important in robustness evaluation purposes.

3. Spatial Hybrid Downscaling Approach

3.1. General Methodology

This method is meant to use both climate model outputs and statistical advantages of the conditional simulations to provide the spatial distribution of local-scale variables at any location of a given region, by using both large-scale and local-scale information. Performing conditional simulation needs conditioning points at which data are available (e.g., observed values). When only large-scale climate outputs are available instead of observed values, the idea is to simulate data at some local-scale conditioning points based on a univariate downscaling of the climate model outputs. Accordingly, the methodology of the proposed SHD method can be divided into two steps:

Calibration. (1) Establish a univariate statistical link (called “1d-link function”), calibrated on the calibration period, between the large-scale information and some selected local-scale conditioning points; (2) choose the spatial model; model the marginal (i.e., univariate) distributions with covariates including latitudes and longitudes; and (3) estimate the parameters of the spatial model based on the selected conditioning points.

Simulation. (1) Based on the 1d-link functions, generate downscaled projections (DS-PRJ) at the selected conditioning locations from the large-scale outputs on the evaluation period and (2) perform a conditional simulation algorithm of the variable of interest according to the spatial model with the DS-PRJ as conditioning values.

This approach enables us to provide simulated data for any location within the region. By repeating several times those conditional simulations, we can also provide an empirical distribution of the variable of interest at any location, even in the future when no observation is available but only climate models outputs.

Many aspects of this method have to be detailed. First, as in this work we are interested in seasonal extremes; in Calibration Step 2, a max-stable model has been chosen, more precisely an extremal t process [Padoan, 2011]. It has to be noted that this step adds some extra information in comparison to classical downscaling techniques. Indeed, the parameters of the simulated max-stable process are estimated from observations (see section 3.3). This brings information about the spatial characteristics of the phenomenon. The gain compared to more classical statistical downscaling methods is that this approach takes explicitly into account both the spatial and the extreme components in the downscaling procedure. The question of the choice of the 1d-link function is treated in section 3.2. The descriptions of the max-stable extremal t process and of the margins modeling (including latitudes and longitudes as covariates) are then provided in section 3.3. The algorithm used to perform the conditional simulations is presented in Appendix B. The selection of the number and of the locations of the conditioning points is discussed in section 5.2.

3.2. 1d-Link Functions

Two 1d-link models commonly used in downscaling contexts have been tested here in Calibration Step 1 and Simulation Step 1. The first method, called “Cumulative Distribution Function - transform” (CDF-t) and initially proposed by *Michelangeli et al.* [2009], consists in comparing the CDF of the local-scale variable Z_{LS} and that of the large-scale variable Z_{HS} . Let us note $F_{LS,c}$ the CDF of the local-scale variable and $F_{HS,c}$ the CDF of the large-scale variable on the calibration period. The CDF-t approach looks for a transformation T from $[0, 1]$ to

[0, 1] such that:

$$T(F_{HS,c}(x)) = F_{LS,c}(x) \quad x \in \mathbb{R}. \quad (1)$$

The underlying assumption is that equation (1) remains valid when $F_{LS,c}$ and $F_{HS,c}$ are replaced by, respectively, $F_{LS,f}$ and $F_{HS,f}$, the CDFs of the local-scale and the large-scale variables in the future. A solution consists in choosing T as

$$T(u) = F_{LS,c}(F_{HS,c}^{-1}(u)) \quad u \in [0, 1], \quad (2)$$

and then we obtain the relationship

$$F_{LS,f}(x) = F_{LS,c}(F_{HS,c}^{-1}(F_{HS,f}(x))) \quad (3)$$

in replacing u by $F_{HS,f}(x)$ in equation (2). Every CDF is empirically modeled from the historical observations or from the large-scale data of the calibration or evaluation time periods. Then, the transformation (equation (3)) is performed for each margin and gives us one DS-PRJ for the selected locations by applying a quantile mapping-based technique between the large- and local-scale distributions of the evaluation period. More theoretical and technical details can be found in *Vrac et al.* [2012].

The other tested 1d-link function consists in a linear regression at each location where the large-scale precipitation is used as an explanatory variable of the local scale. In our application, after different tests (not shown here), the best model consists in explaining the local-scale precipitation by $\overline{Z_{HS}^{9\text{cells}}}$, the mean of the large-scale precipitation over the nine RCM grid cells around the SAFRAN location of interest as shown by equation (4):

$$Z_{LS} = a \times \overline{Z_{HS}^{9\text{cells}}}. \quad (4)$$

Generally, more than the initial 12 WRF-MED grid cells are needed to perform the regression functions. For instance, if there is one conditioning point per grid cell, up to 30 grid cells may be necessary: the 12 initial cells plus 18 others around those. In order to get more observations than the number of maxima for the calibration of the 1d-link functions, all 1d-links are calibrated on the daily observations, not the maxima. As each 1d-link function is univariate, they are different from one conditioning location to another (i.e., for example, the parameter a in equation (4) is estimated separately for each conditioning location). The spatial model is calibrated on 40 fall seasons (1960–1999), corresponding to $40 \times 91 = 3640$ daily values, for each conditioning point. Then, once the daily 1d-link projections are performed over the evaluation period (2000–2007) corresponding to $8 \times 91 = 728$ daily values, the seasonal maximum daily values are retained for each conditioning point to drive the algorithm of conditional simulation.

3.3. Max-Stable Processes and Conditional Simulations

The spatial model chosen for Calibration Step 2 and Simulation Step 2 is described here. In order to model fields of extremes, the max-stable processes [De Haan, 1984] are common tools. The extremal t process [Padoan, 2011] is a particular max-stable process. *Opitz* [2013] proposed a spectral construction of this process which leads to the general expression

$$\mathbf{Z} = \{Z(x)\} = \left\{ m_\alpha^{-1/\alpha} \max_{i \geq 1} \xi_i W_i(x) \right\}, \alpha > 0, x \in S \subset \mathbb{R}^2, \quad (5)$$

with S a Borel set, $m_\alpha = \sqrt{\pi}^{-1} 2^{(\alpha-2)/2} \Gamma((\alpha+1)/2)$, where Γ is the Gamma function, ξ_i a Poisson point process on $(0, +\infty)$ with intensity measure $d\Lambda(t) = \alpha t^{-(\alpha+1)} dt$ and W_i independent copies of a stationary standard Gaussian process on \mathbb{R}^2 with correlation function ρ . The margins of \mathbf{Z} are α -Fréchet distributed. It can be remarked that for $\alpha = 1$, the representation in equation (5) leads to a Schlather process. Hence, the extremal t process can be considered as a generalization of the Schlather process, with a greater flexibility. The correlation function $\rho(\cdot)$ determines the spatial dependence structure, and the parameter α characterizes the spatial asymptotic dependence that is the dependence between extreme values.

Before being modeled by an extremal t process, data must be stationarized in space by transforming the margins into α -Fréchet distributions. As we are interested in maxima of precipitation values, one-dimensional marginal distributions are in the GEV family. To be able to perform the reverse transformation for any location

in the region and to deal with spatial nonstationarity, the parameters μ and σ of the GEV are modeled as functions of the spatial coordinates latitudes and longitude (hereafter referred to as LAT and LON) while the parameter ξ is assumed to be constant over the whole region:

$$\begin{aligned}\mu &= \exp[\mu_0 + \mu_1 \times \text{LAT} + \mu_2 \times \text{LON}] \\ \sigma &= \exp[\sigma_0 + \sigma_1 \times \text{LAT} + \sigma_2 \times \text{LON}] \\ \xi &= \xi_0\end{aligned}\quad (6)$$

The exponential formulation is employed for μ and σ as σ has to be positive and μ is chosen to be positive because of the nature of the variable of interest (precipitation). As the studied area is not very large, and because estimation of the shape parameter ξ is known to be an issue, ξ has been assumed constant in order to facilitate the inference, even though it is likely that it may slightly change over the region.

The nine parameters of the model, i.e., $\mu_0, \mu_1, \mu_2, \sigma_0, \sigma_1, \sigma_2$, and ξ are the parameters for the margin model; α is the extremal t parameter; and θ is the parameters of the correlation function ρ , have to be estimated from the data, i.e., the autumnal maxima over the time period 1960–1999 for the conditioning locations (480 values if 12 conditioning locations are retained). Nevertheless, for a dimension higher than 2, the multivariate density of the extremal t process is not tractable and usual likelihood maximization methods are not available. In this paper we used a full likelihood maximization based on partial censoring of components above a high margin threshold proposed by *Thibaud and Opitz* [2013]. Furthermore, the correlation function chosen here is the exponential one.

In Simulation Step 2, in order to provide distributions at any location in the region of interest, conditional simulations are used. These tools provide simulations of random fields even at locations without observation for calibration, given observations in fixed (i.e., conditioning) locations, according to a chosen spatial model. The simulated fields are required to fit the observed values at the fixed locations. In contrast to simple simulations for which simulated values follow a target distribution, the conditional simulations take into account constraints expressed by the observations. Furthermore, when other techniques provide deterministic values or indicators, conditional simulations provide distributions at any location, which allows us to get numerous statistical quantities as, e.g., quantiles and dispersion indicators. The conditional simulation techniques are then powerful tools. For some spatial models, such as Gaussian or Boolean, conditional simulations algorithms based on kriging techniques are quite straightforward [*Lantuéjoul*, 2002], but for max-stable models, this kind of techniques is not applicable. Then, *Dombry et al.* [2013] proposed a general framework for conditional simulations of max-stable processes. This algorithm, originally used on Schlather and Brown-Resnick processes [*Dombry et al.*, 2013] can be applied to extremal t processes [*Ribatet*, 2013; *Bechler et al.*, 2014]. In this paper, we use a version of the algorithm proposed by *Bechler et al.* [2014] and detailed in Appendix B, which is computationally intensive but shows good results and allows to consider a large number of conditioning points.

4. Methods and Scores

4.1. Downscaling Methods

Different versions of the SHD approach are described here and compared to some basic downscaling techniques.

1. Two standard downscaling techniques have been performed. “*Interpolation*” is a bilinear interpolation of the RCM outputs values taken at the center of each grid cell. “*Linear downscaling + Interpolation*” consists in performing a linear regression between the large-scale and local-scale data sets to compute values at the conditioning points (same as in section 3.2) and then a bilinear interpolation of these DS-PRJ to get values at other locations.
2. We test four versions of the SHD approach. Two of them use *CDF-t* or *linear regression* (described section 3.2) as 1d-link functions. For comparison purposes, we perform two extreme cases of the 1d-link functions. In the “*raw*” method, no 1d-link function is used (i.e., the RCM outputs are taken directly as conditioning values). In the “*optimal*” method, the values from the SAFRAN data set are taken as conditioning values. It is the virtual case where the 1d-link function is perfect and is able to rebuild exactly the observations. This optimal method should not be overcome by any other SHD methods.

All these techniques are summarized in Table 1.

Table 1. Summary of the Different Methods Tested to Downscale Fields of Extreme Precipitation

Methods	1d-Link	Conditional Simulation	Description
Interpolation	NO	NO	Bilinear interpolation of the RCM outputs values
Linear downscaling + interpolation	YES	NO	Linear regression between the local-scale variable and the large-scale variable
Raw	NO	YES	Use directly RCM outputs as conditioning values
CDF-t	YES	YES	Build a bias correction by comparing the CDF of the large-scale variable and the one of the local-scale variable
Linear regression	YES	YES	Same as <i>Linear downscaling</i> but only at the conditioning points
Optimal	NO	YES	Use directly the observations as conditioning values

4.2. Evaluation Scores

Different scores are proposed to compare the performances of these methods (for details about these scores, see Appendix C). In addition to the Continuous Rank Probability Score (CRPS) and the 95% Quantile Score (QS_{95}) [Gneiting and Raftery, 2007], the multivariate energy score (ES), which generalizes the CRPS [Gneiting et al., 2008], is computed to assess the spatial coherence of the simulations. Each of the marginal score is calculated at every location, and global scores are obtained by taking the mean of the individual scores over the region. ES is a multivariate score calculated for each simulated field. Hence, a global score is obtained by taking the mean over every simulation. The lower these scores, the better.

Then, the Kolmogorov-Smirnov test statistics are also computed to compare the concatenated distribution of the SAFRAN values over the eight test years to the distributions obtained by the different methods. Also, the Cramer-Von Mises criterion, which assesses the quadratic mean gap between the CDFs (in contrary to the K-S that evaluates the biggest gap between the distributions), is used. For these two tests, a statistic far from 0 leads to the rejection of the H_0 hypothesis that the two samples come from the same distribution.

Finally, we compute the variograms on the simulations obtained with the different methods. We assess the spatial variability coherence by comparing, for each method, the variograms of the numerous simulated fields to the one from the observations. To do so, the root-mean-squared error (RMSE) between the variograms of the observations and those from the simulations is calculated every year for each method.

For comparison purposes, we focus on the skill scores related to these presented scores. The mathematical formulation of those skill scores is given in Appendix C.

5. Results

We first investigate the results given by the different approaches when the conditioning points are chosen by the colocalization technique, i.e., by taking the points which are the closest to the center of each grid cell.

5.1. SHD With Colocalization

Table 2 shows the skill scores associated to the different scores, which correspond to the percentage of improvement of a method in terms of score with respect to a reference method (here the reference method is the Linear downscaling + Interpolation method).

It appears that concerning the marginal distribution (i.e., CRPS, QS, K-S, and CVM scores), the different versions of the SHD approach present better results than the reference one. More precisely, the linear regression 1d-link function seems to be the most relevant choice. This 1d-link function (combined with conditional simulations) remains also better than the other methods when looking at the spatial coherence (ES), while the best 1d-link function in SHD is CDF-t in terms of spatial variability ($RMSE_v$). However, when looking at the variograms obtained from the different methods for 2003 in Figure 2, we can note that the

Table 2. Skill Scores (See Appendix C), With Linear Downscaling + Interpolation as Reference, of the Different Downscaling Methods From the WRF-MED Data Set, Based on 12 Conditioning Points Chosen by Colocalization Technique^a

Models	CRPSS	QSS ₉₅	ESS	K-S SS	CVM SS	RMSE _v SS
No simulations						
Interpolation	-20.1%	-54.8%	-17.7%	-19.8%	-64.4%	-0.4%
Linear downscaling + Interpolation	0%	0%	0%	0%	0%	0%
Conditional simulations						
Raw	14.0%	44.4%	14.6%	10.4%	-11.7%	20.2%
CDF-t	14.1%	60.6%	17.0%	12.3%	-8.9%	40.4%
Linear regression	20.1%	53.4%	22.8%	15.0%	1.0%	22.5%
Optimal	24.2%	70.7%	29.5%	18.7%	3.9%	52.9%

^aNumbers in italics indicate the best scores in each column, among the non-optimal methods.

tested SHD approaches are not able to reproduce the global spatial variability of the data. We can also see on that figure that the CDF-t 1d-link function is the technique which degrades the least the spatial variability of the data, which is consistent with the fact that it has the lowest RMSE_v. The two variograms of the basic downscaling methods are far from the observed one.

Repeating this variogram comparison for every year of the evaluation period (not shown), it appears that, depending on the year, the spatial variability may be very different. In all cases, the global features present in Figure 2 are still present whatever the year, and moreover, as expected, the optimal method behaves slightly better than the other tested versions of SHD.

Figure 3 shows the evolution of the spatial mean over the region of simulation for the different models, as well as the evolution of the spatial mean of the observations, for 2000–2007. Each year is simulated independently, and the temporal evolution is only driven by the large-scale data. The four versions of the SHD approach are able to reproduce the global variations of the observations and also the magnitude even if the *Raw* method seems to underestimate the values. Interpolation and Linear downscaling + Interpolation follow the variations but provide very biased values. The performances of each method are consistent with the results shown in Table 2.

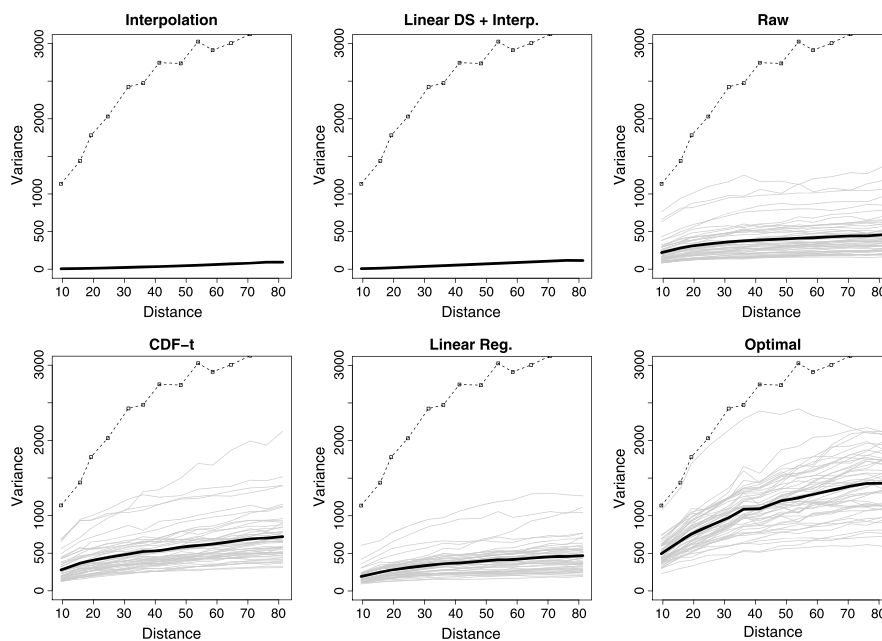


Figure 2. Variograms of the values generated by the six methods. In dotted line, from the SAFRAN data, in grey from the different conditional simulations, and in black from the mean of all conditional simulations (for the year 2003, distance in km).

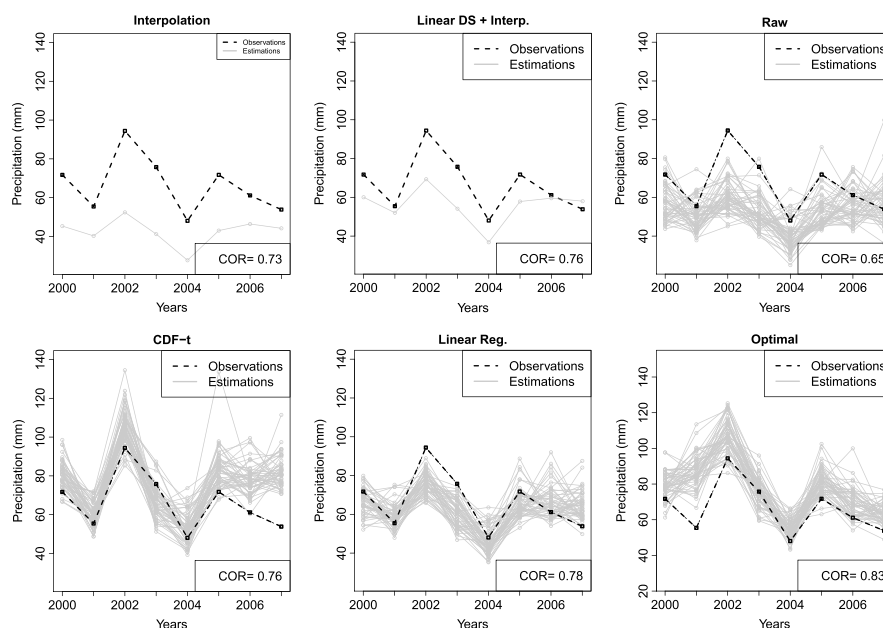


Figure 3. Spatial annual means obtained by the six methods. SAFRAN data are in black, and each trajectory in grey corresponds to eight conditional simulations from 2000 to 2007. COR is the Pearson's correlation coefficient value. See the text for details.

The Pearson correlation coefficient between observed spatial means and simulated spatial means for each simulation has also been calculated. Then, an average of this coefficient over all simulated fields leads to one correlation indicator for each method. Results (given in Figure 3) are globally the same between the SHD approaches and the basic downscaling techniques, except for the *Raw* method which shows a lower correlation than the others. The absence of 1d-link function has major consequences on the quality of the conditional simulations. It also appears that the two basic downscaling techniques are able to reproduce well the general interannual variations of the observations but underestimate the values.

We can also remark that the test period contains very heterogeneous years which is a good feature in context of evaluation of the robustness of our results. Hence, the method is able to reproduce the interannual variations, even when the interannual variability is very large.

5.2. Alternative Choice of the Conditioning Points

We investigate now how the number and the selection of the conditioning points influence the results of the SHD approach.

To find a set of conditioning points that are independent of the RCM, we applied a clustering technique on the SAFRAN data set and keep in each cluster a representative location as conditioning point. The "Partitioning Around Medoids" (PAM) algorithm is a clustering algorithm which represents each cluster by a representative point called medoid (while the *k*-means algorithm represents a cluster via a mean of points). We use a version of PAM designed by *Bernard et al.* [2013] for maxima based on the *F*-madogram [Cooley et al., 2006]. The optimal number of clusters, determined according to a silhouette criterion [Rousseeuw, 1987], gives us 57 medoids. Performing the conditional simulations with the 57 medoids as conditioning points shows that the general ranking of the different methods is more or less the same (i.e., with similar skill scores), but there is no improvement (except for the optimal model) in the different scores (not shown). Many medoids are near the boundaries and have been chosen by the PAM algorithm because of their atypicalness (and therefore may be alone in a class). These features lead to a difficult calibration of the 1d-link functions, complicate the drawing of the partition step, and provide less accurate conditional simulations than with the previous choice of the conditioning points (not shown). However, it is very interesting to remark that the 57 obtained medoids cover exactly the 57 SAFRAN climatologically homogeneous zones present in our region of interest. In light of this fact, we tried to exploit this feature in applying the PAM algorithm to the 57 zones, represented either by their maximum values, their mean values, or the medoid values. Those tests did not bring any improvement since the PAM algorithm chooses only two locations—and more precisely, the same two locations—for the three

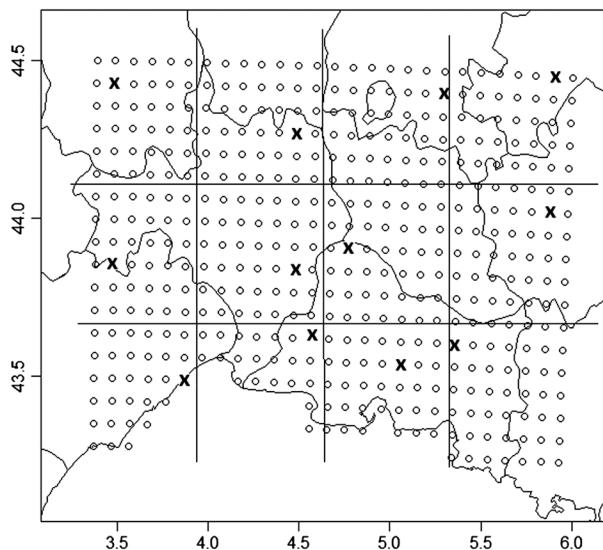


Figure 4. Locations of the 12 conditioning points obtained by a PAM algorithm applied to the SAFRAN data into each RCM grid cell separately.

points found by the algorithm. Table 3 shows the skill scores obtained when using those 12 conditioning points in SHD. This table shows that, again, the different versions of the SHD approach give better results than the basic downscaling methods, but this time it is the CDF-t method that is globally the most accurate. Nevertheless, the scores are sometimes not as good as those obtained with the colocalization technique, and the difference between the optimal method and the SHD method has increased in favor of the optimal method. This would mean that the quality of the 1d-link has weakened for these specific locations. In order to avoid choosing conditioning locations and to have results more independent of those conditioning points, in the next section, we randomly change them for each simulation.

5.3. Stochastic Hybrid Method

For each simulation, one conditioning point into each RCM grid cell is drawn randomly. This methodology, hereafter referred to as stochastic Spatial Hybrid Downscaling (stochastic SHD), is coherent with the fact that an RCM value is an aggregate value over a grid cell [e.g., Skelly and Henderson-Sellers, 1996; Osborn, 1997]. Nevertheless, it raises two major issues. First, it is computationally impossible to estimate the parameters of the model through likelihood maximization for each drawn ensemble of 12 points. Therefore, the parameters are estimated only once by maximization of the full likelihood built with all the available observations (i.e., at the 457 locations). The same estimated parameter set is used for each margin transformation (equation (A4)), whatever the randomly selected conditioning points. Second, in order to use the conditional simulation algorithm, we need to draw partitions of the 12 points. For a matter of computational time, it is not possible to run the Gibbs sampler for each set of 12 points. Therefore, we perform a Data Augmentation algorithm proposed by Lantuéjoul and Bel [2014], which is a deteriorated but still accurate algorithm

approaches (not shown), and these two conditioning points are obviously not enough to perform the SHD procedure properly, specially for inferring the spatial structure. Nevertheless, a major conclusion is that parsimony in the number of conditioning points is an important choice that may have critical consequences on the quality of the SHD results. In other words, increasing the number of conditioning points does not necessarily improve the quality of the simulations.

In order to limit the number of conditioning points, we use the PAM algorithm on maxima to find some representative points among the SAFRAN data set by limiting the search to one representative point per RCM grid cell. Figure 4 shows the conditioning

Table 3. Skill Scores (With Linear Downscaling + Interpolation as Reference) of the Different Methods With 12 Conditioning Points Chosen by PAM Applied Into Each Grid Cell^a

	Models	CRPSS	QSS ₉₅	ESS	KS SS	CVM SS	RMSE _v SS
No simulations	Interpolation	-6.5%	-22.6%	-4.7%	-19.8%	-20.2%	1.9%
	Linear downscaling + interpolation	0%	0%	0%	0%	0%	0%
Conditional simulations	Raw	9.6%	26.9%	12.2%	-7.7%	-9.4%	12.4%
	CDF-t	22.5%	58.9%	26.8%	17.6%	35.7%	34.2%
	Linear regression	18.4%	38.8%	18.9%	4.6%	15.2%	11.8%
	Optimal	34.6%	72.4%	39.3%	23.6%	43.3%	61.5%

^aPAM, Partitioning Around Medoids. Numbers in italics indicate the best scores in each column, among the non-optimal methods.

Table 4. Skill Scores (With Linear Downscaling + Interpolation as Reference) of the Different Methods With 12 Randomly Chosen and Switching Conditioning Points^a

	Models	CRPSS	QSS ₉₅	ESS	KS SS	CVM SS	RMSE _v SS
No simulations	Interpolation	-29.2%	-78.4%	-26.4%	-27.4%	-88.1%	-3.2%
	Linear downscaling + interpolation	0%	0%	0%	0%	0%	0%
Conditional simulation	Raw	-8.3%	-4.9%	-6.1%	-9.5%	44.9%	7.8%
	CDF-t	12.2%	39.8%	14.0%	15.8%	50.4%	23.8%
	Linear regression	10.1%	31.6%	11.6%	8.8%	35.5%	14.2%
	Optimal	36.7%	75.8%	37.3%	18.1%	56.2%	67.6%

^aNumbers in italics indicate the best scores in each column, among the non-optimal methods.

drawing the partitions much faster than a Gibbs sampler. Those two adaptations make the stochastic SHD method possible.

Table 4 displays the associated skill scores. This method provides a distribution at every point of the map, while the other techniques only give the downscaled value. As all the locations may be conditioning points, the optimal method may be slightly overestimated since, at those points, all the scores are zeros. Although the skill scores are not always as good as before, all scores are improved (not shown here) and the hierarchy between the methods remains the same with the hybrid approach being the most accurate method. In addition, CDF-t seems to be the best 1d-link function.

Finally, Table 5 summarizes the skill scores obtained with the different SHD methods for the choice of the conditioning points. Here we only compare the SHD approaches and the reference is the colocalization method. Based on the previous results, the 1d-link function considered is CDF-t. It appears that PAM applied per grid cell does not necessarily improve the quality of the conditional simulations and that stochastic SHD presents generally the best results. Nevertheless, depending on the criterion, the colocalization method provides quite proper results that are not overtaken by more complicated methods. Then, the choice of the conditioning points will largely depend on the density of the available data. Stochastic SHD can be used only when the data set is dense for validation purposes, for instance, in other cases colocalization techniques can still be used and present good results.

The variograms of the optimal SHD approaches with the different choices of points are also presented Figure 5, for the year 2003. It appears that the PAM approach and even stochastic SHD allow more spatial variability than the colocalization technique in the conditional simulations. The results are the same for the other years (not shown). This variability largely depends on the conditioning values and on their locations. Indeed, conditioning points close to points where simulations are provided reduce the variability of the simulations at those points, while spaced enough conditioning points bring more spatial variability between the points since they let the spatial model more free. In the same idea, the colocalization technique leads to uniformly spaced points, and this can partially explain the low spatial variability of the simulations whereas the other tested techniques allow wider space without conditioning points. Nevertheless, enabling more spatial variability in the simulations does not lead necessarily to accurate prediction in average as it appears in Table 5 showing bad scores for the PAM method.

5.4. Results With Other RCM Outputs

In order to test if the performances of our SHD approaches are sensitive to the large-scale data used as inputs and if they are generalizable to other conditions of application, another RCM data set has also been tested.

Table 5. Skill Scores (With Colocalization as Reference) of the SHD Algorithm (With CDF-t as 1d-Link Function) With the Different Methods to Choose the Conditioning Points^a

Models	CRPSS	QSS ₉₅	ESS	KS SS	CVM SS	RMSE _v SS
Colocalization	0%	0%	0%	0%	0%	0%
PAM into a grid cell	-0.2%	-27.6%	0.2%	9.8%	18.9%	-16.1%
Stochastic SHD	9.7%	-15.9%	5.0%	15.8%	20.9%	-29.1%

^aNumbers in italics indicate the best scores in each column, among the non-optimal methods.

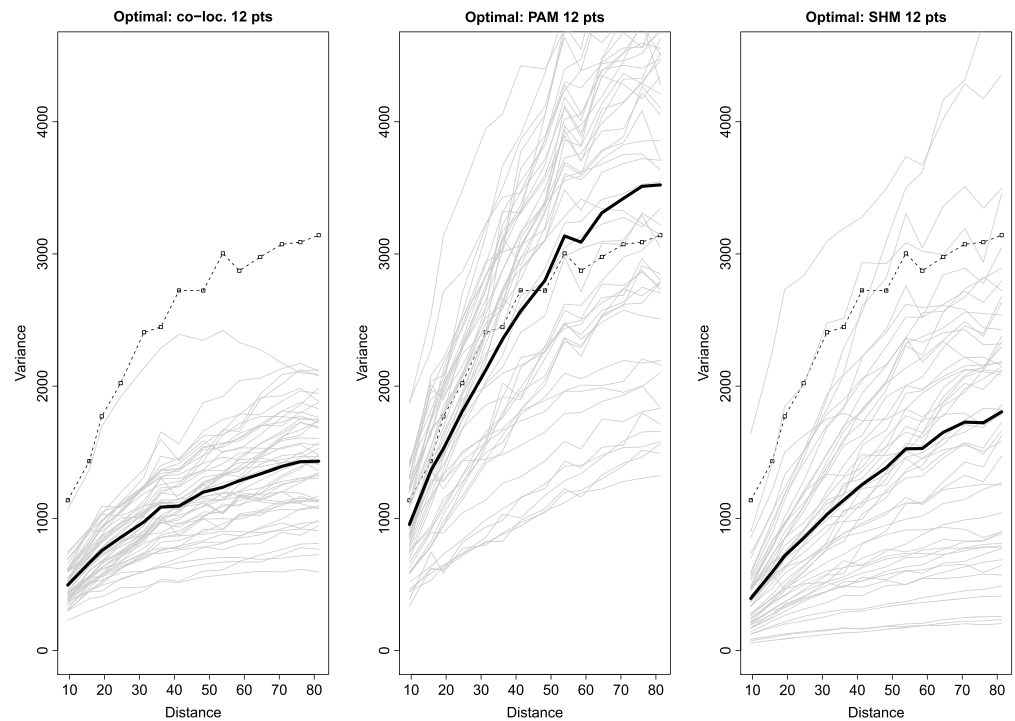


Figure 5. Variograms obtained with the optimal SHD method for different choices of conditioning points. The variogram from SAFRAN is in dotted line, the variograms from the different conditional simulations are in grey, and the variograms of the mean of every conditional simulation are in black (year 2003, distance in kilometers).

The WRF-IPSL-INERIS44 data [e.g., *Gobiet et al., 2012; Vautard et al., 2013*], hereafter referred to as WRF-EURO data, are RCM outputs available from 1989 to 2007. The main difference with the previous RCM (WRF-MED) is that the WRF-EURO model was not nudged. We use the expected daily mean precipitation over a grid with a resolution of 50 km relatively equivalent to the first RCM. The WRF-EURO data cover our region of interest with 15 grid cells, which are not at the same locations as those from the previous RCM (WRF-MED). On each of these grid cells, we build the autumnal maximum of the daily mean precipitation. The period 1989–1999 is considered as the learning period, and the period 2000–2007 is used once again for the evaluation. The conditioning points are selected by colocalization technique, i.e., by taking the points which are the closest to the center of each grid cell. We then obtain 15 conditioning points.

The skill scores results of the different downscaling methods are given in Table 6. We note that although there are more conditioning points, the accuracy of the SHD approaches is not improved compared to the results obtained with the 12 conditioning points provided by the WRF-MED RCM. It may be explained by the fact that this RCM is not nudged, implying less consistency between the large-scale (i.e., RCM) and the local-scale (i.e., SAFRAN) data. Other differences between the WRF-MED and WRF-EURO simulations might also explain the differences in terms of skill scores. For example, the boundaries of the forcing by ERA-Interim reanalysis

Table 6. Skill Scores (With Linear Downscaling + Interpolation as Reference) of the Different Downscaling Methods From the WRF-EURO Data Set With 15 Conditioning Points^a

	Models	CRPSS	QSS ₉₅	ESS	KS SS	CVM SS	RMSE _v SS
No simulations	Interpolation	-35.8%	-62.4%	-29.7%	-25.4%	-72.5%	4.6%
	Linear downscaling + interpolation	0%	0%	0%	0%	0%	0%
Conditional simulations	Raw	18.9%	50.7%	20.3%	16.6%	8.3%	25.6%
	CDF-t	10.8%	62.7%	19.6%	10.0%	-3.0%	35.6%
	Linear regression	24.9%	55.1%	25.3%	13.6%	2.6%	52.9%
	Optimal	36.3%	73.5%	34.8%	20.0%	15.1%	60.5%

^aNumbers in italics indicate the best scores in each column, among the non-optimal methods.

are also different. All these differences have some consequences on the results of the whole SHD algorithm. However, the results and conclusions are globally similar as those for the nudged WRF-MED RCM and the SHD approach remains relevant.

6. Conclusions and Discussion

We have proposed and investigated a new hybrid approach that combines downscaling techniques and statistical conditional simulations to get distributions of extremes at any location. Performing a statistical downscaling only for a given subset of locations and then using these downscaled values as conditioning values for the conditional simulation algorithm enable to provide spatially coherent simulations of extreme fields, even for locations without available observations for calibration.

The application of this approach on the SAFRAN data set and in realistic conditions shows great improvements compared to basic downscaling techniques. Furthermore, different versions of this SHD approach have been tested to highlight the role of the different components of the method. First, the 1d-link functions manage to reduce the biases existing in the RCM outputs. Second, thanks to the inference of the max-stable process from local-scale data, the conditional simulation step brings additional information and spatial coherence compared to a univariate downscaling. It appears that both the statistical downscaling and conditional simulation steps are essential for the quality of the SHD approach. It is also important to note that the quality of the results of this SHD approach largely depends on the quality of the climate model outputs. The biases of these outputs may be corrected by the downscaling step, but the interannual variability remains driven by the climate model. These two statements have been confirmed by the results of the SHD approach on an other RCM.

Besides, the choice of the number and the location of the conditioning points is a major issue. Different selection strategies have been developed, and it appears that increasing the number of conditioning points does not necessarily lead to an improvement of the results. Another way of selection has been proposed, consisting in choosing randomly one conditioning point into each grid cell and to change it at each simulation. It seems to provide the best results even if it presents more technical difficulties. In practice, this choice will depend on the number of available observations, but it has to be kept in mind that the choice of the conditioning points has to be done very carefully. After various tests on the configuration of the conditioning points, we can say that, in general, the hierarchy between the different methods is the same from one configuration to another. In addition, the range of the different scores is also very narrow for the different tested strategies. This shows a robustness in our conclusions, despite the difficulties to highlight one optimal strategy for the choice of the conditioning points. Besides, it is important to remark that the choice of the number of the conditioning points can be seen as a trade-off between the importance given to the 1d-link functions and to the simulated spatial model.

Among the various possible extensions and perspectives of studies, this choice of points constitutes one potential improvement. Other clustering algorithms can be tested to deal with outlier classes (i.e., to prevent point from being alone in a class) as, for example, the superparamagnetic clustering algorithm [Blatt *et al.*, 1996]. Other technical studies can be performed, for example, a comparative study between the conditional simulation algorithm presented here and the data augmentation one performed for time-saving reasons when applying the stochastic SHD method.

Furthermore, the 1d-link (i.e., downscaling) functions performed in this work are not specific to extreme modeling. Some downscaling or bias correction methods adapted to extremes might improve the quality of the downscaling step. For example, an extension of the CDF-t method has been proposed by Kallache *et al.* [2011]. This model, called XCDF-t, based on the extreme value theory could be incorporated into the SHD approach in a future work to see if there is an improvement of the quality of the simulations. In addition, the general methodology of the SHD approach is not specific to extreme events and could be applied to more "regular" events with classical statistical spatial models and at a daily time step. This could be very useful for some hydrological impact models or any model requiring daily fields of climate variables as inputs.

This work enables us to provide projections of future local-scale variables, based on climate model outputs. Hence, naturally, one of the major perspectives consists in describing the local consequences of the different large-scale scenarios provided by the IPCC [2013]. It may also be interesting to compare the results of the SHD approach presented here to those obtained by taking directly the GCM outputs instead of the RCM. Indeed, it may appear that some statistical downscaling techniques on GCM are more efficient than those on RCM

[e.g., *Eden et al.*, 2014]. This obviously depends on many aspects of the chosen method, as the downscaling techniques or the RCM resolution, but the idea of deleting one intermediary—the RCM—may be interesting to reduce the modeling error.

Finally, this SHD approach is meant to enable us to generate fields and time series useful for many impact studies aiming to anticipate the consequences of extreme precipitation events in a climate change context.

Appendix A: Max-Stable Processes

Commonly used tools to model extreme values and specially observation of maxima are max-stable processes. Let us consider a stationary random field $\{Z(x)\}_{x \in \mathbb{R}^d}$ with a given distribution \mathcal{H}_0 and (Z_1, \dots, Z_n) independent copies of Z . If there exists sequences $a_n(x) \in \mathbb{R}$ and $b_n(x) \in \mathbb{R}^+$ such that for all $n \in \mathbb{N}$,

$$\left\{ \frac{\max_{i=1, \dots, n} Z_i(x) - a_n(x)}{b_n(x)} \right\}_{x \in \mathbb{R}^d} \tag{A1}$$

follows the distribution \mathcal{H}_0 , then $\{Z(x)\}_{x \in \mathbb{R}^d}$ is a max-stable process [*De Haan*, 1984].

As a consequence, all finite dimensional marginal distributions are max-stable, and in particular, the univariate marginal distribution $Z(x)$ belongs to the Generalized Extreme Value (GEV) family:

$$\mathbb{P}(Z(x) \leq z) = \begin{cases} \exp \left(- \left[1 + \xi(x) \left(\frac{z - \mu(x)}{\sigma(x)} \right) \right]^{-1/\xi(x)} \right) & \text{if } \xi(x) \neq 0, \\ \exp \left(- \exp \left(\frac{z - \mu(x)}{\sigma(x)} \right) \right) & \text{if } \xi(x) = 0, \end{cases} \tag{A2}$$

where $\mu(x)$, $\sigma(x)$, and $\xi(x)$ are, respectively, the location, scale, and shape parameters at site x . Furthermore, the sign of $\xi(x)$ will determine if the distribution $Z(x)$ belongs to the Fréchet ($\xi > 0$), Gumbel ($\xi = 0$), or Weibull ($\xi < 0$) family and consequently the characteristics of the distribution tail.

Different max-stable models have been proposed and used in the past by several authors, e.g., *Brown and Resnick* [1977], R. L. Smith (unpublished manuscript, 1990), *Schlather* [2002], or *Kabluchko et al.* [2009].

Their margins are Fréchet distributed, i.e.,

$$\mathbb{P}(Z(x) \leq z) = e^{-z^{-1}}. \tag{A3}$$

If Z is distributed according to a $GEV(\mu, \sigma, \xi)$, then Y defined as

$$Y = \left(1 + \frac{\xi(Z - \mu)}{\sigma} \right)_+^{1/\alpha\xi} \tag{A4}$$

follows an α -Fréchet distribution [*Beirlant et al.*, 2004].

Appendix B: Conditional Simulation Algorithm

Dombry et al. [2013] proposed a general framework for conditional simulations of max-stable processes. The algorithm is divided in three steps: *Step 1*: Generate a partition of the conditioning points. *Step 2*: For each class of this partition, generate one extremal function which is equal to the observed values at the points of this class and which is lower to the observed values at the other conditioning locations. *Step 3*: Independently of the two first steps, generate a high number of subextremal functions which are lower to the observed values at all conditioning points.

This algorithm, originally used on the Schlather and Brown-Resnick processes [*Dombry et al.*, 2013], can be applied to the extremal t process [*Ribatet*, 2013; *Bechler et al.*, 2014]. In this paper, the first step is performed thanks to a Gibbs sampler as in *Dombry et al.* [2013]. For the second step, the general methodology of simulation under constrains detailed by *Bechler et al.* [2014], following *Geweke* [1991], is applied and the simulation of subextremal functions is performed via an acceptance-rejection step.

Appendix C: Scores

Let F_n be the predictive cumulative distribution function derived from the n conditional simulations $\{\hat{y}^{(j)}\}_{j=1,\dots,n}$ and y the observation. The continuous ranking probability score is written

$$\text{crps}(F_n; y) = \int_{-\infty}^{+\infty} [F_n(t) - H(t - y)]^2 dt, \quad (\text{C1})$$

where $H(u)$ is the Heavyside function which is equal to 0 for $u < 0$ and 1 otherwise. The steeper the CDF is around the observed value, the better, and a good score is a score close to 0. The calculation of the crps will result in a value with the same unit as the variable.

The quantile score (qs) is defined as

$$\text{qs}_p(F_n; y) = v_p(y - F_n^{-1}(p)), \quad (\text{C2})$$

where p is the probability of interest and $v_p(u) = pu$ if $u \geq 0$ and $v_p(u) = (p - 1)u$ otherwise. As shown by *Friederichs and Thorarinsdottir* [2012], the crps is proportional to the integral of the qs_p over all possible probabilities p . Then, a good qs_p is a qs_p close to 0 as well. By choosing a high probability p , the quantile score focuses on the predictive performance in the upper tail [*Gneiting and Raftery*, 2007].

In addition to these two univariate scores, which assess the marginal qualities of the conditional simulation, we propose to use the energy score [*Gneiting et al.*, 2008] that may be written as

$$ES(\hat{y}^{(1)}, \dots, \hat{y}^{(n)}, \mathbf{y}) = \frac{1}{n} \sum_{i=1}^n \|\hat{y}^{(i)} - \mathbf{y}\| - \frac{1}{2n^2} \sum_{j=1}^n \sum_{i=1}^n \|\hat{y}^{(i)} - \hat{y}^{(j)}\|, \quad (\text{C3})$$

where $\hat{y}^{(i)}$ is one of the n simulated fields and \mathbf{y} is the field of observations. This score enables us to assess the spatial coherence of the simulations.

For comparison of distributions, we also use the statistics of the Kolmogorov-Smirnov test

$$\text{K-SS} = \sup_x |F_n(x) - F(x)| \quad (\text{C4})$$

and of the Cramèr-von-Mises test statistic

$$\text{CVMS} = \int_{-\infty}^{+\infty} [F_n(x) - F(x)]^2 dx, \quad (\text{C5})$$

where F is the empirical cumulative distribution function of the observations. All these scores (except the energy score) are first spatially and then temporally aggregated to obtain global scores (CRPS, QS_p , K-SS, and CVMS).

In order to compare models, skill scores can be built [*Gneiting and Raftery*, 2007]:

$$S^{\text{skill}} = \frac{S^{\text{mod}} - S^{\text{ref}}}{S^{\text{opt}} - S^{\text{ref}}}, \quad (\text{C6})$$

with S^{mod} , S^{ref} , and S^{opt} , respectively, the score of the current model, the score of a reference model, and the optimal score. For models better than the reference model, S^{skill} belongs to $[0, 1]$ and 1 corresponds to the optimal case. For the five scores presented here, we have $S^{\text{opt}} = 0$ and then the corresponding skill scores can be written as

$$\text{CRPSS} = \frac{\overline{\text{CRPS}}^{\text{ref}} - \overline{\text{CRPS}}^{\text{mod}}}{\overline{\text{CRPS}}^{\text{ref}}}, \text{QS}_p = \frac{\overline{\text{QS}}_p^{\text{ref}} - \overline{\text{QS}}_p^{\text{mod}}}{\overline{\text{QS}}_p^{\text{ref}}}, \text{ESS} = \frac{\overline{\text{ES}}^{\text{ref}} - \overline{\text{ES}}^{\text{mod}}}{\overline{\text{ES}}^{\text{ref}}}. \quad (\text{C7})$$

$$\text{K-S SS} = \frac{\overline{\text{K-SS}}^{\text{ref}} - \overline{\text{K-SS}}^{\text{mod}}}{\overline{\text{K-SS}}^{\text{ref}}}, \text{CVM SS} = \frac{\overline{\text{CVMS}}^{\text{ref}} - \overline{\text{CVMS}}^{\text{mod}}}{\overline{\text{CVMS}}^{\text{ref}}} \quad (\text{C8})$$

corresponding to the percentages of improvement of the scores brought by the current model with respect to the reference model.

Acknowledgments

This work has been partially supported by the ANR-project McSIM, the ANR-project StaRMIP, the GICC-project REMedHE, the VW-project PLEIADES, and the ANR-project REMEMBER. This work is part of the Med-CORDEX (<http://www.medcordex.eu/>) and EURO-CORDEX (<http://www.euro-cordex.net/>) initiative. All computations have been made in R. An R package containing functions for the conditional simulations procedure based on the extremal t processes should soon be made available on the CRAN website (<http://cran.r-project.org/>) or upon request to the authors. IPSL-WRF311 data set is available on <http://www.medcordex.eu/medcordex.php>, WRF-IPSL-INERIS44 data set on <http://www.euro-cordex.net/EURO-CORDEX-Data.2613.0.html>, and SAFRAN data set is available upon request to Météo-France.

References

- Beaulant, A.-L., B. Joly, O. Nuissier, S. Somot, V. Ducrocq, A. Joly, F. Sevault, M. Deque, and D. Ricard (2011), Statistico-dynamical downscaling for Mediterranean heavy precipitation, *Q. J. R. Meteorol. Soc.*, *137*(656), 736–748, doi:10.1002/qj.796.
- Bechler, A., L. Bel, and M. Vrac (2014), Conditional simulations of extremal t process: Application to fields of extreme precipitation. [Available at <http://hal-agroparistech.archives-ouvertes.fr/hal-01053605>.]
- Beirlant, J., Y. Goegebeur, J. Teugels, J. Segers, D. D. Waal, and C. Ferro (2004), *Statistics of Extremes. Theory and Applications*, xiii + 490 pp., John Wiley, Hoboken, N. J., doi:10.1002/0470012382.
- Bernard, E., P. Naveau, M. Vrac, and O. Mestre (2013), Clustering of maxima: Spatial dependencies among heavy rainfall in France, *J. Clim.*, *26*, 7929–7937.
- Blanchet, J., and A. C. Davison (2011), Spatial modeling of extreme snow depth, *Ann. Appl. Stat.*, *5*(3), 1699–1725.
- Blatt, M., S. Wiseman, and E. Domany (1996), Superparamagnetic clustering of data, *Phys. Rev. Lett.*, *76*, 3251–3254, doi:10.1103/PhysRevLett.76.3251.
- Boé, J., L. Terray, F. Habets, and E. Martin (2006), A simple statistical-dynamical downscaling scheme based on weather types and conditional resampling, *J. Geophys. Res.*, *111*, D23106, doi:10.1029/2005JD006889.
- Brown, B. M., and S. I. Resnick (1977), Extreme values of independent stochastic processes, *J. Appl. Probab.*, *14*(4), 732–739.
- Buishand, T., L. de Haan, and C. Zhou (2008), On spatial extremes: With application to a rainfall problem, *Ann. Appl. Stat.*, *2*, 624–642.
- 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.
- Coles, S., and J. Tawn (1994), Statistical methods for multivariate extremes—An application to structural design, *J. R. Stat. Soc.*, *43*(1), 1–48.
- Cooley, D., P. Naveau, and P. Poncet (2006), Variograms for spatial max-stable random fields, in *Dependence in Probability and Statistics*, pp. 373–390, Springer, New York.
- Davison, A. C., and M. M. Gholamrezaee (2012), Geostatistics of extremes, *Proc. R. Soc. London, Ser. A*, *468*(2138), 581–608.
- Davison, A. C., S. A. Padoan, and M. Ribatet (2012), Statistical modeling of spatial extremes, *Stat. Sci.*, *27*(2), 161–186, doi:10.1214/11-STS376.
- De Haan, L. (1984), A spectral representation for max-stable processes, *Ann. Probab.*, *12*(4), 1194–1204, doi:10.1214/aop/1176993148.
- De Haan, L., and J. De Ronde (1998), Sea and wind: Multivariate extremes at work, *Extremes*, *1*(1), 7–45.
- Dee, D. P., et al. (2011), The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, *Q. J. R. Meteorol. Soc.*, *137*(656), 553–597.
- Déqué, M. (2007), Frequency of precipitation and temperature extremes over France in an anthropogenic scenario: Model results and statistical correction according to observed values, *Global Planet. Change*, *57*(1), 16–26.
- Domby, C., F. Éyi-Minko, and M. Ribatet (2013), Conditional simulation of max-stable processes, *Biometrika*, *100*(1), 111–124.
- Ducrocq, V., O. Nuissier, D. Ricard, C. Lebeaupin, and T. Thouvenin (2008), A numerical study of three catastrophic precipitating events over southern France. II: Mesoscale triggering and stationarity factors, *Q. J. R. Meteorol. Soc.*, *134*(630), 131–145.
- Eden, J., M. Widmann, D. Maraun, and M. Vrac (2014), Comparison of GCM- and RCM-simulated precipitation following stochastic postprocessing, *J. Geophys. Res. Atmos.*, *119*, 40–53, doi:10.1002/2014JD021732.
- Fawcett, L., and D. Walshaw (2012), Estimating return levels from serially dependent extremes, *Environmetrics*, *23*(3), 272–283, doi:10.1002/env.2133.
- Fowler, H. J., and M. Ekström (2009), Multi-model ensemble estimates of climate change impacts on UK seasonal precipitation extremes, *Int. J. Climatol.*, *29*(3), 385–416, doi:10.1002/joc.1827.
- Fowler, H. J., D. Cooley, S. R. Sain, and M. Thurston (2010), Detecting change in UK extreme precipitation using results from the climateprediction.net BBC climate change experiment, *Extremes*, *13*(2), 241–267.
- Friederichs, P., and A. Hense (2007), Statistical downscaling of extreme precipitation events using censored quantile regression, *Mon. Weather Rev.*, *135*, 2365–2378, doi:10.1175/MWR3403.1.
- Friederichs, P., and T. L. Thorarinsdottir (2012), Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction, *Environmetrics*, *23*(7), 579–594, doi:10.1002/env.2176.
- Fuentes, M., J. Henry, and B. Reich (2013), Nonparametric spatial models for extremes: Application to extreme temperature data, *Extremes*, *16*(1), 75–101, doi:10.1007/s10687-012-0154-1.
- Gaume, J., N. Eckert, G. Chambon, M. Naaim, and L. Bel (2013), Mapping extreme snowfalls in the French Alps using max-stable processes, *Water Resour. Res.*, *49*, 1079–1098, doi:10.1002/wrcr.20083.
- Geweke, J. (1991), Efficient simulation from the multivariate normal and Student- t distributions subject to linear constraints and the evaluation of constraint probabilities, in *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, pp. 571–578, Citeseer, New York.
- Gneiting, T., and A. E. Raftery (2007), Strictly proper scoring rules, prediction, and estimation, *J. Am. Stat. Assoc.*, *102*(477), 359–378.
- Gneiting, T., L. Stanberry, E. Gmit, L. Held, and N. Johnson (2008), Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds, *Test*, *17*(2), 211–235.
- Gobiet, A., D. Jacob, and Euro-Cordex Community (2012), A new generation of regional climate simulations for Europe: The EURO-CORDEX Initiative, in *EGU General Assembly Conference Abstracts, EGU General Assembly Conference Abstracts*, vol. 14, edited by A. Abbasi and N. Giesen, p. 8211, Vienna.
- Huth, R. (2001), Disaggregating climatic trends by classification of circulation patterns, *Int. J. Climatol.*, *21*(2), 135–153, doi:10.1002/joc.605.
- Intergovernmental Panel on Climate Change (IPCC) (2012), *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change*, Cambridge Univ. Press.
- Intergovernmental Panel on Climate Change (IPCC) (2013), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker et al., 1535 pp., Cambridge Univ. Press, Cambridge U. K., and New York.
- Jonathan, P., K. Ewans, and D. Randell (2014), Non-stationary conditional extremes of northern North Sea storm characteristics, *Environmetrics*, *25*(3), 172–188, doi:10.1002/env.2262.
- Kabluchko, Z., M. Schlather, and L. de Haan (2009), Stationary max-stable fields associated to negative definite functions, *Ann. Probab.*, *37*(5), 2042–2065, doi:10.1214/09-AOP455.
- Kallache, M., M. Vrac, P. Naveau, and P.-A. Michelangeli (2011), Non-stationary probabilistic downscaling of extreme precipitation, *J. Geophys. Res.*, *116*, D05113, doi:10.1029/2010JD014892.
- Kirtman, B., et al. (2013), *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*, edited by T. F. Stocker et al., Near-Term Climate Change: Projections and Predictability, Cambridge Univ. Press, Cambridge, U. K.

- Lantuéjoul, C. (2002), *Geostatistical Simulation: Models and Algorithms*, Springer, Berlin.
- Lantuéjoul, C., and L. Bel (2014), Simulation conditionnelle du processus de schlather, in *46èmes Journées de Statistique de la SFDs*.
- Laprise, R. (2008), Regional climate modelling, *J. Comput. Phys.*, 227(7), 3641–3666.
- Lebeaupin, C., V. Ducrocq, and H. Giordani (2006), Sensitivity of torrential rain events to the sea surface temperature based on high-resolution numerical forecasts, *J. Geophys. Res.*, 111, D12110, doi:10.1029/2005JD006541.
- Levvasseur, G., et al. (2011), Present and LGM permafrost from climate simulations: Contribution of statistical downscaling, *Clim. Past*, 7(4), 1225–1246.
- Michelangeli, P.-A., M. Vrac, and H. Loukos (2009), Probabilistic downscaling approaches: Application to wind cumulative distribution functions, *Geophys. Res. Lett.*, 36, L11708, doi:10.1029/2009GL038401.
- Opitz, T. (2013), Extremal t-processes: Elliptical domain of attraction and a spectral representation, *J. Multivariate Anal.*, 122, 409–413, doi:10.1016/j.jmva.2013.08.008.
- Osborn, T. J. (1997), Areal and point precipitation intensity changes: Implications for the application of climate models, *Geophys. Res. Lett.*, 24(22), 2829–2832, doi:10.1029/97GL02976.
- Padoan, S. A. (2011), Multivariate extreme models based on underlying skew-t and skew-normal distributions, *J. Multivariate Anal.*, 102, 977–991, doi:10.1016/j.jmva.2011.01.014.
- Panagoulia, D., P. Economou, and C. Caroni (2014), Stationary and nonstationary generalized extreme value modelling of extreme precipitation over a mountainous area under climate change, *Environmetrics*, 25(1), 29–43, doi:10.1002/env.2252.
- Piani, C., J. Haerter, and E. Coppola (2010), Statistical bias correction for daily precipitation in regional climate models over Europe, *Theor. Appl. Climatol.*, 99(1–2), 187–192, doi:10.1007/s00704-009-0134-9.
- Quintana-Seguí, P., P. Le Moigne, Y. Durand, E. Martin, F. Habets, M. Baillon, C. Canellas, L. Franchisteguy, and S. Morel (2008), Analysis of near-surface atmospheric variables: Validation of the SAFRAN analysis over France, *J. Appl. Meteorol. Climatol.*, 47(1), 92–107.
- Ribatet, M. (2013), Spatial extremes: Max-stable processes at work, *J. Soc. Fr. Stat.*, 154(2), 156–177.
- Richardson, C., E. Barrow, M. Semenov, and R. Brooks (1998), Comparison of the WGEN and LARS-WG stochastic weather generators for diverse climates, *Clim. Res.*, 10(2), 95–107.
- Rousseeuw, P. J. (1987), Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, *J. Comput. Appl. Math.*, 20, 53–65.
- Rummukainen, M. (2010), State-of-the-art with regional climate models, *Wiley Interdiscip. Rev. Clim. Change*, 1(1), 82–96.
- Schlather, M. (2002), Models for stationary max-stable random fields, *Extremes*, 5(1), 33–44, doi:10.1023/A:1020977924878.
- Schlather, M., and J. A. Tawn (2003), A dependence measure for multivariate and spatial extreme values: Properties and inference, *Biometrika*, 90(1), 139–156, doi:10.1093/biomet/90.1.139.
- Skamarock, W. C., J. B. Klemp, J. Dudhia, D. O. Gill, M. Barker, K. G. Duda, X. Y. Huang, W. Wang, and J. G. Powers (2008), A description of the Advanced Research WRF Version 3, *NCAR Tech. Note NCAR/TN-475+STR*, Natl. Cent. for Atmos. Res., Boulder, Colo.
- Skelly, W. C., and A. Henderson-Sellers (1996), Grid box or grid point: What type of data do GCMs deliver to climate impacts researchers?, *Int. J. Climatol.*, 16(10), 1079–1086.
- Thibaud, E., and T. Opitz (2013), Efficient inference and simulation for elliptical Pareto processes, arXiv preprint arXiv:1401.0168.
- Vautard, R., et al. (2013), The simulation of European heat waves from an ensemble of regional climate models within the EURO-CORDEX project, *Clim. Dyn.*, 41(9–10), 2555–2575.
- Vautard, R., et al. (2014), The European climate under a 2°C global warming, *Environ. Res. Lett.*, 9(3), 034006.
- Vidal, J.-P., E. Martin, L. Franchistéguy, M. Baillon, and J.-M. Soubeyroux (2010), A 50-year high-resolution atmospheric reanalysis over France with the Safran system, *Int. J. Climatol.*, 30(11), 1627–1644, doi:10.1002/joc.2003.
- von Storch, H., H. Langenberg, and F. Feser (2000), A spectral nudging technique for dynamical downscaling purposes, *Mon. Weather Rev.*, 128(10), 3664–3673.
- 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., and P. Yiou (2010), Weather regimes designed for local precipitation modeling: Application to the Mediterranean basin, *J. Geophys. Res.*, 115, D12103, doi:10.1029/2009JD012871.
- Vrac, M., P. Marbaix, D. Paillard, and P. Naveau (2007), Non-linear statistical downscaling of present and LGM precipitation and temperatures over Europe, *Clim. Past*, 3(4), 669–682, doi:10.5194/cp-3-669-2007.
- Vrac, M., P. Drobinski, A. Merlo, M. Herrmann, C. Lavaysse, L. Li, and S. Somot (2012), Dynamical and statistical downscaling of the French Mediterranean climate: Uncertainty assessment, *Nat. Hazard. Earth Syst. Sci.*, 12(9), 2769–2784, doi:10.5194/nhess-12-2769-2012.
- Wilks, D. S. (2012), Stochastic weather generators for climate-change downscaling, Part II: Multivariable and spatially coherent multisite downscaling, *Wiley Interdiscip. Rev. Clim. Change*, 3(3), 267–278, doi:10.1002/wcc.167.