Journal de la Société Française de Statistique
Vol. 156 No. 1 (2015)

Stochastic weather generators: an overview of weather type models

Titre: Générateurs stochastiques de condition météorologiques : une revue des modèles à type de temps

Pierre Ailliot[1], Denis Allard[2], Valérie Monbet[3] and Philippe Naveau[4]

Abstract: A recurrent issue encountered in environmental, ecological or agricultural impact studies in which climate is an important driving force is to provide fast and realistic simulations of atmospheric variables such as temperature, precipitation and wind at a few specific locations, at daily or hourly temporal scales. Spatio-temporal dynamics and correlation structures among the variables of interest, as well as weather persistence and natural variability have to be reproduced accurately in a distributional sense. This quest leads to a large variety of so-called stochastic weather generators (WGs) in the literature. Here, we provide an up-to-date overview of weather type WG models. Weather types classically represent daily characteristics of the relevant atmospheric information at hand. There are many ways to build such weather states, either hidden or observed, and to infer their properties. This overview should help statisticians as well as meteorologists and climate product users to understand the probabilistic concepts and models behind weather type WGs, and to identify their advantages and limits.

Résumé : Pour réaliser des études d'impact dans lesquelles le climat est un paramètre d'entrée important, un problème fréquemment rencontré consiste à produire des séries temporelles de variables climatiques telles que températures, précipitation, vent ou humidité relative, en plusieurs sites simultanément, au pas de temps journalier et parfois horaire. Ces séries doivent être faciles à générer. Elles doivent aussi être réalistes en ce sens que les distributions des statistiques liées à la dynamique spatio-temporelle, telles que les corrélation entre variables, la persistence temporelle et les différentes sources de variabilité doivent être correctement reproduites. De nombreux générateurs stochastiques de conditions météorologiques ont été proposés dans ce but. Dans cet article, nous proposons de passer en revue la classe particulière des générateurs stochastiques à base de types de temps. En règle générale, un type de temps est une caractérisation grossière des conditions atmosphériques journalières. Il existe de nombreuses façons de définir les types de temps, qu'ils soient observés ou cachés dans une structure latente, et d'en inférer leur propriétés. Cette revue a pour objet d'aider les statisticiens, les scientifiques du climat et les utilisateurs de produits climatiques à appréhender les concepts et modèles probabilistes utilisés dans les générateurs stochastiques de conditions météorologiques et d'en cerner les avantages et leurs limites.

Keywords: Stochastic Weather Generators, Precipitation, Regime Switching Models, Weather Type

Mots-clés : Générateurs aléatoires de conditions météorologiques, Précipitations, Modèles à Changements de Régimes, Type de temps

AMS 2000 subject classifications: 62-02, 62P12

[1] LMBA, Université de Bretagne Occidentale, Brest, France.
E-mail: pierre.ailliot@univ-brest.fr
[2] INRA, Biostatistique et Processus Spatiaux (BioSP), Avignon, France.
E-mail: denis.allard@avignon.inra.fr
[3] IRMAR, INRIA/ASPI, Université de Rennes 1, Rennes, France.
E-mail: valerie.monbet@univ-rennes1.fr
[4] LSCE, IPSL, CNRS/CEA, Saclay, France.
E-mail: philippe.naveau@lsce.ipsl.fr

 ISSN: 2102-6238

1. Introduction

Stochastic weather generators (WGs) are statistical models that aim at quickly simulating realistic random sequences of atmospheric variables such as temperature, precipitation and wind (Wilks and Wilby, 1999). Ideally, spatio-temporal dynamics and correlation structures among the variables of interest, as well as weather persistence and natural variability, have to be reproduced accurately in a distributional sense by WGs.

At least three features distinguish WGs from numerical global climate models. WGs focus on small spatial scales (typically a few sites within a region extending over few kilometers), they have to be computationally very fast to provide numerous random realizations and those outputs should have the same distributional properties as observed time series, mainly at the daily or subdaily scales. In contrast, climate models have to reproduce the behavior of the whole atmosphere and its interactions with other components of the Earth system (vegetation, oceans, etc.) at the global scale and for a long time period. The price to pay for this inclusiveness is that only few runs can be provided by global climate models and they do not correspond to a specific site but live on large spatial grids. Those differences explain why WGs have been adopted in impact studies as computationally inexpensive tools to generate synthetic daily time series of atmospheric variables at local sites. Such simulated outputs are then fed into process-based models, typically electricity demand models or crop models (e.g., Kolokotroni et al., 2012; Launay et al., 2009). Non-linear interactions in process-based models imply that small variations in weather inputs can lead to large output discrepancies, and to counter intuitive behavior. For example, complex relationships among sowing dates, temperatures, droughts and growth render very complex the assessment of the weather impact on agricultural yields. To investigate the influence of weather conditions on such crop models, it is essential to be able to explore the weather parameter space via simulations. To a certain extent, such a strategy is not new. It has been commonly used in geosciences, where it is referred to as simulators or emulators (e.g., Lantuéjoul, 2002; De Marsily et al., 2005).

Current WGs can be broadly divided into four groups: resampling methods (e.g., Rajagopolan and Lall, 1999; Oriani et al., 2014; Yiou, 2014), Box-Jenkins methodology (e.g., Box and Jenkins, 1976), point process models (e.g., Onof et al., 2000) and hierarchical models. The latter encompass the weather type models which include a discrete variable and multivariate statistical distributions modeling the climatic variables conditional on this discrete variable. This conceptual discrete variable is meant to describe a limited number (typically from 2 to 6) of weather "states", "types" or "regimes". Depending on the problem at hand and depending on the availability of good descriptors of weather patterns, weather states can be considered as observed or latent. They are said to be observed when they are extracted from external variables such as descriptors of large scale synoptic climatological patterns (Bardossy and Plate, 1991, 1992; Wilson et al., 1992). Weather types are considered as latent variables when they are estimated on local variables by means of an *a priori* clustering algorithm (Flecher et al., 2010), or when they are estimated as a hidden variable in the statistical model. Modeling strategies for weather type models will be detailed in Section 2

Quite often, it is possible to relate the latent states to typical weather patterns, a simple example being the straightforward classification in three states corresponding to dry days, days with light rain and days with heavy rain. Even though observed and latent weather states correspond to quite different modeling options, they do constitute a common framework for building stochastic

 ISSN: 2102-6238

weather generators that is today largely prevailing in the literature due to its flexibility and interpretability. This review will focus on this class of models since, based on our experience, it is the most versatile approach for building multisite, multivariate weather generators. In the following we shall use indistinctly “state”, “type” or “regime” to name the latent discrete variable. We will restrict ourselves to daily stochastic generators. Even though subdaily precipitation models are not different to daily ones in essence, subdaily models for variables such as temperature, solar radiation and humidity require a precise modeling of the daily cycle. Adequate models are driven more by physical considerations than by statistical ones, which is beyond the scope of this paper.

Historically, weather generators have been first developed for hydrological application (Gabriel and Neumann, 1962; Todorovic and Woolhiser, 1975). Rainfall occurrences at a single site were described by a two-state Markov chain and their intensities by independent exponential or Gamma random variables, leading to the so-called "chain dependent model" (see also Katz, 1977). In this simple model, weather states correspond to the states of the Markov chain, i.e. to dry and wet states. In a seminal paper Richardson (1981) added the modeling of daily minimum and maximum temperature and solar radiation to the generator in Katz (1977). After removing the seasonal cycle and conditionally on the weather type, residuals of these variables were viewed as a multivariate autoregressive process independent on rainfall amounts. Following those early papers, numerous extensions have been proposed, and they were summarized by two review articles published at the turn of the century. Wilks and Wilby (1999) gave a detailed presentation of Richardson’s model and its extensions, with a discussion of the advantages and drawbacks of these models and some application issues. Srikanthan et al. (2001) provided a quite comprehensive list of models for annual, monthly and daily climate variables at a single site together with some remarks on multisite models. Recently, impact questions with respect to large scale climate changes have spurred a strong interest in linking local and global climate variables, leading the way to the so-called downscaling methods. Maraun et al. (2010) and Wilks (2010, 2012) discussed in detail the strong links between downscaling approaches and WGs, mainly focusing on how to make the connection between circulation patterns and local atmospheric variables at the daily scale. Our review departs from those recent studies by zooming in on multisite weather type models and outlining the principal ideas of the referred papers. Its outline follows the steps required to build a weather type model. In Section 2, different strategies for choosing weather types are discussed. Conditionally on the weather type, statistical models for the weather variables are detailed in Section 3. The last section makes some propositions for future research, regarding the modeling of the weather types, the space-time statistical models for weather variables and the modeling of extreme values in this framework.

2. Modeling weather types

2.1. Defining weather types

As mentioned above, in the early days two weather types were introduced in order to capture the dynamical changes between wet and dry days at a single site. In practice, a day was qualified as wet if the precipitation amount was greater than a chosen level, for instance 0.2 mm of daily total rainfall (Richardson, 1981). Beyond the natural dichotomy between wet and dry events, weather types intend to capture recurrent patterns by breaking spatio-temporal information into a finite

 ISSN: 2102-6238

number of blocks. For example, daily weather patterns over Europe in winter are often linked to the North Atlantic Oscillation (NAO) that consists of two pressure centers in the North Atlantic, one typically located near Iceland, the other one being an area of high pressure over the Azores. Such a configuration can be used to create four weather pattern, classically referred to as NOA+, NOA-, blocking and ridge. By breaking the spatio-temporal information into four blocks, one can assign a weather type for each winter day.

As impact study requirements and datasets at hand moved from one single variable (e.g. precipitation) towards multivariate random vectors (precipitation, temperature, wind, etc.), it was natural to wonder if the definition and the numbers of weather types could benefit from the extended database. Flecher et al. (2010) decomposed the wet and dry contrast into finer nuances. Sub-regimes of wet (respectively dry) days were obtained by running a clustering algorithm on variables such as daily minimum and maximum temperature, radiation and wind speed recorded at the same single site.

Additional large scale information such as pressure fields, synoptic patterns, etc. can also improve the definition of weather types. Any given day can be attached to a specific weather type, or circulation pattern, by running a clustering algorithm on large scale atmospheric variables (e.g. Bogardi et al., 1993; Wilson et al., 1992; Hay et al., 1991; Garavaglia et al., 2010). There is a large variety of possible approaches to classify large scale atmospheric conditions. The most common one is to perform k-means clustering on the first empirical orthogonal functions of geopotential anomaly fields (e.g. Cattiaux et al., 2010). The k-means algorithm is sometimes initialized using hierarchical clustering (e.g. Garavaglia et al., 2010; Guanche et al., 2013). More modern methods have also been implemented, such as fuzzy classification based on mixture models (e.g. Vrac et al., 2007) or simulated annealing optimization (e.g. Bárdossy, 2010; Haberlandt et al., 2014). One advantage of linking weather types with large scale data is that the practitioner can investigate the impact of large scale changes on the weather type distribution. This road has been explored by researchers working on statistical downscaling (e.g. Hughes and Guttorp, 1994; Haberlandt et al., 2014; Wilks, 2012) and sea state condition generators (e.g. Guanche et al., 2013). Time stationarity of this link remains usually a key assumption. Recently, some authors (e.g. Jones et al., 2011) proposed to account for nonstationarity in the context of climate change by re-estimating the parameters of the distributions for future conditions based on corrected observations using a delta change approach with respect to simulations from a regional climate model.

Imposing an *a priori* weather type, although interpretable, may be too restrictive and may not necessarily provide an optimal clustering to capture the stochastic properties of meteorological variables of interest. A natural alternative is to introduce the weather type as a latent variable. Hidden Markov Models (HMM) have been proposed in this context (Zucchini and Guttorp, 1991). With HMMs, the states are optimally fitted to the data given the chosen parametrization. However they may not always have a simple interpretation in terms of weather types. Furthermore, the existence of the hidden variable complicates the statistical inference and, despite recent progress, only relatively simple models can be considered to describe the sequence of weather types and the distribution of the weather variables within weather types. As a consequence, these models may be too simple to reproduce the complexity of the data.

At this stage, one can already figure some issues met by the practitioner, in a nutshell solving the various trade-offs between model complexity, inference efficiency and interpretability. To discuss those points, classical approaches to represent the temporal dynamics among weather

 ISSN: 2102-6238

types (hidden or not) have to be recalled.

2.2. *Temporal models for weather types*

The sequence of weather state is often modeled as a homogeneous first order Markov chain. It leads to simple and interpretable models when the number of states is small. Standard fitting procedures can be used even when the state is introduced as a hidden process (e.g. Zucchini and Guttorp, 1991). However, such an assumption may be too simplistic to catch some important properties of the meteorological data as discussed below.

Meteorological time series are nonstationary with important seasonal and daily components and some possible inter-annual variability. It is usual to treat each season or month independently. This leads to a large number of parameters, which can be a problem for small datasets. To overcome the difficulty of defining limits between seasons, stochastic seasonality with seasons starting at random dates according to a non-homogeneous HHMs have also been studied (Carey-Smith et al., 2014; Sansom et al., 2013). In a Markovian world, the chain can become nonhomogeneous to capture cycles and trends. Transition probabilities can be allowed to depend on time and other covariates via a link function as in the Generalized Linear Model framework (e.g. Katz and Parlange, 1995; Furrer and Katz, 2007; Ailliot and Monbet, 2012). In the same spirit, large scale atmospheric variables or climate indices such as ENSO (El-Niño Southern Oscillation) or NAO may also be introduced in the switching mechanism parameters. It generally improves the description of the inter-annual climate variability and offers a way to link WGs to global climate models (e.g. Hughes and Guttorp, 1994; Hughes et al., 1999; Bellone et al., 2000; Qian et al., 2002; Robertson et al., 2004; Vrac et al., 2007; Zheng and Katz, 2008; Kim et al., 2012).

An other important limitation of first order homogeneous Markov models is that the sojourn time in each weather state is distributed as a geometric random variable, which may not allow reproducing long heat waves or long dry spells (Racsko et al., 1991). For daily rainfall occurrences, working with second or third order Markov chains improves the fit significantly (e.g. Katz and Parlange, 1999; Jimoh and Webster, 1996; Wilks, 1999; Lennartsson et al., 2008; Chen et al., 2012), but this considerably increases the number of parameters. Reduced models may alleviate this issue in some cases (Zucchini and McDonald, 2009). Whenever interpretable and physically realistic constraints can be identified, the risk of overparametrization diminishes in statistical models that can easily allow covariates, e.g. the GLM approach. Semi-Markov models have also been proposed in this context, with sojourn durations in the regimes modeled by parametric (Racsko et al., 1991; Wilby et al., 1998) or semi-empirical distributions (Semenov et al., 1998). Inference for semi-Markov models becomes difficult when the weather type is viewed as a latent variable (e.g. Sansom and Thomson, 2001; Bulla et al., 2010).

Overall, the joint choice of the number of states (hidden or not), the order of the Markov chain and the sojourn time distribution remains a subjective task. In time series analysis it is usual to consider residuals for the task of model selection, but the residuals are not properly defined in mixture models and, although one could construct pseudo-residuals, they are rarely used in the weather generator context. Automatic criteria such as AIC, BIC (e.g. Brockwell and Davis, 2002) can help, but the final choice will also depend on other criteria like interpretability, computing time, robustness and adaptability.

 ISSN: 2102-6238

2.3. *Spatial models for weather types*

In a multisite analysis, one can either view weather types as “local entities" or as a “constant spatial feature". In the latter case, all sites share the same regional weather type at a given day and no spatial model is needed.

In the former case, one has to create local variabilities and spatial dependence in the weather type space while being parsimonious in order to keep a small number of interpretable parameters. This is not an easy task. At each time step, discrete random realizations with a spatio-temporal structure must be drawn at each location to represent the local weather type.

To reach this goal, in Wilks (1998), a collection of single-site chain-dependent models are tied together by drawing correlated random numbers at each time point. The dependence among sites is based on a correlation index obtained from a spatially multivariate Gaussian vector. The weather at a given day is said to be wet at a given site if the corresponding Gaussian coordinate is above a site-dependent threshold. The inference method proposed in (Wilks, 1998) may lead to ill-defined covariance estimates. Alternative inference schemes were discussed in Lee et al. (2010) and Thompson et al. (2007) who reformulated the model as a HMM with the local weather types being dry, light rain or heavy rain.

The idea of censoring a Gaussian vector is mathematically rich because it offers a simple way to generate the space-time evolution of binary variables. For rainfall occurrence modeling, Allard and Bourotte (2014) and Kleiber et al. (2012) followed this approach. Censoring was also used by Khalili et al. (2007, 2009), but a moving average of a white noise with uniform distribution, instead of a Gaussian one, provided spatial dependence. The threshold for censoring can also depend on covariates (Qian et al., 2002). It was also proposed to use the autologistic model (Hughes et al., 1999) to describe the spatial structure of rainfall occurrence.

3. Modeling weather variables conditionally upon weather types

Conditionally upon the weather type, the choice of the distribution describing the meteorological variables of interest is of primary importance and is a complicated task. First, marginal distributions may be hard to model with complex features such as a point mass at the origin for rainfall (corresponding to dry conditions), circular variables (wind direction for example), heavy or bounded tails... Then, the dependence structure among the meteorological variables is generally complex, even within a weather type which corresponds to homogeneous weather conditions. For multisite models, it is also necessary to add the spatial dependence to the model. The family chosen for the joint distribution must be flexible enough to catch such features. Yet, at the same time it must be simple enough in order to yield tractable and interpretable models, especially when the weather type is introduced as a hidden process.

3.1. *Single site models*

Precipitation

Precipitation has always been a key variable of interest in hydrology and climatology, in particular for the first WGs (e.g. Katz, 1977; Richardson, 1981). From a statistical point of view,

 ISSN: 2102-6238

precipitation modeling is complex because it mixes a Bernoulli random variable corresponding to dry or wet events with a positive random variable corresponding to the rainfall intensity, therefore leading to a strong departure from the classical Gaussian framework.

Given the weather type sequence, precipitation amounts have been classically assumed to be conditionally independent in time (e.g. Richardson, 1981; Wilks, 1999). More recent developments propose to take the temporal dependence into account. During wet days, a large class of distributions can be fitted to rainfall amounts. Since the early exponential and Gamma distributions (Todorovic and Woolhiser, 1975; Katz, 1977), researchers have tried to move away from the classical Gamma distribution family for at least two reasons.

A unique Gamma distribution may not be flexible enough to capture all rainfall amount behaviors. Even for non-extreme events, imposing an unique parametric distribution can be viewed as too restrictive, for example at sites where precipitations are heavy-tailed. Alternatives are thus needed to model heavier tails and extreme amounts. A first approach is to simply recognizing that mixtures of conditional Gamma distributions can yield heavier tails than a unique Gamma distribution (Kenabatho et al., 2012). Weather generators based on weather types are thus capable of fitting a relatively large variety of precipitation distribution. As an alternative, mixtures of exponential random variables (Wilks, 1998, and references therein) or semi-parametric distributions (Lennartsson et al., 2008) can also be favored. Gamma distributions, conditional or not, still presents a too light-tailed distribution for very extreme distributions. A second approach is thus to use distributions specifically designed to model extreme values. In Lennartsson et al. (2008), a generalized Pareto distribution (GPD) modeled heavy rainfall above a high level. In Furrer and Katz (2008), a stretched exponential distribution was used as an alternative to the GPD. In Vrac et al. (2007), a dynamic mixture of the Gamma and GPD distributions with a weight depending on the amount of precipitation is proposed.

A second difficulty is the lack of a clear path on how to extend the Gamma distribution to a multivariate and/or spatial setting. This leads to the idea of transforming data into the Gaussian world that offers a simple dependence structure, the covariance matrix. For example, the transfer function can be a power-transform (Katz and Parlange, 1995) or the powered exponential of a truncated Gaussian distribution (Allard and Bourotte, 2014), or indeed any non parametric transform (Chilès and Delfiner, 2012). This is not limited to precipitation, and for instance the square root of the wind intensity is often considered instead of its raw value. Although powerful and flexible, these transformations complicate the assessment of uncertainties and render the interpretability challenging, the measurement unit being lost. Another route consists of modeling the Anscombe residuals, which are very often approximately Gaussian. It is then natural to describe the dependence via a model for their temporal dependence (Chandler and Wheater, 2002; Yang et al., 2005).

Overall, despite all these drawbacks, the Gamma density has still a lot of attractive mathematical properties and remains a strong candidate to capture basic rainfall amount properties at the daily scale, and it should be viewed as an important yardstick.

Dependence between successive rainfall amounts has been modeled by an autoregressive process (Hutchinson, 1995), by a parametric auto-correlation function (Flecher et al., 2010) or a Gaussian copula (Lennartsson et al., 2008). Another extension consists of assuming that the distribution depends on covariates such as the the season Kim et al. (2012).

 ISSN: 2102-6238

Other variables

Besides precipitation, other meteorological variables like minimum and maximum daily temperature, solar radiation, humidity or wind intensity have been generally modeled by a multivariate autoregressive model (Parlange and Katz, 2000). The autoregressive parameters depend on weather types, and thus marginal distributions may not be Gaussian anymore, even for linear models. In most weather generators, precipitation are generated first. Other variables are then simulated conditional on precipitation. This strategy, which seems to be driven by statistical considerations rather by physical reasoning, could be challenged. In Flecher et al. (2010), conditionally on weather types, the multivariate distribution of all variables (thus including precipitation) for two successive days was modeled by using the multivariate closed skew-normal distributions (González-Farías et al., 2004; Gupta et al., 2004). This family of distributions allows a rather flexible modeling of the residual skewness generally observed in climate data. For modeling wind conditions (i.e. wind speed and wind direction), Ailliot et al. (2014) used Markov-switching autoregressive processes. In these processes, weather types define a latent process. An autoregressive model describes wind conditions conditionally to the latent process. Hidden Markov models have also been proposed for wind directions (Holzmann et al., 2006).

3.2. Multisite modeling

If only one single weather type drives a multisite weather generator, a simple multisite modeling strategy is to assume that, given this weather type, the sites are mutually independent in space and time (e.g. Zucchini and Guttorp, 1991; Hughes and Guttorp, 1994; Robertson et al., 2004). In Bellone et al. (2000), an autologistic model was used to describe the spatial structure of rainfall occurrence. Still, rainfall amounts were assumed to be Gamma random variables, conditionally independent in space and time and also independent on the occurrence process. Such assumptions may not be realistic for many datasets, in particular when the network of rainfall stations is dense and when the topography of the area is diverse. Thus, modeling the dependence structure within weather types becomes necessary, but is challenging even when only a unique weather type is considered. Few tractable models for spatial processes exist and Gaussian processes are often considered.

As marginals may not be normally distributed, Gaussian processes cannot be used directly and marginal transformations may be needed. In the literature on multisite WGs with regional weather type, different flavors exist to make the link between a non-Gaussian multivariate random vector and its normally distributed counterpart (e.g. Wilks, 1998; Brissette et al., 2007; Khalili et al., 2007; Thompson et al., 2007; Khalili et al., 2009; Heaps et al., 2015).

For example, Bardossy and Plate (1992) and Ailliot et al. (2009) opted for a censored power-transformed Gaussian distribution for daily rainfall. Negative values of the Gaussian vector correspond to dry days and the power transformation was applied to the positive part of the distribution.

Kleiber et al. (2012) developed a multisite extension of the chain-dependent model where rainfall amount at each site was modeled by Gamma distributions with shape and scale parameters varying according to latent Gaussian fields. In these generators, the spatial structure was described by both the weather type and the covariance of the Gaussian vector.

 ISSN: 2102-6238

4. Trends and challenges

Comparing the various models presented above is a difficult if not impossible task since they have been validated on different datasets and for different scientific purpose: downscaling, prediction or simulation. Building a shared framework of datasets and the associated statistical tools for comparing the different weather generators in various contexts remains a hurdle but is an absolutely necessary step. In this context, it might be useful to refer to the European Union VALUE network (http://www.value-cost.eu/), an example of such a validation initiative.

Despite the recent progress on WGs, as already pointed out, there is still a strong research needed to build multisite, multivariate generators that can accurately capture the observed temporal and spatial coherence in meteorological data and the interrelationships among different weather variables.

Concerning weather types, clustering schemes could be refined to catch additional features of the dependence structure and simplify the modeling of the residual dependence inside the regimes. For example, the comparison of the numerical results obtained on the same data set in Thompson et al. (2007) and Ailliot et al. (2009) indicates that a local weather type models improves local persistence of rainfall occurrence, whereas a regional weather type models gives better results for the spatial distribution of rainfall. This suggests including local weather states within regional weather states. More generally, hierarchical models with several layers of weather types corresponding to different space-time scales or kinds of dependence between different weather variables could be further investigated. To close this paragraph on weather types, one could also challenge the definition of weather types on dry and wet days. Physically, rainfall behavior is rather a consequence of other variables (winds, pressures, temperatures, etc.) than a cause of those variables. Hence, conditioning on dry and wet days may be statistically convenient, but this classical modeling approach could miss important physical links among atmospheric variables. Koch and Naveau (2015) investigated the impact of regional covariates such as humidity and temperature to improve variability in simulated hourly multisite rainfall in northern Brittany.

Another promising avenue to represent dependencies in precipitation modeling is to take advantage of recent advances in copula modeling (e.g. Bárdossy and Pegram, 2009; Serinaldi, 2009; Serinaldi and Kilsby, 2014). Bayesian hierarchical modeling can also offer a flexible framework to integrate different layers of complexity. For example, Fuentes et al. (2008) merged different types of data, rainfall measurements and radar outputs, by assuming hidden processes that drive the spatio-temporal dynamics.

A long-standing open problem concerns the reproduction of extremes by WGs. Extremes can be observed in the intensity of the considered variables but also in the duration of certain events types like long heat waves. WGs with Markovian structure are not able to reproduce exceptionally long sojourns in a weather type, and other modeling approaches have to be considered. Models with nonhomogeneous transitions between the regimes could be investigated since they imply a more flexible dynamical structure and since they can be inferred quite easily (e.g. Ailliot et al., 2014).

Concerning the joint modeling of multivariate extremes, especially for heavy rainfalls, a strong research effort has been undertaken by the Extreme Value Theory community these last decades. Complex models based on max-stable processes exist and have been used to analyze extreme rainfall (e.g. Davison et al., 2012; Thibaud et al., 2013; Bernard et al., 2013). Still, it is not clear

 ISSN: 2102-6238

on how to make the link between the upper tail behavior, either represented by block maxima or excesses above some high threshold, and the bulk of the multivariate distribution. WGs aim at reproducing the full range of observed atmospheric variables. This challenge is open, and a joint effort between statisticians and climatologists is clearly needed here.

Acknowledgments

Two international workshops devoted to stochastic weather generators were organized by the authors of this paper. The first one took place May 2012 in Roscoff. The second one was held in Avignon in September 2014. We would like to thank all participants that have motivated us to write this review.

Part of Philippe Naveau's work has been supported by several projects: INSU-LEFE MULTI-RISK, ANR MOPERA, ANR DADA and ExtremoScope. The authors are very grateful to Richard Chandler (Imperial College), Thomas Opitz (BioSP, INRA) and one anonymous reviewer for their careful reading of the paper and their comments that lead to significant improvements of the initial draft.

References

Ailliot, P., Bessac, J., Monbet, V., and Pène, F. (2014). Non-homogeneous hidden Markov-switching models for wind time series. *Journal of Statistical Planning and Inference*.

Ailliot, P. and Monbet, V. (2012). Markov-switching autoregressive models for wind time series. *Environmental Modelling and Software*, 30:92–101.

Ailliot, P., Thompson, C., and Thomson, P. (2009). Space time modeling of precipitation using a hidden Markov model and censored Gaussian distributions. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 58(3):405–426.

Allard, D. and Bourotte, M. (2014). Disaggregating daily precipitations into hourly values with a transformed censored latent Gaussian process. *preprint*.

Bárdossy, A. (2010). Atmospheric circulation pattern classification for south-west germany using hydrological variables. *Physics and Chemistry of the Earth, Parts A/B/C*, 35(9):498–506.

Bárdossy, A. and Pegram, G. G. S. (2009). Copula based multisite model for daily precipitation simulation. *Hydrology and Earth System Sciences*, 13:2299–2314.

Bardossy, A. and Plate, E. (1991). Modeling daily rainfall using semi-Markov representation of circulation pattern occurrence. *Journal of hydrology*, 122:33–47.

Bardossy, A. and Plate, E. (1992). Space-time model for daily rainfall using atmospheric circulation patterns. *Water Resources Research*, 28(5):1247–1260.

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

Bernard, E., Naveau, P., Vrac, M., and Mestre, O. (2013). Clustering of maxima: Spatial dependencies among heavy rainfall in France. *Journal of Climate*, 26(20):7929–7937.

Bogardi, I., Matyasovsky, I., Bardossy, A., and Duckstein, L. (1993). Application of a space-time stochastic model for daily precipitation using atmospheric circulation patterns. *Journal of Geophysical Research*, 98(D9):16653–16667.

Box, G. and Jenkins, G. (1976). *Time series analysis, forecasting and control (revised edn.)*. Holden-Day, San Francisco.

Brissette, F., Khalili, M., and Leconte, R. (2007). Efficient stochastic generation of multi-site synthetic precipitation data. *Journal of Hydrology*, 345(3):121–133.

Brockwell, P. and Davis, R. (2002). *Introduction to Time Series and Forecasting, second edition*. Springer-Verlag, New York.

Bulla, J., Bulla, I., and Nenadig, O. (2010). hsmm - an R package for analyzing hidden semi-Markov models. *Computational Statistics and Data Analysis*, 54(3):611–619.

 ISSN: 2102-6238

Carey-Smith, T., Sansom, J., and Thomson, P. (2014). A hidden seasonal switching model for multisite daily rainfall. *Water Resources Research*, 50(1):257–272.

Cattiaux, J., Vautard, R., Cassou, C., Yiou, P., Masson-Delmotte, V., and Codron, F. (2010). Winter 2010 in Europe: a cold extreme in a warming climate. *Geophysical Research Letters*, 37(20).

Chandler, R. E. and Wheater, H. (2002). Analysis of rainfall variability using generalized linear models: A case study from the west of Ireland. *Water Resources Ressearch*, 38(10):1192.

Chen, J., Brissette, F., and Leconte, R. (2012). WeaGETS–a Matlab-based daily scale weather generator for generating precipitation and temperature. *Procedia Environmental Sciences*, 13:2222–2235.

Chilès, J. and Delfiner, P. (2012). *Geostatistics: Modeling Spatial Uncertainty, Second Edition*. Wiley.

Davison, A. C., Padoan, S. A., and Ribatet, M. (2012). Statistical modeling of spatial extremes. *Statistical Science*, 27:161–186.

De Marsily, G., Delay, F., Gonçalvès, J., Renard, P., Teles, V., and Violette, S. (2005). Dealing with spatial heterogeneity. *Hydrogeology Journal*, 13(1):161–183.

Flecher, C., Naveau, P., Allard, D., and Brisson, N. (2010). A stochastic daily weather generator for skewed data. *Water Resources Research*, 46:W07519.

Fuentes, M., Reich, B., and Lee, G. (2008). Spatial-temporal mesoscale modeling of rainfall intensity using gage and radar data. *Annals of Applied Statistics*, 2(4):1148–1169.

Furrer, E. and Katz, R. (2008). Improving the simulation of extreme precipitation events by stochastic weather generators. *Water Resources Research*, 44(12).

Furrer, E. M. and Katz, R. W. (2007). Generalized linear modeling approach to stochastic weather generators. *Climate Research*, 34:129–144.

Gabriel, K. and Neumann, J. (1962). A Markov chain model for rainfall occurrence at Tel-Aviv. *Quart. J. R. met. Soc.*, 88:90–95.

Garavaglia, F., Gailhard, J., Paquet, E., Lang, M., Garçon, R., Bernardara, P., et al. (2010). Introducing a rainfall compound distribution model based on weather patterns sub-sampling. *Hydrology and Earth System Sciences Discussions*, 14.

González-Farías, G., Domínguez-Molina, J. A., and Gupta, A. (2004). The closed skew-normal distribution. In *Skew-elliptical distributions and their applications: a journey beyond normality*, pages 25–42. Chapman & Hall/CRC, Boca Raton, FL.

Guanche, Y., Mínguez, R., and Méndez, F. (2013). Climate-based Monte Carlo simulation of trivariate sea states. *Coastal Engineering*, 80:107–121.

Gupta, A. K., González-Farías, G., and Domínguez-Molina, J. (2004). A multivariate skew-normal distribution. *Journal of Multivariate Analysis*, 89(1):181–190.

Haberlandt, U., Belli, A., and Bárdossy, A. (2014). Statistical downscaling of precipitation using a stochastic rainfall model conditioned on circulation patterns–an evaluation of assumptions. *International Journal of Climatology*.

Hay, L., McCabe, G., Wolock, D., and Ayers, M. (1991). Simulation of precipitation by weather-type analysis. *Water Resources Research*, 27:493–501.

Heaps, S., Boys, R., and Farrow, M. (2015). Bayesian modelling of rainfall data by using non-homogeneous hidden Markov models and latent Gaussian variables. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*.

Holzmann, H., Munk, A., Suster, M., and Zucchini, W. (2006). Hidden Markov models for circular and linear circular time series. *Environmental and Ecological Statistics*, 13:325–347.

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

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

Hutchinson, M. (1995). Stochastic space-time weather models from ground-based data. *Agricultural and Forest Meteorology*, 73(3):237–264.

Jimoh, O. and Webster, P. (1996). Optimum order of Markov chain for daily rainfall in Nigeria. *Journal of hydrology*, 222:1–17.

Jones, P., Harpham, C., Goodess, C., and Kilsby, C. (2011). Perturbing a weather generator using change factors derived from regional climate model simulations. *Nonlinear Processes in Geophysics*, 18(4):503–511.

Katz, R. (1977). Precipitation as a chain-dependant process. *Journal of Applied Meteorology*, 16:671–676.

Katz, R. and Parlange, M. (1995). Generalizations of chain-dependent processes: Application to hourly precipitation. *Water Resources Research*, 31(5):1331–1341.

 ISSN: 2102-6238

Katz, R. and Parlange, M. (1999). Overdispersion phenomenon in stochastic modeling of precipitation. *Journal of Climate*, 11:591–601.

Kenabatho, P., McIntyre, N., Chandler, R., and Wheater, H. (2012). Stochastic simulation of rainfall in the semi-arid limpopo basin, botswana. *International Journal of Climatology*, (32):1113–1127.

Khalili, M., Brissette, F., and Leconte, R. (2009). Stochastic multi-site generation of daily weather data. *Stochastic Environmental Research and Risk Assessment*, 23(6):837–849.

Khalili, M., Leconte, R., and Brissette, F. (2007). Stochastic multi-site generation of daily precipitation data using spatial autocorrelation. *Journal of Hydrometeorology*, 8(3):396–412.

Kim, Y., Katz, R. W., Rajagopalan, B., Podestá, G. P., and Furrer, E. M. (2012). Reducing overdispersion in stochastic weather generators using a generalized linear modeling approach. *Climate research*, 53:13–24.

Kleiber, W., Katz, R., and Rajagopalan, B. (2012). Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes. *Water Ressources Research*, 48:W01523.

Koch, E. and Naveau, P. (2015). A frailty-contagion model for multi-site hourly precipitation driven by atmospheric covariates. *Water Resources Research*.

Kolokotroni, M., Ren, X., Davies, M., and Mavrogianni, A. (2012). London's urban heat island: Impact on current and future energy consumption in office buildings. *Energy and buildings*, 47:302–311.

Lantuéjoul, C. (2002). *Geostatistical simulation: models and algorithms*. Springer Verlag.

Launay, M., Brisson, N., Beaudoin, N., and Mary, B. (2009). *Conceptual basis, formalisations and parameterization of the STICS crop model*. Editions Quae.

Lee, D., An, H., Lee, Y., Lee, J., Lee, H., and Oh, H. (2010). Improved multisite stochastic weather generation with applications to historical data in South Korea. *Asia-Pacific Journal of Atmospheric Sciences*, 46(4):497–504.

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

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

Onof, C., Chandler, R., Kakou, A., Northrop, P., Wheater, H., and Isham, V. (2000). Rainfall modelling using Poisson-cluster processes: a review of developments. *Stochastic Environmental Research and Risk Assessment*, 14:384–411.

Oriani, F., Straubhaar, J., Renard, P., and Mariethoz, G. (2014). Simulation of rainfall time-series from different climatic regions using the direct sampling technique. *Hydrology and Earth System Sciences Discussion*, 11:3213–3247.

Parlange, M. and Katz, R. (2000). An extended version of the Richardson model for simulating daily weather variables. *Journal of Applied Meteorology*, 39:610–622.

Qian, B., Corte-Real, J., and Xu, H. (2002). Multisite stochastic weather models for impact studies. *International Journal of climatology*, 22(11):1377–1397.

Racsko, P., Szeidl, L., and Semenov, M. (1991). A serial approach to local stochastic weather models. *Ecological Modelling*, 57:27?41.

Rajagopolan, B. and Lall, U. (1999). A k-nearest neighbour simulator for daily precipitation and other variables. *Water ressources research*, 35(10):3089–3101.

Richardson, C. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, 17(1):182–190.

Robertson, A., Kirshner, S., and Smyth, P. (2004). Downscaling of daily rainfall occurrence over northeast Brazil using a hidden Markov model. *Journal of climate*, 17(22):4407–4424.

Sansom, J. and Thomson, P. (2001). Fitting hidden semi-Markov models to breakpoint rainfall data. *Journal of Applied Probability*, 38:142–157.

Sansom, J., Thomson, P., and Carey-Smith, T. (2013). Stochastic seasonality of rainfall in New Zealand. *Journal of Geophysical Research: Atmospheres*, 118(10):3944–3955.

Semenov, A., Brooks, R., Barrow, E., and Richardson, C. (1998). Comparison of the wgen and lars-wg stochastic weather generators for diverse climates. *Climate Research*, 10:95–107.

Serinaldi, F. (2009). A multisite daily rainfall generator driven by bivariate copula-based mixed distribution. *Journal of Geophysical Research: Atmospheres*, 114(10), D10103.

Serinaldi, F. and Kilsby, C. (2014). Simulating daily rainfall fields over large areas for collective risk estimation. *Journal of Hydrology*, 512:285–302.

Srikanthan, R., McMahon, T., et al. (2001). Stochastic generation of annual, monthly and daily climate data: A review.

 ISSN: 2102-6238

Hydrology and Earth System Sciences Discussions, 5(4):653–670.
Thibaud, E., Mutzner, R., and Davison, A. (2013). Threshold modeling of extreme spatial rainfall. *Water Resources Research*, 49(8):4633–4644.
Thompson, C., Thomson, P., and Zheng, X. (2007). Fitting a multisite rainfall model to New Zealand data. *Journal of Hydrology*, 340:25–39.
Todorovic, P. and Woolhiser, D. A. (1975). A stochastic model of n-day precipitation. *Journal of Applied Meteorology*, 14:17–24.
Vrac, M., Stein, M., and Hayhoe, K. (2007). Statistical downscaling of precipitation through non homogeneous stochastic weather typing. *Climate Research*, 34:169–184.
Wilby, R., Wigley, T., Conway, D., Jones, P., Hewitson, B., Main, J., and Wilks, D. (1998). Statistical downscaling of general circulation model output: a comparison of methods. *Water Resources Research*, 34:2995–3008.
Wilks, D. (1998). Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology*, 210(1):178–191.
Wilks, D. (1999). Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agricultural and Forest Meteorology*, 93:153–169.
Wilks, D. (2010). Use of stochastic weather generators for precipitation downscaling. *Wiley Interdisciplinary Reviews: Climate Change*, 1(6):898–907.
Wilks, D. (2012). Stochastic weather generators for climate-change downscaling, part ii: multivariable and spatially coherent multisite downscaling. *Wiley Interdisciplinary Reviews: Climate Change*, 3(3):267–278.
Wilks, D. and Wilby, R. (1999). The weather generation game: a review of stochastic weather models. *Progress in Physical Geography*, 23(3):329–357.
Wilson, L., Lettenmaier, D., and Skyllingstad, E. (1992). A hierarchial stochastic model of large-scale atmospheric circulation patterns and multiple station daily precipitation. *Journal of Geophysical Research*, 97(D3):2791–2809.
Yang, C., Chandler, R., Isham, V., and Wheater, H. (2005). Spatial-temporal rainfall simulation using generalized linear models. *Water ressources research*, 41(11):W11415.
Yiou, P. (2014). Anawege: a weather generator based on analogues of atmospheric circulation. *Geoscientific Model Development*, 7(2):531–543.
Zheng, X. and Katz, R. (2008). Mixture model of generalized chain-dependent processes and its application to simulation of interannual variability of daily rainfall. *Journal of Hydrology*, 349(1):191–199.
Zucchini, W. and Guttorp, P. (1991). A hidden Markov model for space-time precipitation. *Water Resources Research*, 27:1917–1923.
Zucchini, W. and McDonald, I. (2009). *Hidden Markov models for time series : an introduction using R*. Chapman & Hall/CRC, London.

 ISSN: 2102-6238