Electric Power Systems Research 152 (2017) 9–17

Contents lists available at ScienceDirect

Electric Power Systems Research

journal homepage: www.elsevier.com/locate/epsr

A periodic spatial vine copula model for multi-site streamflow simulation

Guilherme A.A. Pereira [a,*], Álvaro Veiga [a], Tobias Erhardt [b], Claudia Czado [b]

[a] Pontifical Catholic University of Rio de Janeiro, Department of Electrical Engineering, Rua Marquês de São, Vicente 225, 22451-900 Rio de Janeiro, Brazil
[b] Technische Universität München, Department of Mathematics, Garching b. München, Germany

ARTICLE INFO

Article history:
Received 15 March 2017
Received in revised form 5 June 2017
Accepted 21 June 2017

Keywords:
Periodic spatial vine copula model
Multi-site stochastic streamflow simulation
Energy modelling
Energy simulation
Uncertainty modelling

ABSTRACT

Operational strategies of the Brazilian Electric Sector have direct impacts on operating costs, energy prices, planning the expansion of the system, etc. These decisions are taken under a high level of uncertainty, as the future availability of water for energy generation is a stochastic variable. Computational models, routinely based on stochastic optimization, support these decisions. Some of them make use of streamflow scenarios as entries. In this way, the aim of this paper is to develop a sophisticated statistical model for multi-site stochastic streamflow simulation. Our approach is based on the extension of vine copulas for high dimensional spatial applications. The proposed model copes with both temporal and spatial dependencies of streamflows. At the same time, it can simulate numerous sites concurrently. We tested our approach on streamflow data from 39 Brazilian hydroelectric power plants. The results indicate that the model can simulate streamflow scenarios largely preserving the features observed in the recorded streamflow data.

1. Introduction

An electric power system constantly operates under a high level of uncertainty deriving from many of its components. Dealing with this uncertainty is crucial for reliable operation of the system. There are many approaches to represent the uncertainty, such as possibilistic approach, stochastic optimization, information gap decision theory, interval analysis, and robust optimization [1–3]. All of these techniques have their role in the system and are employed to handle uncertainty. For a general overview, readers are advised to refer to [1].

Particularly in the case of the Brazilian Energy Sector (BES), where energy is predominantly produced by hydroelectric power plants (HPPs), streamflow scenarios are fundamental to accurate operations and planning of the sector.

These scenarios represent the uncertainty related to hydrological regimes and are extremely valuable for stochastic optimization models that determine the hydrothermal dispatch. A general overview of stochastic optimization models and its applications in power systems can be found in [2,4].

The accuracy of a simulation model depends on its ability to generate synthetic streamflow data, preserving certain statistical characteristics, as well as temporal and spatial dependence structures observed in historical data. Hence, a sophisticated model for multi-site streamflow simulation must account for all of these features. Further, particularly in the BES where the number of HPPs is huge, the model must account for the dimensionality of the problem.

Traditionally, multivariate time series models based on the Box & Jenkins family (ARMA model and its variants) are employed for this task. See, for example, [4–6]. According to [7], they are built under some rigid assumptions about the form of dependence between flow variables or the underlying marginal or joint distributions. Ref. [7,8] point out some drawbacks of these formulations such as: (a) Streamflow data usually are not normally distributed. This suggests that the Gaussian assumption implicit in the model structure may not be adequate. Moreover, such hypothesis limits the capability of the model in representing non-standard probability density forms. (b) These models cannot capture non-linear dependencies structures. (c) The support of the Gaussian distribution is in the whole set of real numbers. Therefore, simulated scenarios based on Gaussian models may not be realistic, as the scenarios will be in the range $(-\infty, \infty)$.

Recently, the concept of copulas has been used for stochastic simulation of streamflow scenarios [8–12]. One advantage of copulas is their capacity to model any sort of association between random variables. Also, it separates the marginal behaviour of each

* Corresponding author.
E-mail addresses: gaap@ele.puc-rio.br (G.A.A. Pereira), alvf@ele.puc-rio.br (Á. Veiga), tobias.erhardt@tum.de (T. Erhardt), cczado@ma.tum.de (C. Czado).

http://dx.doi.org/10.1016/j.epsr.2017.06.017

random variable from the dependence structure that governs the relationship between variables.

Because of the dimensionality of the BES, the construction of a model that simultaneously considers several HPPs is vital. In such high-dimensional settings, vine copulas are a suitable class of multivariate dependence models. In contrast to ordinary copulas, which in higher dimensions are very restrictive in terms of model flexibility, vine copulas allow to model by far more flexible dependence structures. As vine copulas are composed of a set of bivariate copulas, they are able to model asymmetric dependence structures and tail dependencies. However, for vine copulas, the number of parameters grows quadratically, as the number of variables increases. Hence, such a model can be computationally costly for applications with a large number of variables. The extension of vine copulas for spatial applications introduced by [13] allows to reduce the number of parameters by exploiting available spatial information (exogenous variables).

The aim of this research is to propose a periodic model for stochastic streamflow simulation based on the spatial vine copula approach of [13]. We exploit the relationship between the distance of different locations and the river network with the vine copula parameters. Moreover, the proposed method allows to model periodic differences in the association between HPPs commonly observed in monthly data. The proposed model is applied to monthly streamflow data of 39 Brazilian HPPs. The results indicate that the proposed model is capable of simulating streamflow scenarios that replicate the main features of the historical streamflow time series.

This paper is organized as follows. Section 2 gives a brief introduction to copulas in general as well as to the special class of R-vine copulas. Section 3 describes the developed model, its components, and the data set to which it was applied. Section 4 contains the results and Section 5 presents our conclusions.

2. Copula models

2.1. Copulas

A copula C is d-variate distribution function on $[0, 1]^d$ with all margins being standard uniform distributions. It can be understood as a function that connects marginal distributions $(F_1, \ldots, F_d)$ to form a joint distribution F. The copula C associated with joint distribution F is a distribution function C: $[0, 1]^d \to [0, 1]$ such that, for all $(x^1, \ldots, x^d) \in \mathfrak{R}^d$ it holds that

$$F(x^1, \ldots, x^d) = C(F_1(x^1), \ldots, F_d(x^d)). \quad (1)$$

If the margins $F_1, \ldots, F_d$ are continuous, then the copula C is unique. This theorem [14] says that a multivariate distribution function is a composition of a set of marginal distributions and a copula, which condenses all information about the dependence structure of the random vector. The theorem (Eq. (1)) can also be stated in terms of densities,

$$f(x^1, \ldots, x^d) = c(F_1(x^1), \ldots, F_d(x^d))\prod_{i=1}^{d} f_i(x^i), \quad (2)$$

where c is a d-dimensional copula density obtained by partial differentiation of the copula C.

2.2. Vine copulas

Vine copulas (also known as pair-copula constructions) are beneficial in situations where the number of variables d is high. They are based on the principle of rewriting a multivariate probability density function as a product of $d(d-1)/2$ bivariate (pair-) copula densities and d univariate marginal densities (see [15]). In contrast to the well-known ordinary parametric copula families, they represent a powerful tool to build flexible high dimensional copulas.

The construction methodology behind vine copulas was first considered by [16] and revisited by [17,18] who proposed a graphical representation of the pair-copula decomposition. Later, Ref. [15] addressed statistical inference of vine copulas, comprising estimation and simulation.

The graph structure, that organizes the pair-copula construction, is called R-Vine ([17,18]). It is a nested set of trees $\mathcal{V} = (T_1, \ldots, T_{d-1})$ that satisfies the following rules (see [19,20]):

- T_1 is a tree with node set $N_1 = \{1, \ldots, d\}$ and edge set E_1.
- T_ℓ is a tree with node set $N_\ell = E_{\ell-1}$ and edge set E_ℓ ($\ell = 2, \ldots, d-1$).
- For an edge $\{a, b\} \in E_\ell$ ($\ell = 2, \ldots, d-1$) it must hold that the corresponding edges a, $b \in E_{\ell-1}$ share a common node (*proximity condition*).

A particular notation is often used for labelling the edges and nodes of a vine tree sequence (see [21]). An edge $e \in E_\ell$, $\ell = 1, \ldots, d-1$, is denoted by $\{i(e), j(e); D_e\}$. It depends on two edges $a := \{i(a), j(a); D_a\}$ and $b := \{i(b), j(b); D_b\}$ in $T_{\ell-1}$, that share a common node. These edges labels will represent the indices used for the conditional copula densities.

The elements $i(e)$ and $j(e)$ make up the conditioned set ($C_e = \{i(e), j(e)\}$) and are defined as $i(e) := \min\{k : k \in (\mathcal{A}(a) \cup \mathcal{A}(b)) \backslash D_e\}$ and $j(e) := \max\{k : k \in (\mathcal{A}(a) \cup \mathcal{A}(b)) \backslash D_e\}$. The conditioning set D_e is denoted by $D_e = \{\mathcal{A}(a) \cap \mathcal{A}(b)\}$ where $\mathcal{A}(a) = \{i(a), j(a), D_a\}$ and $\mathcal{A}(b) = \{i(b), j(b), D_b\}$. In the first tree ($T_1$), the conditioning set is empty. Fig. 1 shows a five dimensional R-vine tree sequence ($\mathcal{V} = (T_1, \ldots, T_4)$). To illustrate the notation introduced, consider the unique edge $e \in T_4$. It depends on the edges $a = 13;24$ and $b = 45;23$ in T_3. For these edges, $\mathcal{A}(a) = \{1, 2, 3, 4\}$ and $\mathcal{A}(b) = \{2, 3, 4, 5\}$. Therefore, the conditioning set of the edge e is $D_e = \{2, 3, 4\}$, $i(e) = \{1\}$, and $j(e) = \{5\}$, as can be seen in Fig. 1.

To specify a vine copula distribution of some random vector $\mathbf{U} = (U^1, \ldots, U^d)$ with $U^1, \ldots, U^d \sim \mathcal{U}(0, 1)$, we need to assign bivariate copulas densities to each edge in the vine tree. We define the bivariate copula set $\mathcal{B} = \{C_{i(e),j(e);D_e} : e \in E_\ell, \ell = 1, \ldots, d-1\}$ where $C_{i(e),j(e);D_e}$ is a bivariate copula with density and E_ℓ are the edges of the R-vine tree sequence. Then, a vine copula distribution of $\mathbf{U}$ is specified through bivariate copula densities $c_{i(e),j(e);D_e}$, corresponding to the bivariate copulas $C_{i(e),j(e);D_e} \in \mathcal{B}$, associated to the edges $\mathbf{E} = \{E_1, \ldots, E_{d-1}\}$ of a regular vine tree structure $\mathcal{V}$. In addition, following [13], we define a set $\mathbf{u}^{\mathcal{I}} := \{u^k : k \in \mathcal{I}\}$ where $\mathcal{I}$ is a subset of the index set $\{1, \ldots, d\}$. Accordingly, the corresponding vine copula density of $\mathbf{U}$ can be written as

$$\begin{aligned} c_{1,\ldots,d}(\mathbf{u}) &= \prod_{\ell=1}^{d-1}\prod_{e \in E_\ell} c_{i(e),j(e);D_e} \\ &\quad \times \{C_{i(e)|D_e}(u^{i(e)}|\mathbf{u}^{D_e}), C_{j(e)|D_e}(u^{j(e)}|\mathbf{u}^{D_e})\}. \end{aligned} \quad (3)$$

To evaluate the density, [16] has shown that the conditional distributions $C_{i(e)|D_e}(u^{i(e)}|\mathbf{u}^{D_e})$ and $C_{j(e)|D_e}(u^{j(e)}|\mathbf{u}^{D_e})$ can be calculated as

$$C_{k|\mathcal{J}}(u^k|\mathbf{u}^{\mathcal{J}}) = \frac{\partial C_{kl;\mathcal{J}_{-l}}\{C_{k|\mathcal{J}_{-l}}(u^k|\mathbf{u}^{\mathcal{J}_{-l}}), C_{l|\mathcal{J}_{-l}}(u^l|\mathbf{u}^{\mathcal{J}_{-l}})\}}{C_{l|\mathcal{J}_{-l}}(u^l|\mathbf{u}^{\mathcal{J}_{-l}})}, \quad (4)$$

where $k, l \in 1, \ldots, d$, $k \neq l$, $\{l\} \subset \mathcal{J} \subset \{1, \ldots, d\} \backslash \{k\}$ and $\mathcal{J}_{-l} := \mathcal{J} \backslash \{l\}$.

As outlined above, a vine copula is composed of $d(d-1)/2$ pair-copulas, which means that a considerable number of parameters has to be estimated. [13] introduced an extension of vine copulas for spatial dependence modelling. In their model, they take advantage of the relationship between spatial variables and the vine copula parameters to re-parametrize the vine copula. This can radically reduce the complexity, as it reduces the number of parameters.