Renewable and Sustainable Energy Reviews 94 (2018) 197–213

Contents lists available at ScienceDirect

Renewable and Sustainable Energy Reviews

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

High dimensional dependence in power systems: A review

Edgar Nuño Martinez*, Nicolaos Cutululis, Poul Sørensen

Department of Wind Energy, Technical University of Denmark, Roskilde, Denmark

ARTICLE INFO

Keywords:
Power systems
Multivariate dependence
Wind energy
Solar energy

ABSTRACT

Weather-driven renewable generation is characterized by being uncertain and geographically dependent. In this regard, the recent deployment of wind and solar power has had a significant impact on the operation and planning of modern electricity grids; justifying the need to model high dimensional dependence. It is a relevant topic which is starting to have a significant importance in power systems. This paper presents a general overview on different multivariate dependence modeling techniques, namely parametric, non-parametric and copula functions. In addition, approximated methods based on limited information e.g. some statistical measures or a predefined dependence structure are presented. Autoregressive moving average (ARMA) and Markov models are discussed as general frameworks to reproduce spatio-temporal processes. Moreover, different applications in power systems are discussed in detail, along with a case study exemplifying the importance of a correct dependence modeling of wind generation.

1. Introduction

The world's energy mix has undergone a structural transformation over the past few years. Environmental concerns regarding climate change along with favorable financial incentives have been the breeding ground for the development of renewable energy sources (RES) as an alternative to conventional power generation technologies. Weather-driven RES, namely wind and solar photovoltaic (PV) generation have become a success story and today cover a significant portion of electricity consumption around the world. Fig. 1 presents the countries with the highest combined penetration of RES as a percentage of the net generation. The average penetration in Europe reached approximately 8% in 2012, although some countries such as Denmark, Portugal and Spain presented penetration levels well above 20%. Note that these figures have increased significantly; e.g. 42% of the total electricity consumption in Denmark in 2015 was provided by wind [33]. The primary energy sources in RES, wind and solar irradiation, are stochastic and non-controllable in terms of availability. Hence, RES generation needs to be predicted in advance for an efficient power system operation. Furthermore, RES are driven by a common large-scale meteorological process, which translates into a significant spatial dependence between individual generators located far apart [143,76]. On the other hand, RES generation can also bring significant benefits to the power system. For example, offering faster power control than most other generation technologies, wind and solar PV may add an extra degree of flexibility to accommodate small deviations in the demand as well as contingencies in the grid. Even though the participation of wind power plants in ancillary services does not represent an extended practice, it is already established as a requirement in some specific power systems e.g. [34]. In addition, there is a strong evidence in favor of RES generation pushing wholesale electricity prices down; known as the *merit-of-order* effect [118]. RES present virtually zero operating costs. Therefore, for a constant demand, more expensive generators will be pushed out of the pool as the volume of RES energy increases. This reduction can even exceed the costs added to the consumer side in the form of green support schemes; consequently bringing economic benefits to society.

RES deployment has dramatically changed how power systems are understood today. Traditionally, demand was seen as the main source of uncertainty. It is essentially determined by human patterns and, when aggregated, it is relatively easy to predict. However, RES are stochastic in nature and do significantly increase the level of uncertainty in the system operation and planning. The impact of variable sources in the system can be assessed either (i) using historical data or (ii) relying on simulated data. The first approach is straightforward, although the length and resolution of currently available data sets are relatively limited. In addition, the investigations are only restricted to a reduced set of observations. On the other hand, simulation has become a robust tool used to handle uncertainties. Thanks to the increase in computer power over the past decades, Monte Carlo (MC)-based techniques allow approximating complex analytical problems in an easy and reliable manner. Moreover, power system operation can be reproduced for longer times and the full space of possible realizations can be explored, e.g. including high-impact low-probability (HILP) events.

* Corresponding author.
E-mail address: ednu@dtu.dk (E. Nuño Martinez).

https://doi.org/10.1016/j.rser.2018.05.056
Received 2 March 2017; Received in revised form 18 May 2018; Accepted 21 May 2018

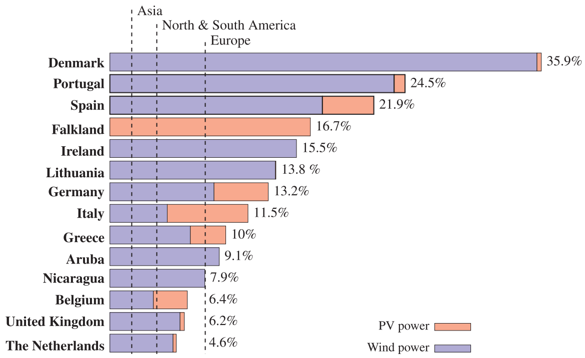

Fig. 1. Cumulated penetration of renewable energy sources with respect to the total net electricity generation in 2012. Source: EIA.

Traditional simulation procedures rely on all quantities of interest or random variables being either fully independent or fully correlated. This can indeed lead to severe inaccuracies when applied to dependent variables. Therefore, there is a need to investigate methods to represent RES in a realistic fashion, capturing the spatial and spatio-temporal dependence observed in real data. As opposed to conventional power plants, renewable generators are characterized by having a small nominal capacity and by exhibiting a high geographical dispersion. Ideally, they could be aggregated per country or per region in order to reduce the mathematical complexity. However, some power system applications require the use of nodal data. Thus, the representation of RES would likely need to be extended to a large dimensional number.

Multivariate dependence is currently becoming a relevant topic in power systems. Nevertheless, methods are usually applied on a per-case basis. The objective of this paper is twofold. First, to make a thorough review of some popular methods to model multidimensional dependence. Second, to exemplify their role in the integration of RES in the power system. The rest of the paper is organized as follows: Section 2 introduces popular methods to model spatial dependence; spatio-temporal dependence is covered in Section 3; their current state-of-the-art in the power system literature is summarized in Section 4; Section 5 presents a simple case study comparing different high dimensional dependence methods; followed by a critical discussion of the review in Section 6, along with some concluding remarks.

2. Overview of dependence modeling methods

The underlying dependence structure between two or more random variables can only be fully described through their joint probability distribution function. Let $\boldsymbol{X} = [X_1,...,X_d]^{\mathrm{T}}$ be a $d \times 1$ random vector. When all the random variables are mutually independent, i.e. $X_i \perp X_j,\ \forall\, X_i, X_j \in \boldsymbol{X}$, their joint probability can be expressed as:

$$\mathbb{P}(\boldsymbol{X}) = \prod_{i=1}^{d} \mathbb{P}(X_i) \tag{1}$$

Not satisfying Eq. (1) implies that there is any form of relationship. In practice, this means that the ith variable can be predicted conditioned on the realization of the jth variable, i.e. $\mathbb{P}(X_i) = \mathbb{P}(X_i|X_j = x_j)$. There are different metrics to quantify the level of dependence between two r.v.'s such as Pearson's rho ρ, Spearman's rho ρ_S or Kendall's tau τ [80,58,94]. However, the definition of their joint distribution needs to be addressed. This section summarizes some relevant approaches to model multivariate dependence, i.e. $d \geq 2$. The list is restricted to the RES literature and readers are referred to [10,67] for a more exhaustive survey. In this paper, 8 different families of methods, presented in Fig. 2 are analysed. They can be classified in two wide categories, based on whether they aim at defining a full joint probability distribution (Section 2.1) or they describe a partial (approximated) joint probability distribution making use of limited information (Section 2.2). In Fig. 2, gray circles represent the different families of methods, black solid lines show examples of those families and dashed lines are reserved to display similarity between related methods, which will become evident through the paper.

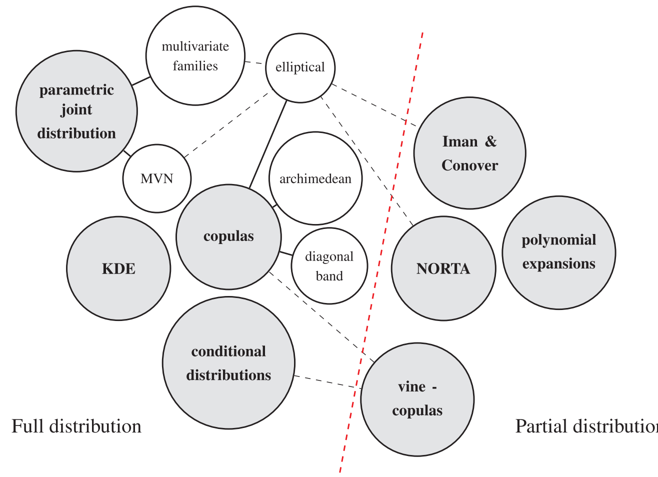

Fig. 2. Taxonomy of multivariate dependence modeling methods.

2.1. *Fully-specified joint distributions*

2.1.1. *Parametric distributions*

A predefined joint distribution model represents the most appealing way of describing multivariate data. Different models have been proposed in literature, specially focusing on multivariate exponential distributions [82]. In general, they are not flexible enough and it is relatively complex to fit and generate samples from such distributions [59]. Nevertheless, the multivariate normal (MVN) distribution still plays a fundamental role in modern statistical practice due to its appealing properties as well as the central limit theorem. A d-dimensional random vector $\boldsymbol{X}$ in $\mathbb{R}^d$ is defined to have a multivariate normal distribution if every non-trivial linear combination of its d-components has a univariate normal distribution. Then, it is uniquely determined by its mean vector $\boldsymbol{\mu} = [\mu_1,...,\mu_d]^{\mathrm{T}}$ such as $\mu_i = \mathbb{E}(x_i)$ and a $d \times d$ covariance matrix $\boldsymbol{\Sigma}$. If matrix $\boldsymbol{\Sigma}$ has full rank d, the density function of $\boldsymbol{X}$ can be formulated as:

$$f(\boldsymbol{x}) = (2\pi)^{-d/2}|\boldsymbol{\Sigma}|^{1/2} \exp\left[\frac{1}{2}(\boldsymbol{X} - \boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{X} - \boldsymbol{\mu})\right] \tag{2}$$

where $|\boldsymbol{\Sigma}|$ is the determinant of the covariance matrix. Alternative multivariate modeling methods heavily rely on the MVN distribution. For example the Johnson translation system [126] allows transforming a random vector $\boldsymbol{X} = [X_1,...,X_d]^{\mathrm{T}}$ with an unknown distribution into the standard normal domain:

$$\boldsymbol{Z} = \boldsymbol{\gamma} + \boldsymbol{\delta}\boldsymbol{g}(\boldsymbol{\lambda}^{-1}\boldsymbol{X} - \boldsymbol{\xi}) \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}) \tag{3}$$

where $\boldsymbol{\gamma} = [\gamma_1,...,\gamma_d]^{\mathrm{T}}$, $\boldsymbol{\delta} = \mathrm{diag}[\delta_1, ..., \delta_d]$, $\boldsymbol{\xi} = [\xi_1,...,\xi_d]^{\mathrm{T}}$, $\boldsymbol{\lambda} = \mathrm{diag}[\lambda_1, ..., \lambda_d]$ are constant parameters, $\boldsymbol{Z}$ is a new standard normal vector in $\mathbb{R}^d$ and $\boldsymbol{g}(\cdot) = [g_1(y_1),...,g_d(y_d)]^{\mathrm{T}}$, where the ith transformation can take the form:

$$g_i(y_i) = \begin{cases} \ln(y_i), & S_L(\text{lognormal})\text{ family} \\ \ln[\,y_i + \sqrt{y_i^2 + 1}\,], & S_U(\text{unbounded})\text{ family} \\ \ln[\,y_i/(1 - y_i)], & S_B(\text{bounded})\text{ family} \\ y_i, & S_N(\text{normal})\text{ family} \end{cases} \tag{4}$$

Once the vector parameters and the transformation family have been fitted for each random variable, realizations from the joint distribution of $\boldsymbol{X}$ can be sampled by inverting Eq. (3).