Electric Power Systems Research 144 (2017) 150–156

Contents lists available at ScienceDirect

Electric Power Systems Research

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

Statistical modeling of the power grid from a wind farm standpoint

Saber Farajzadeh [a,*], Mohammad H. Ramezani [a], Peter Nielsen [b], Esmaeil S. Nadimi [a]

[a] University of Southern Denmark, Campusvej 55, DK-5220 Odense, Denmark
[b] DONG Energy, WTG Electrical, Kraftværksvej 53, DK-7000 Fredericia, Denmark

ARTICLE INFO

Article history:
Received 20 May 2016
Received in revised form 9 September 2016
Accepted 25 November 2016
Available online 7 December 2016

Keywords:
Wind turbine
Wind farm
Principal component analysis
Statistical modeling
Power grid

ABSTRACT

In this study, we derive a statistical model of a power grid from the wind farm's standpoint based on dynamic principal component analysis. The main advantages of our model compared to the previously developed models are twofold. Firstly, our proposed model benefits from logged data of an offshore wind farm over several years which results in the development of a useful model for practical purposes. Secondly, the derived model is computationally inexpensive. Considering an arbitrary wind turbine generator, we show that the behavior of the power grid at the connection point can be represented by 4 out of 9 registered variables, i.e. 3-phase voltages, 3-phase currents, frequency, and generated active and reactive powers. We further prove that the dynamic nature of the system can be optimally captured by a time lag shift of two samples. To extend the derived model of a wind turbine generator to a wind farm, we propose an algorithm that optimally combines the principal components and scores. The principal components are estimated by the optimization of a cost function based on the modeling error while the scores are selected corresponding to the worst case scenario among the wind turbine generators. Our results show that the optimized principal components result in the modeling error less than 5% and the selected scores cover the variance of the data with probability higher than 95% among all generators in the wind farm.

1. Introduction

In order to fulfill the EWEA[1]'s central scenario for 2020, 754 new offshore wind turbines in 15 wind farms were grid-connected during 2015 which is 108% more than in 2014 [1,2]. Considering this rapidly increasing level of wind energy penetration, understanding the interaction between wind turbine generators (WTG) and the power grid is of great importance. This interaction needs to be precisely analyzed prior to grid connection in order to determine the impacts of the WTGs on each other within the wind farm as well as the impact of the newly installed wind farms on the power grid.

To analyze the mentioned interaction, one major requirement is a test facility where a power grid is simulated by utilizing appropriate hardware and software. A wind turbine test facility is an infrastructure to ensure having minimal impact not only from wind turbines on the existing power grid but also from grid disturbances on the wind turbines. The test facility includes a grid simulator to sink/source active and reactive power from/to the WTG. The grid simulator employs a realistic model of the power grid to simulate its behavior from a WTG point of view (Fig. 1).

Considering Fig. 1, there are several approaches to model and simulate a power grid. A power grid is an interconnected network of several subsystems such as generators, transmission lines, and distribution networks. One approach to have a simulated power grid is to model its subsystems separately and then connect them to achieve the whole grid model. Based on this approach, IEEE introduced some benchmarks with specified number of nodes and branches which have been used in many studies [3–5]. Another approach is to consider the power grid as a complex random graph in which the number of nodes and branches vary randomly following a specific distribution [6]. This approach can represent the random nature of the power grid due to abnormalities, accidental and abrupt changes in power generation and consumption [7,8].

Both mentioned approaches provide valuable information regarding the performance and vulnerability of power grids [6]. Although, they represent the interaction among different parts of the power grid, in case of WTG testing, two main disadvantages might be considered. (1) Modeling all details of the power grid makes these approaches too complicated to be used in real-time applications. (2) The interactions between different parts of the grid are not necessary for testing a WTG, since the only information we need from the power grid is the information at the

* Corresponding author.
E-mail address: saber@mmmi.sdu.dk (S. Farajzadeh).

[1] European Wind Energy Association.

http://dx.doi.org/10.1016/j.epsr.2016.11.020
0378-7796/

Nomenclature

a	optimum number of principal components
CPV	Cumulative Percent Variance
$\mathbf{E}$	modeling error
$EIG(w)$	$= \lambda_{m(w-1)+1}$
EIG_N	normalized EIG
$EIGR(w)$	$= \frac{EIG(w)}{EIG(w-1)}$
$EIGR_N$	normalized $EIGR$
F	power grid frequency
I_R, I_S, I_T	three phase currents
$\mathbf{J}$	cost function for $\mathbf{\Psi}^\gamma$ selection
j	index to represent the jth WTG
m	number of measured variables
N_{WTG}	number of WTGs in a wind farm
n	number of samples in the data matrix
n_w	number of samples in the trajectory matrix
P_{ac}	active power
Q_{rea}	reactive power
$Q(t)$	sum of squared residuals at time t of an arbitrary WTG
$Q^j(t)$	sum of squared residuals at time t of the jth WTG
$\mathbf{S}$	scores matrix of an arbitrary WTG
$\mathbf{S}^j$	scores matrix of the jth WTG
$\mathbf{S}_a$	matrix of the first a scores of an arbitrary WTG
$\mathbf{S}_a^j$	matrix of the first a scores of the jth WTG
S_i^j	ith score of the jth WTG
$T^2(t)$	Hotelling's residuals at time t of an arbitrary WTG
$T^{2^j}(t)$	Hotelling's residuals at time t of the jth WTG
V_R, V_S, V_T	three phase voltages
WTG	wind turbine generator
w	time lags in DCPA
$\mathbf{X}$	sample data matrix
$\mathbf{X}^w$	trajectory matrix
$X(t)$	row of sample data matrix at time stamp t
$X^w(t)$	row of trajectory matrix at time stamp t
$\mathbf{\Lambda}$	a diagonal matrix of λ_is of an arbitrary WTG in descending order
$\mathbf{\Lambda}_a$	a diagonal matrix of the first a λ_is of an arbitrary WTG
$\mathbf{\Lambda}^j$	a diagonal matrix of λ_is of the jth WTG
$\mathbf{\Lambda}_a^j$	a diagonal matrix of the first a λ_i of the jth WTG
$\mathbf{\Lambda}^\gamma$	a diagonal matrix of λ_is of the generalized model
λ_i	ith eigenvalue of $\mathbf{\Sigma}$
λ_i^j	ith eigenvalue of $\mathbf{\Sigma}$ of the jth WTG
λ_i^γ	ith eigenvalue of the generalized model
$\mathbf{\Sigma}$	sample covariance matrix
ϕ	cost function for w selection
$\mathbf{\Psi}$	matrix of principal components of an arbitrary WTG
$\mathbf{\Psi}_a$	matrix of the first a principal components of an arbitrary WTG
$\mathbf{\Psi}^j$	matrix of principal components of the jth WTG
$\mathbf{\Psi}^\gamma$	matrix of principal components of the generalized model
Ψ_i^γ	ith principal component of the generalized model
Ψ_i^j	ith principal component of the jth WTG

connection points. A proper solution to have both accuracy and simplicity at the same time is to use a statistical data-driven model. This model extracts the statistical features of the data in a time interval to simulate the behavior of the system in the future. To the best of our knowledge, such a model has not been presented for the power grid at the connection point. In order to implement this idea, one approach would be to consider the power grid at the connection point as a multivariate statistical model. This model should preserve all the statistical properties of the power grid over a specific time period. In comparison with the previous approaches, this method is less complex, hence more suitable for real-time applications.

In this study, we first derive an eigenvector-based multivariate model of a power grid from the WTG's standpoint using dynamic principal component analysis (DPCA). DPCA is one of the well-developed methods for system modeling, fault detection, and fault identification [9–11] which transforms the sample data into a new projected space. Furthermore, the projected variables (scores) are uncorrelated and the method can be used for data dimensionality reduction [12]. The DPCA model has two parameters which have to be estimated: the number of time lags which represents the system's dynamic and the number of selected principal components which depends on the number of uncorrelated variables. We will use two different approaches to find the optimum values of these parameters. The obtained model will be validated by a new data set and observing model residuals.

The next issue that will be addressed in the paper is to propose a generalized model for a wind farm to model the behavior of all WTGs with a certain level of confidence. Given that the generalized model is also in latent space, it is consisting of two elements: the principal components and the scores. For calculating the proper principal components, we will introduce several possible selections and a cost function to choose the best one. The scores of the generalized model are determined by some assumptions on their distributions and selecting the parameters of these distributions by considering WTGs with harsh conditions. We will introduce a criterion to validate the proposed generalized model using a new data set.

The rest of this paper is organized as follows. The nature of the registered data in the wind farm is presented in Section 2. Section 3 represents the statistical modeling of the power grid from WTG's standpoint using DPCA. The generalized model and its corresponding formulation are described in Section 4. Section 5 covers the results of this study in addition to the validation process. Finally, Section 6 provides some concluding remarks.

2. The registered data

DONG Energy as one of the major energy suppliers in Denmark is the key partner in this project. The logged data set of a Danish offshore wind farm that is used for the analysis in this study is provided by DONG energy and due to confidentiality agreements, no further information is available on the farm or the type of wind turbines within the farm. The logged variables are 3 phase voltages and currents, frequency, and active and reactive powers and each sample is a moving average of 10 min sampled data. The data sets, which logged during 2013 and 2014, are used for model identification and validation, respectively. The logged data is normalized prior to method application, considering the estimated mean values and standard deviations of the measured variables.

3. Statistical modeling of the power grid

In multivariate data analysis, PCA is a vector based transformation which transforms a correlated set of sampled data into a linearly uncorrelated data set [12]. Based on the eigenvectors of the sample covariance matrix, PCA also preserves the maximum variance of the data set in a subspace with lower dimensionality. The eigenvalues of the covariance matrix show the variance of the data in the direction of the eigenvectors. The eigenvectors