Available online at www.sciencedirect.com

ScienceDirect

Energy Procedia 80 (2015) 365 – 375

Energy Procedia

12th Deep Sea Offshore Wind R&D Conference, EERA DeepWind'2015

Resampling of Data for Offshore Grid Design based on Kernel Density Estimation and Genetic Algorithm

Vin Cent Tai*, Kjetil Uhlen

Department of Electric Power Engineering, Norwegian University of Technology and Science, NO-7491 Trondheim, Norway

Abstract

Offshore wind power has been a major focus in the renewable energy development in recent years, due to better wind speeds and wind energy are available offshore. Since the development (construction and grid connection) of offshore windfarms is relatively more expensive in nature, careful planning and design are needed to maximise the benefits of the offshore wind projects. Optimisation with only one operational state is not sufficient in grid design as the state of power system is not stationary due to the fluctuations of the wind power and power consumption. Eventually this leads to the fluctuation of the base load power generations. To account for this variability, the optimisation has to be done with many operational states. Historical data of power consumption at each load centre and simulation data of wind power have to be used to describe the system states. Ideally, the complete set of data should be used to describe the power system states but this could also lead to unsolvable case as there are too many unknowns involved in the calculation. To keep the number of states as low as possible to reduce the computation time, selection of smaller number of samples that can represent the whole data set has to be carried out. This involves detail studies of the statistical distributions of the data. This study is therefore dedicated to develop a procedure for selecting a set of statistically sound samples to represent the entire data set for grid design purposes.

Peer-review under responsibility of SINTEF Energi AS

Keywords:
non-parametric statistics, genetic algorithm, wind power integration, offshore grid design, data sampling

1. Introduction

Driven by political wills and out of environmental concerns, the activity in renewable energy development has increased substantially in recent years and is expected to grow exponentially in the near future. Due to higher wind speeds are available offshore, scientists and engineers have been considering to build more offshore wind-farms to tap more renewable energy from the nature. According to the data provided in [1], it is estimated that a total of 97 windfarms will be in operation within the Baltic Sea region by year 2030, with a total installed capacity of 27 GW. The penetration of increasingly large amount of wind power into power grids may pose a challenge to power system plan-

* Corresponding author.
E-mail address: vin.tai@elkraft.ntnu.no

Peer-review under responsibility of SINTEF Energi AS
doi:10.1016/j.egypro.2015.11.441

ning and operation. As the development of offshore windfarms is relatively more expensive than onshore windfarms, careful planning and design are needed to ensure the total benefit the offshore project will bring is maximised.

As the state of power system is always changing due to the dynamic of the power consumption coupled with the stochastic nature of the wind power generation, designing an offshore grid with only one operational state is obviously inappropriate. Wind power data and power consumption data that reflect the variability of the system states are needed to get a more practical design. Ideally, the complete set of data should be used to describe the power system states but this could also lead to unsolvable case as there are too many unknowns involved in the calculation. To keep the number of states as low as possible to reduce the computation time, selection of smaller number of samples that can represent the whole data set has to be carried out. This involves the detail studies of the statistical distributions of the data. Figure 1 shows the probability distribution functions (PDF) of energy consumption of Germany and Estonia. Both the energy consumption data were obtained by using the historical data and scaled to meet the expected annual demand of each country at the point of interest in time (in this study year 2030 has been chosen). Historical hourly data from 2010 to 2013, obtained from the *European Network of Transmission System Operator for Electricity* (ENTSO-E), have been used to generate the energy consumption data. Note that both load consumption distributions are multi-modal. For a variable which is multi-modal (has more than one peak), the conventional parametric statistical method which assumes a known distribution (Gaussian, Poisson, etc.) is no longer appropriate to describe the variable. Non-parametric methods are more suitable to be used in this case as those methods use only data to obtain the underlying statistical properties of the data [2]. The work described in this is dedicated to develop a sampling procedure to obtain a set of samples out of a design variable (power consumption or wind power in this case), which is small in size but still retains the statistical properties of the original data. The rest of the paper is organised in the following manner : Section 2 describes the proposed method, alongside with some existing methods in section 3, and their performance comparisons in section 4. Last but not least, some conclusions are presented in section 5.

2. Methodology

2.1. Kernel Density Estimation

Kernel Density Estimation (KDE) is a non-parametric statistical method to estimate the PDF of a random variable. For a univariate case with n data, $\mathbf{X} = [X_1, X_2, \ldots, X_n]$, the probability density at point x, $f(x)$, is defined as

$$f(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \tag{1}$$

where $K(\cdot)$ is the *kernel function* and h is called the *bandwidth* [3] or *window width* [4]. The kernel function chosen for this study is the Gaussian kernel, with $\hat{u} = (x - X_i)/h$

$$K(\hat{u}) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\hat{u}^2\right) \tag{2}$$

Scott's rule [5] is used to determine the bandwidth for the respective data with standard deviation $\tilde{\sigma}$:

$$h = \left(\frac{4\tilde{\sigma}^5}{3n}\right)^{\frac{1}{5}} \tag{3}$$

2.2. Genetic Algorithm

Genetic Algorithm (GA) [6] was used to select the desired number of samples (n) from the data. In GA, a string of n samples is called an individual (or chromosome) and each sample data an individual holds is called a gene. GA works by generating a number of individuals, make them "interact" with each other, and allow them to "evolve" through a number of "generations". To better explain the working principle of GA, the general procedure of GA to select the samples is listed as follows: