

Stochastic Modeling of Multi-Area Wind Power Production

Anthony Papavasiliou
CORE, UCL
anthony.papavasiliou@uclouvain.be

Shmuel S. Oren
IEOR, UC Berkeley
oren@ieor.berkeley.edu

Ignacio Aravena
CORE, UCL
ignacio.aravena@uclouvain.be

Abstract

In this paper we present a stochastic model for multi-area wind production that is used for planning reserves in transmission-constrained systems with large amounts of integrated renewable power supply. The stochastic model accounts for the inter-temporal and spatial dependencies of multi-area wind power production. Results are presented for two case studies of the California and the German power system.

1. Introduction

The large-scale integration of renewable power supply in power systems has recently motivated researchers to consider stochastic unit commitment policies for committing reserves in order to guarantee the reliable operation of the grid. Such studies include the work of Ruiz et al. [1], Wang et al. [2], Constantinescu et al. [3], Tuohy et al. [4], Morales et al. [5], Bouffard et al. [6] and Papavasiliou et al. [7], [8]. Stochastic unit commitment models explicitly account for uncertainty in the formulation of the unit commitment problem and therefore have the potential to outperform ad-hoc deterministic reserve rules that are used in practice. The formulation of the stochastic unit commitment problem requires explicit modeling of the uncertain parameters in the unit commitment problem in terms of a few appropriately weighted representative scenarios.

Uncertainty in power system operations can be categorized between discrete and continuous disturbances. Discrete disturbances refer to the failure of equipment such as generators and transmission lines. Continuous disturbances include parameters of the unit commitment problem that vary smoothly such as electricity demand and renewable power production.

Transmission constraints strongly affect the optimal rule for allocating reserves in each area of the network. In order to account for transmission constraints, operators often use ad-hoc import constraints for determining locational reserve requirements. Import constraints can be categorized between ‘bubble’ constraints and intertie constraints. ‘Bubble’ constraints limit the total amount of power that can

flow into a load pocket in order to ensure that the unit commitment schedule reserves sufficient transfer capability on the lines in order to protect against the possibility of generation capacity failure within the load pocket. On the other hand, intertie constraints limit the amount of power that can flow over interties in order to protect the system against the failure of major corridors that bring significant amounts of power from outside the system. Both types of constraints are formulated on an ad-hoc basis, in the sense that there is no formal methodology for determining the set of lines belonging to an import constraint and the limit on the amount of power that can flow on the import set. Various authors, including Arroyo and Galiana [9], Galiana et al. [10] and Bouffard et al. [11], have demonstrated the complexity of committing reserves in the presence of transmission constraints. Beyond their influence on reserve requirements, transmission constraints also affect the cost of operating the system. This is due both to the fact that transmission constraints reduce the flexibility of dispatching conventional generators in the system, and also due to the fact that they result in the waste of renewable energy supply.

The inclusion of transmission constraints in the unit commitment model necessitates the development of a multi-area wind production model. Moreover, in order to assess the impact of wind power production on power system operations over an entire year, it is necessary to account for the non-stationary (seasonal and diurnal) patterns of wind power production. This paper presents a multi-area stochastic wind production model that captures the seasonal and diurnal patterns of wind power production, accounts for the temporal and spatial correlations of the original data set and accurately reproduces the marginal distribution of wind power production at each location of the network. Moreover, the proposed model is applied to two case studies of wind power integration: the California power system and the German power system. The remainder of the paper is structured as follows. In Section 2 we summarize relevant literature and present the methodology for calibrating the multi-area wind production model and for simulating the process. In

Section 3 we present results for the two case studies. In Section 4 we summarize the conclusions of our study.

2. Methodology

The nonlinear dependence of wind power production on wind speed raises challenges in the statistical modeling of wind power production. It is therefore common in the wind power modeling literature to model wind speed and use a static power curve to calculate the corresponding wind power production.

The task of modeling wind speed consists of removing seasonal and daily patterns from wind speed data, fitting the resulting process to a parametric or non-parametric distribution, and fitting an appropriate time series model to the underlying noise in order to capture the strong temporal correlation of wind speed time series. Brown et al. [12] performed early work on wind power modeling. The authors list various parametric distributions for fitting wind speed data, such as the Weibull, inverse Gaussian and exponential distribution. The authors use an exponential function to transform their data to an approximately Gaussian data set. They remove hourly means and estimate the order of an appropriate autoregressive model and they use the Yule-Walker equations [13] to estimate the parameters of the autoregressive model. Torres et al. [14] follow the same methodology as Brown et al. [12]. The authors use autoregressive moving average models and find that these more general models provide a more satisfactory fit.

Transmission constraints have recently prompted researchers to develop multi-area wind production models. Moreover, diurnal and seasonal patterns of wind power production need to be accounted for in order to assess the impact of wind integration on power system operations over an entire year. Morales et al. [15] develop a multi-area wind speed model by using a noise vector that drives a vector autoregressive process. In order to simplify the calibration of the model, the authors assume a diagonal matrix of autoregressive coefficients, which implies that spatial correlations among wind speed in various locations are captured fully by the underlying noise vector. The calibration and simulation model that we use extends the approach of Morales et al. [15] in order to account for seasonal and diurnal wind speed patterns.

2.1 Calibration

Given a multi-area wind speed data set y_{kt} , where k indexes location and t indexes time, the first step of the calibration procedure is to remove diurnal and seasonal patterns. We normalize the data by

subtracting the hourly mean and dividing by the hourly standard deviation in order to obtain a stationary data set y_{kt}^S for each location. Systematic patterns can be monthly or seasonal. In each case, the appropriate portion of the data set should be chosen for estimating the mean and variance. In the present study, the data is partitioned by month.

We next filter the data set in order to obtain an approximately Gaussian data set y_{kt}^{GS} . Brown et al. [12], Torres et al. [14] and Morales et al. [15] use this approach for transforming Weibull-distributed wind speed data to Gaussian data, and Callaway [16] uses a non-parametric transformation. In the single-area wind integration study of Papavasiliou et al. [7], the authors find that the inverse Gaussian distribution provides a satisfactory fit for the data set. For the multi-area wind integration study presented in this paper and used in Papavasiliou and Oren [8], no single parametric distribution provides a close fit for the observed data in all locations, therefore we fit a non-parametric distribution $\hat{F}_k(\cdot)$ to the data of each location k .

The resulting time series y_{kt}^{GS} can be modeled by an autoregressive model:

$$y_{k,t+1}^{GS} = \sum_{j=0}^p \hat{\phi}_{kj} y_{k,t-j}^{GS} + \hat{\omega}_{kt}, \quad (1)$$

where $\hat{\omega}_{kt}$ is the estimated noise and $\hat{\phi}_{kj}$, $j \in \{1, \dots, p\}$, are the estimated coefficients of the autoregressive model. The estimated noise at each location is not auto-correlated, but it might be cross-correlated among different locations. This can occur, for example, by weather fronts that impact neighboring areas with a time lag. The calibration process is summarized in the following steps:

Step (a). Remove systematic seasonal and diurnal effects:

$$y_{kt}^S = \frac{y_{kt} - \hat{\mu}_{kmt}}{\hat{\sigma}_{kmt}}, \quad (2)$$

where y_{kt} is the data, y_{kt}^S is the transformed non-seasonal data, and $\hat{\mu}_{kmt}$ and $\hat{\sigma}_{kmt}$ are the sample mean and standard deviation respectively for location k , epoch (e.g. month or season) m and hour t .

Step (b). Transform the data in order to obtain a Gaussian stationary data set:

$$y_{kt}^{GS} = N^{-1}(\hat{F}_k(y_{kt}^S)), \quad (3)$$

where y_{kt}^{GS} is the transformed stationary data that follows a Gaussian distribution, $N^{-1}(\cdot)$ is the inverse of the cumulative distribution function of the normal distribution and $\hat{F}_k(\cdot)$ is the cumulative function of the (parametric or non-parametric) fit for the data in location k . Following Morales et al. [15], we assume

that the transformed wind speed data y_{kt}^{GS} closely follows a multivariate Gaussian distribution and can be modeled using a vector autoregressive model with a diagonal autoregressive coefficient matrix.

Step (c). Use the Yule-Walker equations [13] to estimate the autoregressive parameters $\hat{\phi}_{kj}$ and the covariance matrix $\hat{\Sigma}$, as well as the significant lagged cross-correlations $\hat{\Gamma}$ of the residual noise obtained from Eq. (1).

2.2 Simulation

In order to simulate multi-area wind power production, we assume that the process is driven by an autoregressive ‘noise’ vector. For K locations and p periods of lag the model is:

$$Y_{k,t+1} = \sum_{j=0}^p \phi_{kj} Y_{k,t-j} + \omega_{kt}, \quad (4)$$

where $\Phi_j = (\phi_{kj})$, $k \in \{1, \dots, K\}$, $j \in \{1, \dots, p\}$, is the matrix of autoregressive parameters and (ω_{kt}) , $k \in \{1, \dots, K\}$, are multivariate Gaussian random variables with mean 0, covariance matrix Σ and lagged cross-correlation Γ . The simulation of the multi-area process can then be summarized in the following steps:

Step (a). Generate autoregressive noise of order p by using the estimated autoregressive parameters, variance and significant cross-correlations:

$$Y_{k,t+1}^{GS} = \sum_{j=0}^p \hat{\phi}_{kj} Y_{k,t-j}^{GS} + \omega_{kt}, \quad (5)$$

Residual noise samples ω_{kt} can be constructed by sampling a standard Gaussian distribution and multiplying these samples by the Cholesky decomposition of the lagged cross-covariance matrix [15].

Step (b). Transform the resulting process such that it obeys the non-Gaussian distribution of the transformed stationary data:

$$Y_{kt}^S = \hat{F}_k^{-1}(N(Y_{kt}^{GS})), \quad (6)$$

where Y_{kt}^S is the stationary, non-Gaussian process, $N(\cdot)$ is the cumulative distribution function of the normal distribution and $\hat{F}_k^{-1}(\cdot)$ is the inverse of the cumulative function of the data for each location.

Step (c). Transform Y_{kt}^S by its seasonal and hourly mean and variance:

$$Y_{kt} = \hat{\sigma}_{kmt} Y_{kt}^S + \hat{\mu}_{kmt}, \quad (7)$$

where Y_{kt} is the resulting process that is non-stationary and distributed according to the original data for each location.

Step (d). Use an approximation $\hat{P}_k(\cdot)$ of the aggregate power curve¹ for each location to simulate wind power production:

$$P_{kt} = \hat{P}_k(Y_{kt}), \quad (8)$$

where P_{kt} is the simulated wind power production process for each location.

3. Results

We use the multi-area wind production model to simulate the large-scale integration of wind power in California and Germany.

3.1 Data for California

We use wind speed and wind power production data from the 2006 data set of the National Renewable Energy Laboratory (NREL) Western Wind and Solar Integration Study (WWSIS), described by Potter et al. [18]. We study two wind integration cases. The first represents a moderate energy integration level for wind power corresponding to the 2012 integration target of California, and the second case represents a deep integration level corresponding to the 2020 integration target. Ex post we have estimated that the moderate integration case corresponds to approximately 7% wind energy penetration, while the deep integration case corresponds to approximately 14% wind energy penetration. In the subsequent analysis we will refer to these cases as moderate and deep integration respectively.

In order to collect data for each case, we examined the interconnection queue of the California ISO until 2020 (see [19]), and placed individual wind generators in our model by matching the geographical locations of planned wind power installations with the corresponding wind park data in the WWSIS data set. In Table 1 we present the location of existing wind generation capacity, as well as capacity for the moderate and deep integration cases.

Table 1. Current and projected capacity of wind power installations in California (MW).

County	Existing	Moderate	Deep
Altamont	954	954	1,086
Clark	-	-	1,500
Imperial	-	-	2,075
Solano	348	848	1,149
Tehachapi	1,346	4,886	8,333
Total	2,766	6,688	14,143

¹ The power curve maps wind speed to wind power production. The aggregated power curve approximation $\hat{P}_k(\cdot)$ can be constructed using historical wind speeds and historical wind power registers.

In Fig. 1 we present a schematic diagram of the WECC model. The dashed boxes represent load and generation pockets. The thick solid lines represent the import constraints discussed in the introduction. Each solid red line intersects a set of transmission lines over which the total amount of power cannot exceed a certain limit. The wind generators of Table 1 are located in the five buses that are depicted as solid black circles. In order of appearance from top to bottom, these wind sites are Solano, Altamont, Tehachapi, Clark and Imperial.

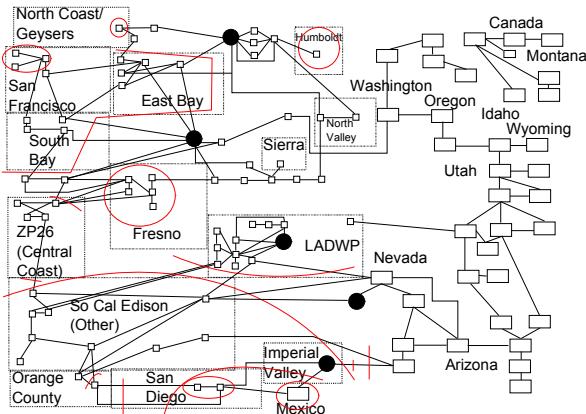


Figure 1. A schematic of the WECC model.

3.2 Data for Germany

We use 2012 wind power production and wind speed data for both calibrating and testing our model. Wind power production for each Transmission System Operator (TSO) in Germany was obtained from the Transparency platform of the European Energy Exchange [20]. The four TSOs operating the Germany network and the 2012 wind power capacity installed in their control zones are:

- 50Hertz: 12,200MW of installed wind capacity,
- Amprion: 5,314MW of installed wind capacity,
- TenneT: 11,958MW of installed wind capacity; and
- TransnetBW: 569MW of installed wind capacity.

Wind speed measurements at meteorological stations were obtained from the German National Meteorological Service [21]. Meteorological stations were assigned within each control area by cross-referencing the geographical location of stations with the geographical jurisdiction of each TSO, as shown in Fig. 2.

Even though wind speed data is available at more than 70 stations across Germany, wind power data is only available at TSO level. We therefore consider a four-area model, with each area corresponding to a TSO. The average wind speed for each TSO control area is computed as a weighted average of the wind speed over the corresponding meteorological stations, where the weights have been chosen so as to maximize the fit between the observed power production at each TSO and the power obtained by passing the wind speed data through a generic wind power curve [22, Table 2.2].

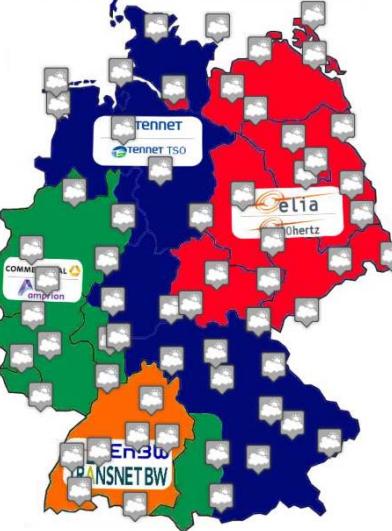


Figure 2. TSOs and meteorological stations (gray markers) in Germany.

3.3 Results

Figures 3-8 show the results for the California case study. In Fig. 3 we compare the auto-correlation and cross-correlation functions of the Gaussian noise data y_{kt}^{GS} with the simulated autoregressive noise Y_{kt}^{GS} , for various significance threshold levels for the lagged cross-correlations². Fig. 4 presents the approximate power curve and the fit of the complementary cumulative probability distribution of wind output to the data for the Altamont area for the moderate and deep integration cases. The corresponding results for Solano and Tehachapi are presented in Figs. 3, 4 respectively. In Figs. 7, 8 we present results for the deep integration case for Clark County and Imperial

² The observed correlation in the sample data has to be superior to $Z_{1-a/2}/\sqrt{N}$ in absolute value to be considered significant and be reproduced in the simulated noise. $(1 - a)$ is the significance level and N is the size of the sample. The level of significance influences the number of estimated cross-correlation parameters, with higher significance resulting in principle in a more accurate model but requiring the estimation of a larger number of parameters.

Valley respectively. The results for Germany are presented in Figures 10-13.

For the California case study, the auto-correlation function of the simulated noise Y_{kt}^{GS} deviates from the data only for lags greater than 6 hours at the Clark County, Imperial and Tehachapi areas. The complementary cumulative probability distribution deviates from the data only for high wind output levels for the Tehachapi area, and to a lesser extent for the Solano area. From the power curve of the Tehachapi area we note that the scatter plot of wind speed to wind power exhibits a significant spread. This is due to the fact that Tehachapi covers a wide geographic area with wind parks located in most regions of the area. As a

result, the power curve cannot reproduce the high-power results observed in the data. In order to alleviate this problem, we experimented with further partitioning the Tehachapi area in smaller regions. However, this introduced greater inaccuracy to the model due to the higher dimensions of the correlation matrix Σ and the lagged cross-correlations Γ . As a result, we chose to model five areas as the best compromise between capturing locational dependencies and retrieving marginal wind speed distributions at each location. Similarly, the Solano area exhibits a noticeable spread in the scatter diagram between wind speed and wind power production.

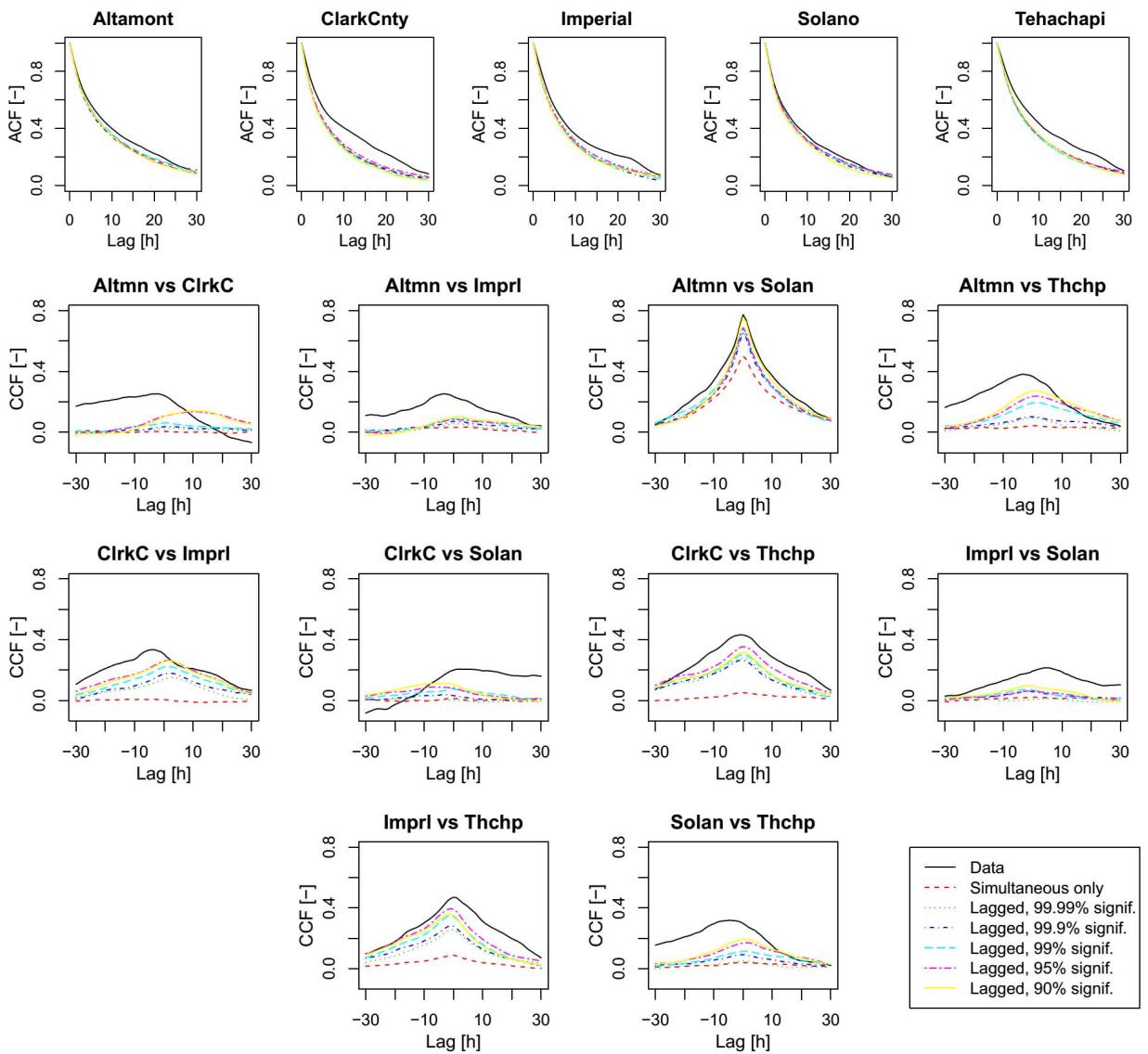


Figure 3. Auto-correlation (ACF) and cross-correlation functions (CCF) of the Gaussian noise y_{kt}^{GS} (transformed data) and simulated autoregressive noise Y_{kt}^{GS} , considering various significance threshold levels for the lagged cross-correlations. California deep integration case study.

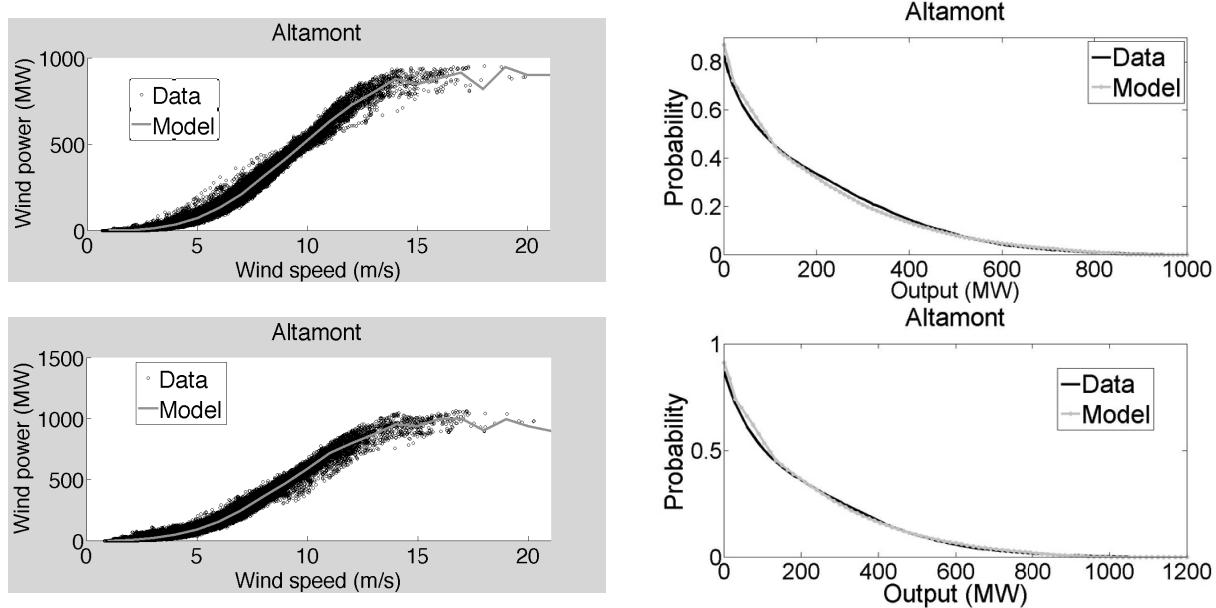


Figure 4. Power curves (left) and complementary cumulative probability distribution of wind output (right) for Altamont for the moderate (up) and deep (down) integration study.

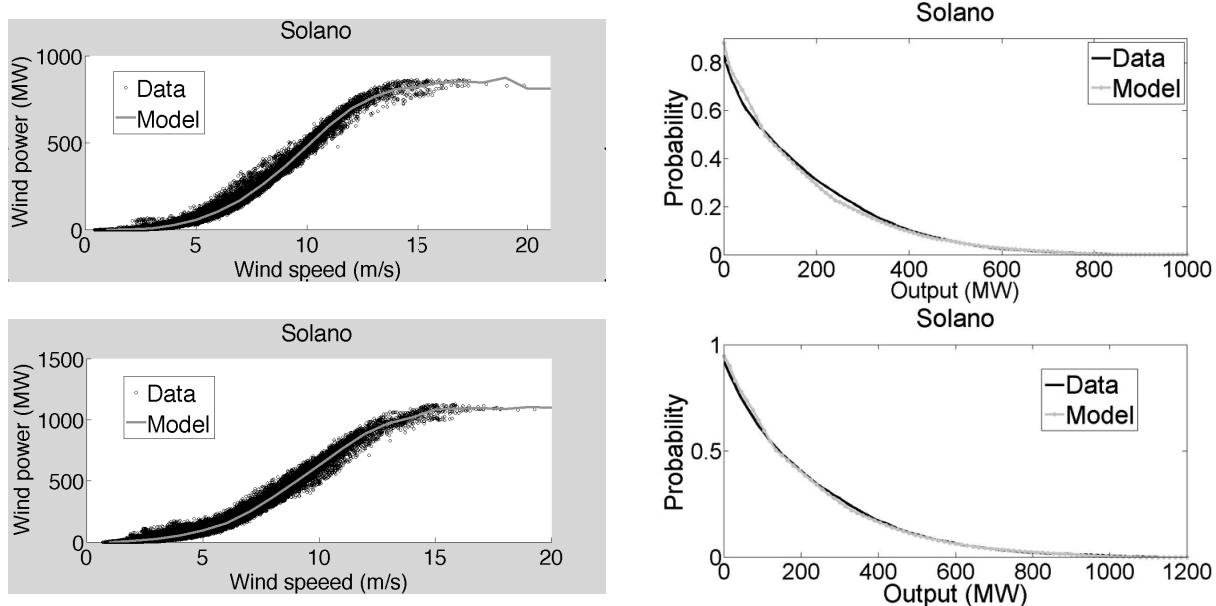


Figure 5. Power curves (left) and complementary cumulative probability distribution of wind output (right) for Solano for the moderate (up) and deep (down) integration study.

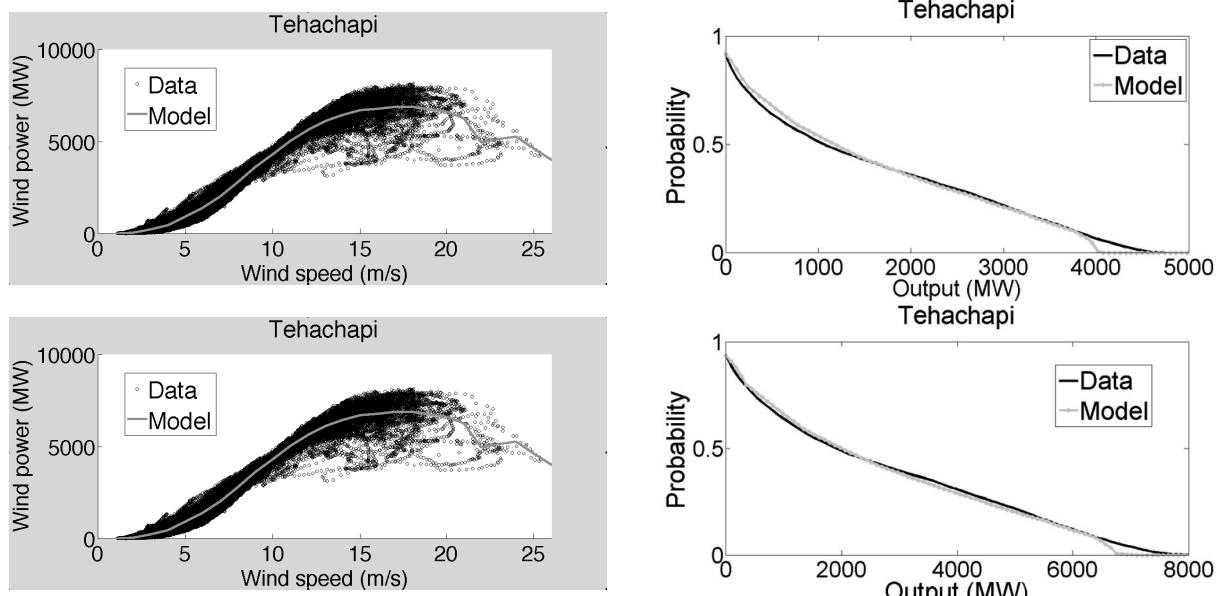


Figure 6. Power curves (left) and complementary cumulative probability distribution of wind output (right) for Tehachapi for the moderate (up) and deep (down) integration study.

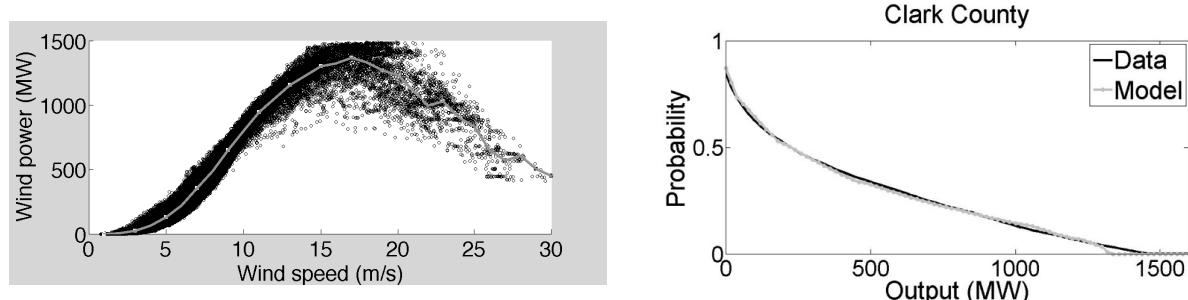


Figure 7. Power curves (left) and complementary cumulative probability distribution of wind output (right) for Clark County for the deep integration study.

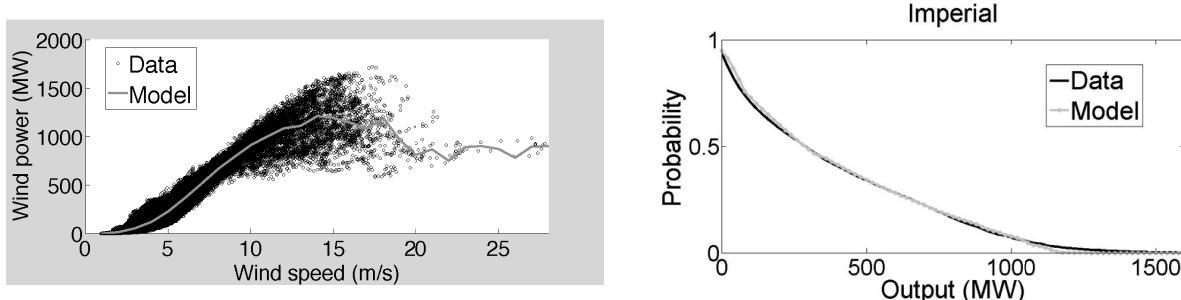


Figure 8. Power curves (left) and complementary cumulative probability distribution of wind output (right) for Imperial Valley for the deep integration study.

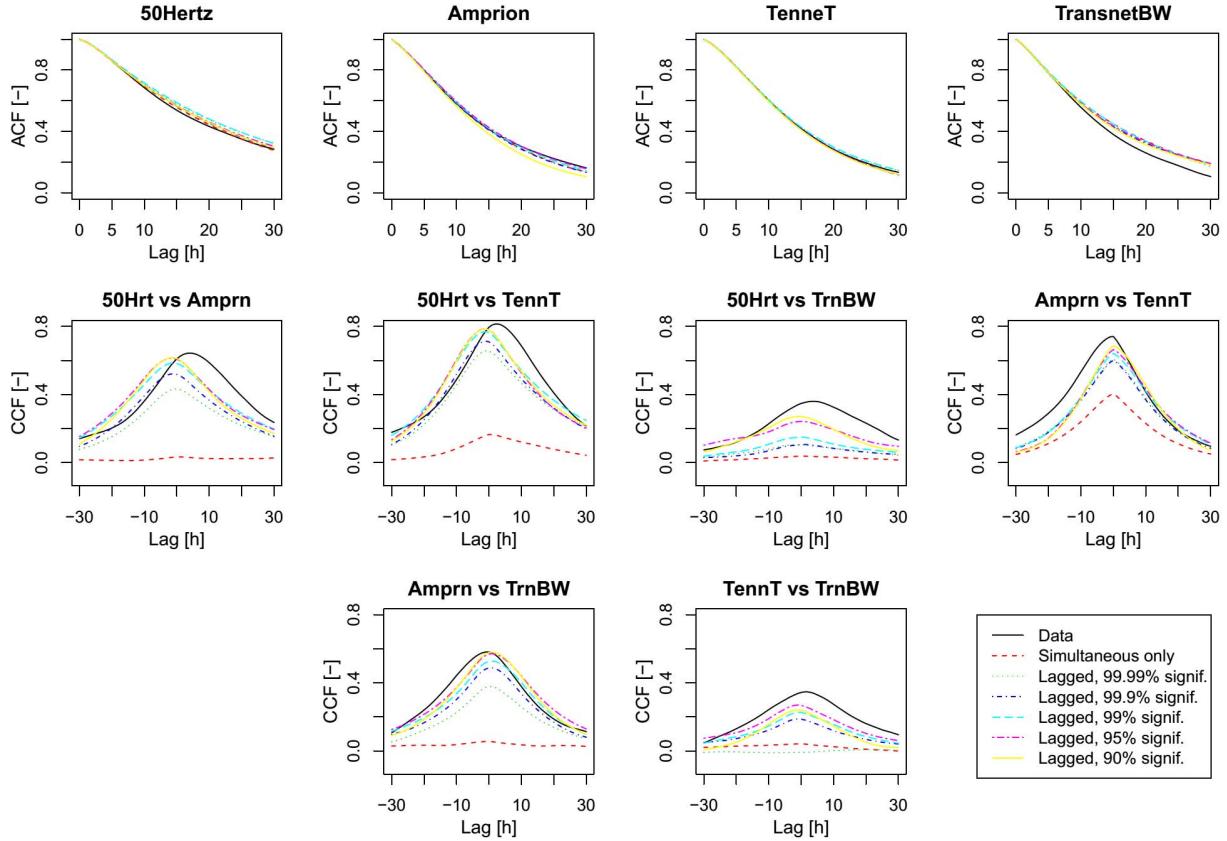


Figure 9. Auto-correlation (ACF) and cross-correlation functions (CCF) of the Gaussian noise y_{kt}^{GS} (transformed data) and simulated autoregressive noise Y_{kt}^{GS} , considering various significance threshold levels for the lagged cross-correlations. Germany case study.

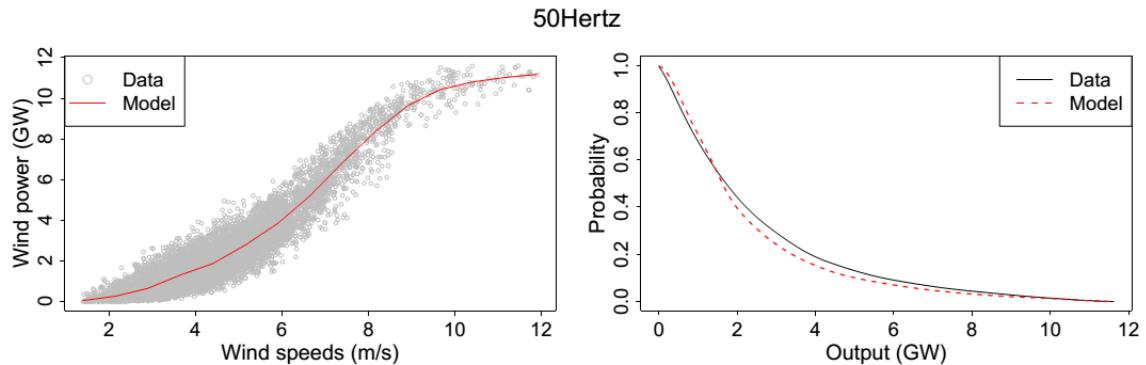


Figure 10. Power curves (left) and complementary cumulative probability distribution of wind output (right) for 50Hertz area.

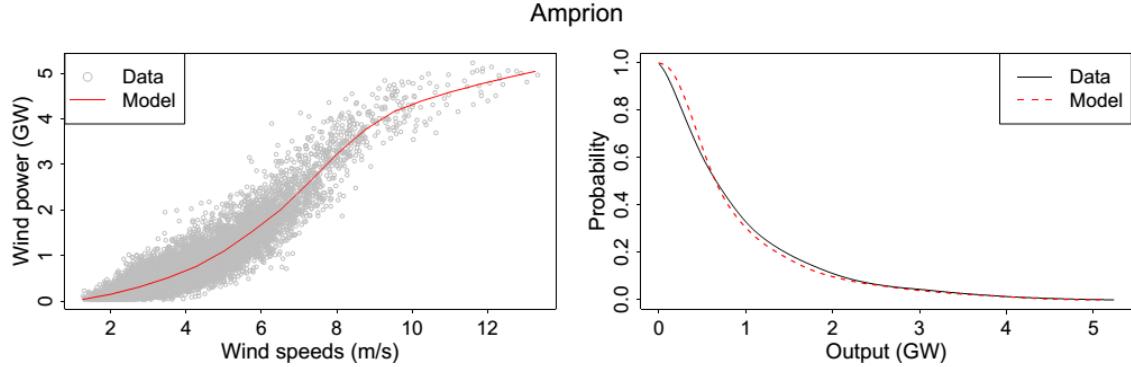


Figure 11. Power curves (left) and complementary cumulative probability distribution of wind output (right) for Amprion area.

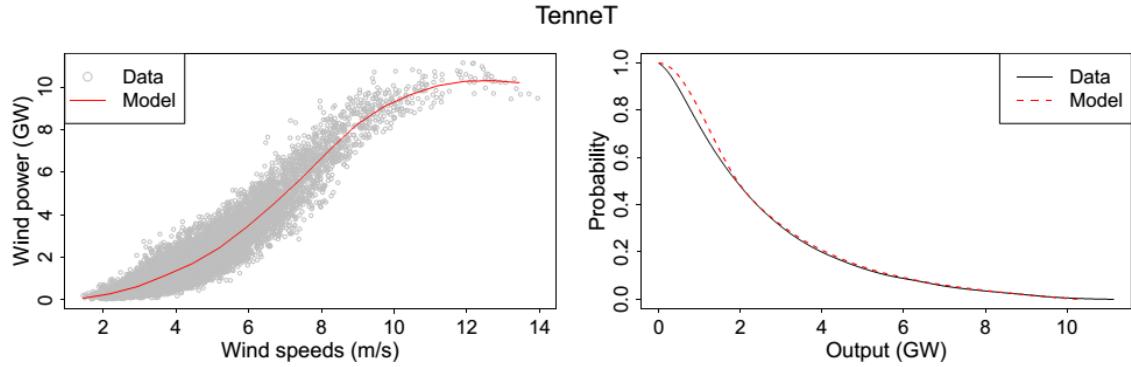


Figure 12. Power curves (left) and complementary cumulative probability distribution of wind output (right) for TenneT area.

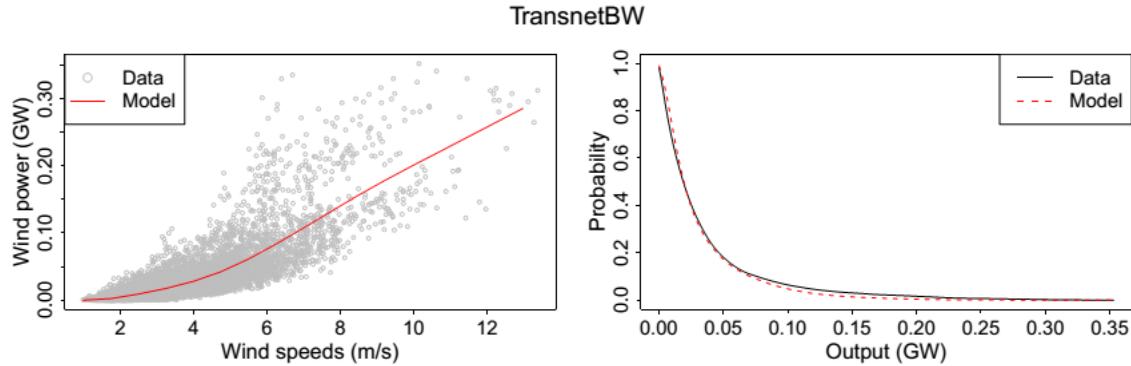


Figure 13. Power curves (left) and complementary cumulative probability distribution of wind output (right) for TransnetBW area.

In the German case study, the auto-correlation function of the simulated noise Y_{kt}^{GS} closely replicates the transformed data y_{kt}^{GS} , while the cross-correlations are partially captured for all pair of areas, with an apparent shift in the case of ‘50Hertz vs. Amprion’ and ‘50Hertz vs. TenneT’. Minor deviations in the fit of the complementary cumulative probability distribution can be observed in the middle range of the power output for 50Hertz, while for Amprion and TenneT minor differences arise in the low output range. No major differences are observed in TransnetBW. Wind power generation in Germany is

highly distributed, which explains the dispersion of the scatter plots in Figs. 10-13.

In Figs. 3 and 9 we observe that accounting for more lagged cross-correlations improves the accuracy of simulated time series with respect to the transformed data. The number of total cross-correlation parameters estimated for each significance threshold level is presented in Table 2.

Table 2. Total number of estimated cross-correlation parameters for different significance threshold levels (including all areas and lags)

Significance	California	Germany
99.99%	33	32
99.9%	49	45
99%	80	59
95%	132	81
90%	172	96

In both case studies the primary source of discrepancy in the model is the approximate speed-power curve. The aforementioned drawback is acceptable in the context of unit commitment studies of wind integration. Wind power variability affects reserve requirements due to the fact that wind power production often reaches a near-zero level for extended periods of time. We note from the complementary cumulative probability distribution of wind output that the behavior of wind power production is depicted accurately at low wind production levels. Wind production ramping also has the potential of affecting reserve requirements. Our model accounts for the inter-temporal fluctuations of wind power supply by isolating monthly and diurnal patterns and by using a time series model for wind speed.

4. Conclusions

We have presented a stochastic model of multi-area wind production that can be used in stochastic unit commitment studies of renewable energy integration. We fit a time series model of wind speed and use a piecewise linear approximation of the regional power curve to simulate wind power. We account for monthly and diurnal patterns of wind speed and use a time series model for reproducing temporal correlation. We represent spatial correlation by introducing a correlation matrix in the noise that drives the vector autoregressive process. We present simulation results for two wind integration case studies: the California power system (2012 and 2020 wind integration targets) and the German power system (2012 wind integration levels). We observe that the fitness of the model to the data depends largely on the accuracy of the piecewise approximation of the power curve. Increasing the number of regions improves the accuracy of the power curve approximation, at the cost of increasing the size of the correlation matrix that drives the wind speed process.

5. References

- [1] Ruiz, P. A.; Philbrick, R. C.; Sauer, P. W. Wind Power Day-Ahead Uncertainty Management through Stochastic UC Policies. In Power Systems Conference and Exposition, March 2009, pp. 1-9.
- [2] Wang, J.; Shahidehpour M; Li, Z. Security-Constrained Unit Commitment with Volatile Wind Power Generation. IEEE Transactions on Power Systems. August 2008, 23:3, pp. 1319-1327.
- [3] Constantinescu, E. M.; Zavala V. M.; Rocklin, M.; Lee S.; Anitescu M. A Computational Framework for Uncertainty Quantification and Stochastic Optimization in Unit Commitment with Wind Power Generation. IEEE Transactions on Power Systems. February 2011, 26:1, pp. 431-441.
- [4] Tuohy, A.; Meibom P.; Denny, E.; O'Malley M. Unit Commitment for Systems with High Wind Penetration. IEEE Transactions on Power Systems. May 2009, 24:2, pp. 592-601.
- [5] Morales, J. M.; Conejo, A. J.; Perez-Ruiz J. Economic Valuation of Reserves in Power Systems with High Penetration of Wind Power. IEEE Transactions on Power Systems. May 2009, 24:2, pp. 900-910.
- [6] Bouffard, F.; Galiana, F. D. Stochastic Security for Operations Planning with Significant Wind Power Generation. IEEE Transactions on Power Systems. May 2008, 23:2, pp. 306-316.
- [7] Papavasiliou, A.; Oren, S. S.; O'Neill, R. P. Reserve Requirements for Wind Power Integration: A Scenario-Based Stochastic Programming Framework. IEEE Transactions on Power Systems. November 2011, 26:4, pp. 2197-2206.
- [8] Papavasiliou, A.; Oren, S. S. Multi-Area Stochastic Unit Commitment for High Wind Penetration in a Transmission Constrained Network. Operations Research. May/June 2013, 61:3, pp. 578-592.
- [9] Arroyo, J. M.; Galiana, F. D. Energy and Reserve Pricing in Security and Network-Constrained Electricity Markets. IEEE Transactions on Power Systems. May 2005, 20:2, pp. 634-643.
- [10] Galiana, F. D.; Bouffard, F.; Arroyo J. M.; Restrepo J. F. Scheduling and Pricing of Coupled Energy and Primary, Secondary, and Tertiary Reserves. Proceeding of the IEEE. November 2005, 93:11, pp. 1970-1983.
- [11] Bouffard, F.; Galiana, F. D.; Conejo A. J. Market-Clearing with Stochastic Security. IEEE Transactions on Power Systems. November 2005, 20:4, pp. 1827-1835.
- [12] Brown, B. G.; Katz, R. W.; Murphy A. H. Time Series Models to Simulate and Forecast Wind Speed and Wind Power. Journal of Climate and Applied Meteorology. 1984, 23, pp. 1184-1195.
- [13] Box, G. E. P; Jenkins, G. M. Time Series Analysis: Forecasting and Control; Publisher: Holden-Day, CA, USA, 1976.
- [14] Torres, J. L.; Garcia, A.; Blas M. D.; Francisco A. D. Forecast of Hourly Wind Speed with ARMA Models in Navarre (Spain). Solar Energy. July 2005, 79:1, pp. 65-77.

- [15] Morales, J. M.; Minguez, R.; Conejo A. J. A Methodology to Generate Statistically Dependent Wind Speed Scenarios. *Applied Energy*. 2010, 87, pp. 843-855.
- [16] Callaway, D. Sequential Reliability Forecasting for Wind Energy: Temperature Dependence and Probability Distributions. *IEEE Transactions on Energy Conversion*. June 2010, 25, pp. 577-585.
- [17] Yu, N.-P.; Liu, C.-C.; Price, J. Evaluation of Market Rules Using a Multi-Agent System Method. *IEEE Transactions on Power Systems*. February 2010, 25:1, pp. 470-479.
- [18] Potter, C. W.; Lew D.; McCaa, J.; Cheng, S.; Eichelberger, S.; Grimit, E. Creating the Dataset for the Western Wind and Solar Integration Study (U. S. A.). In 7th International Workshop on Large Scale Integration of Wind Power and on Transmission Networks for Offshore Wind Farms, Madrid, Spain, May 2008.
- [19] California Independent System Operator, “The California ISO Controlled Grid Generation Queue as of January 8, 2010”, 2010.
- [20] European Energy Exchange AG, Transparency in Energy Markets, [online] <http://www.transparency.eex.com/>
- [21] Deutscher Wetterdienst (National Meteorological Service, Germany), WebWerdis (Weather Request and Distribution System), [online] <https://werdis.dwd.de/>
- [22] European Wind Energy Association, “WP2.6 – Equivalent Wind Power Curves”. Prepared by J R McLean, July 2008.