Stochastic Modeling of
Multi-Area Wind Power Production

Anthony Papavasiliou
Department of Industrial Engineering and
Operations Research
University of California at Berkeley
Berkeley, CA 94709
Email: tonypap@berkeley.edu

Shmuel S. Oren
Department of Industrial Engineering and
Operations Research
University of California at Berkeley
Berkeley, CA 94709
Email: oren@ieor.berkeley.edu

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 a case study of the California power system.

Keywords - **Wind power generation, time series analysis, power system simulation.**

I. 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]. 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 inter-tie 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, inter-tie

constraints limit the amount of power that can flow over inter-ties 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. The complexity of committing reserves in the presence of transmission constraints has been demonstrated by various authors, including Arroyo and Galiana [8], Galiana et al. [9] and Bouffard et al. [10]. 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 a detailed dataset of the California wind power resources corresponding to the 2012 and 2020 Renewable Portfolio Standards. The remainder of the paper is structured as follows. In Section II we summarize relevant literature and present the methodology for calibrating the multi-area wind production model and for simulating the process. In Section III we present a case study of the California power system. In Section IV we summarize the conclusions of our study.

II. 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. Early work on wind power modeling was performed by Brown et al. [11]. 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 [12] to estimate the parameters of the autoregressive model. Torres et al. [13] follow the same methodology as Brown et al. [11]. 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. In recent work, Morales et al. [14] 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. [14] in order to account for seasonal and diurnal wind speed patterns.

A. Calibration

Given a multi-area 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, seasonal, or may even vary between weekdays and weekends as is the case for load data. 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 stationary data set y_{kt}^{GS}. Brown et al. [11], Torres et al. [13] and Morales et al. [14] use this approach for transforming Weibull-distributed wind speed data to Gaussian data, and Callaway [15] 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, no single parametric distribution provides a close fit for the observed data in all locations, therefore we fit an empirical 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:

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

where $\hat{\omega}_{kt}$ is the estimated noise and $\hat{\phi}_{kj}$, $j \in \{1, \ldots, p\}$, are the estimated coefficients of the autoregressive model. The calibration process is summarized in the following steps:

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

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

where y_{kt} is the data, y_{kt}^S is the transformed stationary 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:

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

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 $\hat{F}_k$ is the cumulative function of the (parametric or non-parametric) fit for the data in location k.

Step (c). Use the Yule-Walker equations [12] to estimate the autoregressive parameters $\hat{\phi}_{kj}$ and covariance matrix $\hat{\Sigma}$ of the residual noise obtained from Eq. (1).

B. 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:

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

where $\Phi = (\phi_{kj}), k \in \{1, \ldots, K\}, j \in \{1, \cdots, p\}$, is the matrix of autoregressive parameters and $(\omega_{kt}), k \in \{1, \cdots, K\}$, are independent, identically distributed, multivariate Gaussian random variables with mean 0 and covariance matrix Σ. 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 and variance.

$$Y_{k,t+1}^{GS} = \sum_{j=0}^{p} \hat{\phi}_{kj} Y_{k,t-j}^{GS} + \omega_{kt}, \tag{5}$$
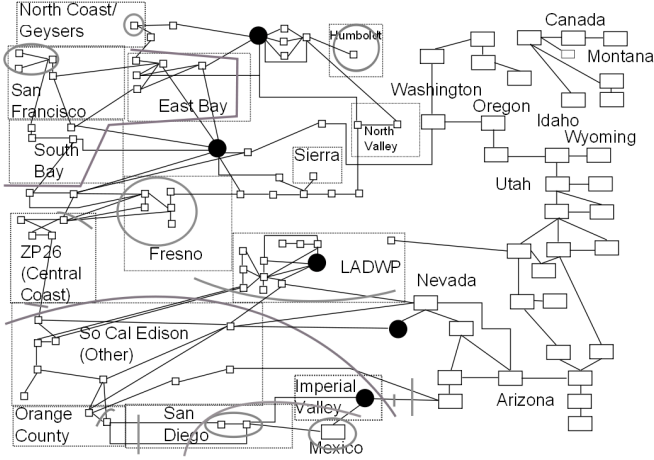
Fig. 1. A schematic of the WECC model.

TABLE I
CURRENT AND PROJECTED CAPACITY OF WIND POWER INSTALLATIONS
(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

where $\omega_{kt} = (\hat{L}\omega)_k$, ω are independent standard normal random vectors with K entries, $\hat{L}$ is the Cholesky factorization of $\hat{\Sigma}$ and Y_{kt}^{GS} is the Gaussian stationary autoregressive process for location k.

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

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

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

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

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 curve for each location to simulate wind power production:

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

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

III. RESULTS

We use the multi-area wind production model to study the impacts of large-scale renewable energy integration in the California power system. We use a model of the California ISO with imports from the Western Electricity Coordinating Council (WECC) that is also used by Yu et al. [16] and is described in detail by Papavasiliou et al. [7].

A. Data

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. [17]. 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 [18]), 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 I we present the location of existing wind generation capacity, as well as capacity for the moderate and deep integration cases.

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 thick solid line intersects a set of transmission lines over which the total amount of power cannot exceed a certain limit. The wind generators of Table I 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.

B. Results

In Fig. 2 we present the approximate power curve and the fit of the complementary cumulative probability distribution of wind output to the data for the Altamont area for both integration studies. The corresponding results for Solano and Tehachapi are presented in Figs. 3, 4 respectively. In Figs. 5, 6 we present results for the deep integration case for Clark County and Imperial Valley respectively.

As the figures indicate, the primary source of discrepancy in the model is the approximate power curve. Note that 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
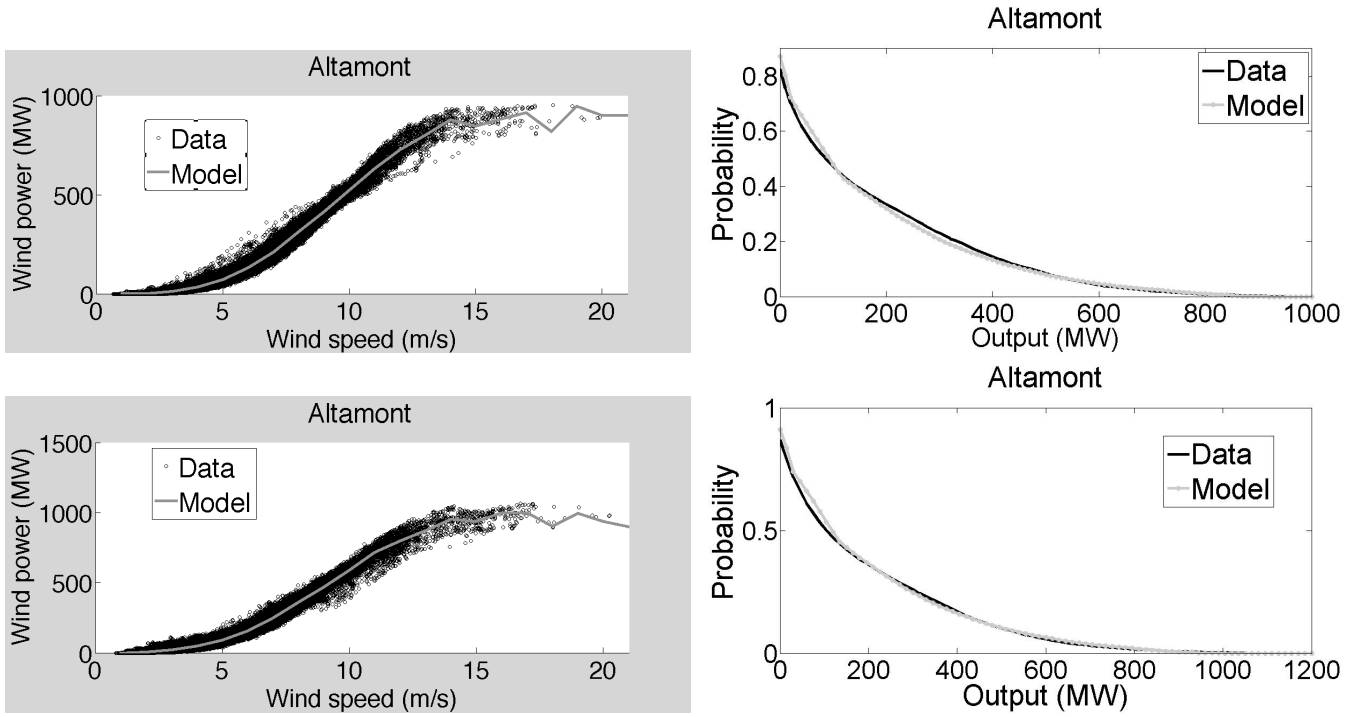
Fig. 2. In reading order: power curves (left) and complementary cumulative probability distribution of wind output (right) for Altamont for the moderate (up) and deep (down) integration study.
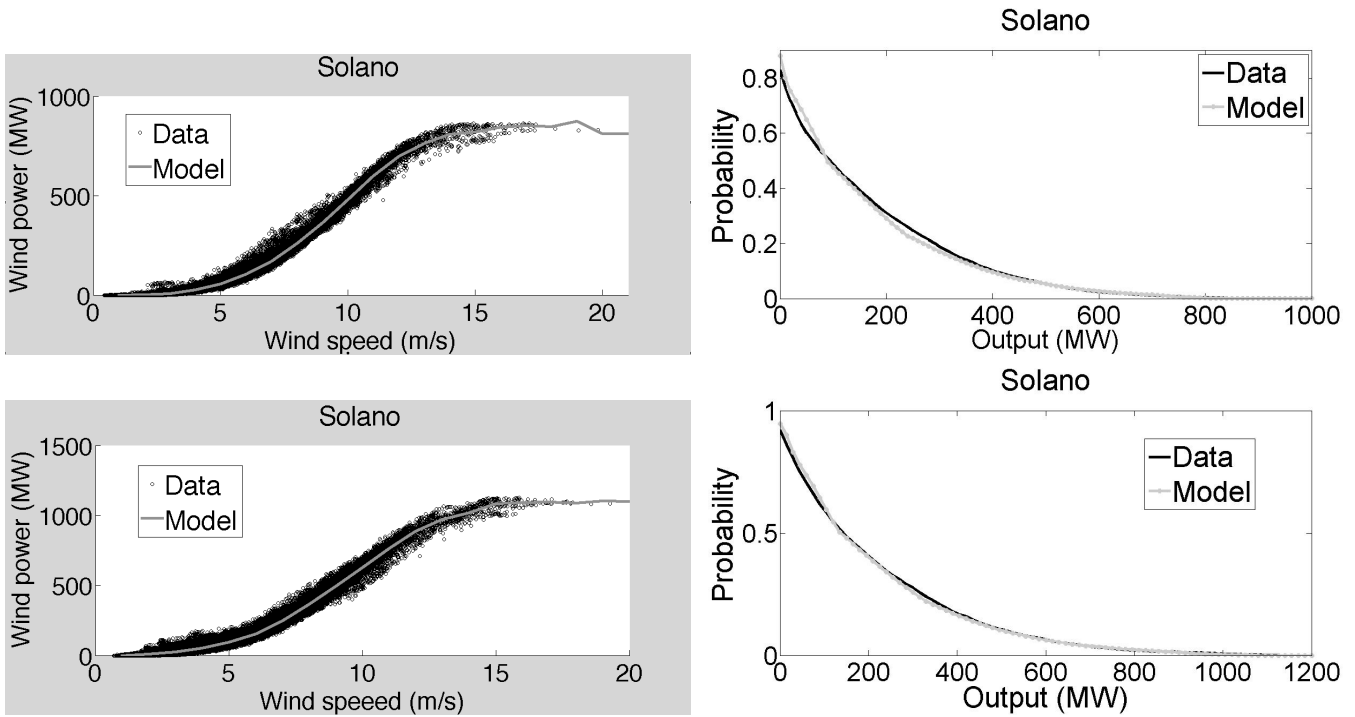

Fig. 3. In reading order: power curves (left) and complementary cumulative probability distribution of wind output (right) for Solano for the moderate (up) and deep (down) integration study.
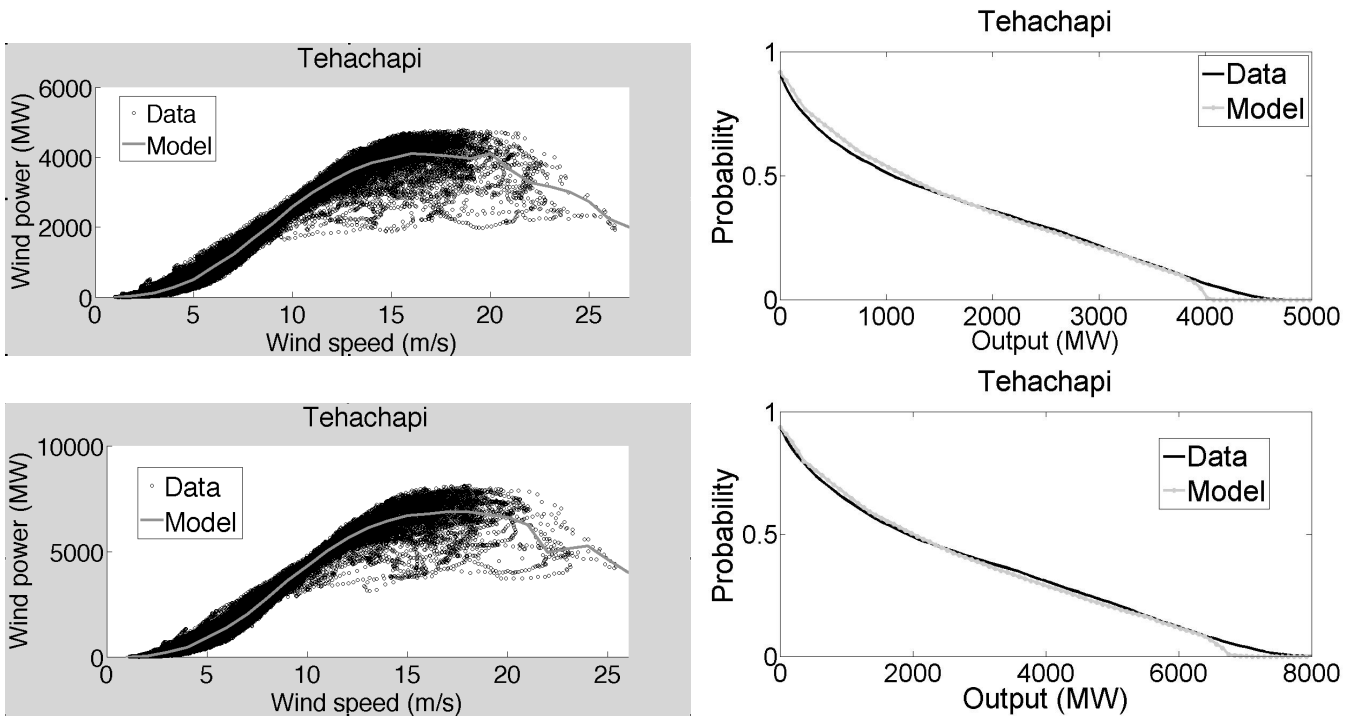
Fig. 4.   In reading order: power curves (left) and complementary cumulative probability distribution of wind output (right) for Tehachapi for the moderate (up) and deep (down) integration study.
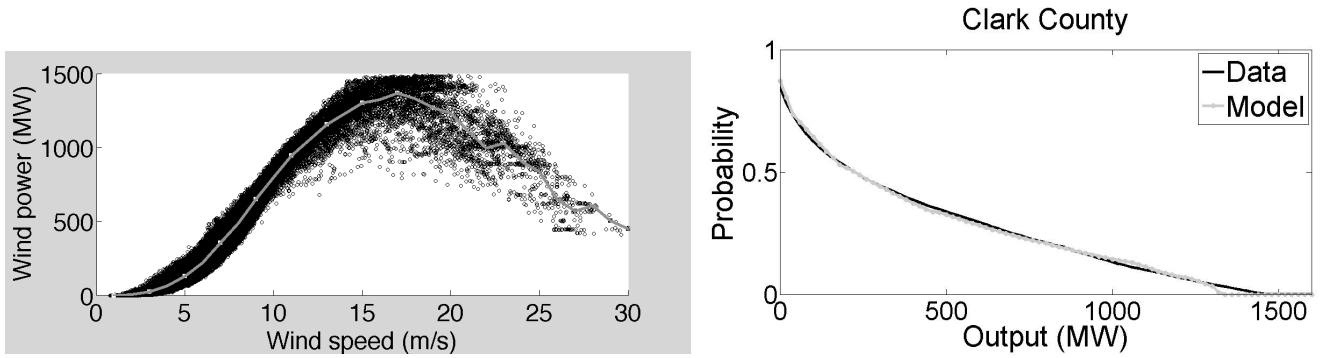

Fig. 5.   In reading order: power curves (left) and complementary cumulative probability distribution of wind output (right) for Clark County for the deep integration study.
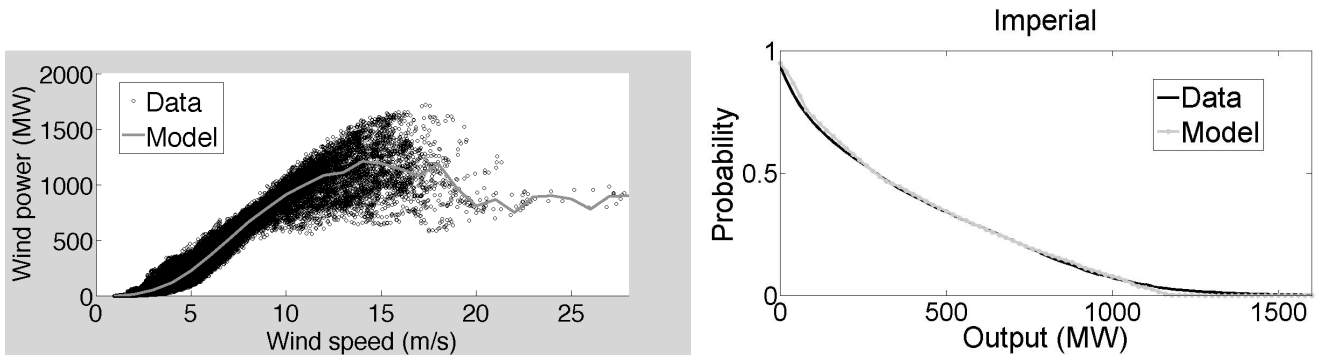

Fig. 6.   In reading order: power curves (left) and complementary cumulative probability distribution of wind output (right) for Imperial Valley for the deep integration study.

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. 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.

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 accurately depicted 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.

IV. 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 studies of the California power system that correspond to the wind integration targets of 2012 and 2020. We observe that the fitness of the model to the data depends largely on the accuracy of the piecewise linear 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.

REFERENCES

[1] P. A. Ruiz, R. C. Philbrick, and P. W. Sauer, "Wind power day-ahead uncertainty management through stochastic uc policies," in *Power Systems Conference and Exposition*, March 2009, pp. 1–9.

[2] J. Wang, M. Shahidehpour, and Z. Li, "Security-constrained unit commitment with volatile wind power generation," *IEEE Transactions on Power Systems*, vol. 23, no. 3, pp. 1319–1327, August 2008.

[3] E. M. Constantinescu, V. M. Zavala, M. Rocklin, S. Lee, and M. Anitescu, "A computational framework for uncertainty quantification and stochastic optimization in unit commitment with wind power generation," *IEEE Transactions on Power Systems*, vol. 26, no. 1, pp. 431–441, February 2011.

[4] A. Tuohy, P. Meibom, E. Denny, and M. O'Malley, "Unit commitment for systems with high wind penetration," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 592–601, May 2009.

[5] J. M. Morales, A. J. Conejo, and J. Perez-Ruiz, "Economic valuation of reserves in power systems with high penetration of wind power," *IEEE Transactions on Power Systems*, vol. 24, no. 2, pp. 900–910, May 2009.

[6] F. Bouffard and F. D. Galiana, "Stochastic security for operations planning with significant wind power generation," *IEEE Transactions on Power Systems*, vol. 23, no. 2, pp. 306–316, May 2008.

[7] A. Papavasiliou, S. S. Oren, and R. P. O'Neill, "Reserve requirements for wind power integration: A scenario-based stochastic programming framework," *IEEE Transactions on Power Systems*, vol. 26, no. 4, pp. 2197–2206, November 2011.

[8] J. M. Arroyo and F. D. Galiana, "Energy and reserve pricing in security and network-constrained electricity markets," *IEEE Transactions on Power Systems*, vol. 20, no. 2, pp. 634–643, May 2005.

[9] F. D. Galianaa, F. Bouffard, J. M. Arroyo, and J. F. Restrepo., "Scheduling and pricing of coupled energy and primary, secondary, and tertiary reserves," *Proceedings of the IEEE*, vol. 93, no. 11, pp. 1970–1983, November 2005.

[10] F. Bouffard, F. D. Galiana, and A. J. Conejo., "Market-clearing with stochastic security," *IEEE Transactions on Power Systems*, vol. 20, no. 4, pp. 1827–1835, November 2005.

[11] B. G. Brown, R. W. Katz, and A. H. Murhpy, "Time series models to simulate and forecast wind speed and wind power," *Journal of Climate and Applied Meteorology*, vol. 23, pp. 1184–1195, 1984.

[12] G. E. P. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. San Francisco, CA: Holden-Day, 1976.

[13] J. L. Torres, A. Garcia, M. D. Blas, and A. D. Francisco, "Forecast of hourly wind speed with ARMA models in Navarre (Spain)," *Solar Energy*, vol. 79, no. 1, pp. 65–77, July 2005.

[14] J. M. Morales, R. Minguez, and A. J. Conejo, "A methodology to generate statistically dependent wind speed scenarios," *Applied Energy*, vol. 87, pp. 843–855, 2010.

[15] D. Callaway, "Sequential reliability forecasting for wind energy: Temperature dependence and probability distributions," *IEEE Transactions on Energy Conversion*, vol. 25, pp. 577–585, June 2010.

[16] N.-P. Yu, C.-C. Liu, and J. Price, "Evaluation of market rules using a multi-agent system method," *IEEE Transactions on Power Systems*, vol. 25, pp. 470–479, February 2010.

[17] C. W. Potter, D. Lew, J. McCaa, S. Cheng, S. Eichelberger, and E. Grimit, "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.

[18] CAISO, "The California ISO controlled grid generation queue as of January 8, 2010," 2010.