**Proceedings of the ASME 2010 4th International Conference of Energy Sustainability**
**ES2010**
**May 17-22, 2010, Phoenix, Arizona, USA**

# ES2010-90441

# MODELING OF WIND POWER PRODUCTION FORECAST ERRORS FOR WIND INTEGRATION STUDIES

**Jason J. Kemper**
MS Candidate,
Mechanical Engineering
Northern Arizona University
Flagstaff, AZ, USA

**Mark F. Bielecki**
MS Candidate,
Mechanical Engineering
Northern Arizona University
Flagstaff, AZ, USA

**Thomas L. Acker**
Professor of Mechanical Engineering
Northern Arizona University
Flagstaff, AZ, USA

## ABSTRACT

In wind integration studies, accurate representations of the wind power output from potential wind power plants and corresponding representations of wind power forecasts are needed, and typically used in a production cost simulation. Two methods for generating "synthetic" wind power forecasts that capture the statistical trends and characteristics found in commercial forecasting techniques are presented. These two methods are based on auto-regressive moving average (ARMA) models and the Markov random walk method.

Statistical criteria are suggested for evaluation of wind power forecast performance, and both synthetic forecast methods proposed are evaluated quantitatively and qualitatively. The forecast performance is then compared with a commercial forecast used for an operational wind power plant in the Northwestern United States evaluated using the same statistical performance measures. These quantitative evaluation parameters are monitored during specific months of the year, during rapid ramping events, and at all times.

The best ARMA based models failed to replicate the auto-regressive decay of forecast errors associated with commercial forecasts. A modification to the Markov method, consisting of adding a dimension to the state transition array, allowed the forecast time series to depend on multiple inputs. This improvement lowered the artificial variability in the original time series. The overall performance of this method was better than for the ARMA based models, and provides a suitable technique for use in creating a synthetic wind forecast for a wind integration study.

## INTRODUCTION

As utilities and system planners consider adding more wind power to the electrical grid system, wind integration studies are performed in order to understand and predict the influence and economic impacts of wind power on power system operations. For such studies, the wind power is typically modeled using numerical weather prediction (NWP) techniques. The creation of a corresponding synthetic wind power forecast, which captures the elements of a commercial forecast, is challenging, and a technique that permits some control over key forecast performance parameters is desired.

The uncertainty in predicting the wind exacerbates the intrinsic variability and uncertainty already present in planning and operating a power system, increasing the need for flexible generation resources and reserves in order to balance generation with load. Load-serving entities (LSEs) are learning how to accommodate this form of power generation. Since the costs associated with integrating wind do not scale linearly with increasing penetration and these costs differ depending on both the balancing area's system attributes (e.g., generation mix, load profile) and the proposed system of wind power attributes (e.g., capacity, geographic diversity, turbine model and characteristics), responsible LSEs seek to quantify these costs and to ensure the necessary steps are being taken to obviate the challenges associated with wind power integration. The challenge to integrating wind power is two-fold: more so than conventional generation, generating electricity from wind power produces a variable and uncertain power signal. The costs and challenges associated with managing this variability and uncertainty are collectively known as *integration costs*. As recent developments have proven, methods for quantifying these costs are important topics of debate that will shape the future of wind power development[1].

[1] The Bonneville Power Administration in the United States has promised to revisit the methodology used in the decision to charge wind generators penalties for persistent deviations of four or more consecutive hours and $5.70/MWh for intra-hour balancing in the 2010 rate case. The NYISO, MISO, PJMISO, ERCOT, and CAISO have all begun to address the uncertainty issue, either by establishing centralized forecasting or requiring wind generators to supply wind forecasts to scheduling entities.

There are generally two hindrances to estimating the integration costs: realistic dispatch simulation *modeling* and realistic *inputs* to these models. Unit commitment models are used in wind integration studies to establish the generation resources and the level of reserves necessary to serve the power system load which is typically estimated, first the day before, and then a few hours before their actualization to ensure that generation and load are balanced in real-time. These increasingly sophisticated models use (typically proprietary) algorithms that determine which units should be committed to run and may be called on to provide more (or less) energy as needed. If insufficient flexibility in the power system is scheduled, power system reliability is jeopardized; conversely, if a generator is unnecessarily started, then avoidable costs are incurred. Since all forms of generation have unique constraints, such as ramping rates, it is critical that these and dispatch models be capable of representing their characteristics. Most models have very refined algorithms that can portray the complexities of the scheduling process, at least when applied to conventional generators. However, incorporating variable and uncertain energy sources, such as wind, in these models is an active area of research; improvements in developing these models to that end are identified and discussed in [1].

Among the inputs necessary to simulate power system operations containing wind energy are chronologies of expected (forecasted) and actualized (modeled or historical, actual) wind power, hydropower, load, and market prices for energy transactions. Since most of these can be correlated with observations of weather, especially of temperature, it is important that they be temporally synchronized so that cross-correlations between them are realistically portrayed. In a wind integration study, the actual wind power production from proposed wind facilities is typically modeled using NWP techniques to estimate wind speeds which are then converted to estimated power. Since NWP techniques are deterministic and the historically observed initial conditions have already been used in the power output model, using NWP techniques to *also* generate a wind forecast, dubbed a "hind-cast", that possesses similar error characteristics to a typical commercial wind forecast is difficult. By using a different set of boundary conditions to update the NWP models, generating a hind-cast becomes possible. This is the approach described in [2]. However, the computational effort and time to create a wind power simulation based upon an NWP is not trivial, and the use of an NWP to create a hind-cast increases this. Furthermore, whether the difference between two deterministic model-runs using different input data sets actually captures the uncertainty associated with a state-of-the-art forecast rather than perpetuating or nullifying systematic errors inherent in such approximations merits further research. Therefore, an alternative approach that permits some control over key forecast performance parameters is desired.

One approach to quantifying the uncertainty component of wind integration costs is to determine a range of integration costs associated with a range of wind forecast errors [3]. An upper limit to forecast error may be assumed to result from a persistence forecast whereby future realizations of wind production are expected to be identical to present observations; a lower limit of no error would result from a perfect wind forecast whereby production is known in advance with no uncertainty. This approach requires running a dispatch model multiple times. The cost of integrating wind utilizing a state-of-the-art forecast would never be estimated but a range of integration costs would be established.

As mentioned, the computational effort to model wind using NWP techniques for large regions is quite high. A description of the efforts to model the Eastern US is telling: storage requirements measured in terabytes and computational time in weeks [4]. Therefore, an alternative approach was adopted where a synthetic forecast was used in the Eastern Wind Integration and Transmission Study (EWITS[2]). This approach, based on a Markov chain, is briefly described in [5].

The Markov chain approach was first applied to wind for use in generating wind speed time series data [6]. This was also described in [7] with a note of caution that such an approach may not capture seasonal or diurnal patterns without modifications. The key to producing a simulation with this approach is developing a probability matrix whereby a desired distribution, or set of observations, relates columns representing future wind speed bins to rows representing current wind speed bins. This probability matrix is converted to a cumulative probability matrix by summing entries horizontally. This cumulative probability matrix can then be used with a random number generator to initiate a "random walk" where the randomly selected number is matched to the appropriate row and column to determine the future wind speed bin.

## COMMERCIAL WIND POWER FORECASTING

There are generally two approaches to forecasting the wind: the physical approach and the statistical approach. More comprehensive reviews of the literature are found in [8;9]; however, a brief summary of these approaches follows.

The physical approach begins with interpolating and downscaling very coarse resolution rawinsonde observations which are then used as boundary conditions. These observations are measured and disseminated by national weather prediction centers. If available, additional data from other sources, such as meteorological masts, can be brought in at this stage as well. The major constraints with the radiosonde data are the infrequent releases (every 12 hours) and very coarse spatial resolution of these releases. With this input, a discretized mesh portrayal of the atmosphere is iteratively stepped forward in time as determined by numerical approximations to the solutions to the governing equations for convective flows: the Navier-Stokes equations. The output is saved every several time steps (e.g., hourly or every 10 minutes), and the NWP process proceeds until the next series of boundary condition updates become available. The final step in this approach is converting wind speed to wind power. Evolving, site-specific, proprietary approaches to ensemble

[2] See http://www.nrel.gov/wind/systemsintegration/ewits.html

combinations of NWP models and use of modified output statistics (MOS) can improve these forecasts further.

In contrast to the deterministic output of unaltered (i.e., prior to MOS and ensemble combinations) NWP results, there is the statistical approach. There are several types of statistical forecasts and variations on each. Kalman filters, ARMA models, and neural networks are some statistical methods that have been used in wind forecasting.

ARMA models, also called Box-Jenkin's models, are a linear and deterministic class of statistical forecasts. There are many different types of Box-Jenkin's models [10] that can be used to extract some of the embedded patterns in a time series. These models typically begin with an autoregressive term weighted by a correlation coefficient at the relevant time lag and then add a Gaussian white noise term as well as other terms such as a moving average of residuals. The challenge in forecasting with ARMA models is determining the best model order (p,q) to use and then determining the coefficients. The model order refers to the number of recent observation terms to include for both the auto-regression of past observations (order p) and the moving average of residuals or past errors (order q); therefore, ARMA models are said to be of order (p,q), where p and q are integers. The model coefficients can be determined by one of numerous methods such as a least squares method. The ARMA equation for predicted power, $\hat{P}$, is

$$\hat{P}_t = \sum_{i=1}^{p} a_i P_{t-i} + \sum_{j=0}^{q} b_j e_{t-j} \quad (1)$$

where e is the error or residual term and a and b are the coefficients for autoregressive terms and moving average terms, respectively.

In [11], several ARMA models were tested at sites in Minnesota, Iowa, and Oregon, and the performance of each model was found to be sensitive to the season or month of the year. Thus, adjusting the model order and corresponding coefficient values depending on which performs best for each "training period" is prudent. Variation of the most appropriate coefficients on an even shorter (diurnal) cycle and the possibility of stochastically variable coefficients are discussed in [12].

The commercial wind power forecasting process is synergistic, relying primarily on statistical approaches for the very short term and primarily on NWP for forecasts several (i.e., 6-8+) hours in advance. This transition can be clearly seen in graphs of forecast metrics as a function of forecast horizon [13] and is called the "best intersection point" in [14]. For intermediate horizons, the forecasting process will involve a combination of the two approaches.

## METHODOLOGY

In order to describe the characteristics of wind power uncertainty and build a synthetic forecast, two sets of data were analyzed. These were time series histories of the power output from a wind power plant in the Northwestern United States, and a commercial wind power forecast for the same wind power plant. The time series reflect the amount of wind energy reaching the point of interconnection with the balancing area to where the wind power was delivered. The installed rated capacity of the wind power plant was in excess of 60 MW from turbines in excess of 1 MW each.

The forecasts for the production of this wind power were produced by a leading commercial provider of wind power forecasting. Hourly averages of the forecast wind power production were provided for analysis.

Of special interest to any wind integration study is the day-ahead forecast and its implications at the unit commitment time frame. Such a forecast is generated about one day prior to the forecast period for which it is intended, typically at an early morning cut-off hour. This allows a utility to finalize purchase agreements and schedule generation based on the expected load demands for the following day. The period of time between the cut-off hour and the beginning of the next day's operations are known as the forecast closure window.

As used here then, the term "day-ahead" is not synonymous with 24 hours ahead. The day-ahead forecasts used in this study were generated once a day at 6AM for the following day. The day-ahead forecast is a forecast for hourly averages of power output rather than 24 hour blocks of energy.

A similar forecast closure window is needed to resolve generation scheduling and energy transactions for next hour dispatch optimization. Depending on the LSE, wind and load forecasts for "next hour" dispatch decisions are needed 2-6 hours ahead.

Determination of coefficients for ARMA models was accomplished using Matlab and the supplemental econometrics toolbox. Evaluation criteria of resulting models of various orders were then compared for important forecast horizons. The best performing models for each forecast horizon were selected and the deterministic time series were recorded for future analysis or use in a synthetic forecast ensemble. However, prior to saving the time series, a sub-program was used to reduce predictions that exceeded the nameplate capacity of the wind power plant and increase predictions that fell below zero.

Since commercial forecasts outperform ARMA models that are produced with equivalent lead times, attempts at generating a synthetic forecast from an ARMA model required a weighted average combination of ARMA forecasts from different horizons. For example, a six-hour ahead synthetic forecast was produced by combining the three-hour ahead ARMA (1,0) model with the six-hour ahead (10,1) model. At longer horizons ARMA models tend towards predicting the mean, and combining different models compounds this range constraint. To counter this in the synthetic day-ahead forecast, the forecast values were expanded about the mean. Thus, values above the mean would be increased by an amount proportional to the difference between the largest forecast value and nameplate capacity.

The other approaches to synthesizing a forecast were Monte Carlo simulations containing many of the properties of a Markov chain. The first of these was built by determining how probability histograms of forecast power were functions of the actual power, past values of power, and the forecast horizon. Since this method for building a probability array contained a

dimension that was based on past observations, it violates the Markov property that future realizations depend only on the current state; therefore, this approach may be considered "pseudo-Markovian". The second of these was built by determining how probability histograms of future forecasts were functions of the current forecast, current actual power, and the forecast horizon. In Figure 1 a probability transition array divided coarsely into quintiles is shown.

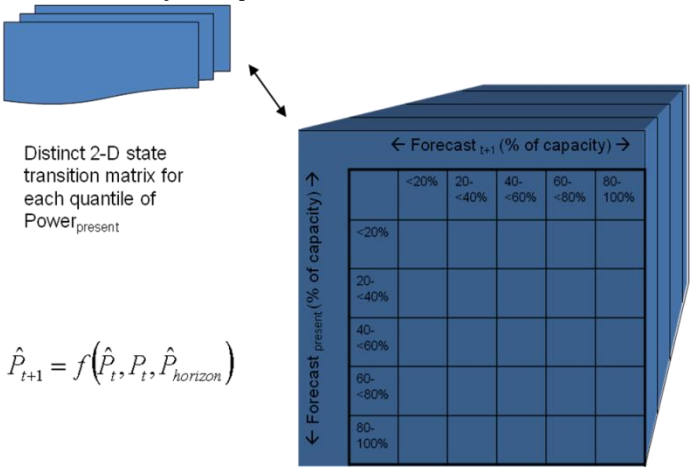


**Figure 1: 3-Dimensional Markov Transition Array**

Since the layout of these transition tables and their probability histograms are also functions of quantiles of current actual power, it is appropriate to conceive of the transition array as being 3-dimensional with the 3rd dimension determined by the current actual power. The dimensions of the array used in building the synthetic forecast were determined empirically based on model performance but were much finer than shown above (e.g., 30x30x30). A different matrix was constructed for each forecast horizon, conceivably adding another dimension.

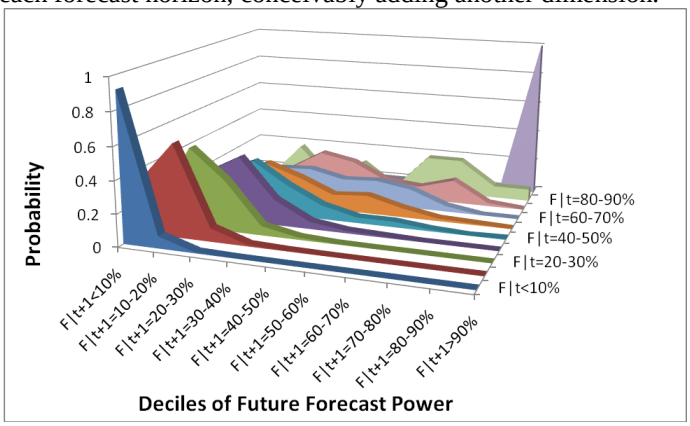


**Figure 2: Possible Probability Curves**

After these multi-dimensional arrays were built, they were used in a random walk process to generate a contiguous time series: the synthetic forecast. This was achieved by iteratively selecting a cumulative probability curve or row from this array. As an example, all of the measured power data was divided into current power quintiles and all of the 2-hour forecasts associated with these actual power measurements were divided into current forecast deciles. Then, if the current power production was in the 1st quintile (i.e., <20% of capacity), the current forecast decile would determine the selection of the possible probability curves shown in Figure 2. Note the strong shift in the skewness of the probability curves depending on the decile of current forecast. If the current forecast lay in the fifth decile (i.e., 40-50%) then the probability that a forecast for the next hour would belong to a particular decile is shown in Table 1. This converts to a cumulative probability as shown in Table 2. For this example, 33% of the next hour's forecasts were between 20-30% of capacity and 91% of the next hour's forecasts were below 50% of capacity.

**Table 1: Selected Probability Row**

| Deciles of Forecast Power for Next Hour | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1st | 2nd | 3rd | 4th | 5th | 6th | 7th | 8th | 9th | 10th |
| 5% | 24% | 33% | 20% | 9% | 4% | 4% | 1% | 0% | 1% |

**Table 2: Selected Cumulative Probability Row**

| Deciles of Forecast Power for Next Hour | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 1st | 2nd | 3rd | 4th | 5th | 6th | 7th | 8th | 9th | 10th |
| 5% | 29% | 63% | 82% | 91% | 95% | 98% | 99% | 99% | 100% |

Another step was necessary before proceeding to the next iteration to prevent this process from producing an artificially random series. Between 13-19% of the time (depending on the forecast horizon) the synthetic forecast value would be averaged with the actual power value. This step was not necessary with the pseudo-Markovian algorithm, which utilized different independent variables in the transition array. However, the pseudo-Markovian synthetic model required several extra post-processing steps to "smooth" the forecast output. These steps were relaxation methods that often caused other unintended consequences. More details of both methods are provided in [15].

## EVALUATION CRITERIA

**Table 3: Evaluation Criteria**

| | | | | |
|---|---|---|---|---|
| Primary Criteria | Quantitative | Qualitative | Secondary Criteria (Ramp Specific) | i) # Ramps |
| | 1a) MAE | 1) Auto-Regression of Errors | | ii) Mean Ramp Duration |
| | 1b) RMSE | | | |
| | 2) R | 2) Distribution of Power | | iii) Ramp Forecast Accuracy Parameters |
| | 3a) $\sigma_\Delta$ | | | |
| | 3b) Mean $\lvert\Delta\rvert$ | 3) Distribution of $\Delta$ | | |
| | 4a) Mean Bias | 4) Distribution of Errors | | |
| | 4b) $\bar{\hat{P}}$ | | | |
| | 5) $\sigma_p$ | 5) Auto-Regression | | |

Whatever approach is used to generate a synthetic forecast, it is important that the auto-regression (i.e., level of persistence) of wind power errors and cross-correlation with load forecast errors be reproduced. As shown in [16], two synthetic wind forecasts with identical mean absolute errors (MAE) can produce drastically different estimates of wind integration costs. For this reason, efforts at generating a synthetic forecast were evaluated for temporal trends as well as by other evaluation metrics: some qualitative in nature and some focused on ramp performance criteria. Table 3 lists the primary quantitative and qualitative evaluation criteria used, as well as secondary criteria related to ramp events.

R is the correlation coefficient between the forecast and measured wind power. In the following criterion definitions, N is the total number of hours in the time series.

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{P}_i - P_i\right| \quad (2)$$

$$Root\ Mean\ Square\ Error(RMSE) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{P}_i - P_i)} \quad (3)$$

$$Mean\ Bias = \frac{1}{N}\sum_{i=1}^{N}(\hat{P}_i - P_i) \quad (4)$$

$$\sigma_{\hat{P}} = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(\hat{P}_i - \bar{\hat{P}})^2} \quad (5)$$

By replacing the mean forecast power, i.e., $\bar{\hat{P}}$, with the step change from the previous period ($\Delta$), $\sigma_\Delta$ can be inferred from equation (5) with the following definition of a step change:

$$\Delta = \hat{P}_t - \hat{P}_{t-1} \quad (6)$$

Two approaches were used to evaluate the performance of the forecasts during critical times. First, the MAE was averaged during each month and hour to identify temporal trends in the errors. Secondly, ramp event analysis of the number, duration and accuracy of ramps (rapid changes in production) was performed. Since no standard exists for the definition of a rapid ramp event, this analysis required a ramp identification algorithm. The method used to discover ramps identified just less than one ramp per day. Once a ramp was identified, it was defined to continue until the change in hourly production ($\Delta$) switched signs. The ramp identification process used is explained in more detail in [13].

In [17] two metrics of the accuracy of ramp event forecasting are introduced. However, their analysis oversimplifies the possible outcomes of a forecasted ramp. An additional possibility not explicitly discussed in [17] is the possibility that a forecasted up ramp is associated with a measured down ramp. This would be an expensive proposition that could compromise system reliability. Therefore, with this inclusion, Table 4 conveys the contingencies of ramp forecasts.

**Table 4: 2x3 Forecast Ramp Contingency Table**

| | Possible Outcomes of Forecast Ramp | | |
|---|---|---|---|
| | Actual Ramp of Opposing Sign | Actual Ramp of Same Sign | No Actual Ramp |
| Forecast Ramp | Misleading Forecast | True Forecast | False Forecast |
| No Forecast Ramp | Missed Ramp | Missed Ramp | No Event |

With minor modifications, the definitions used in [17] are presented here:

$$forecast\ accuracy = true\ forecasts/total\ forecast\ ramps \quad (7)$$

$$ramp\ capture = true\ forecasts/total\ measured\ ramps \quad (8)$$

However, the analysis presented here considered only a 9 hour window (+/-4 hrs) of ramp starts for correlations between forecast and measured ramps as opposed to the generous 25 hour window in [17]. In addition to (7) and (8), two more ramp performance criteria that consider the possibility of a misleading forecast are the ratio of true forecasts to misleading forecasts and forecast uncertainty:

$$forecast\ uncertainty = misleading\ forecasts/total\ forecast\ ramps \quad (9)$$

## RESULTS

By combining ARMA time series using a weighted average of the desired forecast horizon and those from shorter lead times, time series with reasonable quantitative evaluation criterion values were produced. However, after several hours the standard deviation of forecast power was lower than seen in the commercial forecast.

Additionally, the ARMA models produced a synthetic forecast that poorly emulated probability distribution function (PDF) of the commercial forecast. The distribution of ARMA forecast power production was strongly limited at intermediate forecast horizons (more than six hours). Efforts to expand this distribution corrected the issue with the range of forecast power values but left a distribution shape with dissimilar skewness and kurtosis to that of either the commercial forecast or actual power production. This is shown in Figure 3 for the day-ahead forecast horizon. The commercial forecast's distribution is heavily skewed and clearly non-Gaussian.

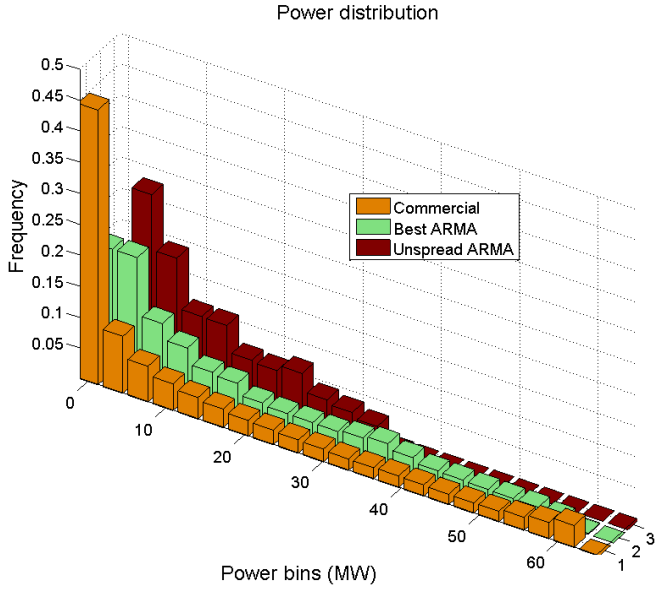


**Figure 3: Distribution of Day-ahead ARMA Forecasts (before and after expansion) vs. Commercial Forecast**

**Table 5: Auto-Regression of 6-Hour Forecast Errors**

| Lag (Hours) | Commercial Errors | ARMA Errors |
|---|---|---|
| 1 | 0.84 | 0.82 |
| 2 | 0.63 | 0.60 |
| 3 | 0.46 | 0.38 |
| 4 | 0.32 | 0.17 |
| 5 | 0.22 | 0.00 |
| 6 | 0.17 | 0.00 |
| 7 | 0.16 | -0.01 |
| 8 | 0.15 | -0.02 |

More importantly, the ARMA models' auto-regression of errors was much lower than in commercial forecasts for even very short horizons. In Table 5, the auto-regression of errors for up to eight hours lag is shown; the rapid decay of correlation after a few hours' lag was not seen in the commercial forecast. Relying on such an input to a production cost simulation would

result in an overly optimistic estimate of reserve costs due to a lack or partial reversal (note slight negative correlation) of trends over time. Due to these issues, ARMA synthetic forecasts were not analyzed further (e.g., for ramp performance).

Next, a 2-dimensional probability matrix relating columns of current forecast quantiles to rows of actual power quantiles was used as the basis for generating a synthetic forecast in a random walk process. The best results were found using a 40x40 matrix. However, the variability was very high.

To correct this, a different matrix was developed for quantiles of past power production:

$$Sum\ of\ Past\ Power\ Observations = \sum_{i=1}^{N} P_{t-N} \quad (10)$$

The number of quantiles used in this third dimension and the number of past power values was optimized for each forecast horizon. For example, 4-hour synthetic forecasts produced using quantiles of such sums yielded reduced variability as evaluated by the resulting average magnitude of step changes (average |Δ|) and the standard deviation of step changes ($\sigma_\Delta$). The optimal number, N, in equation (10) was found to be five for this forecast horizon, as shown in Figure 4.

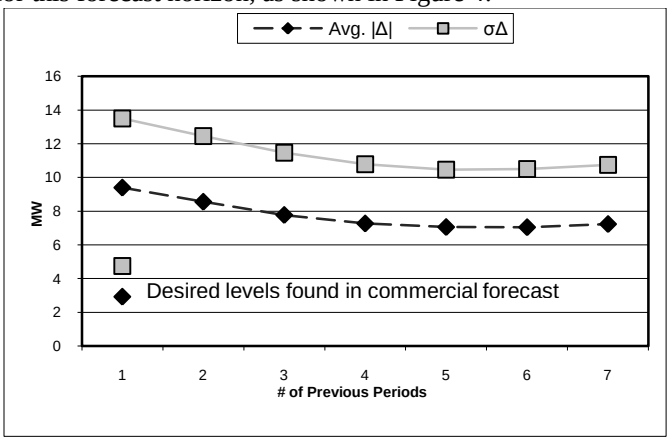


**Figure 4: Optimizing the Number of Past Power Measurements to Reduce Variability in Synthetic Forecast**

Despite this improvement, these variability parameters were still much higher than found in the commercial forecast. This necessitated large reliance on relaxation methods, to reduce variability further. However, these actions transformed the PDFs. Several versions of the synthetic forecast were created and their PDFs were analyzed. The reduction in the variability and results for other parameters for a 4-hour synthetic forecast are shown in Table 6. Since this analysis is, thus far, for intra-site model performance, the dimensional parameters (i.e., all except R) are listed in MW as opposed to percentage of capacity. The results columns listed from left to right are: commercial forecast, 2-D model, 3-D model prior to relaxation adjustments, final (7th) version pseudo-Markovian model, ARMA (1,8) model, and final (best) ARMA-based model. Although some of the unaltered ARMA (1,8) results are closer to the commercial forecast, recall that the best ARMA model was also optimized to improve the auto-regression of errors and distribution of power.

**Table 6: 4-Hour Commercial and Synthetic Forecast Results**

| | Com. | PM40 2-D | PML4-40-10 | PM v7 | ARMA (1,8) | Best ARMA |
|---|---|---|---|---|---|---|
| MAE (MW) | 7.3 | 7.2 | 7.2 | 7.3 | 7.6 | 7.3 |
| $\bar{\bar{P}}$ (MW) | 16.4 | 16.6 | 16.3 | 16.4 | 16.4 | 17.1 |
| $\sigma_P$ (MW) | 17.2 | 17.6 | 17.2 | 16.6 | 17.4 | 18.7 |
| Mean Bias (MW) | -1.3 | -1.1 | -1.4 | -1.4 | -1.4 | -0.7 |
| RMSE (MW) | 11.0 | 11.0 | 11.0 | 10.8 | 11.7 | 11.6 |
| Avg \|Δ\| (MW) | 2.9 | 9.4 | 7.1 | 3.1 | 3.4 | 2.5 |
| $\sigma_\Delta$ | 4.7 | 13.5 | 10.5 | 5.1 | 6.0 | 4.2 |
| R | 0.85 | 0.85 | 0.85 | 0.86 | 0.83 | 0.83 |

The next approach to generating a synthetic forecast was a Markov chain used to simulate a non-Gaussian forecast time series. Referring to Figure 1, in each random walk step, the future forecast value was directly linked to the current forecast value (unlike in the pseudo-Markov algorithm), and the variability was greatly reduced. However, the error was initially high using this approach. As previously mentioned, between 13-19% of the time (depending on the forecast horizon) the synthetic forecast value would be averaged with the actual power value. This was a simple fix nested within the loop used to generate the forecast. Final efforts to reduce variability further relied again on relaxation methods. However, depending on the opinion of the analyst, this step may be unnecessary. The raw time series produced without averaging with actual power values will be referred to throughout as "FFP"[3]; time series produced by averaging with actual power values will be referred to as version 1 or "FFP v1"; time series also utilizing relaxation methods will be referred to as version 2 or "FFP v2". Time series of the actual forecast errors are contrasted with FFP v2 errors for a sample month in Figure 5.

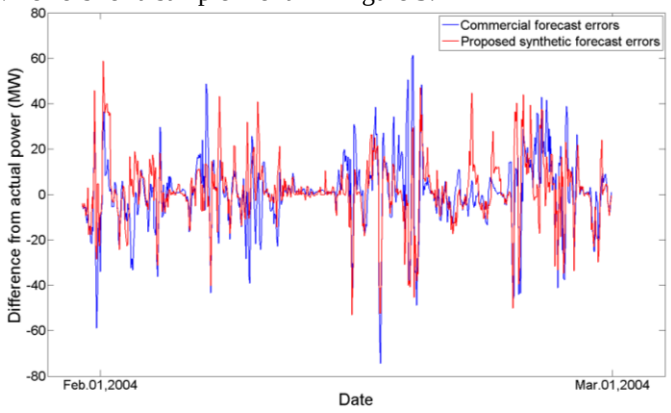


**Figure 5: Time series of Actual vs. FFP v2 errors**

Table 7 shows the performance (again in MW) of these three methods contrasted with the state-of-the-art 4-hour wind power forecast for the Northwestern US wind power plant.

[3] FFP is short for the three dimensions of the transition array: future Forecast, current Forecast, and current Power. This is shown in Figure 1.

**Table 7: Results for a Commercial 4-Hour Forecast Contrasted with Three Markov Models**

| | Commercial | FFP | FFP v1 | FFP v2 |
|---|---|---|---|---|
| MAE (MW) | 7.3 | 8.9 | 7.3 | 7.3 |
| $\bar{\hat{P}}$ (MW) | 16.4 | 16.2 | 16.7 | 16.7 |
| $\sigma_{\hat{P}}$ (MW) | 17.2 | 16.1 | 17.2 | 17.5 |
| Mean bias (MW) | -1.3 | -1.6 | -1.1 | -1.1 |
| RMSE (MW) | 11.0 | 13.0 | 11.1 | 11.2 |
| Avg. $\lvert\Delta\rvert$ (MW) | 2.9 | 3.5 | 3.6 | 2.8 |
| $\sigma_{\Delta}$ (MW) | 4.7 | 5.5 | 5.6 | 4.8 |
| R | 0.85 | 0.78 | 0.85 | 0.85 |

Unlike the ARMA models, the auto-regression of errors and auto-regression of past forecast values for the Markov models closely resembled the auto-regressive traits of commercial forecasts. Table 8 shows the auto-regression of measured power and the auto-regression of the commercial forecast, the FFP v1 forecast, the commercial forecast errors, and the FFP v1 forecast errors. The two columns of Table 8 farthest to the right may be compared with Table 5. The auto-regression results for other forecast horizons, other Markov methods, and pseudo-Markov methods were also adequately similar to commercial forecasts.

**Table 8: Auto-Regression of 6-hour Ahead Forecast and Errors**

| Lag | Actual Power | Comm. Forecast | FFPv1 | Commercial errors | FFPv1 errors |
|---|---|---|---|---|---|
| 1 | 0.95 | 0.97 | 0.95 | 0.84 | 0.83 |
| 2 | 0.89 | 0.93 | 0.90 | 0.63 | 0.65 |
| 3 | 0.83 | 0.88 | 0.85 | 0.46 | 0.51 |
| 4 | 0.78 | 0.83 | 0.80 | 0.32 | 0.41 |
| 5 | 0.73 | 0.79 | 0.75 | 0.22 | 0.32 |
| 6 | 0.68 | 0.75 | 0.71 | 0.17 | 0.25 |
| 7 | 0.64 | 0.71 | 0.67 | 0.16 | 0.19 |
| 8 | 0.60 | 0.67 | 0.63 | 0.15 | 0.14 |

FFP v1 and v2 qualitative metric performance was also good as can be seen in the distribution of forecast power and distribution of step changes (Δ) as shown in Figures 6 and 7, respectively. Figures 6 and 6 were produced from the 4-hour forecast horizon: distributions from other horizons had similar results. The improvement in synthetic forecast PDFs in Figure 7 from the synthetic forecast PDFs in Figure 3 is clear.

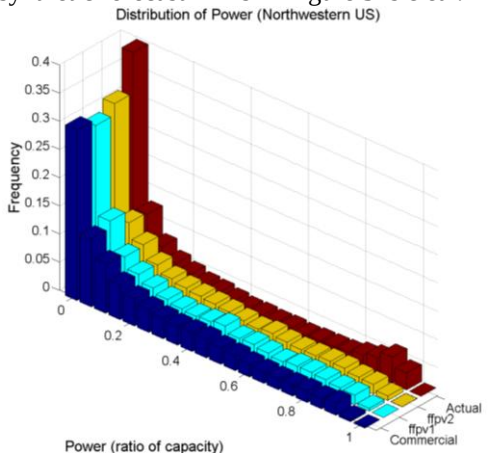


**Figure 6: Distribution of Measured Power Contrasted with 4-hour Commercial and Synthetic Forecasts**

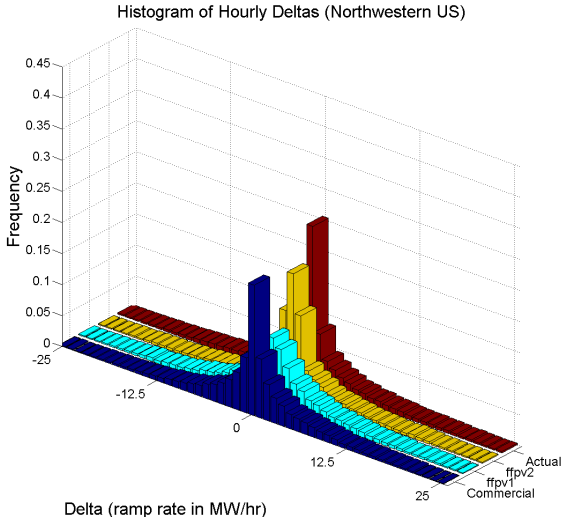


**Figure 7: 4-hour Forecast Distribution of Step Changes**

Eventually, every wind production forecast must be synchronized with a load forecast for use in a production cost simulation. Therefore, it is important that seasonal and diurnal (and ideally weather pattern-specific) variation in actual wind forecast errors be approximated in synthetic forecast errors. In areas where peak load hours are determined by summer cooling demands, it is especially important that wind's contribution to a wind net load forecast be accurately expressed in the peak summer afternoon hours; wind energy production during such hours is usually low. Figure 8 shows that the MAE during these peak hours for the commercial 4-hour forecast is much lower than during other times.

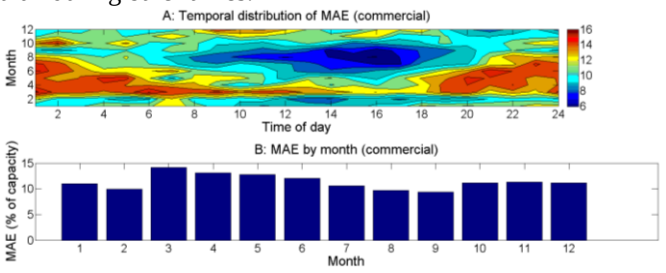


**Figure 8: Temporal MAE Analysis of Commercial Forecast- (A) Contour Plot, (B) Monthly MAE**

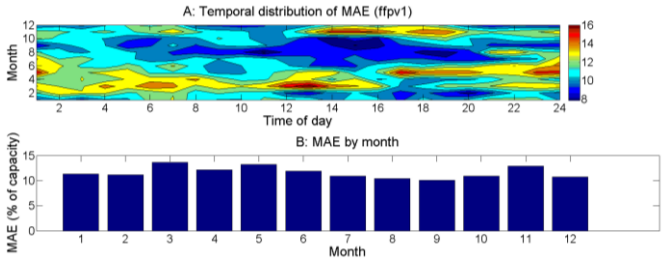


**Figure 9: Temporal MAE Analysis of Synthetic Forecast- (A) Contour Plot, (B) Monthly MAE**

Figure 9 shows the results of identical temporal error analysis for the FFP v1 4-hour synthetic forecast. The lower MAE during these peak hours is approximately reproduced. In both, there is more seasonal variation than diurnal variation (not shown as a bar graph) of error. Results from other forecast horizons and the other probabilistic simulation methods were

similar: error was not randomly distributed temporally but showed reduction during peak hours.

A goal of this analysis was to be able to use this information from the wind power plant and apply it to modeled (simulated) wind power sites to build synthetic forecasts for proposed installations. Ideally, a new transition matrix would be developed for different classes of sites (e.g., based on geography or capacity factor, defined as the fraction of energy produced over the course of a year divided by the product of the capacity times the number of hours in a year). To this end, the transition matrices developed with the wind power plant's actual power and commercial forecast data were applied to time series of theoretical power production from a randomly selected site from EWITS. These time series, which were produced with NWP models as described in [4], were treated as the "actual" power produced in the random walk process. The resulting synthetic forecast was compared with the synthetic forecast produced by AWS Truewind for the forecast horizons available for that study[4]. It should be noted that there are several differences between Northwestern US wind power plants and the EWITS site that make such a comparison of synthetic forecasts dubious: different hub heights, topographies, and resource characteristics[5]. Given these constraints, even reasonably similar performance with AWS' synthetic forecast for this site should serve as proof of concept for the model.

The algorithms used to build synthetic forecasts for the wind power plant in the Northwestern US, where the desired results could be quantified exactly, were then adjusted for application at EWITS Site 6521. Since this site does not actually exist, the modeled power that was produced with NWP techniques was treated as actual, measured power. However, the following analysis is *not* an invitation to compare two very different data sets from different regions. Future research from a site with actual wind power production data, (NWP) modeled production data, commercial forecast data, and multiple approaches to synthetic forecasting would be appropriate.

**Table 9: 4-Hour Synthetic Forecast Results**

| | AWS | FFPv1 | FFPv2 | PM-v7 |
|---|---|---|---|---|
| MAE (%) | 13.3 | 12.4 | 12.4 | 12.6 |
| $\bar{\hat{P}}$ (%) | 31.5 | 29.8 | 29.5 | 28.6 |
| $\sigma_{\hat{P}}$ (%) | 24.2 | 24.7 | 24.9 | 23.2 |
| Mean Bias (%) | 0.0 | -1.6 | -2.0 | -2.9 |
| RMSE (%) | 17.7 | 17.6 | 17.6 | 17.2 |
| Avg $\lvert\Delta\rvert$ (%) | 6.3 | 6.2 | 4.9 | 5.4 |
| $\sigma_{\Delta}$ (%) | 8.6 | 9.1 | 7.7 | 8.3 |
| R | 0.80 | 0.80 | 0.81 | 0.82 |

The quantitative analysis results are summarized in Table 9 for the 4-hour horizon. Dimensional parameters are presented as percentages of capacity in contrast with Tables 5 and 6, which were presented in MW. The performances were similar, although there was more variability in the AWS and FFP v1 synthetic forecasts than for the FFP v2 or Pseudo-Markov synthetic forecasts, which used relaxation techniques.

Quantitative results for the 6-hour horizon (not shown here) were similar to the 4-hour results. However, the synthetic forecasts proposed here had slightly more error than AWS' synthetic forecast at the 6-hour horizon. Results for the day-ahead forecast horizon showed a disparity. In comparison with the 4-hour and 6-hour forecasts, there was little change in AWS' quantitative criteria for the day-ahead horizon but this is likely unintentional[6]. Since it is difficult to analyze many of the quantitative parameters at the day-ahead horizon due to these unresolved issues, the comparisons are not shown here.

The auto-regression of (synthetic) forecasts and (synthetic) forecast errors were very similar to Table 9. All synthetic forecasts and their errors persisted in a realistic manner. Therefore, this analysis is also omitted here.

Revisiting Figures 6 and 7, one can see that measured power and commercial forecast PDFs did not substantially differ at the operational wind plant. Therefore, in lieu of actual production and commercial forecast data, one can entertain the conjecture that modeled power PDF is the best available benchmark. With this in mind, it is interesting to consider the shapes of the PDFs from Site 6521 in Figure 10. In contrast to Figure 6, the positions of actual (or modeled) and FFP v1 results have been switched in Figure 10. The apex of the "spike" in the modeled step change row lies between FFPv1 and FFPv2. AWS' synthetic forecast does not have such a feature and is generally more leptokurtic (fat-tailed) than the others.

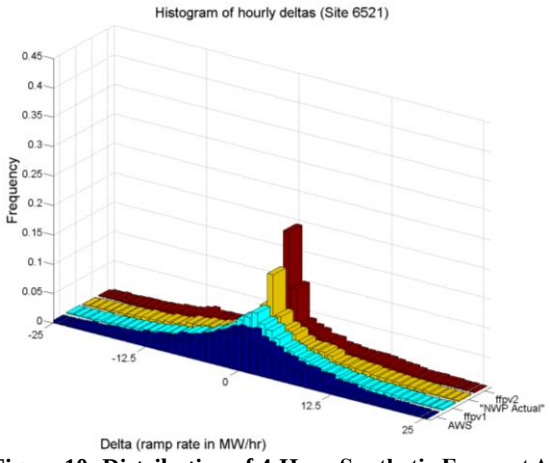


**Figure 10: Distribution of 4-Hour Synthetic Forecast Δs Contrasted with "Actual" Δs**

Again using the modeled power result as a benchmark, one can see that FFP v1 and FFP v2 distributions of power in Figure 11 are close semblances to the benchmark. In contrast, AWS' synthetic forecast appears to have some of the same

[4] AWS Truewind produced 4-hour, 6-hour, and day-ahead synthetic forecasts for every one of the potential wind sites that would be needed to produce up to 30% of the US' energy from wind.

[5] The EWITS comparison site, Site 6521 located in New York State, had a similar capacity factor and rated capacity as the Northwestern US wind plant. The AWS Truewind approach to applying probability matrices was to randomly select one developed from one of three validation sites inside the EWITS study area (MN, IA, OK) or one validation site outside the study area (TX).

[6] The offshore EWITS sites' forecasts were adjusted so that the MAE from each forecast horizon was not as close. This adjustment was not made for the onshore sites (such as Site 6521), but this step would be superfluous as the day-ahead forecast was the only one actually used in the dispatch model.

issues as the best ARMA model did in Figure 3: too few occurrences of forecast power in the smallest production bin, a preponderance of occurrences of forecast power between 15-50% of capacity, and the highest frequency of forecast power is skewed to the right. This could be the result of the previously mentioned random selection of probability matrix process, as opposed to a selection process that chooses a probability matrix from the most similar validation site.

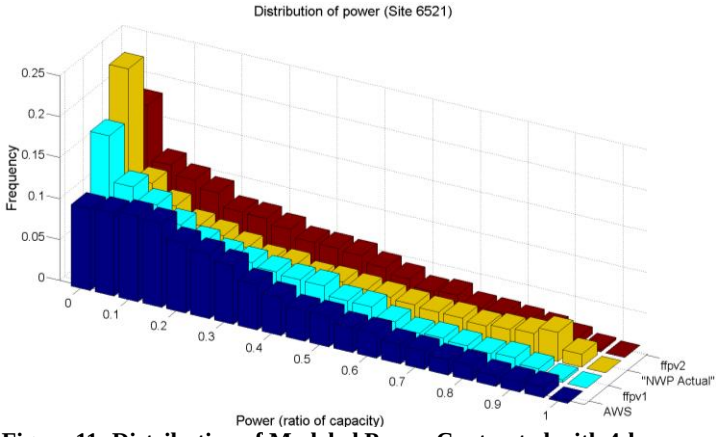


**Figure 11: Distribution of Modeled Power Contrasted with 4-hour Synthetic Forecasts**

Finally, the secondary criteria were analyzed for both the actual and theoretical wind sites. These results are presented in Tables 10-14. At the operational wind power plant, 348 measured ramp events/year were discovered. Referring to the last column in Table 10, there were fewer forecast ramp events than measured ramp events. Using the same ramp definition applied to the NWP modeled time series, there were 320 ramp events identified at Site 6521 for 2004. In Table 11, AWS' synthetic forecast produced more forecast ramp events (closer to 320/year), but of shorter duration.

**Table 10: Ramp Counts and Duration from Northwest**

| | | Forecast Horizon | Avg. Ramp Duration | # Ramp Events/ Year |
|---|---|---|---|---|
| Actual Wind Plant | Comm. (Actual) Forecast | 4 | 3.39 | 116 |
| | | 6 | 4.15 | 92 |
| | | day | 4.83 | 183 |
| | FFPv1 | 4 | 3.44 | 174 |
| | | 6 | 3.48 | 188 |
| | | day | 3.31 | 259 |

**Table 11: Synthetic Forecast Ramp Counts and Duration from New York**

| | | Forecast Horizon | Avg. Ramp Duration | # Ramp Events/ Year |
|---|---|---|---|---|
| Site 6521 | AWS | 4 | 3.25 | 120 |
| | | 6 | 3.25 | 109 |
| | | day | 3.34 | 140 |
| | FFPv1 | 4 | 3.32 | 204 |
| | | 6 | 3.66 | 176 |
| | | day | 3.38 | 259 |

Initial inspection of Tables 12 and 13 indicates much lower ramp forecast accuracy and ramp capture than in [17]. This is due to the shorter window for correlating between forecast and actual ramps and the different thresholds used to identify ramps. In Table 12, the FFP v1 method resulted in high estimates of ramp forecast accuracy and ramp capture for the 4 and 6-hour horizons. In Table 13, FFPv1 had lower results than AWS' synthetic forecast for these metrics. Furthermore, AWS' ramp forecast uncertainty and ratio of true to misleading forecasts deviated by a full order of magnitude from the commercial forecast's results.

**Table 12: Ramp Forecast Accuracy Measures from Northwest**

| | | Forecast Horizon | Forecast Accuracy | Ramp Capture | Forecast Ramp Uncertainty | Ratio of True/ misleading |
|---|---|---|---|---|---|---|
| Nine Canyon | Comm. Forecast | 4 | 24.2% | 7.4% | 15.9% | 1.52 |
| | | 6 | 22.8% | 6.8% | 9.6% | 2.36 |
| | | Day | 21.3% | 14.7% | 9.7% | 2.20 |
| | FFP v1 | 4 | 42.3% | 19.7% | 6.0% | 7.11 |
| | | 6 | 32.9% | 16.8% | 5.8% | 5.66 |
| | | Day | 19.8% | 13.2% | 7.9% | 2.49 |

**Table 13: Synthetic Forecast Ramp Accuracy Measures from New York**

| | | Forecast Horizon | Forecast Accuracy | Ramp Capture | Forecast Ramp Uncertainty | Ratio of True/ misleading |
|---|---|---|---|---|---|---|
| Site 6521 | AWS | 4 | 38.2% | 11.9% | 1.8% | 21.3 |
| | | 6 | 33.1% | 9.3% | 2.5% | 13.0 |
| | | Day | 25.5% | 9.5% | 4.3% | 5.95 |
| | FFP v1 | 4 | 32.4% | 17.5% | 4.7% | 6.88 |
| | | 6 | 24.4% | 12.5% | 5.0% | 4.91 |
| | | Day | 14.6% | 10.2% | 6.5% | 2.25 |

In [13] it was shown that the number of forecast ramps declines rapidly with increasing lead times before recovering and leveling at a forecast horizon of 8+ hours when NWP models are used exclusively. The ratio of true to misleading forecast hours also followed this trend: declining from 10 at the hour-ahead horizon to 1.5 for the 4-hour horizon and then rising to a steady 2-2.5 for 8+hour horizons (including day-ahead). Thus, 4-hour commercial forecasts predict the fewest ramp events and these forecast ramp events are the least reliable. The Markov method presented here did not replicate this trend. However, comparing the FFP v1 to the AWS results, an expectation of a forecast ramp being 5-7 times more likely to be a true rather than misleading forecast (when the observed ratio is only 1.5-2.3) is a large improvement over an expectation of 13-21. Likewise in comparing the forecast ramp uncertainty criterion, an expectation of 5% of the forecast ramps for a particular horizon to be misleading ones (when the observed percentage is 10-16%) is a large improvement over the expectation of 1.8-2.5%. This could lead to large underestimates of integration costs caused from ramps. Results for other methods, including FFP v2 and the pseudo-Markovian synthetic forecast were similar to the results for FFP v1.

## CONCLUSIONS

Four methods were presented for generating a synthetic forecast that can be used in a wind integration study. These methods attempted to replicate the trends and characteristics associated with a state-of-the-art wind forecast.

The first of these methods, based on ARMA time series modeling, failed to replicate the persistence of the errors, the shape of the distribution of forecast power bins, and the day-ahead version's standard deviation of forecast power was lower than in the commercial forecast. The second of these methods was a stochastic, pseudo-Markovian probabilistic Monte Carlo simulation that relied heavily on relaxation techniques to reduce the standard deviation of step changes and average magnitude of step changes to realistic levels. Such efforts when adjusted for a particular site might have unintended consequences on qualitative criteria (PDFs) when applied to another site.

The other methods were based on a novel 3-dimensional approach to probabilistic Markov simulations. One version relied partially on relaxation techniques and the other did not (although it had slightly more variability than found in commercial forecasts). These models replicated the temporal trends in errors, persistence of errors, and all of the quantitative criteria of commercial wind power forecasts as well as an existing, proprietary synthetic forecast model. Furthermore, these models produced superior PDFs and better ramp contingency analysis results in comparison to an existing synthetic forecast model. However, the existing synthetic forecast model produced ramp counts more in line with those found in commercial forecasts.

Overall, the preponderance of evaluation metrics showed that a 3-dimensional state transition array can be used to temporally link a realistic distribution of forecast errors with a time series produced from NWP techniques. In comparison to generating a forecast using NWP models, the overall process is computationally simple. In conclusion, the probabilistic Markov method appears to be the best choice for an accurate and computationally efficient method to generate a synthetic forecast for a wind integration study.

## ACKNOWLEDGMENTS

The National Renewable Energy Laboratory is gratefully acknowledged for their support of this work under subcontract XXL-7-77283-01.

## REFERENCES

[1] K. Dragoon and M. Milligan (2003) *Assessing Wind Integration Costs with Dispatch Models: A Case Study of PacifiCorp*, Proc. of WINDPOWER 2003, Austin, Texas.

[2] C. Potter, et al., (2008) *Creating the Dataset for the Western Wind and Solar Integration Study (U.S.A.)* Wind Engineering, Volume 32, No. 4.

[3] R. Zavadil (2007) *Avista Corporation Wind Integration Study*, Enernex Corporation, Knoxville, Tennessee. http://www.uwig.org/AvistaWindIntegrationStudy.pdf

[4] M. Brower (2008) *Development of Eastern Regional Wind Resource and Wind Plant Output Data Sets*, AWS Truewind. http://wind.nrel.gov/public/EWITS/AWST_EWITS_Final_Technical_Report_Draft.pdf

[5] R. Piwko, et al., (2005) The Effects of Integrating Wind Power on Transmission System Planning, Reliability, and Operations, New York State Energy Research and Development Authority.

[6] G. McNerney and P. Veers (1985) *A Markov Method for Simulating Non-Gaussian Wind Speed Time Series*, Sandia National Laboratory.

[7] Y. Wan (2005) *A Primer of Wind Power for Utility Applications*, National Renewable Energy Laboratory Technical Report.

[8] G. Griebel, G. Kariniotakis, and R. Brownsword (2003) *State of the Art on Short-term Wind Power Prediction*, Project ANEMOS Report D1.1.

[9] C. Monteiro, et al., (2009) *Wind Power Forecasting: State-of-the-Art 2009*, Argonne National Lab Report. http://www.dis.anl.gov/pubs/65613.pdf

[10] G. Box and G. Jenkins (1970) *Time series analysis: Forecasting and control*, San Francisco: Holden-Day.

[11] M. Milligan (2004) *Statistical Wind Power Forecasting for U.S. Wind Farms*, Proc. of the 17th Conference on Probability and Statistics in the Atmospheric Sciences.

[12] G. Reikard (2008) *Using Temperature and State Transitions to Forecast Wind Speeds*, Wind Energy, Volume 11, No. 5.

[13] M. Bielecki, et al., (2010) *A Methodology for Comprehensive Characterization of Wind Power Production Erros*, Proc. of the ASME 2010 4th International Conference on Energy Sustainability.

[14] A. Costa, et al., (2004) *Mathematical and Physical Wind Power Forecasting Models*, Proc. of Second Joint Action Symposium on Wind Forecasting Techniques, IEA.

[15] J. Kemper (2010) *Applications and Modeling of Wind Power Production Forecast Errors.* Master's Thesis, Northern Arizona University.

[16] R. Zavadil (2008) *Wind Integration Study for Public Service Colorado Addendum: Detailed Analysis of 20% Wind Penetration.*

[17] B. Greaves, et al., (2009), *Temporal Forecast Uncertainty for Ramp Events*, Wind Engineering, Volume 33, No. 4.