# 1.3 STATISTICAL WIND POWER FORECASTING FOR U.S. WIND FARMS

Michael Milligan, Consultant*
Marc Schwartz and Yih-Huei Wan
National Renewable Energy Laboratory, Golden, Colorado

## ABSTRACT

Electricity markets in the United States are evolving. Accurate wind power forecasts are beneficial for wind plant operators, utility operators, and utility customers. An accurate forecast allows grid operators to schedule economically efficient generation to meet the demand of electrical customers. The evolving markets hold some form of auction for various forward markets, such as hour ahead or day ahead. While the longer-term forecasting relies on numerical weather models, this paper describes several statistical forecasting models that focus on short-term forecasts that can be useful in hour-ahead markets. The purpose of the paper is not to develop forecasting models that can compete with commercially available models. Instead, we investigate the extent to which time-series analysis can improve simplistic persistence forecasts. This project applied a class of models known as autoregressive moving average (ARMA) models to both wind speed and wind power output. The ARMA approach was selected because it is a powerful, well-known time-series technique and has been used by the California Independent System Operator in some of its forecasting work. The results from wind farms in Minnesota, Iowa, and along the Washington-Oregon border indicate that statistical modeling can provide a significant improvement in wind forecasts compared to persistence forecasts.

## 1. INTRODUCTION

In some electricity markets, wind is becoming a significant source of energy. As the contribution of wind power plants continues to grow, the impact of wind on various aspects of power system operation receives greater scrutiny. Because wind is an intermittent power source, these operational impacts are unlike those of other power plants. This intermittent characteristic of wind power generation means that efficient power system operation will depend in part on the ability to forecast available wind power.

This paper examines the use of a standard class of statistical time-series models to predict wind power output up to 6 hours in advance. The purpose is not to develop models to compete with commercial forecasting models, such as those described in Bailey (1999) or Landberg (1997). Rather, our goal is to investigate the feasibility of relatively inexpensive statistical forecasting models that do not require any data beyond historical wind power generation data. This may limit the ability and the usefulness of this type of forecasting model, but for small wind farms that are unable to participate in formal forecasting projects, it might be desirable to use a statistical model that can be developed and used at lower cost.

For this project, we used data collected from operating wind power plants in Minnesota and Iowa, along with Bonneville Power Administration's (BPA's) Stateline project along the Washington-Oregon border. We found some differences in model performance at the three sites, which was expected because of the different climatic regions. Wan (2002) described some of the data used for this project. A more complete report on our forecasting results can be found in Milligan (2003).

## 2. STATISTICAL MODELING FRAMEWORK

Although many time-series methods could be applied to this problem, a general class of models known as autoregressive integrated moving average (ARIMA) models is applied in this paper. Similar models have been applied by Makarov

* *Corresponding author address:* Michael Milligan, National Wind Technology Center, National Renewable Energy Laboratory, 1617 Cole Blvd., Golden, CO 80401; e-mail: Michael_Milligan@nrel.gov.

(2002), Nielsen (1999), and Kariniotakis (1997). ARIMA models have up to three components: autoregressive, integrated, and moving average. When the second component provides no significant explanatory power in the model, it is dropped. We found the integration term to be unimportant in this analysis, so it was dropped. The resulting model is known as ARMA, and can be characterized as

$$X_t = \sum_{j=1}^{p} a_j X_{t-j} + \sum_{k=0}^{q} b_k e_{t-k}$$

The equation states that a realization of the time-series X at time t depends on a linear combination of past observations of X plus a moving average of series e, which is a white-noise process characterized by zero mean and constant variance. The time series X is known as an ARMA(p,q) process, where p is the order of the autoregressive process of X on itself and q is the order of the moving-average error term. The AR term allows us to capture the previous *p* wind power values, and the MA term involves the use of the prior *q* error terms to help improve the forecast.

Interested readers can consult Box (1976) for details of well-known methods for applying ARMA models in practice. After estimation, the model can be checked with several diagnostics. It is not uncommon for a promising model identification to lead to a poorly performing model, so the diagnostic phase is important because it weeds out models that do not work well. We found many such models during this project. We chose to present results for similar models whenever the model forecasting performance improved upon the persistence model.

## 3. METRICS

A well-known method of forecasting wind is the simplistic persistence method. This approach uses the past hour wind speed (or wind power) as the forecast for the next hour. As a forecasting technology, this method is not impressive, but it is nearly costless and can perform surprisingly well for 1 to 3 hours ahead. Therefore, any forecast method first should be measured by the extent it can improve on persistence forecasts. This approach is applied in this paper. The persistence forecast can offer a range of forecasting accuracy, depending on the wind regime and the number of periods to be forecast. In this project, we calculated the root mean square error (RMSE) of each forecast over the relevant time period to compare our methods with the persistence model. A lower RMSE implies that the forecast is more accurate, whereas a high RMSE value implies less accuracy.

We are interested in how well ARMA models can forecast more than 1 hour in advance. We applied our forecasting techniques up to 6 hours ahead, which is approximately the limitation of purely statistical forecasting methods.

## 4. HOURLY POWER FORECASTS

We evaluated the model performance on different data than were used to train (fit the parameters) the model in each of the forecasting cases we examined. A number of different training periods were applied so that we could see how this would affect the forecasting performance. In all cases discussed in this section we refer to the training period as a fixed-window period. The next section discusses the use of a dynamic training period with a moving window that is adjusted for each hour.

We used actual wind power data from Lake Benton 2 (LB), Minnesota (located in the southwest part of the state), as our primary source in this project. We also compared some of the LB forecasts with forecasts at Storm Lake (SL), Iowa (located in northwestern Iowa), and from BPA's Stateline project on the eastern Washington-Oregon border. Data from LB and SL are collected at the National Renewable Energy Laboratory (NREL) and are described more fully in Wan (2002). Data from LB and SL are from 2001, and data from the BPA site are from 2002.

Decisions must be made regarding how to develop statistical hourly power forecasts to analyze their accuracy. One of the most important decisions is the selection of the training period for the forecasting algorithm. In some of our initial screening work using the LB data, we calculated separate time series for 10 months, training the model for the first 2 weeks and calculating forecasts for the remainder of the month. The statistical evidence showed a significant difference in the ARMA model specifications that were calculated in each of the training periods. This finding is not surprising because it provides evidence of statistically significant variations in wind generation patterns at different times of the

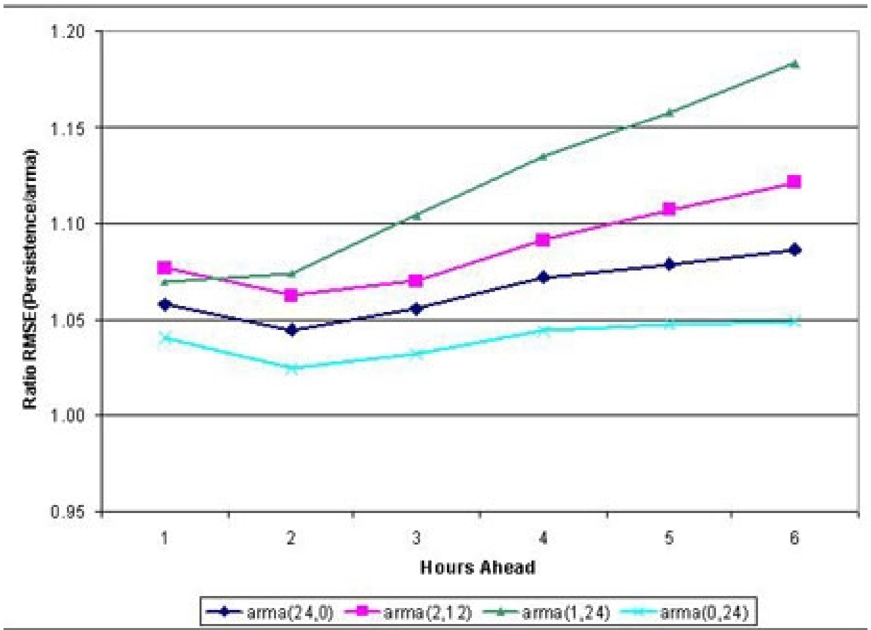


**Figure 1. Lake Benton 2 kW forecasts: January/February 2001.**

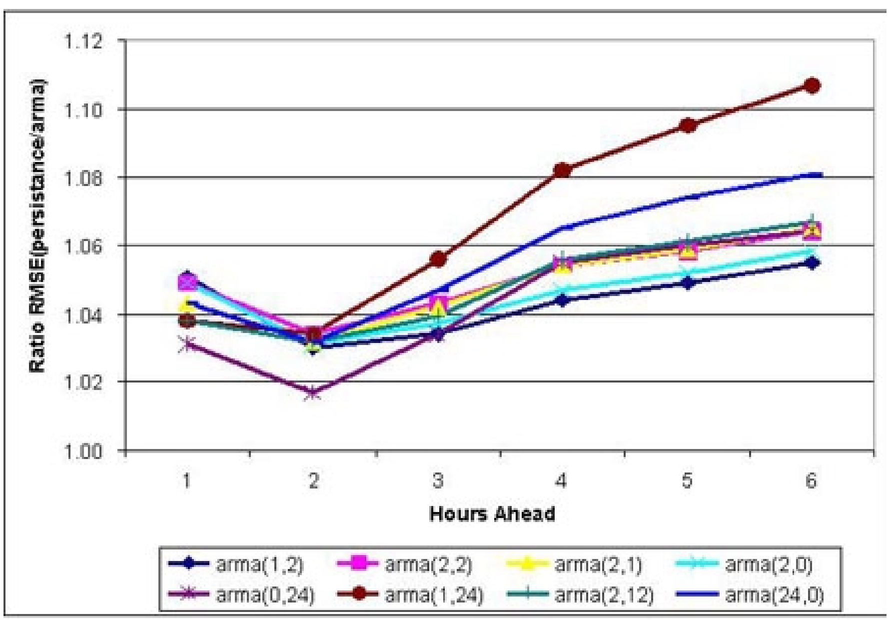


**Figure 2. Storm Lake kW Forecasts: January/February 2001.**

year, and these differences are important in developing a forecasting model.

The first set of cases used wind power data from January-February 2001. The data from January were used to train the model, and hourly forecasts were developed for the month of February. Each forecast predicted the 6-hour period ahead. The model results were compared with the persistence model, which was also applied to the same time periods. In many cases, the ARMA forecasting results were similar for different model specifications. We report results for forecast models that did improve over persistence.

Figure 1 illustrates the forecasting cases for LB. The graph shows the ratio of the persistence RMSE to the ARMA RMSE for several alternative ARMA model specifications. Higher ratios indicate better ARMA performance, and ratios less than 1 would indicate that persistence does a better job of forecasting than the ARMA model. The best overall model specification is the ARMA(1,24), which improves on persistence by approximately 7% in the first hour, increasing to approximately 18% in the sixth hour. However, the ARMA(2,12) model does slightly better in the first hour.

A similar set of forecasting models was applied to the same time period at SL. The same model specification, the ARMA(1,24), performed the best overall. However, it is worth noting that the ARMA(1,2) model was best in the 1-hour forecast, whereas the ARMA(1,24) model was better from 2 hours out to 6 hours. This suggests the possibility of using an ensemble of ARMA models, depending on the forecast performance of different model specifications for different forecast horizons. The SL results appear in Figure 2.

Unfortunately we were not able to complete forecasts for all sites for the same time period. Figure 3 shows the results from Stateline using January and February data, but from 2002. In contrast to the other sites, the best overall forecasting performance was obtained from the ARMA(0,24) case, although the ARMA(1,12) specification performed slightly better for the first hour.

The ARMA forecasts at all wind sites eroded significantly as the number of forecast periods increased, although the SL results were slightly better than the LB results. As measured by the RMSE, the forecast error for LB nearly tripled from the 1-hour forecast to the 6-hour forecast, and the SL results were similar. For Stateline, the forecast errors were between those at the other two sites, and the degradation in RMSE is also apparent. Figure 4 depicts the ratio of RMSE to rated capacity as a function of the length of the forecast period. In all cases the model with the best improvement over persistence was chosen for the graph. The model specification for LB and SL are both ARMA(1,24), whereas the Stateline model is ARMA(0,24). When the results from Figures 1, 3 and 4 are compared, it is also apparent that RMSE(arma) at Stateline is better than the RMSE(arma) at LB, yet the ARMA forecast model at LB provides a more significant improvement over persistence than the ARMA model at Stateline.

We also investigated additional periods during 2001 and changed the lengths of the training and testing intervals. Figure 5 shows an example of LB for April 2001. For this case, a 3-week training period was used, and the remaining hours of the month were used to test the forecast. For this period, it is apparent that the suite of tested models did improve on persistence, but not to the same degree as the January-February cases.

## 5. DYNAMIC TRAINING PERIODS

Because of the apparent forecast model sensitivity to time of year, we also applied a dynamic training period to some of the LB cases. A more complete study than ours might look at a wide range of dynamic training periods, ranging from perhaps 1 week to several weeks. Although our original study used several time periods and training periods, we report on two cases here. The first case used a sliding 2-week training period. A set of 6 hourly forecasts was calculated. Then the training period was shifted forward by 1 hour, dropping the first hour and adding the most recent hour. With this new training period the model was recalculated and the next 6 hours were forecast. The process continues until forecasts were completed thru the month of April. A similar process was applied to the same time period, but we used a 3-week training period instead of a 2-week training period.

Several alternative model specifications improved upon persistence. For clarity, we selected the best-performing model from each training period. Coincidently, the best-performing models both have ARMA(2,1) specifications. The results appear in Figure 6. The results indicate that the 2-

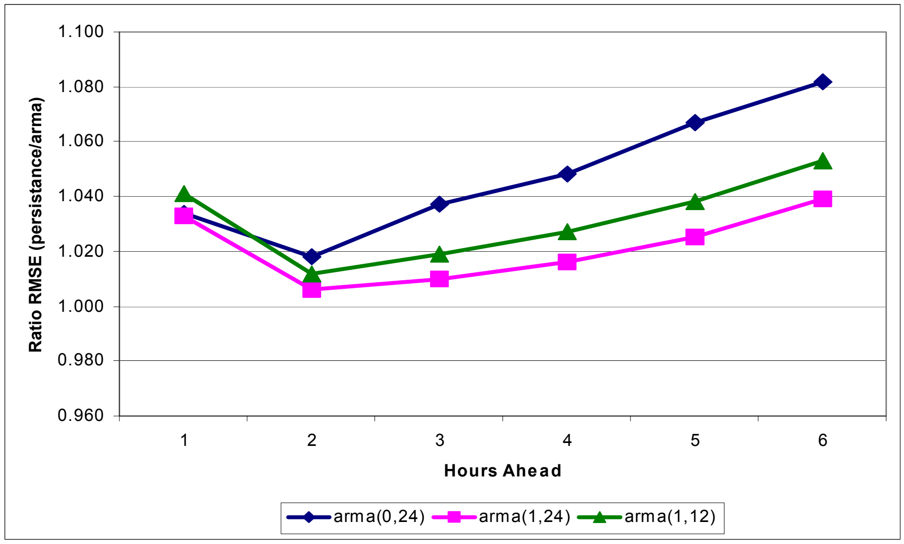


**Figure 3. Stateline kW forecasts: January/February 2002.**

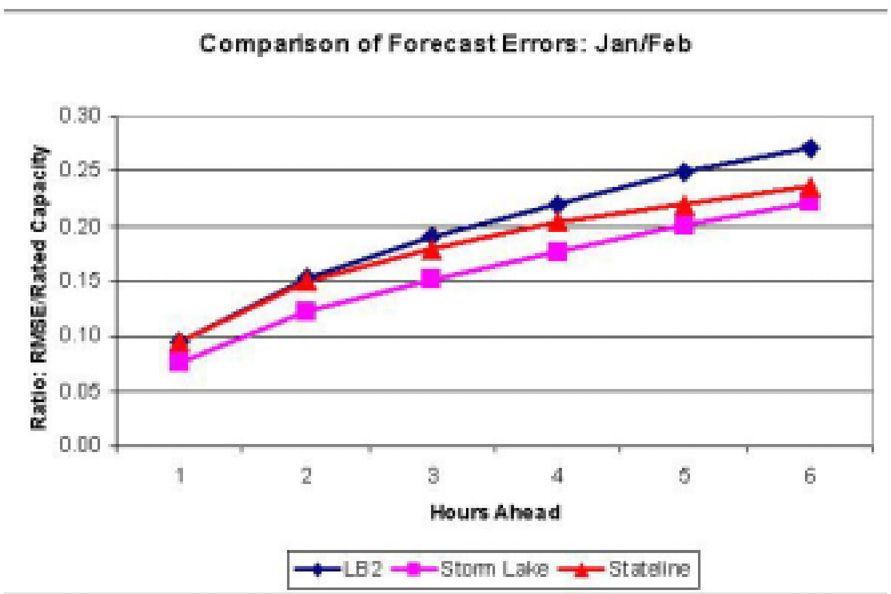


**Figure 4. Forecast degradation: January/February 2001.**

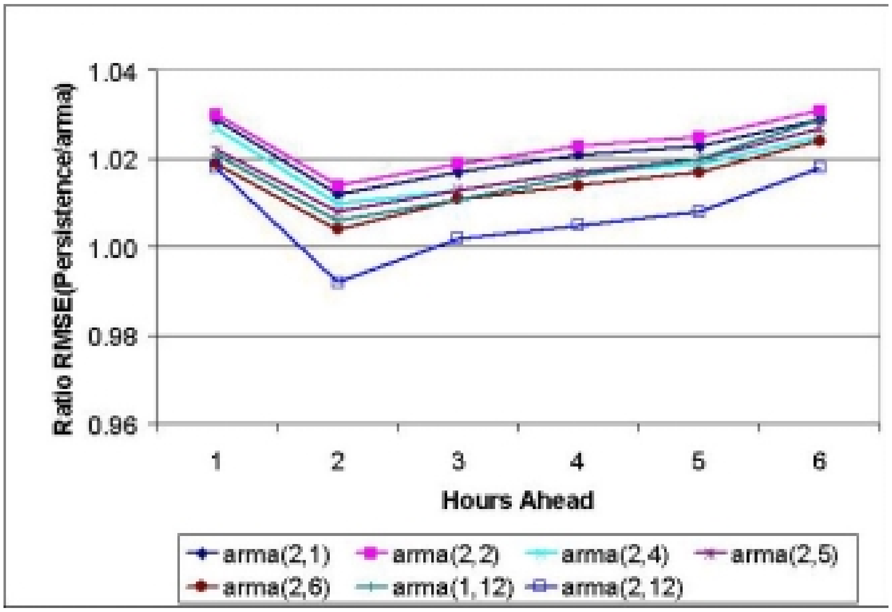


**Figure 5. Lake Benton 2 kW forecasts: April 2001.**

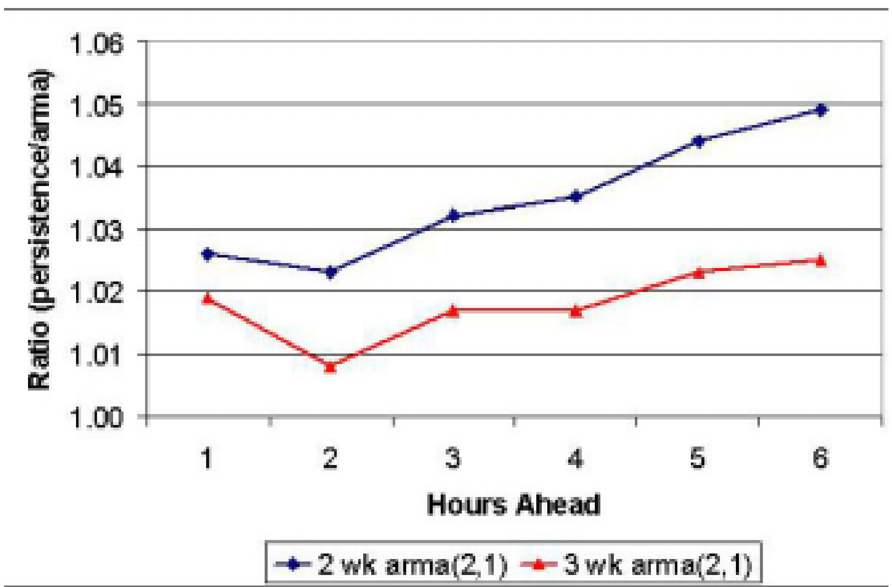


**Figure 6. Lake Benton 2 kW forecasts: dynamic training, March-April 2001.**

week training period offers a modest improvement over the April-only forecasts that were based on a fixed window (as shown in Figure 5 above). The shorter training period also appears to outperform the longer period.

Aside from the statistical performance of the models, different training period lengths have different tradeoffs between the information provided to the model and the constraints that are imposed by the model parameters. For example, a single ARMA model that is fit over a period such as 1 year will potentially have the ability to use a great deal of information that is embedded in the wind signal. Because different physical mechanisms can affect the level of wind resource at different times of the year, the variability of the wind can also have different statistical properties during these periods. This implies that a large number of model parameters may be needed to fully describe the process, or that the parameters could be a function of time. Partitioning the year into a number of distinct time periods allows the model parameters to be re-fit to account for different climatological properties during different times of the year. Choosing too short a period for model training could omit some important information that would help the model's forecasting accuracy. The ideal training period would pick up the important drivers and patterns for different times of the year. Parameters based on one set of climatic drivers should not be imposed on other time periods if it is known that another set of climatic drivers affects the wind resource. When model training and application span a distinct seasonal boundary, for example, we expect the forecasting performance to suffer.

The results from using the dynamic training periods are complex and not open to easy interpretation. Further investigation into this issue is certainly warranted.

## 6. CONCLUSIONS

The ability of ARMA forecast models clearly differs when applied to different time periods. Our forecasting models for LB improved on persistence more in January-February 2001 than it was did March-April 2001. Some of the models developed in this paper offer a significant improvement over the persistence model. Some ARMA models have difficulty reducing the forecast RMSE significantly as compared to persistence. In some cases, the model that provides the best forecast from 1 to 2 hours out is eclipsed by another model for longer forecast horizons. This raises the possibility of simple ensemble forecasts that can offer better forecasting over the horizon than a single model specification.

Comparison of the Stateline results with the other sites also suggests that improvements over persistence do not necessarily correspond to overall model accuracy. We found that the same model specification for the January/February period appeared to work best for LB and SL, with an alternative specification for Stateline. Because our data are from different time periods, we are unable to attribute the model difference to location or time.

In several cases, we found many alternative ARMA models that did a good job forecasting over the testing time frame. We also found it difficult to determine the proper identification for several models, and that led to the evaluation of alternative specifications. It is apparent that a one-size-fits-all approach may not be the best one, based on the apparent sensitivity of the model performance to the length of the training period.

## 7. ACKNOWLEDGEMENTS

This paper was written at NREL in support of DOE under contract number DE-AC36-99-GO10337.

## 8. REFERENCES

Bailey, B., M. Brower, J. Zack (1999). Short-Term Wind Forecasting: Development and Application of a Mesoscale Model. European Wind Energy Conference, Nice, France.

Box, G. E. P., and G.M. Jenkins (1976). Time Series Analysis: Forecasting and Control. San Francisco, Holden-Day.

Kariniotakis, G., N.E. Nogaret, G. Stavrakakis (1997). Advanced Short-Term Forecasting of Wind Power Production. European Wind Energy Conference, Dublin Castle, Ireland, EWEA.

Landberg, L. (1997). Predicting the Power Output From Wind Farms. European Wind Energy Conference, Dublin Castle, Ireland, EWEA.

Makarov, Y., D. Hawkins, E. Leuze, J. Vidov (2002). California ISO Wind Generation Forecasting Service Design and Experience. Windpower 2002, Portland, OR, AWEA.

Milligan, M., M Schwartz, Y. Wan (2003). Statistical Wind Power Forecasting Models: Results for U.S. Wind Farms. Windpower 2003, Austin, TX, AWEA.

Nielsen, T. H. M. (1999). Experiences With Statistical Methods for Wind Power Prediction. European Wind Energy Conference, Nice, France, EWEA.

Wan, Y., Bucaneg, D. (2002). Short-Term Power Fluctuations of Large Wind Power Plants. 21st AMSE Wind Energy Symposium, Reno, Nevada, NREL/CP-500-30747 (preprint).