A Consensus Forecasting Approach for Improved Turbine Hub Height Wind Speed Predictions

William Myers
National Center for Atmospheric Research
Boulder, CO, USA

Gerry Wiener
National Center for Atmospheric Research
Boulder, CO, USA

Seth Linden
National Center for Atmospheric Research
Boulder, CO, USA

Sue Ellen Haupt
National Center for Atmospheric Research
Boulder, CO, USA

OVERVIEW
The National Center for Atmospheric Research (NCAR) has developed a wind prediction system for Xcel Energy, the power company with the largest wind capacity in the United States (Johnson et al. 2011). The wind power forecasting system includes advanced modeling capabilities, data assimilation, nowcasting, and statistical post-processing technologies. The system ingests both publicly available and specialized model data and weather and wind farm observations. NCAR produces a deterministic mesoscale wind forecast of hub height winds on a very fine resolution grid using the Weather Research and Forecasting (WRF) model, run using the Real Time Four Dimensional Data Assimilation (RTFDDA) system (Liu et al. 2008). In addition, a 30 member ensemble system is run to both improve forecast accuracy and provide an indication of forecast uncertainty. The deterministic and ensemble model output plus data from various global and regional models are ingested by NCAR’s Dynamic Integrated Forecast System (DICast®), a machine learning system. DICast® produces forecasts of wind speed for each wind turbine. These wind forecasts are then fed into a power conversion algorithm that has been empirically derived for each Xcel power connection node. This basic system has consistently improved Xcel’s ability to optimize the economics of incorporating wind energy into their power system.

INTRODUCTION
As the wind capacity of a utility grows, it becomes more difficult to effectively integrate this variable resource into the power mix. Wind power forecasting can significantly improve that integration by improving reliability in a manner that can minimize costs (Ela and Kemper 2009). Xcel Energy, with nearly 10% of its capacity in wind energy, has reached the point where it is economically beneficial to forecast wind power in order to better plan integration into their power mix. This is necessary for day ahead trading as well as for real-time decisions regarding the power levels of alternative units and transmission. Therefore, it is important to have accurate predictions of hub height wind speeds and the resulting power production. The details of a specific flow field are difficult to predict with accuracy, however, due to the chaotic nature of atmospheric flow.

NCAR has configured and built a full wind power forecasting system for Xcel Energy. Advanced modeling, data assimilation, nowcasting, and statistical post-processing technologies comprise this system. Xcel Energy has service areas in three distinct regions of the US, which are roughly described as Minnesota, Colorado, and North Texas/East New Mexico. In these three regions, there is roughly 4.8 GW of installed wind power capacity from over 3500 wind turbines.

System components are diagrammed in Figure 1. The system ingests both publicly available and specialized model data and weather and wind farm observations. The observations range from routine meteorological surface and upper air observing networks to data from the wind farm itself. In addition, the system uses wind farm data including wind speeds measured on the turbine Nacelles, generator power, and turbine availability.

Fig. 1. Flowchart of NCAR’s Xcel Energy Power Prediction System.

The strength of the wind power forecasting system lies in the integration of the various components to produce a power forecast that can be used by Xcel Energy to integrate this variable resource into the electricity grid.

The DICast® system is responsible for this integration. DICast® was developed to emulate the human forecast process. The human considers a number of data sources, e.g. NWP models and weather observations, and makes an improved forecast for each based on past experience. This human-
optimized set of forecasts must be reduced to one final forecast. The human does this through a subjective process which attempts to capture which model performs better at each specific location and lead time. DICast® performs each of these steps in an automated fashion, first statistically optimizing forecasts from each model, and then combining the optimized model-specific forecasts by objectively evaluating the past performance of each. This process provides an optimized consensus hub height wind speed forecast at each turbine. This optimized forecast of wind speed at hub height is used to generate more accurate turbine specific power forecasts that can be aggregated into farm or connection node power forecasts.

In the version of DICast® developed for Xcel Energy, new wind forecasts are generated every hour to take advantage of all the latest model and observational data. These forecasts have hourly resolution and run out to seven days. Downstream from DICast®, in the power conversion module, the latest observations are used to generate power forecasts at 15 minute resolution on a 15 minute update cycle.

The following sections describe the DICast® system in more detail. The next section describes the observational data used by DICast® to optimize its forecasts. The following one discusses the Dynamic Model Output Statistics (DMOS) process that is responsible for generating optimized forecasts from each individual input NWP model. This is followed by a section describing the integration of the DMOS optimized forecasts into a consensus final forecast. The final section provides a summary and conclusions.

OBSERVATIONAL DATA

As an automated learning system, DICast® depends on an observational data set to determine which components are performing better than others. DICast® uses data from the Nacelle anemometers that are mounted atop the turbine hubs (Fig. 2). These anemometers measure something different from the free atmosphere wind speed at hub height because the flow is disturbed by the turbine blades rotating just upstream from the sensors. The value of forecasting for these sensors is that there is a large empirical database that relates the wind speed measured by the Nacelle anemometer to the power generated by that turbine. Therefore, we focus on these sensors rather than those on a possibly very distant meteorological tower sensor, which would not have the same close relationship to the power production.

In working in other wind power regions, we have found that we can statistically improve upon the models' forecasts for “nearby” meteorological tower sensors. However, the derived statistical relationships for the towers do not well represent the nearby farms’ turbines. Observations should be as specific to the critical predictand as possible. In this case, that is the turbine power. The Nacelle sensor is the measurement that is most closely related to turbine power.

To achieve success with this approach, significant numbers of the turbines must be instrumented with these Nacelle anemometers. Also, their data must be accurate, reliable, and regularly transmitted to a data aggregation center from which it can be securely accessed over the internet. Xcel has acquired a real-time data feed that includes the nacelle anemometer data for approximately 90% of the turbines from which they obtain wind energy. They have also developed the communications and database infrastructure to make these data available to the weather forecasting processes.

![Fig. 2. Typical mounting location of the Nacelle anemometer on a wind turbine.](image)

The raw Nacelle observations are available at high frequency, usually more than 2 samples per minute from every turbine. Due to the gusty nature of wind, these observations are often rather noisy; that is, the values tend to jump around rapidly in time. It is difficult, if not impossible, to forecast this variable nature of the winds within short time ranges using NWP model data due to its relatively coarse temporal resolution. Instead, a more representative signal must be derived from this observational time series in order to effectively reduce the error in each of DICast®’s multiple optimization steps. To achieve this, 15-minute averages of the measured wind speed are derived and stored for each turbine. These derived wind speeds are more statistically representative of the wind and power produced during that period.

DYNAMIC MODEL OUTPUT STATISTICS (DMOS)

Dynamic MOS is the first optimization step in the DICast® forecast process. The goal is to optimize the forecast from each input NWP model based on available observation verification data. DMOS operates on each model individually, much like a human forecaster examining a model, for example, NCEP’s Global Forecast System (GFS) model, and attempting to make a stand-alone forecast from that model. To do this the human uses subjective memories of which model variables to consider and how to make use of them to generate an optimized forecast for a particular location and meteorological variable. DMOS does this through an objective process similar to the National Weather Service’s (NWS) Model Output Statistics (MOS) process (Glahn and Lowry 1972). Unfortunately, the NWS
does not provide optimized MOS products at locations or heights relevant to wind energy forecasting. DMOS generates an optimized forecast of the Nacelle anemometer’s measurements from each input NWP model.

Many choices are possible in the configuration of NWP models. These include horizontal and vertical resolutions, physics packages, and initialization techniques. The result is that each model behaves differently. DMOS learns how to relate variables in each model’s output to the Nacelle observations at that location.

One of the main difficulties with using NWP model data in the prediction of hub height wind speed is that models usually do not directly predict the wind speed at hub height. Two main classes of model output exist. In the first, data above the surface is provided on specific pressure levels. The data from US National Center for Environmental Prediction (NCEP), the Canadian Meteorological Center (CMC) and the European Center for Medium Range Forecasting (ECMWF) are provided in this format. The spacing between these pressure levels varies but is usually 25 millibars or more near the surface. This translates to roughly 250 m between relevant data points in the vertical. In these models, that will often mean that, at some sites, the first pressure level may actually be below the model’s earth surface. This is an interesting complexity since some of the models, e.g. the North American Model (NAM), forecast only above the surface and yet their output is modified to match earlier versions of the model’s output. The coarse vertical resolution is mitigated somewhat by the existence of a 10 m above ground level (AGL) wind forecast variable in these models. This model variable adds significant value to the DMOS forecast optimization process. In this class of models, very significant reductions in forecast error are possible.

The second class of models is termed terrain following. The output of these models is provided at sigma levels, i.e. specific levels above the ground. These models are run outside the national centers, often specifically for wind energy forecasts. They are configured to have several closely spaced levels near the surface, e.g. 60 m and 100 m. The number and density of levels is determined by trading off vertical resolution with compute speed. Models with output in this format require less interpolation in the vertical. Since this output is more specific to the height of the hub, the forecast error reductions realized by applying DiCast® are less than for the first class of models.

For each of the output classes, many different methods are available to estimate the wind speed at hub height. Any of these estimation methods (called predictors) could be used as a standalone method of forecasting the hub height wind speed based on a model’s data. In the version of DiCast® developed for Xcel Energy, these hub height wind speed predictors are all derived from the pressure (or sigma) level and/or the 10 m wind speed variables. For example, one predictor would be generated by interpolating between the 10m wind speed and the wind speed at the first pressure level above the hub height. Other methods only use one nearby wind speed and extrapolate up or down to the hub height. These predictors require knowledge of the surface roughness at the turbine location.

Each of these predictors has its strengths and weaknesses. The effectiveness of each predictor will vary with location, time of day and lead time. DMOS determines the best multivariate regression equation that relates these predictors to the Nacelle observation. That is, DMOS optimizes the forecasts of individual models with the goal of reducing the Root Mean Squared Error (RMSE). To accomplish this, it takes advantage of the predictors that work well for a prediction specific to a location, lead time, and forecast generation time. As it turns out, at any particular turbine, the predictors that are best for making afternoon forecasts may be suboptimal for nighttime forecasts. Thus, for each model, there is a different regression generated for each turbine, for each lead time, and for each model run time.

The regressions are generated using only a relatively short history. DMOS uses up to 90 days of past forecast predictors and observation pairs to determine the best relationship. Experience has shown that less than 90 days provides less stable regression equations, while longer training periods provide limited marginal benefits and sometimes introduces unwanted seasonal effects.

Forecasts can, in fact, be produced with less than a full 90 day history. After roughly 30 days, the regression-based forecasts are at least as good as any of the individual predictors. This is actually an advantage of the relatively short training period of DMOS when compared to the NWS MOS approach of using a long history and developing one set of equations that are used indefinitely. That is, a new turbine’s observations can be brought into the system and within a few weeks, DMOS can be producing tuned forecasts that continue to improve with time.

If, during the DMOS regression calculation process, it is determined that there is an insufficient model predictor/observational history, or that none of the potential regression equations has a sufficiently high correlation, forecasts can still be made by using one of the predictors derived from the model. The predictor used is the one that has performed best for the region. This provides robustness in case of quality control issues that allow bad observations to pass through to the DMOS regression calculation process.

A new set of regression equations is automatically calculated each week that considers only the most recent 90 days of forecast/observation pairs. In this way, as models evolve, DMOS adapts naturally to produce a better forecast from the current state of the model. Often this evolution is not related to changing the model, but instead can be as simple as integrating a new observational data set into the model’s analysis that has the effect of modifying the model’s influence. DMOS handles these transitions seamlessly. Any static statistical post-processing system would eventually drift from optimal relationships between model and observations.

When the most recent model data arrives, the hub height wind speed predictors are extracted from the model data and stored. To make the DMOS forecast for that particular model,
the appropriate predictors for each site, lead time, and forecast generation time are paired with the already-calculated regression equation. While the regression calculation can be a computationally expensive procedure, this application of the regression equations to the current data is very fast.

The goal of the whole DMOS process is to make improvements on all of the predictors. The predictors are common algorithms that a meteorologist might use to directly make a forecast. If each of these predictors is used as a standalone forecast method, the error characteristics of each can be compared.

Figure 3 plots the forecast errors of individual predictors from the 12Z NAM run for the 1976 turbines operating during the month of January 2010. The diurnal cycle can clearly be seen in the errors as well as a general increasing trend in error with lead time. Each line on Fig 3 represents the forecast error of one of the NAM hub height wind speed predictors. Some predictors' forecasts are much more accurate during the day yet are less accurate at night. Others are consistent throughout the entire forecast period. The NAM 10 m wind speed (number 2304) is particularly poor for direct use as the hub height wind speed forecast. This highlights the difference in wind speeds at 80-100 m compared to the traditionally-forecast near-surface wind speeds.

While these 1976 sites represent a statistically significant sample, it is important to remember that although a predictor may perform poorly in this overall sense, it may be excellent at some sites and, is thus of significant value at these sites. Each predictor has its strengths and weaknesses. No predictor should be removed from the process unless it does not significantly contribute at any sites. In fact, even though the 10 m wind speed is a poor standalone predictor, it is less correlated with the pressure-level winds, and thus often used in combination with the other predictors during the DMOS process.

The error for the DICast® DMOS forecast is plotted as the black line below the other error lines in Fig 3. Thus, it is seen to be statistically better than any of the individual predictors at all lead times out to 72 hours.

Each month the relative skills of the predictors change. That is, the best predictor one month may not be best in the following month. In spite of this constantly changing skill in the predictors, the DMOS forecast is, month after month, statistically better than the individual “predictor of the month”. One can average the errors over a range of forecast lead times to obtain a single number describing the forecast methods’ performance. Typically, the DMOS forecast’s errors are around 5-10% lower than the best individual predictor for that month for the day ahead forecast.

FORECAST INTEGRATION

Once optimized forecasts have been generated from the individual forecast models, DICast® combines these forecasts to produce a consensus forecast. For every point- and time-specific forecast, some models will have performed better than others in the recent past. With the assumption that this skill will continue, these models should be given more consideration in any combination. This is analogous to the human’s decision making process about which model(s) to give more credence or weight in the generation of the final forecast.

![Figure 3. Root Mean Squared errors for individual wind speed predictors compared to the NAM DMOS forecast made in January 2010 for 1976 turbines. The NAM DMOS forecast is the black line with consistently lowest RMSE.](image)

At each forecast generation time, the consensus process must consider which models’ forecasts have shown themselves to be better in the recent past. The result of this evaluation will vary for each site and lead time. Also, since the model’s data arrives asynchronously, the best model may vary from hour to hour. Thus, a “fresh” model is often better than other, older models. Therefore, with the assumption that the model arrival times are predictable, development of weights specific to each model’s skill at a specific forecast generation time is merited.

The Xcel Energy version of DICast® currently uses seven input models. The publicly available models include NCEP’s Global Forecast System (GFS), North American Mesoscale (NAM) model, and the Rapid Update Cycle (RUC) models as well as the CMC’s Global Environmental Multiscale (GEM) model. In addition, a high resolution (3.3 km) Weather Research and Forecast (WRF) model, using the Real Time Four Dimensional Data Assimilation (RTFDDA), is also included. Additionally, the means from each of the two 15 member model ensembles are included in DICast®. These two ensembles use the WRF model and the Penn State/NCAR Mesoscale Model (MM5), each run at 10 km resolution and also using the RTFDDA technology. Only the most recent ensemble run is used. This means that, for example, once the 12Z GFS run has arrived, the previous 06Z GFS run is no longer used in the real-time forecasting process.
DICast® creates its consensus through a simple approach. Each integrated forecast is generated through a bias-corrected weighted sum of input forecasts.

\[
F = \left( \sum w_i f_i \right) / \left( \sum w_i \right) + \text{Bias} \quad (1)
\]

Missing forecasts are removed from the consensus. This is a computationally simple, yet effective, forecast combination method. More complex combination schemes are possible, but in our experience, the marginal improvement does not merit the additional complexity.

The DICast® integrator learns how to make the best forecast over time, just as the human does. Each day the integrator examines how well it did on the previous day’s forecast and uses that information to nudge its weights toward an improved combination. Because of the models’ spatial and temporal performance differences, the integrator’s combination will be different for each site, lead time, and forecast generation time.

By design, the DICast® system is robust in its ability to produce quality forecasts despite the unavailability of one or more of the input forecast models. When one model’s data has not arrived at its expected arrival time, the last available run is used until new data arrives. This may lead to a small degradation in forecast quality that is rectified at the next model run cycle.

A variety of weight calculation approaches are possible. These are compared and summarized in papers by Young (2002) and Gerding and Myers (2003). The adaptive learning approach taken in the DICast® integrator was chosen for several reasons. First, it is computationally simple and robust. Secondly, it can easily adapt to the addition of new input forecast models or the removal of obsolete models. Finally, new sites can be added to the system and, with an initial default weight set, forecasts can be made almost immediately. These weights will rapidly evolve toward a near optimal solution.

The weights are modified in the direction of the gradient in weight space. That is, the vector of weights is nudged in the direction of steepest descent of the error (the difference between the verification \( V \) and the forecast values):

\[
\Delta w_i = S \times (\partial / \partial w_i ) \{ (V - F)^2 \} \quad (2)
\]

The step length \( S \) is a parameter determined by the user to affect how quickly the system adapts. The choice of \( S \) effectively trades off the initial discovery of the optimal combination against the daily update magnitude. Unless the step size is too large, the updated weights would have, by design, led to a forecast with a smaller error had they been used for the previous day’s forecast. There is also a cap on the magnitude of any change so that one day’s missed forecast does not completely alter a set of weights that work reasonably well.

In this way, the DICast® integrator never attempts to directly calculate the optimal set of weights. Instead it takes an approach of pursuing the location of the minimum error. The location of the minimum is rarely stationary. It changes daily. In a larger sense, the nexus of the optimal vector changes seasonally to capture the variability in the models’ skills. Other weight calculation approaches that examine a longer history require more computational resources. Due to the daily variability in the model skill and observational representativeness, this additional computational cost is not merited.

The value of this consensus, multi-model approach and the implementation choices made within the DICast® integrator can be seen in Figures 4 and 5. In the figures, the errors of the individual DMOS optimized forecasts are compared to the errors of the DICast® integrated forecast. The errors are shown for the first 72 forecast hours during two different periods. The first shows errors during the fall and early winter months of 2010. The second shows errors for the first months of 2011. These statistics are over the same 1976 turbines used for Fig. 3.

The errors from the individual DMOS modules are analogous to the DMOS error line in Figure 3, i.e. the lowest error line. During the seasons shown, each of these model’s DMOS forecasts have their unique error characteristics. Some are better during the night than during the day. Others tend to be relatively better at longer lead times. No matter the lead time, the DICast® integrated forecast made using these ingredients statistically outperforms every one of the ingredients by a substantial margin. During any given month, the integrated forecast typically outperforms the best “model of the month” with a reduction in forecast error of 10-15%. Each month there is a reshuffling of the best models, yet the integrated forecast learns rapidly enough that it always has the best error characteristics. Studies of errors in wind power production indicate that this reduction in wind speed error...
directly translates into a similar reduction in wind power generation error.

The results demonstrated in Figs. 4 and 5 argue strongly for a multi-model solution rather than using any single model solution in order to reduce the RMSE (or MAE) of the hub height wind speed forecast. The linkage between accuracy of wind speed forecasts and power forecasts further establishes the benefits of consensus forecasting for hub height wind speeds. It is worth noting that DICast® produces similar reductions in error for other forecast variables, e.g. air temperature and dewpoint temperature.

CONCLUSIONS

NCAR has configured a system for Xcel Energy that integrates high resolution and ensemble modeling with artificial intelligence methods to produce a state-of-the-science wind power forecasting system. The DICast® system is an integral part of this wind forecasting system. DICast® ingests multiple NWP forecast models and optimizes each of them individually through the DMOS forecast process. These optimized forecasts are then combined into an intelligent consensus forecast that statistically outperforms all of the ingredients. This makes a compelling case for post-processing NWP models and generating a consensus forecast.

DICast® is a flexible and extendible system that can easily scale to more forecast locations or different NWP forecast models. Also, since DICast® is not dependent upon a single NWP model, it handles data outages in a robust and graceful manner without a substantial degradation in forecast quality. It provides a completely automated, objective forecasting approach. The result is a repeatable high level of forecast accuracy. That is, DICast® forecasts do not display the subjectivity of the on duty human forecaster.

ACKNOWLEDGMENTS

NCAR thanks Xcel Energy for support of this work, particularly Keith Parks. Many other NCAR employees have contributed to the DICast® development and verification effort, including Julia Pearson, Brice Lambi, Jim Cowie, and William Mahoney.

REFERENCES


