

An Overview of AWS Truewind's Wind Power Production Forecasting Methods for the AESO Wind Power Forecasting Pilot Project

Submitted To:

Alberta Electric System Operator
2500, 330 – 5th Avenue SW
Calgary, Alberta, Canada
T2P 0L4
Attention: Darren McCrank, P. E.

Submitted By:

AWS Truewind, LLC
255 Fuller Road, Suite 274
Albany, NY 12203-3656
Tel: 518-437-8660
Fax: 518-437-8659

Submitted On:

September 11, 2007



1. Introduction

AWS Truewind, LLC (AWST) has been contracted to provide 1 to 48 hour-ahead forecasts of the hourly average wind speed and direction, temperature and pressure as well as the hourly average power production for 12 existing and future Wind Generation Resources (WGRs) in Alberta for a one-year period. The power production forecasts for the existing sites are intended to predict the actual power production as reported by each existing WGR. The power production forecasts for the future sites are intended to provide predictions of an estimated power production based on an assumed wind farm scale power production algorithm that provides a power production estimate from the measured wind speed.

The AWST forecasts are generated by a customized configuration of AWST's wind power production forecasting system, which is called eWind. Section 2 of this document provides a high level overview of the eWind system. Section 3 presents a modestly detailed description of the configuration of the eWind forecast system that is being used to produce the AESO forecasts. Section 4 provides a high level summary of the customized forecast system described in Section 3.

2. eWind System Structure Overview

eWind is a state-of-the-art automated forecasting system developed specifically to meet the wind energy industry's need for accurate plant- or regional-scale power production forecasts from 10 minutes to seven days in advance. eWind is composed of four basic elements: (1) a set of physics-based high-resolution Numerical Weather Prediction (NWP) models; (2) a suite of advanced statistical prediction techniques; (3) a set of plant output models; and (4) a forecast quality control and delivery system. The conceptual design of the eWind system is depicted in Figure 1.

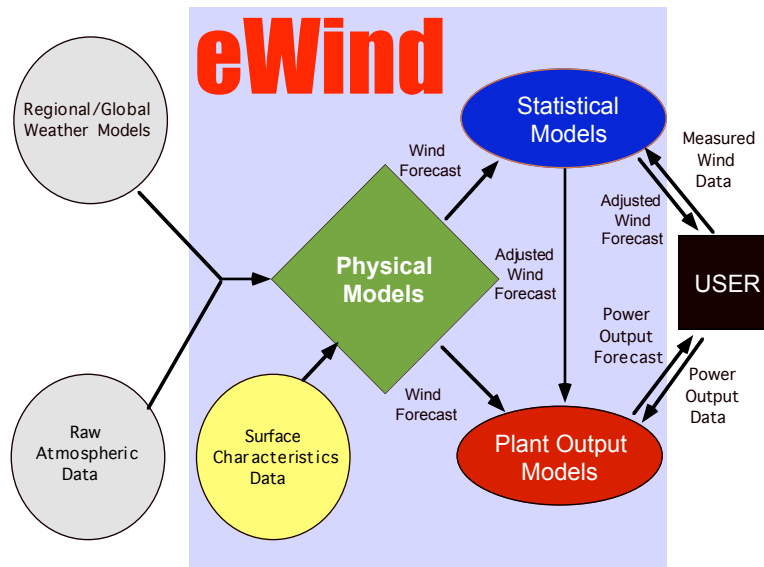


Figure 1. A schematic depiction of the major components and data flow of the eWind wind power production forecast system.

The eWind system utilizes four basic types of input data: (1) the grid point output from regional-scale and global-scale physics-based NWP models executed at government-sponsored forecast centers such as the National Centers for Environmental Prediction (NCEP) in the United States or Environment Canada's operational NWP center; (2) the measurement data from a diverse set of meteorological sensors that provide information about the regional and global structure of the atmosphere at a particular time; (3) high-resolution geophysical data that is used to specify the physical properties of the earth's surface such as the terrain height, the roughness height and heat capacity; and (4) meteorological and power production data from the WGRs (the "user" in Figure 1) for which forecasts are to be produced. In a typical application, data types 1, 2 and 4 are updated many times per day as new measurements are made or new NWP model simulations are executed at a government forecast center. Data type 3 generally consists of slowly changing parameters that do not need to be updated frequently. High resolution databases of the geophysical parameters are compiled by AWST from a variety of sources and reside in a central database in the AWST forecast operations center. This geophysical data is closely reviewed at the start of a forecasting application for a new region to ensure that it is representative for all locations in the region.

While the data from the government-sponsored NWP models provides a general picture of regional weather patterns, such forecasts are almost always too coarse in scale to provide accurate and detailed information for a specific WGR. Therefore, eWind utilizes a set of higher resolution NWP models to simulate atmospheric processes in the vicinity of a wind generation resource. A number of physics-based NWP models are available for use in eWind applications. These include the MASS, MM5, WRF, FOREWIND, and OMEGA models. Typically, several models and initialization datasets are used to create an ensemble of high-resolution simulations for the region surrounding a WGR.

The data from the ensemble of eWind NWP forecasts is input into a "potential predictor" database along with data from a variety of meteorological sensors and meteorological data from the WGR itself. A suite of advanced statistical prediction techniques is then used to transform the potential predictor database into specific predictions of the atmospheric variables at the WGR meteorological tower site. The eWind statistical tools include Screening Multiple Linear Regression (SMLR), Artificial Neural Networks (ANN), Support Vector Regression (SVR), Fuzzy Logic Clustering (FLC) and Principal Components Analysis (PCA). FLC and PCA are often used to define regimes that are utilized to stratify the training samples for SMLR, ANN or SVR. All of the tools are not generally used in an individual application. Typically, an ensemble of predictions is generated by employing different prediction methods as well as variations in the method configuration and the characteristics of the training sample. The ensemble of predictions generated by the statistical procedures is then transformed into a single probabilistic or deterministic atmospheric variable forecast for the WGR site by another statistical model, often referred to as an ensemble compositing model.

The next step is to transform the atmospheric variable predictions for the WGR site into a power production forecast. The eWind system does this with a statistical plant output model that is derived from a sample of concurrent atmospheric and power production data from the WGR. The final step is to perform a quality control check on the forecast data and deliver the forecast to the user via web site display, ftp, email or other methods.

3. eWind Configuration for AESO Forecasting

3.1 Physics-Based Model Simulations

A small ensemble of physics-based model simulations is employed in the configuration of eWind used to produce the AESO power production forecasts. Two different physics-based models are employed: (1) the Mesoscale Atmospheric Simulation System (MASS); and (2) the Weather Research and Forecasting model (WRF). MASS (Kaplan et al, 1982) is a proprietary physics-based model that has been developed by MESO, Inc. (one of the AWST partners) and licensed to AWST. A version of MASS has been customized specifically for wind power production forecasting by improving the atmospheric boundary layer submodel, refining the surface property databases and other modifications. The WRF model (Michalakes et al, 2001, Michalakes et al, 2004) is a community model developed jointly by the US National Center for Atmospheric Research (NCAR) and the US National Centers for Environmental Prediction (NCEP). WRF is an “open source” model and universities and research laboratories around the world have made contributions to it but none have made modifications specifically for the purpose of wind power production forecasting.

In the AESO application, the MASS and WRF models are run from four different initialization and boundary condition datasets produced by larger scale NWP models executed at US or Canadian operational weather prediction centers. These datasets are from: (1) the US North American Model (NAM), which was also previously referred to as the Eta model; (2) the US Global Forecast System (GFS); (3) Environment Canada’s GEM model; and (4) the US Rapid Update Cycle (RUC) model.

A total of 5 sets of simulations are used in the forecast production process. The initialization times and simulation lengths for a reference day (Day 0) are depicted in Figure 2. The first set (light blue in Figure 2) is based on a MASS model simulation employing initialization and lateral boundary condition data from the NAM. The NAM is initialized and run out to 84 hours every 6 hours (00 UTC, 06 UTC, 12 UTC, 18 UTC). An updated NAM dataset is available every 6 hours at approximately 2 to 3 hours after the initialization time. Therefore, the NAM-based MASS simulation are initialized every 6 hours and run out to 66 hours to provide data for the entire 48-hr AESO forecast interval for all forecasts until the next set of NAM-based MASS output data is available. These simulations employ a grid with a horizontal grid cell size of approximately 8 km.

The second set of simulations is also generated by the MASS model. However, the initialization and boundary condition data for these simulations is from Environment Canada’s GEM model. The GEM model is initialized and executed at 12-hour intervals (00 UTC and 12 UTC) by Environment Canada and thus updated output data is available only every 12 hours. The data is available approximately 3 to 4 hours after the initialization time. Therefore, the GEM-based MASS simulations are executed only twice per day. These GEM-based MASS simulations are executed for a period of 72 hours after the initialization time to ensure that output data is available for all 48-hour forecast periods until the next GEM-based MASS simulation is completed. The length of the simulation is longer than for the NAM-based simulations because the time interval

until the next GEM dataset is available is 6 hours longer than the interval between NAM datasets. Thus, the GEM-based MASS simulation data tends to be older than other NWP datasets in the “predictor pool”. The MASS model grid used for these simulations is identical to the grid use for MASS simulations initialized from the NAM data.

The third set of simulations is generated by the WRF model. The initialization and boundary condition data for this set of WRF simulations are from the data produced by the NAM model. As noted previously, new NAM datasets are available 4 times per day and a new NAM-based WRF simulation is executed as soon as each new data set becomes available. These WRF simulations employ a grid with a horizontal grid cell size of 10 km.

The fourth set of physics-based numerical simulations are produced by the WRF model as well. These simulations utilize the output data from the GFS for initialization and lateral boundary conditions. The GFS is initialized and used by the US National Weather Service to produce a 15-day (384 hours) forecast every 6 hours (00 UTC, 06 UTC, 12 UTC and 18 UTC). Therefore, the GFS-based WRF simulations are initiated at 6-hour intervals as soon as the GFS output data from the most recent GFS production cycle is available. The GFS-based WRF simulations are executed on the same grid as that used for the NAM-based WRF simulations.

The fifth and final set of simulations also employs the WRF model. These simulations use the data from the RUC model. A RUC model simulation is initialized and executed every hour. Every third hour (00 UTC, 03 UTC, 06 UTC, 09 UTC, 12 UTC, 15 UTC, 18 UTC, 21 UTC) the RUC simulations have a length of 12 hours. In the intervening hours, the RUC simulations only extend to 9 hours after the initialization time. The RUC-based WRF simulations used in the AESO forecast production procedure are executed every 3 hours using the 12-hour RUC datasets. Each RUC-based WRF simulation is executed on a 4 km horizontal grid and has a simulation length of 12 hours (to match the extent of the data available from RUC). These simulations are intended to provide frequent updates of the anticipated evolution of the small-scale atmospheric features that are responsible for a substantial portion of the wind variations on the 1 to 12 hour time scale.

The data from each of these sets of numerical simulations flows into a database as soon as the simulation is completed. This data is available to the statistical forecast production process for all of the forecast production hours after each numerical simulation is completed, and is used in the statistical component of the forecast process until the next simulation from the same set is available. Thus, the numerical simulation data from each simulation set becomes available to the forecast production process at different times, depending on what time the initialization and boundary condition data is available and how long it takes to execute the MASS or WRF numerical simulation. The availability of the input data (i.e. the NAM, GFS, GEM or RUC data) varies from day to day due to a variety of factors, most of which are beyond the control of AWST, and the time required to execute the MASS and WRF simulations also varies because the number of computations required for a simulation depends on the atmospheric processes that are active during the simulation. For example, a case in which condensation and precipitation occurs in a simulation requires more calculations, than one in which there is no condensation or precipitation.

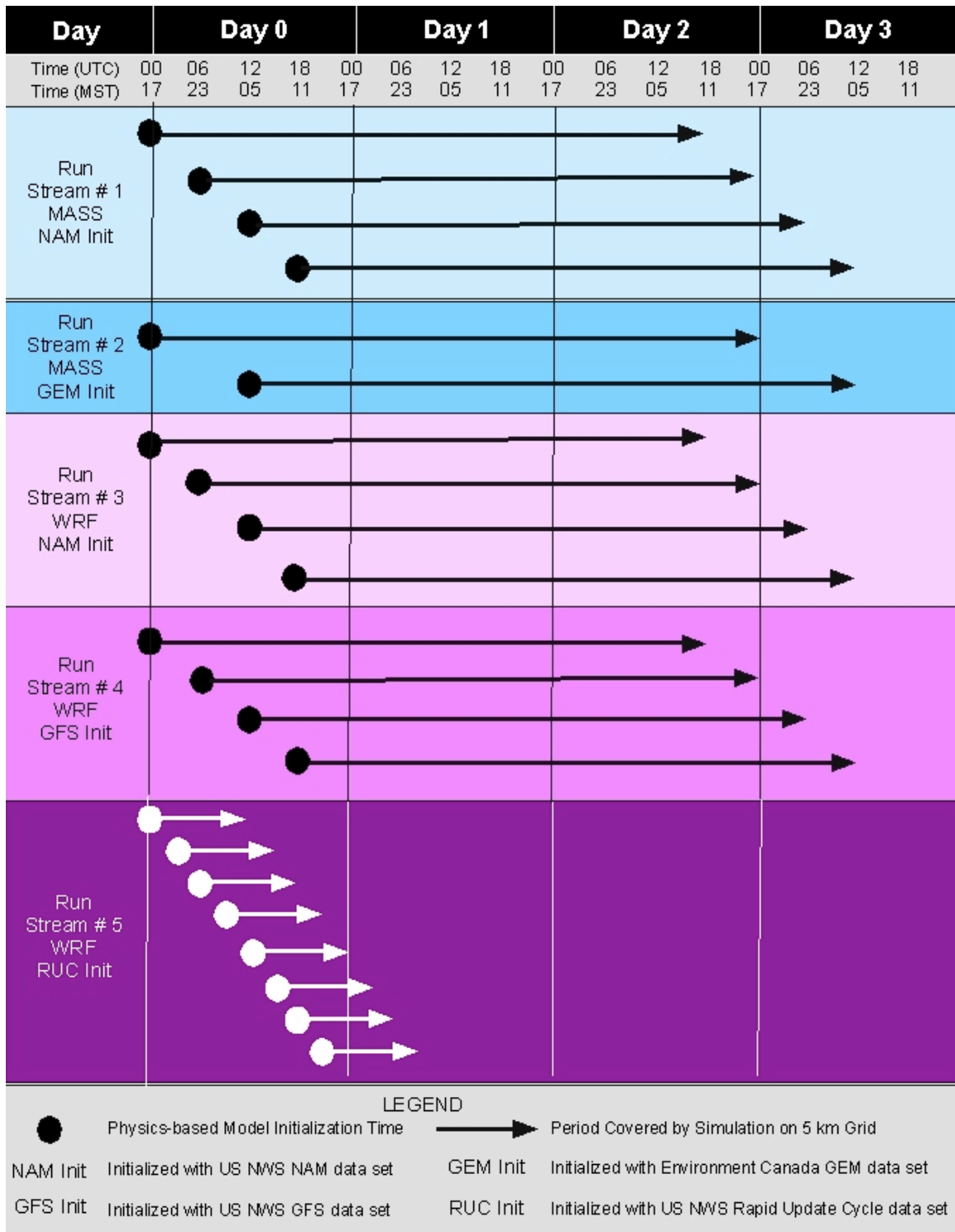


Figure 2. A timeline depiction of the eWind physics-based (NWP) model simulations employed to produce the power production forecasts for the AESO WGR sites.

3.2 Statistical Models

Two fundamental types of statistical prediction procedures are used in the eWind AESO forecast production process: a Screening Multiple Linear Regression (SMLR) method and an Artificial Neural Network (ANN) scheme. An ensemble of statistical forecasts is created through the use of these two procedures by varying the characteristics of the training sample. A separate set of statistical models is created for each forecast look-ahead hour. This enables different predictors to be selected for each look-ahead hour. Typically, predictors based on recent measurements from the WGR are selected for the early look-ahead hours (< 6 hrs) while predictors based on NWP model output are usually dominant for the longer-term look-ahead periods (> 12 hrs).

The master training sample consists of output data from the physics-based numerical simulations described in the previous section, as well as measured meteorological data from the WGRs and other sites in the region. Four different training sample types are employed. The first is a rolling 30-day sample. A new set of statistical forecast equations is created each day by using a training sample consisting of the most recent 30 days. Each day, the data from the oldest day in the sample is deleted and the most recent day's data is added. The second type of training sample is a rolling sample of the data from previous years. This sample consists of data for a 45-day period centered on the forecast production day. One year or multiple years of data may be used depending on the data availability. This sample provides a better representation of the current season than the 30-day trailing sample since it is centered on the forecast production date. The third type of sample is a stratified sample that is divided into "sample bins" based on a physics-based regime classification scheme. Separate statistical prediction equations (SMLR and ANN) are generated for each bin. The bin classification is done through the use of the physics-based model output data. This enables each hour in a forecast period to be classified with the model forecast data. For the AESO application, the bins are derived from the classification of southern Alberta wind regimes proposed in the 1996 paper by Gary Browne (Browne, 1996). This classification system primarily considers the direction and strength of the middle tropospheric flow relative to the orientation of the western Alberta mountains, but also considers the proximity of trough or ridges in the flow. The fourth type of sample is a stratified sample based on the synoptic weather regime. This classification system is based on a PCA of the lower atmospheric wind and pressure patterns. The sample bins are derived from the patterns that explain most of the variance in the wind and pressure fields over Alberta.

The combination of four training sample types and two statistical prediction methods yields a total of 8 statistical forecast schemes. If each statistical scheme is separately applied to the data from each of the 5 sets of physics-based simulations, a total of 40 forecasts can be produced for each forecast hour in each forecast cycle. However, limitations in the availability of data prevent all possible combinations of the statistical and physics-based models from being employed. For example, the regime classification schemes can't be used until a sufficient sample of data is present in each bin. These restrictions have prevented some of these methods from being used early in the forecast test period. These methods will be brought online as a sufficient sample of both model output and measured data is accumulated during the forecast evaluation period. In

addition, the data from the RUC-based simulations is available only for 12 hours after the time of initialization. Therefore, the statistical schemes that employ the RUC-based simulations will only yield results during this time frame and thus cannot be employed for the entire 48-hr forecast interval.

An additional statistical model is used to combine the predictions from the ensemble of individual statistical forecasts into a composite prediction. This model employs an ANN-based scheme to optimally combine the individual forecasts into a single composite forecast for each forecast hour. Initially, the ANN-based forecast compositing scheme is trained through the use of a rolling 30-day trailing sample. This typically means that the forecast methods which have been performing well over the previous 30-days will be assigned a heavier weight when constructing the composite forecast, but in some cases the ANN-scheme will construct complex non-linear combinations of the individual forecasts. Once a sufficient amount of forecast performance data for all of the methods has been accumulated, the method used to composite the forecasts may be switched to a regime-based approach in which relative weighting of the component forecasts will depend upon what wind regime is projected to be present during the forecast hour and the performance characteristics of each method for that regime.

The final output of the ensemble compositing model is a deterministic and a probabilistic forecast of the meteorological parameters (wind speed, wind direction, temperature and pressure) at the meteorological tower associated with each WGR.

3.3 Wind Plant Output Models

The power production forecast for each wind generation resource is constructed by using the forecasts of the meteorological variables for the met tower site associated with each WGR as input into a wind plant output model. The plant output model can be applied to the meteorological forecasts from an individual ensemble member or the composite forecast. In the early months of the AESO application, the final power production forecast has been produced by applying the plant output model to the ensemble composite forecast of the meteorological variable for each forecast hour. However, power production forecasts have also been produced from each of the ensemble members. These forecasts have then been composited to create a single forecast. The compositing of the ensemble of power production forecasts often produces different results than the creation of a power production forecast from the ensemble-composite meteorological forecast. This is typically attributable to the non-linear nature of the transformation between wind speed and power production.

Two types of plant output models are used in the eWind configuration employed for the AESO application. The first is a basic relationship between the average wind speed at the hub height anemometers associated with each WGR and the corresponding power production for the same time interval. The second type of plant output model employs an ANN-based model with inputs that include wind speed and direction, wind speed variability (as a measure of turbulence), and atmospheric stability (i.e. the rate of change of temperature in the vertical direction). The use of this more complex plant output model depends on the availability of a large sample of high quality data from the WGR.

4. Summary

A customized configuration of AWST's eWind forecast system was implemented to produce hourly 1 to 48 hour ahead forecasts of the wind power production for a total of 12 existing and future WGRs in Alberta. The previous section (Section 3) provides a more detailed description of the individual components of the forecast system. As a summary of the Section 3 description, a schematic overview of the eWind configuration used for the AESO forecasting application is presented in Figure 3. The light gray circles depict input data sources. The dark gray ovals denote the intermediate and final output from the forecast system. The rectangular boxes depict forecast models. The dark blue boxes represent physics-based models. The lighter blue boxes represent statistical models. The final output of the system, the wind power production forecast, is denoted by a black oval.

The top row of circles in Figure 3 represents the output data from external (to eWind) NWP models that are run at government forecast centers. As noted in Section 3, four different types of external NWP data are ingested into the eWind forecast process for the AESO application. This data, along with the raw regional atmospheric data (light gray circle on the left side of Figure 3), are used to run eWind's own set of NWP models. These models employ higher horizontal and vertical resolution than the government center models and in some cases also include physics-based formulations that are more customized for low-level wind forecasting than those in the government center models. An ensemble of 5 different types of physics-based simulations are executed in the eWind configuration used in the AESO application. These models produce 3-D forecasts of meteorological variables on a relatively high-resolution grid. The output from the physics-based simulations, as it becomes available from each physics-based model cycle, goes into a "potential predictor" database along with the raw regional atmospheric data and the meteorological data from the WGRs.

The continuously updated composite NWP and observational database is used to train the statistical models to produce forecasts of atmospheric variables at the meteorological tower sites. An ensemble of these forecasts are produced by using two different statistical prediction procedures (SMLR and ANN) and a number of different training sample sizes, contents and stratification bins. The result of this process is an ensemble of forecasts for the atmospheric variables at the meteorological tower sites. This ensemble is converted into a single deterministic or probabilistic forecast for each variable and forecast hour by the ensemble composite model. This ANN-based model is trained on historical forecast performance data and essentially weights each forecast according to its recent performance or its performance in previous occurrences of the anticipated weather regime.

The hourly forecasts of atmospheric variables at the meteorological tower sites are converted to a power production forecast by "the plant output models". These models are typically trained with measured atmospheric variable and power production data although simulated atmospheric variable data may be used for those variables that cannot be computed with the available measured data. The output from the plant output models is a deterministic and probabilistic power production forecast for each forecast hour.

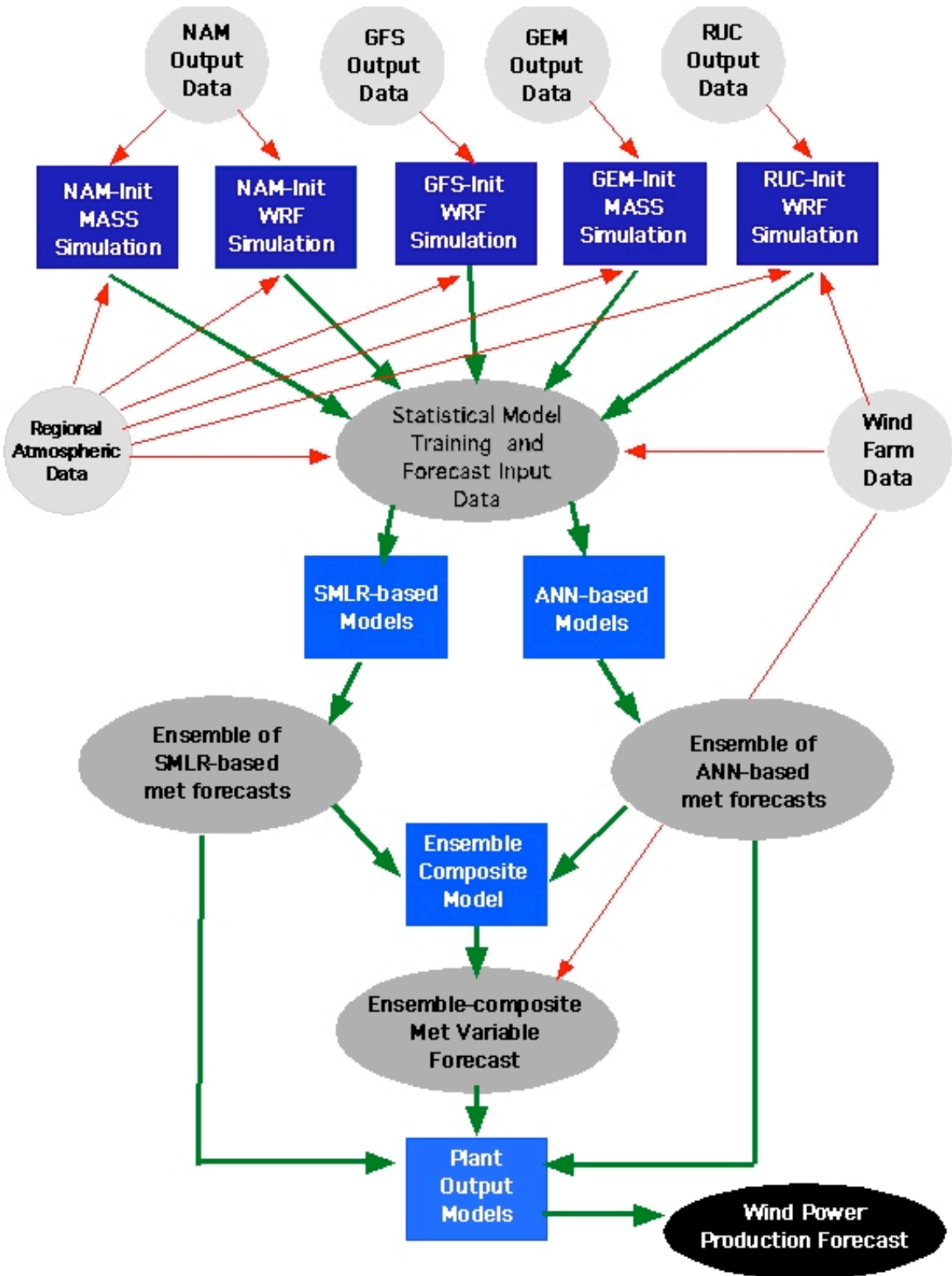


Figure 3. A schematic depiction of the data flow and computational process for the configuration of the eWind forecast system used for the AESO forecasting application.

References

- Browne, G., 1996: Forecasting Winds in Chinook Country of Southern Alberta. Paper prepared by Southern Alberta Environment Services Centre, Atmospheric Environmental Branch, Environment Canada, Calgary, Alberta.
- Kaplan, M. L., J. W. Zack, V. C. Wong and J. J. Tuccillo, 1982: Initial results from a mesoscale atmospheric simulation system and comparisons with an AVE-SESAME I data set., *Mon. Wea. Rev.*, **110**, 1564-1590.
- Michalakes, J., S. Chen, J. Dudhia, L. Hart, J. B. Klemp, J. Middlecoff, and W. C. Skamarock, 2001: Development of a Next Generation Regional Weather Research and Forecast Model. *Developments in Teracomputing: Proceedings of the Ninth ECMWF Workshop on the Use of High Performance Computing in Meteorology*, W. Zwiefelhofer and N. Kreitz, Eds., World Scientific, 269–276.
- Michalakes, J., J. Dudhia, D. Gill, T. Henderson, J. Klemp, W. Skamarock, and W. Wang, 2004: "The Weather Research and Forecast Model: Software Architecture and Performance," to appear in proceedings of the 11th ECMWF Workshop on the Use of High Performance Computing In Meteorology, 25-29 October 2004, Reading U.K. Ed. George Mozdzynski.