Powering Up with Space-Time Wind Forecasting

Amanda S. Hering[1] and Marc G. Genton[2]

May 28, 2009

Abstract

The technology to harvest electricity from wind energy is now advanced enough to make entire cities powered by it a reality. High-quality short-term forecasts of wind speed are vital to making this a more reliable energy source. Gneiting et al. (2006) have introduced a model for the average wind speed two hours ahead based on both spatial and temporal information. The forecasts produced by this model are accurate, and subject to accuracy, the predictive distribution is sharp, i.e., highly concentrated around its center. However, this model is split into nonunique regimes based on the wind direction at an off-site location. This paper both generalizes and improves upon this model by treating wind direction as a circular variable and including it in the model. It is robust in many experiments, such as predicting at new locations. We compare this with the more common approach of modeling wind speeds and directions in the Cartesian space and use a skew-t distribution for the errors. The quality of the predictions from all of these models can be more realistically assessed with a loss measure that depends upon the power curve relating wind speed to power output. This proposed loss measure yields more insight into the true value of each model's predictions.

Some key words: Circular variable, power curve, skew-t distribution, wind direction, wind speed.

Short title: Space-Time Wind Forecasting

[1]Assistant Professor, Department of Mathematical and Computer Sciences, Colorado School of Mines, Golden, CO 80401-1887, USA.

[2]Professor, Department of Statistics, Texas A&M University, College Station, TX 77843-3143, USA.
E-mail: {mhering, genton}@stat.tamu.edu
This research was partially supported by NSF grants DMS-0504896, CMG ATM-0620624, and Award No. KUS-C1-016-04, made by King Abdullah University of Science and Technology (KAUST).

1 Introduction

1.1 Wind Energy Background

The history of harnessing the power in wind for the benefit of man is long and diverse, yet wind energy's current role is evolving rapidly. Throughout the world, the number of installed megawatts increased in 2008 from 2007 by 29%. More facts and information on the role of statistics in wind power can be found in Genton and Hering (2007) and the references therein. Wind farms capable of powering many thousands of homes are springing up both on land and sea. Since the cost of a kilowatt (kW) of wind powered electricity is now nearly the same as a kW produced by coal or nuclear energy, many users are switching to this green energy that produces no greenhouse gases or harmful byproducts. Uneven heating of the earth's surface by the sun produces wind and guarantees that this natural resource will never be diminished or depleted.

Despite its many advantages, utilizing wind energy also presents its share of challenges. The windiest places tend to be the most remote, requiring transmission lines to carry electricity to populated areas. Some complain that the wind turbines ruin the scenery of pristine lands and interfere with bird migration. But by far, the biggest challenges are: (1) the wind is not a steady, constant supply of energy, and (2) no cost-effective method for storing its power currently exists. Its intermittent nature can create a problem for those managing the electrical grid, which is where the supply and demand of electricity meet and must be balanced. Electrical demand is easily predictable based on weather patterns, daylight hours, and holidays or work days. Usually, an equal amount of electricity is ordered to meet this demand from traditional sources. Wind-powered electricity must be used as soon as it enters the electrical grid, so the amount of additional electricity to order from traditional sources becomes unpredictable. Ordering too much or too little electricity can carry severe penalties and fines in utility markets.

Making accurate predictions of the future wind speed reduces the variability and risk that the electrical grid faces once it accepts wind energy as a source (Smith et al. 2007). For a range of wind speeds, the amount of energy that can be produced from a wind turbine is proportional to the cube of wind speed, so small improvements in predicting wind speed lead to larger im-

provements in predicting wind energy. Predictions of wind energy could be made directly, but these are highly dependent upon the types, sizes, and number of wind turbines in operation. A prediction of wind speed, on the other hand, can be used to derive a prediction of wind energy for a given wind farm. The typical forecast horizon needed for scheduling transmission and dispatch is two to four hours. Longer horizons, such as two to three days, are useful for scheduling maintenance of the turbines, and numerical weather prediction models are best for this purpose.

Statistical models, especially those that incorporate expert knowledge of wind characteristics and geography, are unmatched in making short-term predictions (Giebel, Brownsword, and Kariniotakis 2003). However, this area of application has not been exhaustively explored by statisticians (Kestens and Teugels 2002). Gneiting, Larson, Westrick, Genton, and Aldrich (2006) have recently proposed several models for predicting the two-hour ahead average wind speed near a wind farm in northern Oregon. Their best model, called the Regime-Switching Space-Time Diurnal (RSTD) model, accounts for the diurnal, non-negative, and volatile nature of wind speed. It takes advantage of the topography of the Columbia River Gorge in which winds are generally channeled in either an easterly or westerly direction to define two regimes. The regimes switch based on whether the wind direction at a point west of the wind farm is blowing from the west or from the east.

1.2 New Models and Evaluation Tools

In this paper, two new models are introduced that eliminate the RSTD regimes, a loss measure to assess the quality of the predictions in terms of power is proposed, and experiments demonstrate the robustness of the new models. The two new models highlight differences in how the wind speed and direction variables may be treated—either in polar coordinates or in Cartesian coordinates. In the first model, the Trigonometric Direction Diurnal (TDD) model, the wind direction is not simply used to determine the regimes. It is incorporated directly into the predictive mean function of the RSTD model by treating it as a circular variable and using its sine and cosine. Weisberg (2005, pp. 226-230) found that including the sine and cosine of wind direction did not improve wind speed prediction, but his model building approach is not systematic. The TDD

model is more general than the RSTD model and has similar predictive ability.

The second model is called the Bivariate Skew-T (BST) model and uses the 2-dimensional Cartesian wind vector at different locations and lags in time to model the wind speed at the location of interest. The errors in this bivariate regression model are not distributed according to a normal distribution but to a skew-t distribution which is normal in a special case; see the review paper by Azzalini (2005). The skew-t distribution has additional parameters that are flexible for capturing skewness and heavy tails.

Predictions of wind speed are ultimately for the purpose of predicting power; thus, assessing the quality of wind speed predictions should link speed and power (Lange 2005, Lange and Focken 2005). Typical measures such as Root Mean Squared Error (RMSE) or Mean Absolute Error (MAE) for gauging the quality of predictions do not make this link. Power curves describe the relationship between speed and power, and we develop a new loss measure that depends upon this curve. For various ranges of wind speeds, the power output is either constant or proportional to the cube of wind speed. Using a wind power curve for a standard turbine, penalties are assigned to each prediction in terms of power output. Finally, empirical evidence has shown that underestimating wind power averages a higher economic cost than overestimating it does (Pinson, Chevallier, Kariniotakis 2007). Therefore, the penalties are weighted based on the ratio between costs for overproducing versus underproducing, and the effect of the weight on model performance is investigated.

We investigate the robustness of these new models with various experiments. We examine the RSTD predictions when another site besides the most westerly one is chosen to determine the regimes. In fact, choosing a different site with northerly/southerly regimes produces predictions that are as good as those produced with the RSTD model, and choosing a poor set of regimes deteriorates the predictions. This example illustrates that complex decisions involved in selecting regimes can impact the predictions. We also rebuild each model to make predictions at other sites in the dataset and find that the TDD model can perform significantly better than the RSTD model. Finally, the models are rebuilt on data observed at the ten-minute scale instead of data that have been aggregated to the hourly scale. These data are more variable, but the TDD model

performs significantly better than the RSTD model.

This paper is organized as follows. In Section 2, we describe the RSTD, TDD, and BST models in detail. Section 3 introduces the power curve loss measure. Predictive performance of each model and robustness in several experiments are reported in Section 4. We conclude in Section 5.

2 Predictive Wind Models

2.1 Data Description

The data used in this study were collected at 3 meteorological towers near the Columbia River which runs along the Oregon-Washington border. The wind speed and direction were recorded every ten minutes. Vansycle, Oregon is located near the Stateline Wind Energy Center and is the location where prediction is desired. Goodnoe Hills, Washington lies 146 km west of Vansycle, and Kennewick, Washington lies 39 km northwest of Vansycle. Figure 1 shows the approximate relative locations of the three stations. The time series of wind speed and direction are simultaneously recorded at all 3 locations for 55 days from September 4, 2002 to October 28, 2002 (used for training) and also for 279 days from February 25, 2003 to November 30, 2003 (used for testing). Wind speed and direction densities for the 2002 training data averaged for each hour are in Figure 1. Each point on the circular histograms represents an observed wind direction. A point at the 0 angle indicates that the wind is blowing from the east to the west, a $\pi/2$ observation means the wind is blowing from the north toward the south and so on. For complete details on the dataset and site information, the reader is referred to Gneiting et al. (2006).

Many characteristics of the wind vector must be considered in building a model. Inherent in this dataset is spatial correlation. As weather systems move through the area, the site upwind of the others will be affected first, and the current wind conditions at that site will soon prevail at the other sites (Alexiadis, Dokopoulos, and Sahsamanoglou 1999). Of course, which site is upwind of the others will change depending on the orientation of the weather system, but this

can be addressed in the modeling. Strong temporal correlation is also present in the data with significant correlations in both the speed and direction lasting for over 24 hours. The wind speed and wind direction are also strongly linked. Martin, Cremades, and Santabárbara (1999) note the strong correlation between wind speed and direction but then ignore it and model the two variables separately. There is a diurnal pattern in the wind speeds, and seasonal differences do exist (Klink 1999) but are more difficult to model with this limited amount of data. Finally, the wind speed variance varies in time as wind speeds change rapidly and with high frequency, which will be referred to as conditional heteroscedasticity.

2.2 Regime-Switching Space-Time Diurnal Model

The best model that Gneiting et al. (2006) build incorporates many of the variable characteristics discussed in Section 2.1. This particular model will be presented briefly here for clarity, but the reader should see the original paper for the most complete description. In this model, the ten-minute observations of wind speed are averaged over each hour to yield a single hourly observation. The hourly wind speed at Vansycle is modeled with the truncated normal distribution, $N^+(\mu, \sigma^2)$, whose mean and α-quantile are given by

$$\mu^+ = \mu + \sigma \cdot \phi\left(\frac{\mu}{\sigma}\right) / \Phi\left(\frac{\mu}{\sigma}\right) \tag{1}$$

and

$$z_\alpha^+ = \mu + \sigma \cdot \Phi^{-1}\left[\alpha + (1-\alpha)\Phi\left(-\mu/\sigma\right)\right], \tag{2}$$

respectively, where ϕ and Φ denote the density and distribution function of a standard normal random variable. The key to the RSTD model is in choosing a structure for the predictive center, μ, and for σ, the predictive spread. The direction that the wind is blowing during the last ten-minute observation of the hour is used to switch the regimes. When the wind at Goodnoe Hills is blowing from the west to the east (i.e., the wind is westerly or in the westerly regime), the

mean hourly wind speed at a particular location, D_s, is regressed on two pairs of harmonics as

$$D_s = d_0 + d_1 \sin\left(\frac{2\pi s}{24}\right) + d_2 \cos\left(\frac{2\pi s}{24}\right) + d_3 \sin\left(\frac{4\pi s}{24}\right) + d_4 \cos\left(\frac{4\pi s}{24}\right),$$

for $s = 1, 2, \ldots, 24$. Then the least squares fit from the wind speed series at each location is removed, resulting in residual series without a diurnal cycle. The V_t^r, K_t^r, and G_t^r denote the residual series at time t for Vansycle, Kennewick, and Goodnoe Hills, respectively. Then, the predictive center is modeled by

$$\mu_{t+2} = D_{s+2} + \mu_{t+2}^r. \tag{3}$$

The D_{s+2} is the fitted diurnal component at Vansycle, and the μ_{t+2}^r is a linear combination of the present and past values of the residual series at the three sites:

$$\mu_{t+2}^r = a_0 + a_1 V_t^r + a_2 V_{t-1}^r + a_3 K_t^r + a_4 K_{t-1}^r + a_5 G_t^r. \tag{4}$$

When the wind is easterly (blowing from the east to the west) at Goodnoe Hills, removing the diurnal variability from the wind speed series does not result in improvement, so the predictive center is modeled as

$$\mu_{t+2} = a_0 + a_1 V_t + a_2 K_t, \tag{5}$$

where V_t and K_t are the original time series.

For the westerly regime, the conditional heteroscedasticity is incorporated by modeling σ as a linear function of the volatility value with

$$\sigma_{t+2} = b_0 + b_1 v_t. \tag{6}$$

The coefficients b_0 and b_1 are constrained to be non-negative, and the volatility value, v_t, is

$$v_t = \left(\frac{1}{6}\sum_{i=0}^{1}\left((V_{t-i}^r - V_{t-i-1}^r)^2 + (K_{t-i}^r - K_{t-i-1}^r)^2 + (G_{t-i}^r - G_{t-i-1}^r)^2\right)\right)^{1/2}. \tag{7}$$

This reflects the magnitude of the most recent changes in the wind speed. In the easterly regime,

the residual series in Equation (7) are replaced by the original wind series. The parameters in Equations (4), (5), and (6) are estimated numerically by minimizing the Continuous Ranked Probability Score (CRPS) for a truncated normal distribution (Gneiting and Raftery 2007).

The 2002 data is used for building and developing the predictive mean structures in Equations (4) and (5), and the model is tested during the last 214 days of the 2003 series. A window of days in 2003 is used to estimate the parameters in the model before making the first prediction, and this window is rolled ahead by one observation after each two-hour prediction is made, the parameters are estimated again, and so on. Based on experiments performed by Gneiting et al. (2006), the window length that yields the best predictions is 45 days.

2.3 Trigonometric Direction Diurnal Model

Much of the structure of the RSTD model is retained in the TDD model, but Figure 2 clearly shows that the distribution of wind directions at Goodnoe Hills changes from the spring to fall months. It is less clear for months such as October and November if two regimes are sufficient. If not, it is even more difficult to determine how many regimes would be necessary and where the boundaries for these regimes would be. Instead of making a subjective decision about the number and position of the regimes, the TDD model eliminates the regimes but includes the wind direction, possibly at all three locations, as a covariate in the predictive mean function. Since wind direction is a circular variable, we include it in the model as the sine or cosine of the wind direction, following the suggestion by Mardia and Jupp (2000, p. 257). We also use the hourly average of the ten-minute observations of wind direction instead of the last observed wind direction of each hour.

We build the predictive mean function from the pool of variables listed in Table 1 with the Bayesian Information Criterion, or BIC (Schwarz 1978). Only lags up to three hours are shown since none greater are selected with this criterion. Using the 2002 data to build the model, the wind speed at Vansycle two hours ahead is regressed on the first variable, Vansycle's wind speed at the current time. If the BIC of this model is less than the model including only an intercept, then V_t is retained in the model. Then, V_{t-1} is added to the regression. If the BIC is reduced, then

it is also added to the model. If BIC increases, then we do not include V_{t-1} in the model and skip the remaining lags of Vansycle wind speed. Next, both the sine and cosine of the current wind direction at Vansycle are added simultaneously to make the model invariant with respect to the axes, and they are retained if their addition reduces the BIC. This process is repeated with the remaining variables in Table 1. The wind speed variables selected by this process are the same as the ones included in the RSTD westerly regime in Equation (4). In addition, several wind direction components are also included—both the sine and cosine of the current Vansycle wind direction, the sine and cosine of the current Kennewick wind direction, and the sine and cosine of the current Goodnoe Hills wind direction. We denote the wind direction at these locations and times as $\theta_{V,t}$, $\theta_{K,t}$, and $\theta_{G,t}$.

Removing the diurnal component of the wind speed was helpful in the RSTD model, and a strong diurnal component in wind direction is also detected. In Figure 3, the fitted values of a linear model regressing the hourly mean for speed and the hourly circular mean for direction (Fisher 1993, p. 31) on a pair of harmonics are plotted against the hour of the day for each location. If there were no diurnal trend, then the lines would be flat. All three locations show a clear cyclical pattern in the wind direction, so the fitted hourly mean direction is subtracted from each of the wind direction series. Thus, the predictive mean is modeled as in Equation (3), where D_{s+2} is still the fitted diurnal component of the wind speed at Vansycle, and

$$\begin{aligned}\mu^r_{t+2} &= a_0 + a_1 V^r_t + a_2 V^r_{t-1} + a_3 K^r_t + a_4 K^r_{t-1} + a_5 G^r_t + a_6 \sin(\theta^r_{V,t}) + a_7 \cos(\theta^r_{V,t}) \\ &\quad + a_8 \sin(\theta^r_{K,t}) + a_9 \cos(\theta^r_{K,t}) + a_{10} \sin(\theta^r_{G,t}) + a_{11} \cos(\theta^r_{G,t}). \qquad (8)\end{aligned}$$

The scale of the truncated normal distribution is modeled as a linear function of the volatility value as in Equation (6).

2.4 Bivariate Skew-T Model

The BST model differs substantially from either the RSTD or TDD models. Instead of using hourly wind speed and hourly direction directly, these variables are converted into Cartesian

components with $x = r\cos(\theta)$ and $y = r\sin(\theta)$ for r a wind speed and θ a wind direction. Let $\mathbf{V}_t = (V_{t,x}, V_{t,y})'$ denote the Cartesian components of the wind vector at Vansycle at time t. Here, $V_{t,x}$ is the east-west component, and $V_{t,y}$ is the north-south component. Let $\mathbf{K}_t$ and $\mathbf{G}_t$ denote similar vectors of values at Kennewick and Goodnoe Hills. The diurnal cycle is again removed from each component at each location by fitting a pair of harmonics to each set of hourly means, denoted by $D_{s,x}$ and $D_{s,y}$. Then, each component is standardized by dividing by an overall standard deviation computed at each location, denoted σ_x and σ_y (Brown, Katz, and Murphy 1984). For example at Vansycle, the series is transformed by

$$\mathbf{V}_t^r = \left(V_{t,x}^r, V_{t,y}^r\right)' = \left(\frac{V_{t,x} - D_{s,x}}{\sigma_x}, \frac{V_{t,y} - D_{s,y}}{\sigma_y}\right)'.$$

These centered and standardized residual series will be denoted as $\mathbf{V}_t^r$, $\mathbf{K}_t^r$, and $\mathbf{G}_t^r$.

The residual series at time $t+2$ at Vansycle is modeled by

$$\mathbf{V}_{t+2}^r = \mathbf{A}_0 + \mathbf{A}_1\mathbf{V}_t^r + \mathbf{A}_2\mathbf{V}_{t-1}^r + \mathbf{A}_3\mathbf{K}_t^r + \mathbf{A}_4\mathbf{K}_{t-1}^r + \mathbf{A}_5\mathbf{G}_t^r + \boldsymbol{\epsilon}_t, \tag{9}$$

where $\mathbf{A}_0$ is a 2-dimensional vector of constants, $\mathbf{A}_i$ is a 2×2 matrix of coefficients for $i = 1, \ldots, 5$, and $\boldsymbol{\epsilon}_t$ follows a bivariate skew-t distribution. Then the random vector $\mathbf{V}_{t+2}^r$ follows a skew-t distribution whose location parameter is $\boldsymbol{\xi} = \mathbf{A}_0 + \mathbf{A}_1\mathbf{V}_t^r + \mathbf{A}_2\mathbf{V}_{t-1}^r + \mathbf{A}_3\mathbf{K}_t^r + \mathbf{A}_4\mathbf{K}_{t-1}^r + \mathbf{A}_5\mathbf{G}_t^r$, with scale matrix $\boldsymbol{\Omega}$, shape parameters $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)'$ to model skewness, and degrees of freedom ν to model kurtosis (Azzalini 2005). In short, $\mathbf{V}_{t+2}^r \sim ST_2(\boldsymbol{\xi}, \boldsymbol{\Omega}, \boldsymbol{\alpha}, \nu)$. The variables in the model in Equation (9) are selected using a BIC procedure similar to that used for the TDD model. The parameters are estimated using maximum likelihood estimation with the `R` package `sn` (Azzalini 2006). Figure 4 shows that the skew-t distribution for the errors is a much better fit to the 2002 training data than the normal distribution for the errors is.

Then, the predicted vector of Cartesian components at Vansycle two hours ahead is given by

$$\hat{\mathbf{V}}_{t+2} = \hat{\boldsymbol{\Sigma}}\hat{\mathbf{V}}_{t+2}^r + \mathbf{D}_{s+2},$$

where $t = 1, 2, 3, \ldots$ and $s = ((t-1) \mod 24) + 1$. The $\hat{\mathbf{A}}_i$, $i = 0, 1, \ldots, 5$, in $\hat{\mathbf{V}}^r_{t+2}$ are estimated from a 45 day window of data before the desired two-hour ahead prediction; $\hat{\boldsymbol{\Sigma}}$ is a matrix with the standard deviations of the x components and the y components estimated from the 45 day window on the diagonal and zeroes on the off-diagonal; and $\mathbf{D}_{s+2} = (D_{s+2,x}, D_{s+2,y})'$ is the fitted diurnal mean of the x and y components at Vansycle. Thus, the linear transformation of $\mathbf{V}^r_{t+2}$ gives $\mathbf{V}_{t+2}$ a $ST_2\left(\boldsymbol{\Sigma\xi} + \mathbf{D}_{s+2}, \boldsymbol{\Sigma\Omega\Sigma}', \boldsymbol{\Sigma}^{-1}\boldsymbol{\alpha}, \nu\right)$ distribution (Azzalini and Capitanio 2003). The predictive distribution of the wind speed requires taking the norm of $\mathbf{V}_{t+2}$, so the norm of 50,000 observations drawn from a skew-t distribution with parameters estimated from each 45 day window of data is taken as the simulated predictive distribution. A large number of observations can be simulated quickly and easily and ensures that the behavior in the tails of the distribution is accurately characterized. The 45 day window length is chosen since it yields slightly better predictions than 30 and 60 day windows. This window length also makes the BST model easier to compare with the RSTD and TDD models, which also use 45 day windows.

3 Power Curve Loss Measure

Wind speed predictions from different models are commonly compared with RMSE and MAE, but these are not necessarily the appropriate loss functions in the wind forecasting paradigm. A better loss function should relate predicted wind speeds to the wind power since predicting power is the ultimate goal (Madsen, Pinson, Kariniotakis, Nielsen, and Nielson 2005). The power depends on several factors, such as the air density ρ, the radius swept by the turbine blades r, and the wind speed v as follows

$$P = \tfrac{1}{2}\,\alpha\rho\pi r^2 v^3, \tag{10}$$

where α is an efficiency constant. As a baseline power curve, we use the GE 1.5 megawatt (MW) manufacturer's power curve (black dots in Figure 5) with fixed air density. The relationship between speed and power is not perfectly predictable, possibly even depending on the wind direction (Potter, Gil, and McCaa 2007), but for practical purposes, we assume here that it is.

Four zones of the power curve are defined by the cut-in speed, the rated speed, and the cut-

out speed. The cut-in speed is the speed at which the turbine blades begin to rotate. The rated speed is the lowest wind speed at which the maximum power output of the turbine is achieved. The cut-out speed is the speed at which the blades stop rotating to protect the turbine from damage. Zone 2 in Figure 5 is where the relationship in Equation (10) holds, and the solid curve in this region is a nonparametric Nadaraya-Watson type of estimate (Nadaraya 1964; Watson 1964) fitted with bandwidth $h = 0.025$. Small changes in the wind speed here can result in large differences in power output since power depends on the wind speed through a cubic function.

When both the observed and forecasted wind speeds are in Zone 1, 3, or 4, either no power output occurs or the maximum power output occurs. For example, if both the forecasted and the observed wind speeds are in Zone 3, then the power output is the same regardless of whether the wind speed forecast is close to the observed speed or not. No penalty would be assessed in terms of power for any differences in the observed and forecast speeds. When both the predicted and observed wind speeds are in Zone 2, small differences in forecasting wind speed will result in greater differences in forecasting wind power. As a result, discrepancies between the observed and forecasted wind speeds should receive greater penalties in this region.

We define $g(\cdot)$ to be the nondecreasing function that maps speed to power. The power curve is not a nondecreasing function, but only four of the 5,136 wind speeds in the testing dataset are greater than the cut-out speed, so we ignore these cases. Precise power output data is not available since the power generated at the wind farm near Vansycle is proprietary information. Instead, an estimate of the true power is obtained with $g(V_{t+2})$ that will be compared to the predicted power output based on the forecast wind speed, $g(\hat{V}_{t+2})$. Thus, a loss function that is of the Generalized Piecewise Linear form is defined to be

$$L(V_{t+2}, \hat{V}_{t+2}) = \begin{cases} \gamma(g(V_{t+2}) - g(\hat{V}_{t+2})), & \hat{V}_{t+2} \leq V_{t+2} \\ (1-\gamma)(g(\hat{V}_{t+2}) - g(V_{t+2})), & \hat{V}_{t+2} > V_{t+2} \end{cases}, \tag{11}$$

where γ is a weight between 0 and 1 and allows underestimates to be penalized differently than overestimates.

Empirical data from the Dutch electricity market in 2002 suggests that $\gamma = 0.73$, penalizing

underestimates more strongly than overestimates (Pinson, Chevallier, Kariniotakis 2007), which may at first seem counterintuitive. However when viewed from a holistic system perspective, an underestimate of wind power will cause the system operator to order too much electricity from traditional sources to meet the demand. In this case, the system operator now has a surplus of electricity, and down-regulation (when generation must be reduced) tends to be more expensive than up-regulation (when generation must be increased). The Power Curve Error, PCE, averages the penalties in Equation (11) over all forecasts and will be directly related to the energy produced by a wind farm (Madsen et al. 2005).

The optimal forecast that minimizes a particular loss function is given by

$$\hat{V}_{t+2} = \arg\min_{v_{t+2}} \; E_F \left[L(v_{t+2}, V_{t+2}) \right],$$

where F is the predictive distribution. In the simple cases where the loss function is squared error or absolute error, the optimal forecast is the mean or the median, respectively. For the error in Equation (11), the γth quantile minimizes PCE (Gneiting 2008). Thus, the mean, median, and γth quantile of each model's predictive distribution will be used to compare the forecasts. The mean of the truncated normal distribution in Equation (1) and the median and γth quantile from Equation (2) are extracted from the RSTD and TDD models. The mean, median, and γth quantile are computed numerically from the simulated predictive distribution generated from the BST model.

4 Model Robustness

4.1 Comparing Model Performance on Testing Data

A simple baseline forecast is the persistence model. The persistence forecast for the average wind speed at Vansycle two hours ahead is simply the current wind speed at Vansycle. The mean of the predictive distributions of the RSTD, TDD, and BST models is used to compute the RMSE, the median is used for the MAE, and the γth quantile is used for the PCE. A measure called

the continuous ranked probability score (CRPS), which essentially measures the spread of the predictive distribution subject to calibration is also computed. The CRPS can be computed explicitly for the predictive truncated normal distribution as given in Gneiting et al. (2006), and the CRPS value for the BST model is computed using the approximation in Equation (3) from Grimit, Gneiting, Berrocal, and Johnson (2006). Table 2 lists the results on the training data. The model with the lowest of each value in each column is bolded.

Overall, the TDD model has the smallest value or one of the smallest values for RMSE, MAE, PCE, and CRPS. It has the advantage over the RSTD model of being more general but retains the RSTD's predictive ability. In terms of PCE, the TDD model has the lowest values through the majority of the months and does as well as or better in terms of the other measures through the fall months. The BST model does not do as well as TDD and RSTD in any of the overall measures, but it does have the smallest RMSE, MAE, and CRPS in May and July and the smallest PCE in May. The BST model, like any robust fitting technique, fits to the majority of the data in the fitting window and is insensitive to unusually large or small values (Azzalini and Genton 2008). Thus, its forecast for unusually high wind speeds tends to be poor. The wind speeds in May and July have the smallest standard deviations of any of the months, so the BST model does well during these months.

The differences among the models may seem small, but small differences are still important from a practical perspective. To test if these differences are significant, the large sample test introduced by Diebold and Mariano (1995) for comparing the forecast accuracy of competing models can be applied to check for significant differences between functions of the errors of two models. We test the null hypothesis that there is no significant difference between the overall MSE, MAE, or PCE of two models. With 5,136 two-hour ahead hourly forecasts, the p-value to test for significant differences between the MSE of the RSTD and TDD models is 0.3337, and the p-value for the test of significant differences between their MAE's is 0.8713. Thus, we do not have evidence that the TDD model is significantly different in terms of squared or absolute errors. Both the TDD and RSTD models are significantly better than the BST model in terms of MSE and MAE. The p-value to test for a significant difference between the PCE of the TDD

and RSTD models is 0.8457 and between the RSTD and BST models is 0.4375, neither of which is strongly significant.

A better sense of the difference between the two models in terms of wind power over the testing set is given in Figure 6. For each observation, the difference in accumulated PCE penalties between the RSTD and the TDD models (top), between the RSTD and the BST models (middle), and between the BST and the TDD models (bottom) is plotted for all predictions made up until that observation. It should be noted that what is plotted is not PCE but the sum of the differences in penalties assigned by the PCE function for each prediction and has not been averaged. A similar graphical approach is taken in de Luna and Genton (2005) and serves to compare the cumulative forecasting ability of two models over a given time period and the gains or losses that would result. Based on this, the RSTD model makes steady improvements over the TDD model from May to the middle of July, from the middle of August to mid-September, and then for the first few days in November. However, the TDD model makes large gains in the beginning of August, middle of October, and end of November that leave it with a better accumulated PCE at the end of the testing period. When comparing the RSTD and TDD models with the BST model in the bottom two panels, except for the short periods in May and July, the RSTD and TDD models dominate the BST model in terms of PCE.

In all three models, the parameter estimates change with each new forecast, but to give a sense of their values, the averages over all forecasts for μ_{t+2} and σ_{t+2} in the RSTD model are 7.02 and 1.70, respectively. The average parameter estimates in the TDD model are 7.00 and 1.74, which are quite similar to the RSTD values. In the BST model, the average of the estimated values of the skewness parameter $\boldsymbol{\alpha}$ is $(-0.17, 0.01)'$, an indication that there is very little skewness in the distribution of the x and y components. The most interesting parameter in the BST model is the degrees of freedom, ν, which averages 5.26 and is always between 3.69 and 7.66, indicating that the distribution has very heavy tails.

The predictive distributions of the three models can look quite different, depending upon the forecast, as shown in Figure 7. The top panel shows the predictive distribution of all 3 models when the TDD model produces the best forecast. The RSTD distribution is very similar, but

the BST model is centered incorrectly and is more concentrated. However, when the TDD model produces a poor forecast in the bottom panel of Figure 7, it also can be centered incorrectly. The RSTD model, in this case, produces a good forecast, but the predictive distribution is very widely spread. The BST model is not only centered closer to the forecast, but it is also very tightly distributed. Over all forecasts, the 90% predictive intervals based on the upper 95% and lower 5% quantiles of these distributions have mean width 5.44, 5.52, and 5.96 for the RSTD, TDD, and BST models, respectively, with empirical coverages of 89.43%, 89.99%, and 91.59%. The TDD model has slightly wider intervals than the RSTD model and also slightly better empirical coverage. The BST has the widest intervals, and the coverage is a bit higher than the stated level.

4.2 Alternate Regime Selection

Some justification for using Goodnoe Hills as the site where the regimes are determined for the RSTD model is given in Gneiting et al. (2006), but Kennewick does not seem to have been considered as a potential site for the regimes to switch. We refit the RSTD model using Kennewick to determine the regimes. First, an easterly/westerly set of regimes is tested and then also a northerly/southerly set of regimes since Kennewick's main mode is nearer $\pi/2$ than it is to π, see Figure 1. The TDD and BST models do not need to be refit. Table 3 shows the results for the RSTD model for both the east/west regimes and the north/south regimes. The TDD and BST model results and the original RSTD model outcomes are also displayed for comparison.

First of note is that using an east/west set of regimes switching at Kennewick deteriorates the RSTD predictions as compared to using Goodnoe Hills as the regime indicator. However, what is remarkable is that the north/south regimes at Kennewick can produce very good results with some values of RMSE, MAE, and PCE being smaller than those for the original RSTD model. The north/south regime still does not have smaller overall PCE than the TDD, but in three months it does produce the best PCE values. This illustrates the fact that unless all possible regimes and stations are tested, it may be impossible to empirically choose the site and regimes that yield the best predictions. If more stations with wind speed and direction data become

available, this would only complicate the selection of a site at which to determine the regimes. In fact, the regimes may not depend on a single site only but on a possibly nonlinear combination of several sites. It seems reasonable to avoid such a selection when possible.

4.3 Predictions at Kennewick and Goodnoe Hills

To test the ease of adaptation and the performance of these models in new locations, the variables are reselected to make predictions at the other two locations in the dataset, Kennewick and Goodnoe Hills. When predicting at Kennewick and Goodnoe Hills, the best choice of regimes for the RSTD model may change, but the model is applied "blindly" in the sense that we want to see how applicable it is to a new location. Variables are reselected for the RSTD predictive mean functions, but the easterly/westerly regimes that switch at Goodnoe Hills are held fixed.

The results in Tables 4 and 5 show that the TDD and BST models have smaller summary measures than the RSTD model at Kennewick, but RSTD is difficult to beat at Goodnoe Hills. The TDD model has a significantly lower MSE at Kennewick than the RSTD model does (p-value = 0.0052), and both TDD and BST have the smallest RMSE, MAE, or PCE in various months. Predicting at Kennewick is more difficult due to the more highly variable wind speeds observed there, which is also reflected in Kennewick's larger PCE values. Goodnoe Hills is the one location situated directly in the Columbia River Gorge, so the regime-switching model best captures the wind flow pattern. Goodnoe Hills also has the fewest unusually large wind speeds, which is evidenced by the lower RMSE and MAE values. In this situation, RSTD has the lowest overall PCE, but it is not significantly different from that of TDD (p-value = 0.7550).

4.4 Finer Scale Data

One final experiment on the models returns us to the full dataset with wind speed and direction measured every ten minutes. This finer scale of data exhibits more variability and is not as predictable as the hourly averaged wind speed. Two approaches are tested in which models are rebuilt to forecast at Vansycle using either the full dataset or the ten-minute observations that occur on the hour. For models built on all ten-minute observations, a twelve-step forecast horizon

is needed to arrive at the two-hour prediction. Predictions are made for $5,136 \times 6 = 30,816$ time-steps. The predictions made on the hour are reserved to compare with the model built from the ten-minute observations that occur on the hour. In that model, a two-step forecast is the two-hour forecast, and only $5,136$ predictions are made.

Questions of interest in these models include whether using the full set of ten-minute observations will improve the two-hour forecast and whether the models will have similar results to those in Table 2. In Table 6, it is shown that models built with all of the ten-minute observations have very little predictive improvement compared to the models using only the ten-minute observations on the hour. However, the TDD model appears stronger relative to the RSTD model than it does in Table 2. In fact, it is significantly better than the RSTD model in terms of MSE for both the full set of observations and the ten-minute observations on the hour (p-values 0.0031 and 0.0000, respectively) and also in terms of MAE (p-values 0.0059 and 0.0088).

4.5 Underestimation Penalty

The weight that is given in Section 3 for the Power Curve Error, $\gamma = 0.73$, deserves some attention. The purpose of this weight is to penalize underestimation more strongly than overestimation of wind power. However, it is not a fixed value. In the Dutch market over the course of the year, the value of γ ranges from 0.51 to 0.98 through the 4 quarters of the year, and it varies from 0.14 to 0.96 over the 12 months of the year (Pinson et al. 2007). Markets with different sets of rules can also affect the value. In addition, a single wind farm usually does not produce enough energy to affect electricity prices, but the larger the penetration of wind energy, the more significantly γ would be affected.

We have used $\gamma = 0.73$ as an example up to this point, but in Table 7, we show the value of PCE for the three models based on hourly data when $\gamma = 0.73$ is replaced with a range of values. We want to determine if the results from PCE are influenced by the value of γ, and in each case, the optimal γth forecast is used in the computation of PCE. With the smallest and largest values of γ, no one model has a consistently smallest PCE over the months. When $\gamma = 0.10$, BST has more small monthly values of PCE than the other models, and when $\gamma = 0.90$, the TDD model

appears to be favored. When $\gamma = 0.50$, both BST and TDD have the smallest PCE for each of 3 months. This experiment demonstrates that no single model is routinely favored over the others for every possible value of γ, so PCE should be used only after a relatively stable estimate of γ for a given market has been determined.

5 Conclusion

The importance of conserving natural resources and exploiting the clean electricity provided by wind energy will only continue to grow in the future. One goal of this paper has been to present model-building strategies for short-term wind speed predictions when both the wind speed and direction information are available over space and time. Wind farms with different terrain and different numbers of nearby meteorological stations can use the TDD or BST modeling approaches to fit similar predictive mean functions, whereas the RSTD model is limited to few locations and known physics. Additionally, speed and direction are often converted to the Cartesian coordinate system, but models like TDD demonstrate the benefit of treating wind direction as a circular variable instead. To conclude, the TDD model produces forecasts that are as good as the RSTD model for this dataset while maintaining more generality. The BST model does not perform as well in terms of PCE on this data, but it does have the added feature of producing a wind direction forecast, which the other two models cannot do.

In comparing models, the power curve error assigns a greater penalty to wind speeds predicted to be in the region where power is roughly proportional to the cube of speed and also penalizes underestimates more strongly than overestimates. Attributing loss in this way directly exploits the nonlinear relationship between power and speed and puts wind power into the larger context of the entire utility system. PCE can easily be adapted for different turbines and different markets and can be averaged over several wind farms to get a more stable estimate. Finally, it may not be reasonable to assume that an error made at a low power has the same economic cost as the same error made at a higher power. An investigation into the effect that the magnitude of wind power for a given error has on the associated loss would need to be conducted.

The work done here could be extended in several ways. Future tests of these models should

incorporate year-round observations so that model performance can be assessed in every season. Including additional covariate information, such as numerical weather prediction model output or pressure differences east and west of Vansycle, should also improve predictions. The optimality of the forecasts can continue to be evaluated with tests such as those introduced in Patton and Timmermann (2007).

While the focus in this work has been on point forecasts, having uncertainty estimates of the forecasts that include uncertainty about the parameter estimates and variable selection would also be of interest. Either model-free bootstrapping techniques (Alonso, Peña, and Romo 2006) or using a fully Bayesian analysis (Wikle, Milliff, Nychka, and Berliner 2001) could be interesting approaches to obtain such intervals. Finally, wind farms with dominant weather patterns that differ from those of the Pacific Northwest and with varying numbers and locations of off-site observations would be interesting applications for the TDD and BST models. The TDD and BST models' predictions for this data give promise that these flexible models could work well with new datasets.

Note: All circular plots were plotted using the `circular` package in `R` by Lund and Agostinelli (2006).

Acknowledgments: The authors would like to thank the Editor and two referees for their helpful comments and suggestions, as well as Tilmann Gneiting for providing the data and computer code for the RSTD model. Stel Walker of Oregon State University's Energy Resources Research Laboratory and Bonneville Power Administration provided the ten-minute data. We also thank Michael Stein for helpful comments made on an earlier version of this work.

References

Alexiadis, M. C., Dokopoulos, P. S., and Sahsamanoglou, H. S. (1999) "Wind speed and power forecasting based on spatial correlation models," *IEEE Transactions on Energy Conversion*, 14, 836-842.

Alonso, A. M., Peña, D., and Romo, J. (2006) "Introducing model uncertainty by moving blocks bootstrap," *Statistical Papers*, 47, 167-179.

Azzalini, A. (2005) "The skew-normal distribution and related multivariate families (with discussion by Marc G. Genton and a rejoinder by the author)," *Scandinavian Journal of Statistics*, 32, 159-200.

Azzalini, A. (2006) "`sn`: The skew-normal and skew-t distributions," `R` package version 0.4-2.

Azzalini, A. and Capitanio, A. (2003) "Distributions generated by perturbation of symmetry with emphasis on a multivariate skew-t distribution," *Journal of the Royal Statistical Society, Series B*, 65, 367-389.

Azzalini, A. and Genton, M. G. (2008) "Robust likelihood methods based on the skew-t and related distributions," *International Statistical Review*, 76, 106-129.

Brown, B. G., Katz, R. W., and Murphy, A. H. (1984) "Time series models to simulate and forecast wind speed and wind power," *Journal of Climate and Applied Meteorology*, 23, 1184-1195.

de Luna, X., and Genton, M. G. (2005), "Predictive spatio-temporal models for spatially sparse environmental data," *Statistica Sinica*, 15, 547-568.

Diebold, F. X. and Mariano, R. S. (1995) "Comparing predictive accuracy," *Journal of Business and Economics Statistics*, 13, 253-263.

Fisher, N. I. (1993) *Statistical Analysis of Circular Data*, Cambridge University Press: Cambridge.

Genton, M. G. and Hering, A. S. (2007) "Blowing in the wind," *Significance*, 4, 11-14.

Giebel, G., Brownsword, R., and Kariniotakis, G. (2003) "The state-of-the-art in short-term prediction of wind power; a literature overview," Technical report, Project ANEMOS public document, *http://anemos.cma.fr/download/ANEMOS_D1.1_StateOfTheArt_v1.1.pdf*

Gneiting, T. (2008) "Quantiles as optimal point predictors," *University of Washington, Technical Report*, no. 538.

Gneiting, T., Larson, K., Westrick, K., Genton, M. G., and Aldrich, E. (2006) "Calibrated probabilistic forecasting at the Stateline wind energy center: The regime-switching space-time method," *Journal of the American Statistical Association*, 101, 968-979.

Gneiting, T. and Raftery, A. E. (2007) "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, 102, 359-378.

Grimit, E. P., Gneiting, T., Berrocal, V. J., and Johnson, N. A. (2006) "The continuous ranked probability score for circular variables and its application to mesoscale forecast ensemble verification," *Quarterly Journal of the Royal Meteorological Society*, 132, 2925-2942.

Kestens, E. and Teugels, J. L. (2002) "Challenges in modelling stochasticity in wind," *Environmetrics*, 13, 821-830.

Klink, K. (1999) "Climatological mean and interannual variance of United States surface wind

speed, direction, and velocity," *International Journal of Climatology*, 19, 471-488.

Lange, M. (2005) "On the uncertainty of wind power predictions—Analysis of the forecast accuracy and statistical distribution of errors," *Journal of Solar Energy Engineering*, 127, 177-184.

Lange, M. and Focken, U. (2005) *Physical Approach to Short-Term Wind Power Prediction*, Springer Verlag: Berlin.

Lund, U. and Agostinelli, C. (2006) `circular`: Circular Statistics, `R` package version 0.3-6.

Madsen, H., Pinson, P., Kariniotakis, G., Nielsen, H. Aa., and Nielson, T.S. (2005) "Standardizing the performance evaluation of short-term wind power prediction models," *Wind Engineering*, 29, 475-489.

Mardia, K. V. and Jupp, P. E. (2000) *Directional Statistics*, John Wiley and Sons: London.

Martin, M., Cremades, L. V., and Santabárbara, J. M. (1999) "Analysis and modelling of time series of surface wind speed and direction," *International Journal of Climatology*, 19, 197-209.

Nadaraya, E. A. (1964) "On estimating regression," *Theory of Probability and its Applications*, 9, 141-142.

Patton, A. J. and Timmermann, A. (2007) "Testing forecast optimality under unknown loss," *Journal of the American Statistical Association*, 102, 1172-1184.

Pinson, P., Chevallier, C. and Kariniotakis, G. N. (2007) "Trading wind generation from short-term probabilistic forecasts of wind power," *IEEE Transactions on Power Systems*, 22, 1148-1156.

Potter, C. W., Gil, H. A., and McCaa, J. (2007) "Wind power data for grid integration studies," *Proceedings of the IEEE/PES General Meeting*, Tampa Bay, US, Paper Number: 07GM0808.

Schwarz, G. (1978) "Estimating the dimension of a model," *Annals of Statistics*, 6, 461-464.

Smith, J. C., Parsons, B., Acker, T., Milligan, M., Zavadil, R., Schuerger, M., and DeMeo, E. (2007) "Best practices in grid integration of variable wind power: Summary of recent US case study results and mitigation measures," Presented at European Wind Energy Conference 2007, Milan, Italy. May 2007.
http://www.wapa.gov/ugp/PowerMarketing/WindHydro/EWEC07paper.pdf.
Accessed Dec. 18, 2008.

Watson, G. S. (1964) "Smooth regression analysis," *Shankya Series A*, 26, 359-372.

Weisberg, S. (2005) *Applied Linear Regression*, Wiley-Interscience: New Jersey.

Wikle, C. K., Milliff, R. F., Nychka, D., and Berliner, L. M. (2001) "Spatiotemporal hierarchical Bayesian modeling: Tropical ocean surface winds," *Journal of the American Statistical Association*, 96, 382-397.

Table 1: This table contains correlations between the variables listed and the hourly wind speed two hours ahead at Vansycle. They are based on the 2002 training data and are used to build the TDD model. V, K, and G indicate the hourly wind speed at one of the three locations, and θ_V, θ_K, and θ_G represent the corresponding hourly wind direction for each location. Values in bold correspond to variables selected in the TDD model.

	Time Lag			
Variable	t	$t-1$	$t-2$	$t-3$
V	**0.90**	**0.85**	0.80	0.75
$\cos(\theta_V)$	**−0.55**	−0.53	−0.51	−0.48
$\sin(\theta_V)$	**−0.21**	−0.20	−0.18	−0.16
K	**0.74**	**0.72**	0.69	0.66
$\cos(\theta_K)$	**−0.63**	−0.63	−0.62	−0.61
$\sin(\theta_K)$	**−0.02**	−0.01	−0.00	0.01
G	**0.60**	0.60	0.58	0.56
$\cos(\theta_G)$	**−0.33**	−0.33	−0.34	−0.35
$\sin(\theta_G)$	**−0.45**	−0.43	−0.42	−0.41

Table 2: Root mean squared error (RMSE), mean absolute error (MAE), power curve error (PCE), and continuous ranked probability score (CRPS) for 2-hour point forecasts of hourly average wind speed at Vansycle in May through November 2003, in m/s. CRPS is not given for the persistence model. The "Overall" column gives the measure over all forecasts from May through November.

Measure	Forecast	May	Jun	Jul	Aug	Sep	Oct	Nov	Overall
	Persistence	2.14	1.97	2.37	2.27	2.17	2.38	2.11	2.21
RMSE	RSTD	1.73	**1.56**	1.69	**1.78**	1.77	2.07	1.87	1.79
	TDD	1.74	**1.56**	1.68	**1.78**	**1.75**	**2.03**	**1.86**	**1.78**
	BST	**1.69**	1.59	**1.64**	1.81	1.85	2.09	2.00	1.82
	Persistence	1.60	1.45	1.74	1.68	1.59	1.68	1.51	1.61
MAE	RSTD	1.31	1.19	1.32	**1.31**	1.36	**1.48**	**1.38**	**1.34**
	TDD	1.34	**1.18**	1.31	1.33	**1.33**	**1.48**	**1.38**	**1.34**
	BST	**1.26**	1.19	**1.27**	1.37	1.42	1.51	1.50	1.36
	Persistence	99.33	72.85	114.59	94.33	75.48	92.19	59.22	87.10
PCE	RSTD	69.45	**48.19**	73.21	63.39	56.31	71.62	48.89	61.73
	TDD	70.17	48.42	**72.70**	**63.14**	**56.13**	**70.24**	**47.13**	**61.28**
	BST	**67.51**	50.46	73.42	66.90	61.57	73.83	50.98	63.65
	RSTD	0.95	**0.85**	0.94	**0.95**	0.97	1.08	**1.00**	**0.96**
CRPS	TDD	0.97	**0.85**	0.93	0.96	**0.95**	**1.07**	**1.00**	**0.96**
	BST	**0.92**	0.86	**0.91**	0.98	1.01	1.10	1.08	0.98

The model with the lowest of each measure in each column is in bold.

Table 3: RSTD model outcomes when easterly/westerly and northerly/southerly regimes are defined by the wind direction at Kennewick. The original RSTD (with the regimes determined by the direction at Goodnoe Hills), the TDD, and the BST model results are also given.

Measure	Forecast	May	Jun	Jul	Aug	Sep	Oct	Nov	Overall
	RSTD-KW-EW	1.77	**1.56**	1.75	1.83	1.79	2.07	1.89	1.82
	RSTD-KW-NS	1.75	**1.56**	1.69	**1.77**	**1.74**	2.04	1.88	**1.78**
RMSE	RSTD-GH-EW	1.73	**1.56**	1.69	1.78	1.77	2.07	1.87	1.79
	TDD	1.74	**1.56**	1.68	1.78	1.75	**2.03**	**1.86**	**1.78**
	BST	**1.69**	1.59	**1.64**	1.81	1.85	2.09	2.00	1.82
	RSTD-KW-EW	1.36	1.19	1.36	1.37	1.37	1.52	1.42	1.37
	RSTD-KW-NS	1.34	**1.18**	1.32	1.33	1.34	1.50	**1.38**	**1.34**
MAE	RSTD-GH-EW	1.31	1.19	1.32	**1.31**	1.36	**1.48**	**1.38**	**1.34**
	TDD	1.34	**1.18**	1.31	1.33	**1.33**	**1.48**	**1.38**	**1.34**
	BST	**1.26**	1.19	**1.27**	1.37	1.42	1.51	1.50	1.36
	RSTD-KW-EW	70.91	49.27	76.82	65.05	57.21	71.90	48.59	62.98
	RSTD-KW-NS	70.73	**47.30**	73.46	64.04	**54.76**	**68.90**	49.19	61.35
PCE	RSTD-GH-EW	69.45	48.19	73.21	63.39	56.31	71.62	48.89	61.73
	TDD	70.17	48.42	**72.70**	**63.14**	56.13	70.24	**47.13**	**61.28**
	BST	**67.51**	50.46	73.42	66.90	61.57	73.83	50.98	63.65

The model with the lowest of each measure in each column is in bold.

Table 4: RSTD, TDD, and BST model outcomes for predictions made at Kennewick.

Kennewick	Forecast	May	Jun	Jul	Aug	Sep	Oct	Nov	Overall
	RSTD	2.34	1.96	2.09	2.17	2.13	2.36	2.34	2.21
RMSE	TDD	**2.32**	**1.94**	**2.08**	**2.15**	2.11	2.36	2.30	**2.19**
	BST	2.37	2.03	2.18	2.23	**2.05**	**2.28**	**2.23**	2.20
	RSTD	1.82	1.44	1.60	**1.58**	1.60	1.77	1.66	1.64
MAE	TDD	**1.79**	**1.43**	**1.59**	1.60	1.59	1.76	1.63	1.63
	BST	1.80	1.45	1.64	1.61	**1.51**	**1.72**	**1.54**	**1.61**
	RSTD	87.45	65.18	**82.96**	83.78	67.53	74.52	78.60	77.24
PCE	TDD	**85.63**	**64.93**	83.81	**83.19**	66.91	**71.31**	80.51	**76.69**
	BST	92.35	70.32	84.49	86.58	**66.47**	72.34	**73.91**	78.18

The model with the lowest of each measure in each column is in bold.

Table 5: RSTD, TDD, and BST model outcomes for predictions made at Goodnoe Hills.

Goodnoe Hills	Forecast	May	Jun	Jul	Aug	Sep	Oct	Nov	Overall
	RSTD	**1.69**	**1.51**	**1.38**	**1.55**	**1.68**	**1.87**	1.75	**1.64**
RMSE	TDD	**1.69**	1.55	1.40	**1.55**	**1.68**	**1.87**	**1.73**	1.65
	BST	1.76	1.64	1.43	1.56	1.70	1.98	1.78	1.70
	RSTD	**1.31**	**1.16**	**1.06**	**1.18**	**1.25**	**1.37**	1.31	**1.23**
MAE	TDD	**1.31**	1.19	1.08	1.20	1.26	**1.37**	**1.28**	1.24
	BST	1.38	1.29	1.09	1.19	1.27	1.45	1.34	1.28
	RSTD	**81.69**	**61.18**	**67.36**	68.96	63.66	70.83	**56.78**	**67.30**
PCE	TDD	82.54	64.33	68.52	69.17	64.67	**69.31**	56.90	68.01
	BST	86.46	67.47	68.99	**68.30**	**63.24**	76.11	61.13	70.33

The model with the lowest of each measure in each column is in bold.

Table 6: RSTD, TDD, and BST model outcomes for the two types of models built on the ten-minute data.

All Ten-Min	Forecast	May	Jun	Jul	Aug	Sep	Oct	Nov	Overall
	RSTD	1.95	1.77	1.90	1.99	1.96	2.23	**2.13**	2.00
RMSE	TDD	1.90	**1.72**	1.84	**1.98**	**1.93**	**2.22**	**2.13**	**1.97**
	BST	**1.85**	1.73	**1.76**	2.00	2.02	2.32	2.35	2.02
	RSTD	1.48	1.37	1.50	1.52	1.50	**1.62**	**1.57**	1.51
MAE	TDD	1.45	1.32	1.44	**1.49**	**1.47**	1.63	1.60	**1.49**
	BST	**1.39**	**1.31**	**1.36**	1.51	1.55	1.68	1.77	1.51
	RSTD	79.42	56.21	83.57	**69.30**	60.85	75.39	52.38	68.32
PCE	TDD	78.10	**54.86**	79.34	69.64	**59.53**	**75.26**	**51.60**	**67.07**
	BST	**74.16**	55.61	**75.74**	74.47	65.86	79.73	63.07	69.92

Hourly Ten-Min	Forecast	May	Jun	Jul	Aug	Sep	Oct	Nov	Overall
	RSTD	1.92	1.77	1.90	1.99	1.96	2.22	2.14	1.99
RMSE	TDD	1.90	**1.73**	1.84	1.98	**1.93**	**2.21**	**2.13**	**1.97**
	BST	**1.86**	1.74	**1.76**	**1.97**	2.01	2.27	2.30	2.00
	RSTD	1.46	1.37	1.49	1.51	1.50	**1.61**	**1.59**	1.51
MAE	TDD	1.44	1.33	1.44	**1.48**	**1.47**	1.62	1.61	**1.48**
	BST	**1.39**	**1.32**	**1.36**	**1.48**	1.53	1.64	1.74	1.49
	RSTD	78.80	56.24	81.84	70.37	61.11	75.45	**52.39**	68.19
PCE	TDD	77.65	**54.83**	78.50	**70.31**	**60.07**	**74.05**	53.11	**67.08**
	BST	**75.89**	55.23	**74.50**	72.56	65.07	78.49	59.50	68.87

The model with the lowest of each measure in each column is in bold.

Table 7: RSTD, TDD, and BST model PCE results for varying penalties on underestimation versus overestimation. A value of γ less (more) than 0.50 penalizes overestimates more (less) heavily than underestimation.

γ	Forecast	May	Jun	Jul	Aug	Sep	Oct	Nov	Overall
	RSTD	4.44	**5.05**	4.20	**3.92**	3.69	11.06	**6.02**	5.49
0.01	TDD	**4.40**	5.13	**4.15**	4.04	3.86	**10.59**	6.14	**5.48**
	BST	4.76	5.72	4.47	4.15	**3.47**	13.02	6.72	6.05
	RSTD	**32.80**	**25.16**	33.55	27.55	25.03	42.10	26.10	30.40
0.10	TDD	33.07	25.49	32.94	27.67	25.17	41.24	**25.85**	30.27
	BST	33.28	26.88	**32.48**	**26.45**	**23.99**	**40.31**	26.90	**30.10**
	RSTD	77.28	**57.24**	84.96	69.28	63.21	81.65	55.23	70.00
0.50	TDD	78.60	57.48	83.77	**68.54**	**62.69**	80.99	**53.00**	**69.46**
	BST	**75.24**	59.56	**82.60**	70.58	65.29	**80.64**	57.64	70.35
	RSTD	**41.36**	**26.61**	40.55	36.38	33.90	48.24	35.37	37.56
0.90	TDD	42.12	27.14	**40.29**	**36.16**	**33.26**	**45.87**	**31.26**	**36.67**
	BST	43.83	30.84	41.16	40.80	38.52	50.10	35.19	40.14
	RSTD	7.76	4.65	**5.95**	7.79	**7.01**	29.28	42.14	14.90
0.99	TDD	**7.65**	**4.61**	6.44	**6.97**	7.48	26.63	**41.56**	**14.43**
	BST	8.89	6.82	5.97	10.85	8.17	**23.01**	43.39	15.24

The model with the lowest PCE for each value of γ in each column is in bold.

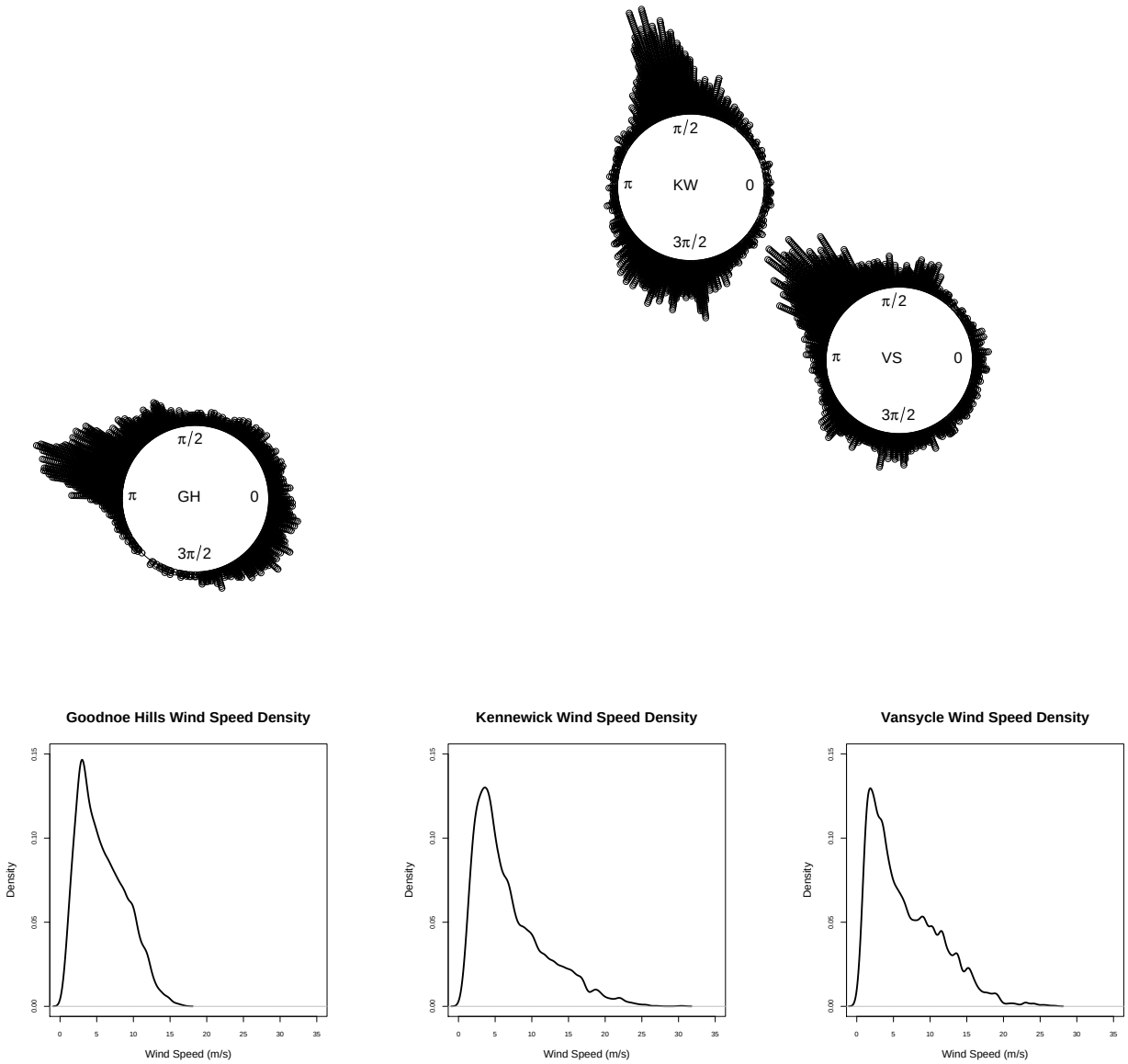

Figure 1: GH, KW, and VS denote Goodnoe Hills, Kennewick, and Vansycle, respectively. The locations of each circle indicate the relative location of each tower to the others. Each point on the circular histograms at the top represents a wind direction from the 2002 training data. For example, at Vansycle the majority of the wind directions blow from the northwest towards the southeast. The bottom panels are nonparametric density estimates of the 2002 wind speed data.

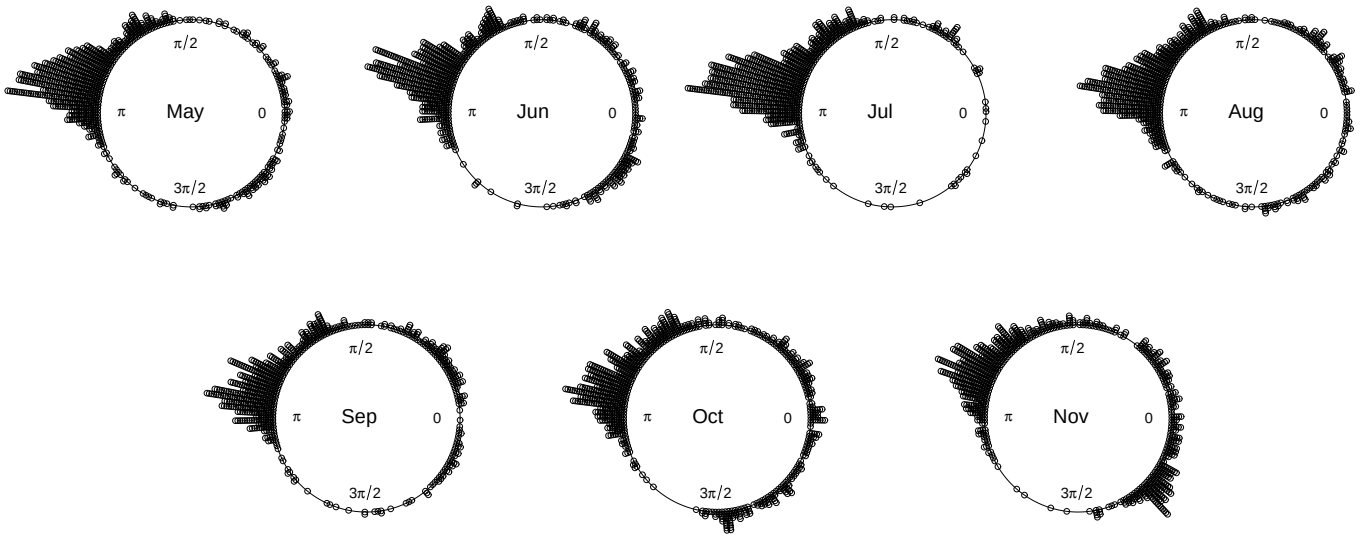

Figure 2: Circular histograms of wind directions at Goodnoe Hills for each month when predictions are made in the year 2003. Easterly winds are defined as those on the right-hand side of the circle between $3\pi/2$ and $\pi/2$. Westerly winds are on the left-hand side.

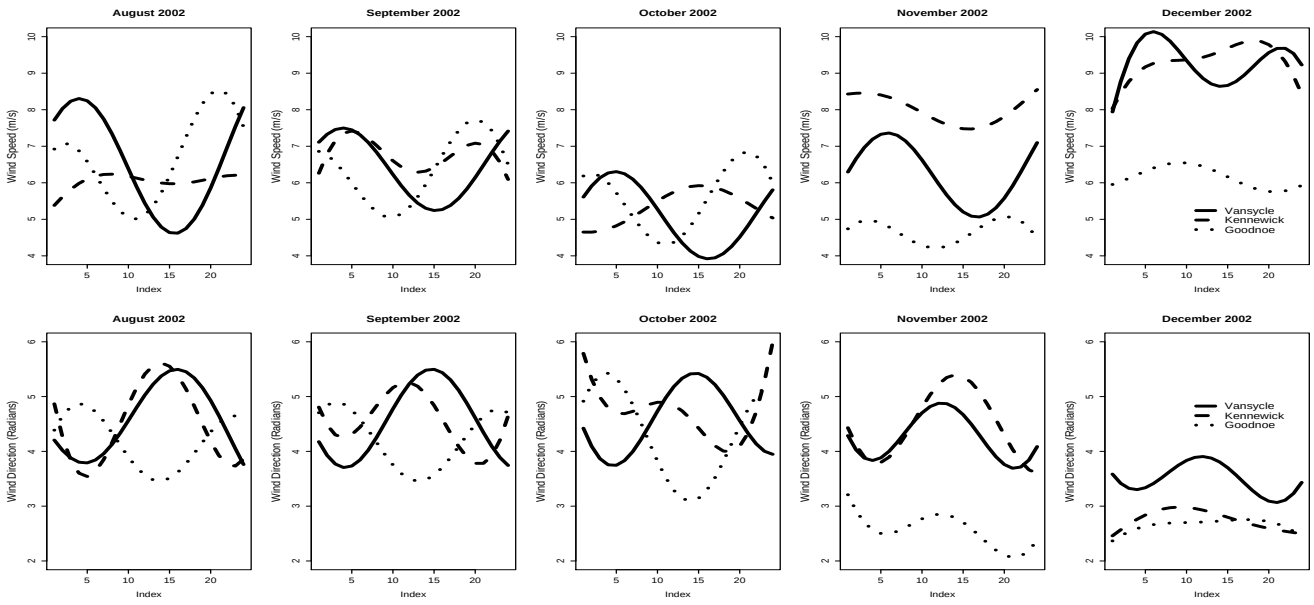

Figure 3: The top panels plot the fitted diurnal model of wind speed at each hour of the day at all 3 sites. It is clearly diurnal in nature for every month. The bottom panels plot the fitted diurnal model of wind direction at each hour of the day at all 3 sites. The diurnal nature of the directions is strong for every month except December which had 10 days of missing data.

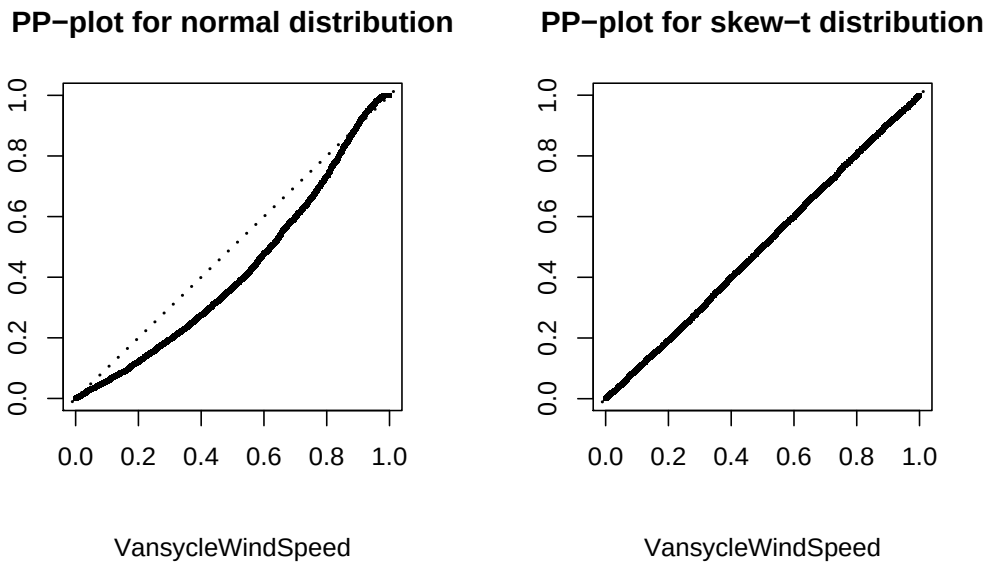

Figure 4: Comparison of the BST with normal errors (left plot) and with skew-t errors (right plot) on the 2002 training data.

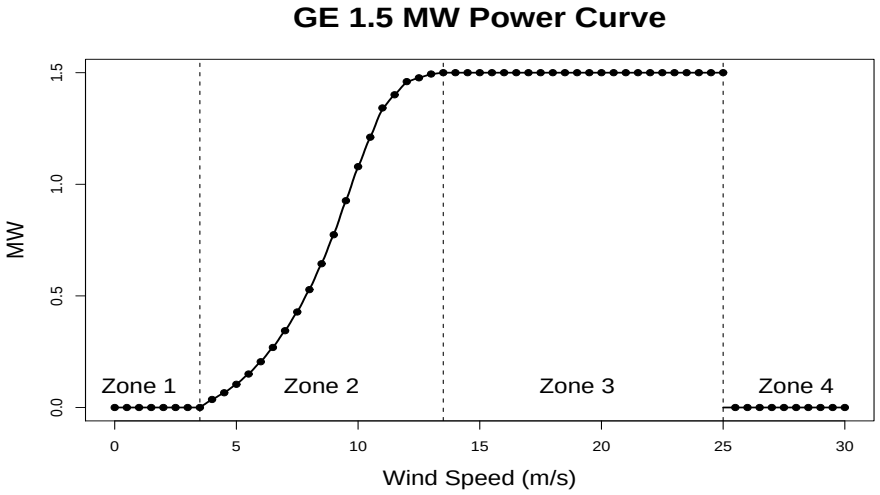

Figure 5: The GE 1.5 MW power curve. The black dots are the manufacturer's data. The solid curve in Zone 2 is a nonparametric fit to those data. It has a cut-in speed of 3.5 m/s, a rated speed of 13.5 m/s, and a cut-out speed of 25 m/s. These values change from one type of turbine to another.

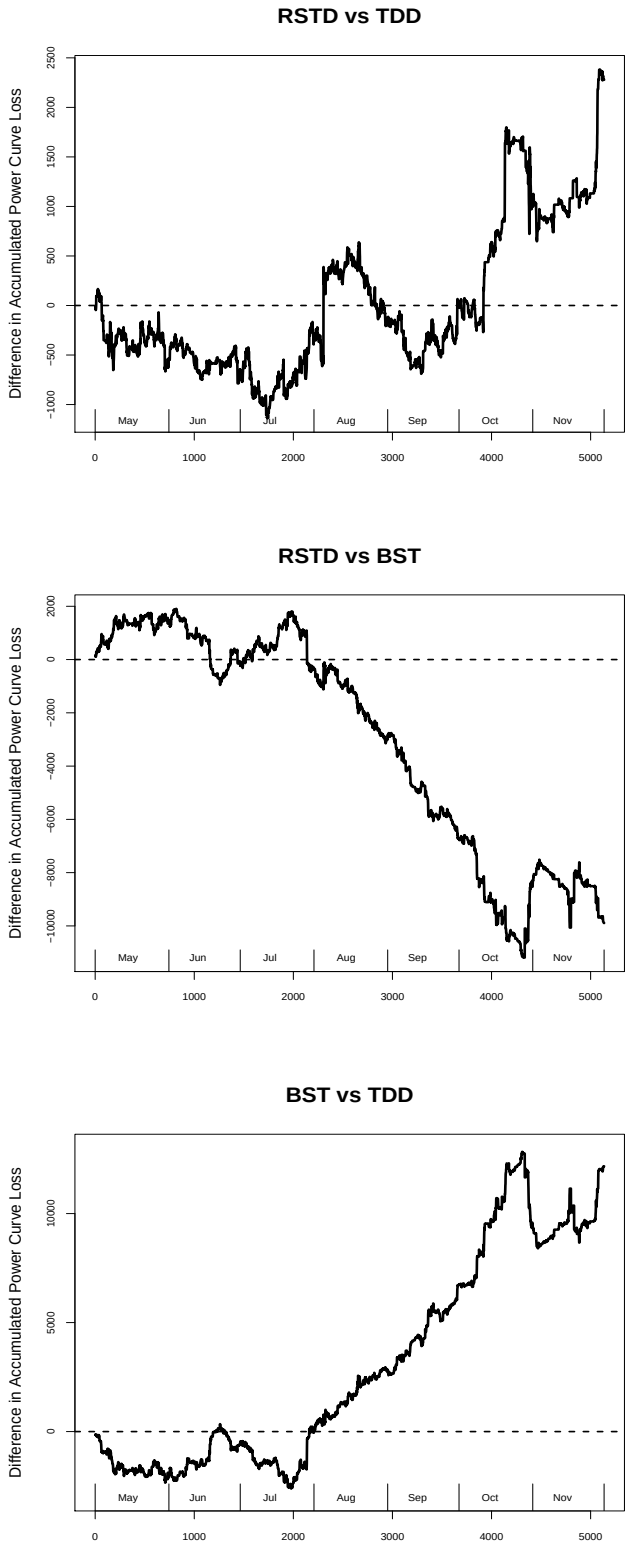

Figure 6: These graphs plot the difference in accumulated PCE (in kW) penalties between the RSTD and TDD models (top), the RSTD and BST models (middle), and the BST and TDD models (bottom). An upward (downward) trend means that the second model is performing better (worse).

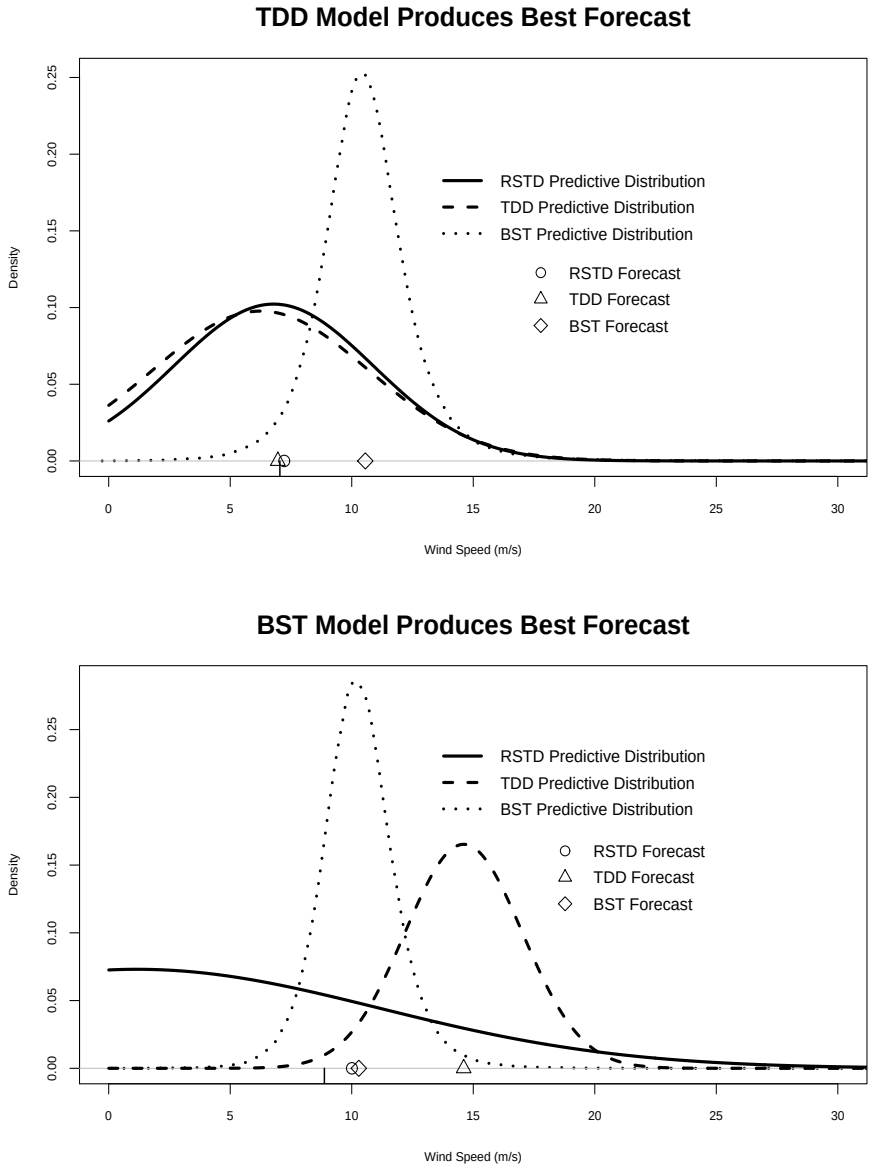

Figure 7: Comparing the predictive distributions for the models when the TDD model produces the best forecast (top panel) and when the BST model produces the best forecast (bottom panel). The small vertical line on the x-axis of each plot represents the observed wind speed.