

# Calibrated Probabilistic Forecasting at the Stateline Wind Energy Center: The Regime-Switching Space-Time (RST) Method

Tilman Gneiting, Kristin Larson, Kenneth Westrick,  
Marc G. Genton and Eric Aldrich\*

Technical Report no. 464  
Department of Statistics, University of Washington

September 2004

## Abstract

With the global proliferation of wind power, accurate short-term forecasts of wind resources at wind energy sites are becoming paramount. Regime-switching space-time (RST) models merge meteorological and statistical expertise to obtain accurate and calibrated, fully probabilistic forecasts of wind speed and wind power. The model formulation is parsimonious, yet takes account of all the salient features of wind speed: alternating atmospheric regimes, temporal and spatial correlation, diurnal and seasonal non-stationarity, conditional heteroscedasticity, and non-Gaussianity. The RST method identifies forecast regimes at the wind energy site and fits a conditional predictive model for each regime. Geographically dispersed meteorological observations in the vicinity of the wind farm are used as off-site predictors.

The RST technique was applied to 2-hour ahead forecasts of hourly average wind speed at the Stateline wind farm in the US Pacific Northwest. In July 2003, for instance, the RST forecasts had root-mean-square error (RMSE) 28.6% less than the persistence forecasts. For each month in the test period, the RST forecasts had lower RMSE than forecasts using state-of-the-art vector time series techniques. The RST method provides probabilistic forecasts in the form of predictive cumulative distribution functions, and those were well calibrated and sharp. The RST prediction intervals were substantially shorter on average than prediction intervals derived from univariate time series techniques. These results suggest that quality meteorological data from sites upwind of wind farms can be efficiently used to improve short-term forecasts of wind resources. It is anticipated that the RST technique can be successfully applied at wind energy sites all over the world.

**KEY WORDS:** Conditional heteroscedasticity; Continuous ranked probability score (CRPS); Minimum CRPS estimation; Predictive distribution; Spatio-temporal; Weather prediction

---

\*Tilman Gneiting is Associate Professor and Eric Aldrich is Ph.D. Student, both at the Department of Statistics, University of Washington, Box 354322, Seattle, WA 98195-4322. Kristin Larson is Senior Research Meteorologist and Kenneth Westrick is Founder and CEO, both at 3 Tier Environmental Forecast Group, Inc., 2825 Eastlake Avenue East, Suite 330, Seattle, WA 98102. Marc G. Genton is Associate Professor, Department of Statistics, Texas A&M University, College Station, TX 77843-3143. In Academic Year 2004/05, Tilman Gneiting is on sabbatical leave at the Soil Physics Group, Universität Bayreuth, Universitätsstr. 30, 95440 Bayreuth, Germany. The authors are grateful to the Washington Technology Center (WTC) for supporting this work through the WTC Research & Technology Development program. Tilman Gneiting also acknowledges support by the National Science Foundation under Award 0134264 and by the DoD Multidisciplinary University Research Initiative (MURI) program administered by the Office of Naval Research under Grant N00014-01-10745. 3 Tier acknowledges the support of PPM Energy, Inc. The data used in this work were obtained from Oregon State University Energy Resources Research Laboratory under the direction of Stel N. Walker. The authors thank Jeff Baars, Barbara G. Brown, Vito Giarrusso, Clifford F. Mass, Adrian E. Raftery, Pascal Storck, Werner Stuetzle and Toby White for helpful discussions.

# 1 Introduction

Wind power is the fastest-growing energy source today. Globally, wind energy has seen an annual average growth rate exceeding 30% during the past decade. Domestically, wind power is a plentiful energy source, particularly in the upper Midwest and in the mountainous Western United States. The rapid recent growth can be chiefly attributed to advances in wind turbine design – that have significantly reduced the cost of wind energy – as well as federally mandated tax credits. Estimates are that within the next 15 years wind energy will fill about 6% of the electricity supply in the United States. In Denmark, wind energy already meets 20% of the country’s total energy needs.<sup>†</sup> However, arguments against higher penetration rates of wind energy have been put forth, often focusing on the perceived inability to forecast wind resources with any degree of accuracy. The development of advanced forecast methodologies helps address these concerns. Increases in the accuracy of wind energy forecasts reduce the requirement for backup energy resulting in increased power grid reliability and monetary savings. One requirement is for the development of improved prediction techniques on the 0- to 3-hour forecast horizon, the typical lead time for transmission scheduling, resource allocation, and generation dispatch. Throughout the paper, the term short-range forecast will refer to the 0- to 3-hour forecast horizon.

Traditionally, short-range forecasts have utilized on-site observations with various persistence based statistical forecast models, including autoregressive time series techniques (Brown, Katz and Murphy 1984) and neural network methodologies (Kretzschmar, Eckert, Cattani and Eggimann 2004). Giebel (2003) surveys the literature on short-range wind speed and wind power forecasting. Forecasts based on numerical weather prediction models outperform statistical forecasts at larger lead times, but they are not competitive at the 2-hour horizon that we consider here. Our paper introduces the regime-switching space-time (RST) method that merges meteorological and statistical expertise to obtain accurate and calibrated, fully probabilistic short-term forecasts of wind energy. The model formulation is parsimonious, yet takes account of all the salient features of wind speed: alternating atmospheric regimes, temporal and spatial correlation, diurnal and seasonal non-stationarity, conditional heteroscedasticity, and non-Gaussianity. The RST method identifies atmospheric regimes at the wind energy site and fits a conditional predictive model for each regime or state. The approach is akin to the nonlinear gated experts technique of Weigend, Mangeas and Srivastava (1995); however, the model selection process is guided by expert knowledge of the local meteorological conditions. The RST method also relates to the space-time model for daily rainfall developed by Bardossy and Plate (1992) that uses multivariate time series models with parameters depending on the atmospheric circulation pattern, and to the non-homogeneous hidden Markov model of Hughes and Guttorp (1994) and Hughes et al. (1999), which postulates an unobserved, discrete-valued weather state.

Whenever appropriate and feasible, the RST method utilizes geographically dispersed meteorological observations in the vicinity of the wind farm. Since changes in wind often propagate with the wind, it is possible to use upwind observations to detect precursors to changes in wind speeds at the wind energy site. The RST technique introduces a new generation of forecast algorithms: it conditions on the forecast regime, makes use of off-site predictors, and provides fully probabilistic forecasts in the form of predictive cumulative distribution functions (CDFs). Alexiadis, Dokopoulos and Sahsamanoglou (1999) considered forecasts of wind speed and wind power at the Thessaloniki Bay, Greece and showed that the use of off-site predictors can improve forecast accuracy. However,

---

<sup>†</sup>Up-to-date information on the domestic and global status of wind energy is available at the American Wind Energy Association website, [www.awea.org](http://www.awea.org).

they did not build a physically interpretable model and did not consider probabilistic forecasts. Kretzschmar et al. (2004, p. 733) refuted the use of off-site observations for forecasts of wind speed in Switzerland, pointing out that upwind may refer to distinct geographic locations in dependence on the atmospheric regime. The RST approach conditions the predictive model on the atmospheric regime and thereby addresses these concerns.

The remainder of the paper is organized as follows. Section 2 gives a concise introduction to probabilistic forecasts, as opposed to deterministic-style or point forecasts. We describe diagnostic tools and scoring rules for the assessment of probabilistic forecasts, and we propose the use of cut-off normal predictive distributions for nonnegative predictands. Section 3 introduces the data used in our case study. We describe hourly time series of wind speed and wind direction at the meteorological towers at Vansycle, Kennewick and Goodnoe Hills in the US Pacific Northwest. The Vansycle tower is adjacent to the Stateline wind energy center at the eastern end of the Columbia River Gorge, right on the border between the states of Oregon and Washington. The Stateline project as-is has 300 MW of wind power capacity although there have been talks of expansion. It is currently the world’s largest single wind energy project. Section 4 describes the various versions of the RST model for 2-hour ahead forecasts of hourly average wind speed at Vansycle. The observations from the towers at Kennewick and Goodnoe Hills are used as off-site predictors. We consider forecasts of wind speed rather than wind power, both because the wind power data are proprietary and because wind speed is of more fundamental scientific interest. Forecasts of wind speed can be transformed into forecasts of wind power using the methods described by Brown et al. (1984). The 2-hour forecast horizon is the shortest that allows for real-time forecasts at the hourly aggregation level.

In Section 5 we assess the predictive performance of the RST method, as compared to more conventional forecast techniques that we use as benchmarks, including the persistence forecast, the reference forecast proposed by Nielsen et al. (1998), forecasts based on univariate time series techniques, and forecasts using state-of-the-art vector time series methods. In July 2003, for instance, the RST forecasts had root-mean-square error (RMSE) 28.6% less than the persistence forecasts. The forecasts based on univariate time series techniques utilized on-site information only and reduced the persistence RMSE by up to 16.0%. In November 2003, the RST and the univariate time series forecasts had RMSE 11.3% and 1.9% lower than the persistence forecasts, respectively. For each month in the test period, the RST forecasts had lower RMSE than the forecasts using vector time series techniques. The RST method yields probabilistic forecasts in the form of predictive distributions, and the predictive CDFs were well calibrated and sharp. The RST prediction intervals were substantially shorter on average than prediction intervals derived from univariate time series techniques. The paper closes with a discussion in Section 6.

## 2 Probabilistic forecasts

### 2.1 The case for probabilistic forecasting

Non-probabilistic or deterministic-style forecasts provide a single numerical value for the predictand, say hourly average wind speed at the wind energy site two hours ahead. Deterministic-style forecasts provide the classical predictive framework and they can be assessed using well-known criteria such as the root-mean-square error (RMSE) and the mean absolute error (MAE). However, deterministic-style forecasts have limited use unless they are accompanied by suitable measures of

the uncertainty associated with them. Consequently, Dawid (1984) argues that forecasts should be probabilistic in nature, taking the form of probability distributions over future events. Indeed, over the past two decades probabilistic forecasts have become routine in applications such as weather prediction (Palmer 2002; Gel, Raftery and Gneiting 2004) and macroeconomic forecasting (Garraat, Lee, Pesaran and Shin 2003). A probabilistic forecast for a real-valued quantity, such as wind speed or wind power, takes the form of a predictive cumulative distribution function (CDF), or simply predictive distribution, say  $F$ , where

$$F(x) = \text{prob}(X \leq x), \quad x \in \mathbb{R},$$

denotes the forecaster’s belief that the predictand will not exceed the threshold  $x$ . Interval forecasts form a special case of probabilistic forecasts, and the construction of prediction intervals from the predictive CDF is straightforward. For instance,  $F(0.05)$  and  $F(0.95)$  form the lower and upper end points of the 90% central prediction interval, respectively. Brown et al. (1984) pioneered the development of wind energy forecasts and applied time series methods to obtain interval forecasts for wind speed and wind power. More recently, there has been a surge of interest in probabilistic forecasts for wind resources. Roulston, Kaplan, Hardenberg and Smith (2003) and Bremnes (2004) studied probabilistic forecasts of wind power at the medium-range, that is, at lead times between one and ten days. Pinson and Kariniotakis (2004) proposed a resampling approach that attaches prediction intervals to deterministic-style forecasts of wind resources.

## 2.2 Assessing probabilistic forecasts

In assessing the quality of probabilistic forecasts, we are guided by the paradigm that probabilistic forecasts strive to maximize the sharpness of the predictive distributions under the constraint of calibration (Gneiting, Raftery, Balabdaoui and Westveld 2003). Calibration refers to the statistical consistency between the predictive distributions and the observations, and is a joint property of the forecasts and the values or events that materialize. Sharpness refers to the spread of the predictive distributions, and is a property of the forecasts only. The more concentrated the predictive distribution, the sharper the forecast, and the sharper the better, subject to calibration.

First, we describe the probability integral transform (PIT) histogram. The probability integral transform is defined as the value that the predictive cumulative distribution function  $F$  attains at the observation  $x$ , that is, a number between 0 and 1. Rosenblatt (1952) studied the probability integral transform and Dawid (1984) proposed its use in assessing calibration. Diebold, Gunther and Tay (1998) introduced the PIT histogram – that is, the histogram of the PIT values – as a diagnostic tool in the evaluation of probabilistic forecasts. Approximately uniform PIT histograms indicate calibration and correspond to prediction intervals that have close to nominal coverage at all levels. For a PIT histogram with 20 equally spaced bins, for instance, the coverage of the central 90% prediction interval corresponds to the area under the 18 central bins and can be read off the histogram. To assess sharpness, we report the average width of the 90% central prediction interval; the shorter, the sharper.

Scoring rules assign numerical scores to forecasts, based on the predictive distribution  $F$  and the value  $x$  that materializes. We take scores to be penalties that a forecaster wishes to minimize. Let  $\mathbf{1}(y \geq x)$  denote the function that attains the value 1 when  $y \geq x$  and the value 0 otherwise. If  $F$  is the predictive CDF and  $x$  materializes, the continuous ranked probability score is defined as

$$\text{crps}(F, x) = \int_{-\infty}^{\infty} (F(y) - \mathbf{1}(y \geq x))^2 dy, \quad (1)$$

which equals the integral of the Brier scores for binary probabilistic forecasts at all real-valued thresholds (Hersbach 2000). Gneiting and Raftery (2004) used a result of Székely (2003) to show that

$$\text{crps}(F, x) = E_F |X - x| - \frac{1}{2} E_F |X - X'|,$$

where  $X$  and  $X'$  are independent copies of a random variable with distribution function  $F$  and finite first moment. Hence, the continuous ranked probability score generalizes the absolute error, to which it reduces if  $F$  is a deterministic-style or point forecast, and can be reported in the same unit as the observations. For a normal predictive distribution,  $\mathcal{N}(\mu, \sigma^2)$ , with mean  $\mu$  and standard deviation  $\sigma$  it is straightforward to show that

$$\text{crps}(\mathcal{N}(\mu, \sigma^2), x) = \sigma \left( \frac{x - \mu}{\sigma} \left( 2 \Phi \left( \frac{x - \mu}{\sigma} \right) - 1 \right) + 2 \phi \left( \frac{x - \mu}{\sigma} \right) - \frac{1}{\sqrt{\pi}} \right),$$

where  $\phi$  and  $\Phi$  denote the probability density function (PDF) and the CDF of the standard normal distribution, respectively. The continuous ranked probability score is proper in the sense that the forecaster maximizes the expected score for an observation drawn from  $F$  if she issues the probabilistic forecast  $F$ , rather than  $G \neq F$ . The logarithmic score (Good 1952; Bernardo 1979; Roulston and Smith 2002) is a proper scoring rule, too, but it is highly sensitive to outliers and difficult to interpret for mixed discrete-continuous predictive distributions. Therefore, we assess probabilistic forecasts by comparing the average value,

$$\text{CRPS} = \frac{1}{n} \sum_{i=1}^n \text{crps}(F_i, x_i), \tag{2}$$

of the continuous ranked probability score over the test set. If each  $F_i$  is a deterministic-style forecast, the CRPS value reduces to the mean absolute error (MAE), and the CRPS and the MAE can be directly compared. Clearly, the smaller the CRPS, the better.

### 2.3 Cut-off normal predictive distributions

Predictive distributions are frequently taken to be Gaussian even though the predictand is a non-negative quantity, such as wind speed, precipitation or pollutant concentrations, or a nonnegative transformation thereof. Brown et al. (1984), Haslett and Raftery (1989) and Carroll et al. (1997), for instance, used normal distributions to model the square root of wind speed and atmospheric ozone concentration, respectively. To address the nonnegativity of the predictand, we replace the normal predictive distribution,  $\mathcal{N}(\mu, \sigma^2)$ , by the cut-off normal predictive distribution,  $\mathcal{N}^0(\mu, \sigma^2)$ , with cumulative distribution function

$$F^0(x) = \begin{cases} 0 & \text{if } x < 0, \\ \Phi \left( \frac{x - \mu}{\sigma} \right) & \text{if } x \geq 0. \end{cases}$$

The normal distribution and the cut-off normal distribution assign the same probability mass to any Borel subset of the positive half-axis. However, the cut-off normal distribution is concentrated on the nonnegative half-axis and assigns point mass  $\Phi(-\mu/\sigma)$  to zero. The cut-off normal distribution,  $\mathcal{N}^0(\mu, \sigma^2)$ , with parameters  $\mu$  and  $\sigma > 0$  has median

$$\mu^+ = \max(\mu, 0) \tag{3}$$

and mean

$$\mu^0 = \mu \Phi\left(\frac{\mu}{\sigma}\right) + \sigma \phi\left(\frac{\mu}{\sigma}\right), \quad (4)$$

where, again,  $\phi$  and  $\Phi$  denote the PDF and the CDF of the standard normal distribution, respectively. The continuous ranked probability score (1) for the cut-off normal distribution  $\mathcal{N}^0(\mu, \sigma^2)$  and an observation  $x \geq 0$  is

$$\begin{aligned} \text{crps}(\mathcal{N}^0(\mu, \sigma^2), x) &= \text{crps}(\mathcal{N}(\mu, \sigma^2), x) \\ &\quad - 2\sigma \phi\left(\frac{\mu}{\sigma}\right) \Phi\left(-\frac{\mu}{\sigma}\right) + \frac{\sigma}{\sqrt{\pi}} \Phi\left(-\sqrt{2}\frac{\mu}{\sigma}\right) + \mu \left(\Phi\left(-\frac{\mu}{\sigma}\right)\right)^2. \end{aligned} \quad (5)$$

Tobin (1958) proposed the use of cut-off normal predictive distributions, and the associated tobit model has been widely used in econometric applications since. Cut-off normal distributions can also be interpreted in terms of latent Gaussian variables (Sansó and Guenni 2000; Allcroft and Glasbey 2003).

### 3 Data

We introduce and describe the data on which our case study is based.

#### 3.1 The meteorological towers at Vansycle, Kennewick and Goodnoe Hills

The meteorological data used hereinafter were collected by Oregon State University for the United States government, represented by the Department of Energy and the Bonneville Power Administration.<sup>‡</sup> We obtained time series of wind speed and wind direction at three meteorological towers in the US Pacific Northwest: Vansycle in northeastern Oregon, in the immediate vicinity of the Stateline wind energy center, and Kennewick and Goodnoe Hills in southern Washington. Data from all three stations simultaneously have been available since August 2002. Figure 1 shows the locations of the three sites along the Columbia River Gorge and the Oregon–Washington border. Goodnoe Hills lies 146 km west of Vansycle, and the meteorological tower at Kennewick is situated 39 km northwest of the Vansycle tower. Table 1 provides further information about the sites. A typical view at the Stateline wind farm is shown in Larson and Gneiting (2004).

The raw data record at the three sites is largely but not entirely complete. We imputed a minimal amount of isolated missing data (less than 0.03% for calendar year 2003) and adopted the quality assurance procedures proposed by Shafer et al. (2000) and Fiebrich and Crawford (2001). Specifically, we applied range, step and persistence tests to the raw data at the 10-minute aggregation level. The persistence test checks whether a parameter undergoes little or no variation, and we performed manual checks for plausibility if three or more subsequent observations of 10-minute average wind speed or wind direction were identical. This test detects anemometer readings that are stuck near zero, typically as a result of freezing rain which is a common hazard in the region. During the cool season, we had to discard various stretches of data that did not pass the persistence test. We aggregated the remaining 10-minute averages of wind speed to obtain hourly averages of wind speed at Vansycle, Kennewick and Goodnoe Hill, and we created hourly series of wind direction at the three sites. For the wind direction series we did not aggregate; we rather used

---

<sup>‡</sup>The data are available online at [me.oregonstate.edu/ERRL/data](http://me.oregonstate.edu/ERRL/data).

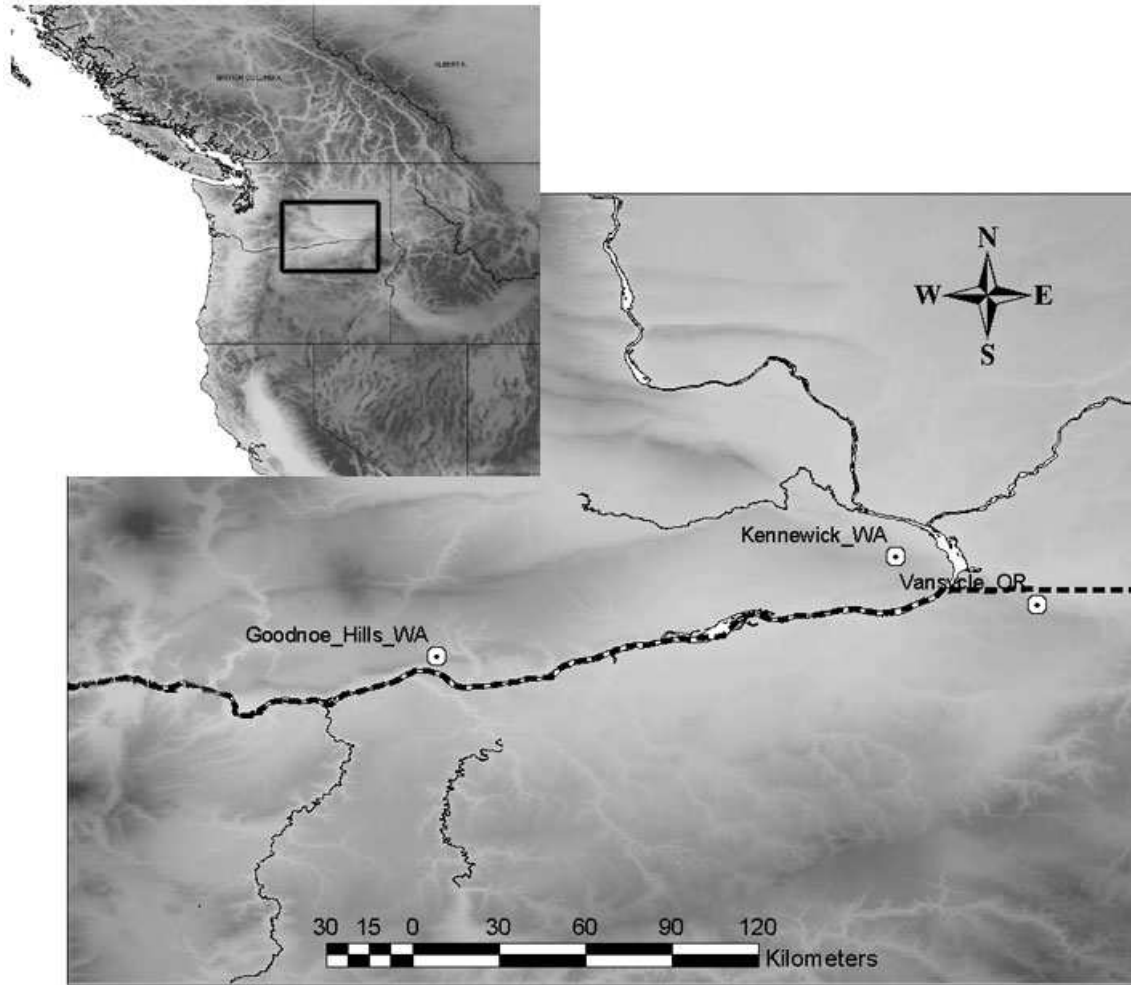


Figure 1: Locations of the meteorological towers at Vansycle, Kennewick and Goodnoe Hills along the Columbia Gorge and the Oregon–Washington border in the US Pacific Northwest. The topography is indicated by different shading: light shading indicates low altitude, and dark shading indicates high altitude.

the value for 9:50am–10:00am to represent the 9:00am–10:00am interval, say. All work reported below is based on the hourly series of wind speed and wind direction. The longest continuous records at all three sites jointly comprise 55 and 279 days and range from 4 September 2002 to 28 October 2002, and from 25 February 2003 to 30 November 2003, respectively. This record spans the range of weather experienced in the region, from strongly synoptically-forced weather systems, prevalent during the winter, to thermally-driven flows – primarily a summer season phenomenon.

### 3.2 Patterns of wind speed

We now describe patterns of wind speed at Vansycle, Kennewick and Goodnoe Hills. Table 2 shows the monthly average wind speed at the three sites in calendar year 2003. Any seasonal effects are weak, although during the summer season there is a much stronger diurnal cycle observed in the wind record. Brown et al. (1984) modeled wind speed at Goodnoe Hills as a square root Gaussian

Table 1: Site information for the meteorological towers at Vansycle, Kennewick and Goodnoe Hills. Latitude and longitude are given in degrees and minutes; elevation and anemometer height are given in meters.

Site	Latitude	Longitude	Elevation	Anemometer Height	Data Record Starting
Vansycle	45° 18' N	118° 41' W	543 m	31 m	August 2002
Kennewick	46° 06' N	119° 08' W	671 m	26 m	June 1976
Goodnoe Hills	45° 48' N	120° 34' W	774 m	59 m	May 1980

Table 2: Monthly average wind speed at Vansycle, Kennewick and Goodnoe Hills in calendar year 2003, in  $\text{m}\cdot\text{s}^{-1}$ .

Site	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Vansycle	6.65	6.74	9.41	7.51	7.10	7.13	7.04	6.63	6.36	7.14	8.55	6.34
Kennewick	7.85	7.65	11.26	9.07	7.80	7.34	6.56	6.78	7.04	8.89	10.59	7.53
Goodnoe Hills	5.22	6.36	8.07	6.65	6.98	7.66	6.79	5.91	5.28	5.81	5.92	4.29

variable. This does not appear to be a defensible assumption at Vansycle. A more detailed analysis shows that the marginal distribution of wind speed at Vansycle is bimodal and can be approximated by a mixture of two square root Gaussian distributions. The mixture components can be physically interpreted, as described below. That said, the predictive distributions that we employ are cut-off Gaussian; this works well, because the conditional forecast errors are approximately normally distributed, as opposed to the wind speed values themselves.

Figure 2 shows the wind speed series at Vansycle, Kennewick and Goodnoe Hills for the period between 3 August 2002 and 9 August 2002. These were the first seven days with a complete record of observations at Vansycle. Temporal and spatial correlation as well as a diurnal trend component can be observed. Figure 3 shows the autocorrelation and cross-correlation functions of wind speed at the three sites, as observed in August–December 2002. The autocorrelation functions show a steady decline with the temporal lag, except for local maxima at lags that are multiples of 24. The local maxima correspond to the diurnal pattern that is more pronounced at Vansycle and Goodnoe Hills than at Kennewick. A noteworthy feature is the asymmetry of the cross-correlation function between Vansycle and Goodnoe Hills, which peaks at lags of two and three hours, and not at lag zero. The asymmetry can be interpreted in terms of the prevailing westerly wind. Goodnoe Hills lies 146 km west of Vansycle, and it therefore seems plausible that wind speeds at Goodnoe Hills tend to lead those at Vansycle. The wind speed series in Figure 2 suggest this, too. Gneiting (2002), de Luna and Genton (2003) and Stein (200x) described similar patterns in the Irish wind data of Haslett and Raftery (1989); and Wan, Milligan and Parsons (2003) reported similar asymmetries and similar lead times in the cross-correlation functions of wind power for wind plants in Minnesota and Iowa.

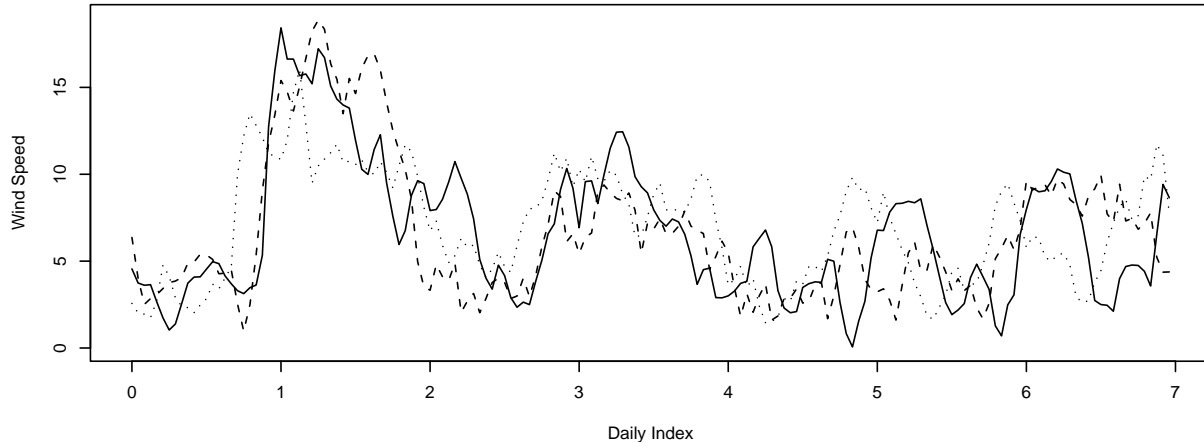


Figure 2: Wind speed at Vansycle (solid line), Kennewick (broken line) and Goodnoe Hills (dotted line) between 3 August 2002 and 9 August 2002, in  $\text{m}\cdot\text{s}^{-1}$ .

### 3.3 Forecast regimes

The winds at Vansycle, Kennewick and Goodnoe Hills are generally dictated by large scale pressure differences between the Pacific Ocean and the interior of Oregon and Washington, and by the channeling effects of the Columbia River Gorge. The Columbia Gorge is the sole near-sea-level passage through the Cascade Mountains and forms the prominent, largely east-west oriented feature in Figure 1. Air movement through gaps and passes due to surface pressure gradients is known as gap flow. The Columbia Gorge gap flow plays a profound role in defining the weather and climate within and near the Gorge, which is one of the windiest places in the Pacific Northwest (Sharp and Mass 2002, 200x). When the surface pressure is higher to the west, the flow within the gorge is normally westerly; conversely, when there is higher pressure to the east, the wind is usually easterly. Westerly winds are more frequent and more persistent in the warm season when the subtropical ridge over the eastern Pacific Ocean moves north, resulting in higher surface pressure offshore. Easterly gap flow is more common during the winter season when high pressure often develops east of the Cascades and low pressure systems frequently approach from the west (Sharp and Mass 2002, 200x). The alternation of westerly and easterly gap flow suggests the postulation of two forecast regimes, a westerly regime and an easterly regime.

Given our goal of 2-hour forecasts of hourly average wind speed at Vansycle, how do we identify the regimes? The Vansycle ridge is located near the easterly end of the Columbia Gorge, where westerly winds prevail. This suggests that the wind direction at the most westerly tower, Goodnoe Hills, is a more useful indicator of the forecast regime than the wind direction at Vansycle. Figure 4 provides strong evidence in favor of this hypothesis. The boxplots on the left-hand side compare wind speeds at Vansycle in August–December 2002, two hours after observing westerly and easterly winds at Goodnoe Hills, respectively. The separation between the two groups is striking and much sharper than the separation based on the on-site flow at Vansycle, which is shown in the right-hand graph. Hence, we distinguish the westerly and the easterly forecast regime based on the current, westerly or easterly, wind direction at Goodnoe Hills. In both regimes, the distribution of wind speed is approximately square root normal, thereby justifying the aforementioned mixture representation for the unconditional marginal distribution. Table 3 shows the seasonal variation

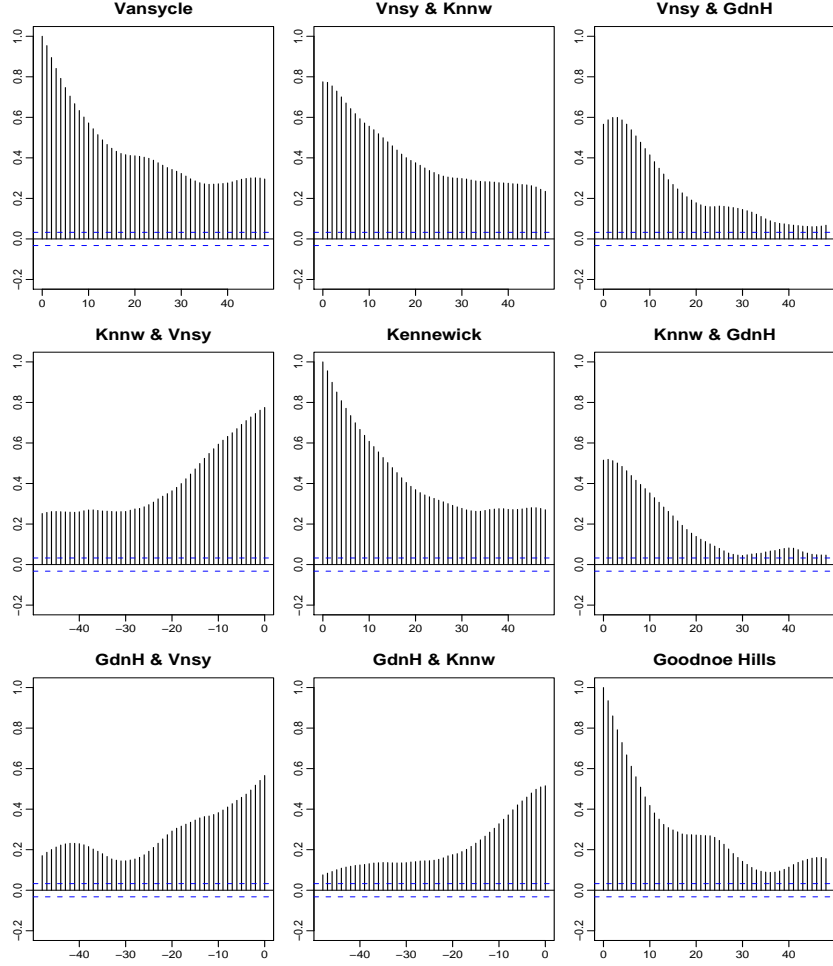


Figure 3: Autocorrelation and cross-correlation functions of hourly average wind speed at Vansycle, Kennewick and Goodnoe Hills in August–December 2002. Goodnoe Hills lies west of Kennewick, and Kennewick is located west of Vansycle. Positive lags indicate observations at the westerly station leading those at the easterly site.

in the relative frequencies of the two regimes during calendar year 2003. In the warm season the westerly regime was clearly dominant. The easterly regime was more common during the cool season, but still occurred less frequently than the westerly regime.

Table 4 illustrates the differences in the spatio-temporal dependence structures between the forecast regimes. Let  $V_t$ ,  $K_t$  and  $G_t$  denote the hourly average wind speed at Vansycle, Kennewick and Goodnoe Hills at time (hour)  $t$ , respectively. The table shows the correlations between wind speed at Vansycle two hours ahead,  $V_{t+2}$ , and present and past wind speeds at the three sites; for all data in August–December 2002, for the westerly regime only, and for the easterly regime only. The correlations associated with the westerly forecast regime were computed from the patterns

$$\mathbf{x}_t = (V_{t+2}, V_t, V_{t-1}, V_{t-2}, K_t, K_{t-1}, K_{t-2}, G_t, G_{t-1}, G_{t-2}) \quad (6)$$

at the times  $t$  with westerly flow at Goodnoe Hills, and similarly for the easterly regime. The spatio-temporal dependence structures in the two forecast regimes were clearly distinct. The westerly

Table 3: Relative frequencies of the westerly and easterly forecast regime, and of rejected or missing data at Goodnoe Hills, in calendar year 2003. The “rejected or missing” category refers to missing values in the raw data record or values rejected by the quality assurance procedures.

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Westerly regime	.35	.60	.88	.78	.83	.75	.92	.81	.65	.66	.59	.54
Easterly regime	.32	.40	.12	.22	.17	.25	.08	.19	.35	.34	.41	.46
Rejected or missing	.32	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00

Table 4: Correlations between hourly average wind speed at Vansycle two hours ahead and present and past values of hourly average wind speed at Vansycle, Kennewick and Goodnoe Hills; for all data in August–December 2002, for the westerly regime only, and for the easterly regime only.

$V_{t+2}$	$V_t$	$V_{t-1}$	$V_{t-2}$	$K_t$	$K_{t-1}$	$K_{t-2}$	$G_t$	$G_{t-1}$	$G_{t-2}$
Unconditional	.90	.84	.79	.76	.73	.70	.60	.60	.59
Westerly regime	.86	.79	.72	.72	.69	.66	.61	.62	.60
Easterly regime	.91	.86	.82	.74	.70	.67	.30	.28	.27

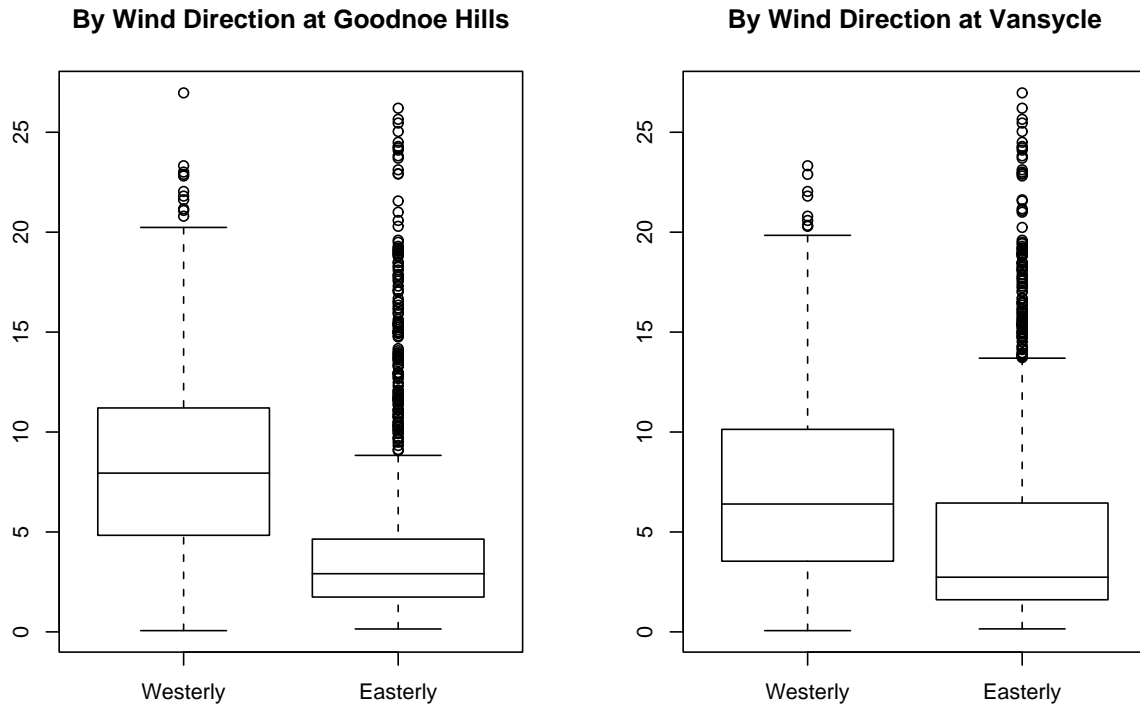


Figure 4: Boxplots for hourly average wind speed at Vansycle in August–December 2002, in dependence on the wind direction at Goodnoe Hills (left) and Vansycle (right) two hours before, in  $\text{m}\cdot\text{s}^{-1}$ , respectively.

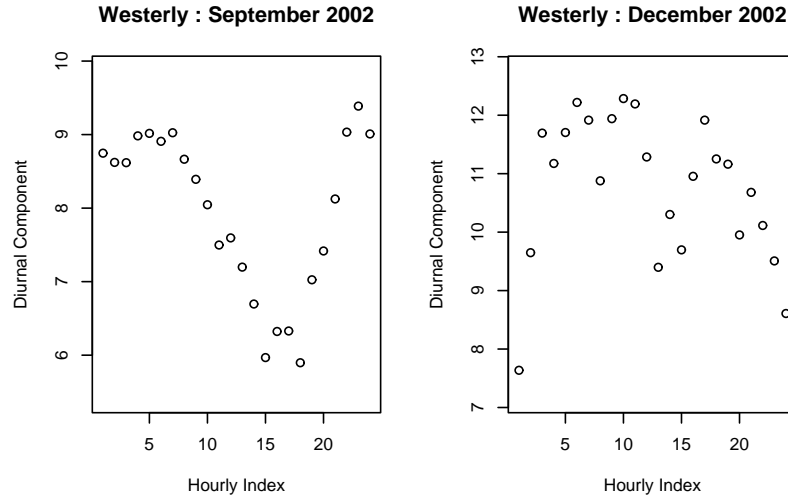


Figure 5: Diurnal component of the wind speed at Vansycle for the westerly regime in September 2002 (left) and December 2002 (right), in  $\text{m}\cdot\text{s}^{-1}$ , respectively. The graphs show the average wind speed observed during the given month and hour of the day, conditional on the wind direction at Goodnoe Hills being westerly two hours before.

regime showed substantially higher correlations between future wind speed at Vansycle and present and past wind speeds at Goodnoe Hills, respectively.

As noted above, the wind record shows a diurnal cycle, particularly in the westerly forecast regime and during the warm season. Figure 5 shows the diurnal component of wind speed at Vansycle for the westerly forecast regime in September 2002 (left) and December 2002 (right), respectively. The circles represent the average wind speed at the given hour of the day in Pacific Standard Time (PST), conditional on the wind direction at Goodnoe Hills being westerly two hours before. There was a pronounced diurnal component in September 2002, with wind speeds that peaked at night. Based on real-time mesoscale weather simulations conducted at 3 Tier Environmental Forecast Group, Inc., the evening peak is at least partially attributable to drainage winds – from the north side of the Blue Mountains in northeastern Oregon – interacting with the mean westerly flow. This nighttime drainage flow was only found in a relatively shallow layer extending a few hundreds of meters in to the atmosphere. Staley (1959) explored the diurnal cycle of surface winds in the Columbia Basin of eastern Washington and Oregon in July and August and found flow away from the lowlands of the Columbia Basin (the north east corner of Figure 1, near Moses Lake, Washington) during the daytime and flow towards the basin at night. Observations at Vansycle ridge were not used in the analysis, though the summer time diurnal cycle shown here is consistent with Staley’s analysis. This type of wind regime is predominant primarily in the warm season, when the absence of strong synoptic weather systems, coupled with much stronger daytime solar heating in the Columbia Basin, allows for the development of the phenomenon. For this reason, we do not see the diurnal signal during December 2002. In the easterly forecast regime the wind speeds tended to be much lower, and both in the warm and in the cool season there was little evidence of a clear-cut diurnal trend component.

## 4 The regime-switching space-time (RST) method

Regime-switching space-time (RST) models merge meteorological and statistical expertise to obtain probabilistic forecasts of wind resources at wind energy sites. The RST approach relies on two key ideas, the identification of distinct forecast regimes, and the use of geographically dispersed meteorological observations as off-site predictors. The regime switches address conditional spatio-temporal dependence structures that cannot be modeled by conventional vector time series techniques or geostatistical space-time methods, with the notable exception of the Huang and Hsu (2004) extension of the space-time model derived by Wikle and Cressie (1999) that allows for flow-dependent, non-stationary covariance structures. The computational requirements of the RST method are modest, and the technique can readily be implemented in real time. The model formulation is parsimonious and physically interpretable.

### 4.1 Overview

We now give an overview of the RST approach to 2-hour forecasts of hourly average wind speed at Vansycle. In Section 3.3, we identified two forecast regimes depending on the current wind direction at Goodnoe Hills, the westerly regime and the easterly regime. This was a natural choice: the regional climatology in the Pacific Northwest is well documented, and the literature distinguishes westerly and easterly flow regimes in the Columbia Gorge (Sharp and Mass 2002, 200x). At other wind energy sites, the identification of two or more forecast regimes might be less straightforward, and may involve substantial amounts of exploratory data analysis as well as local meteorological expertise.

Recent work by Campbell and Diebold (2003) and Cripps and Dunsmuir (2003) suggests the presence of conditional heteroscedasticity, that is, high frequency changes of predictability, in meteorological time series. To investigate this, we consider homoscedastic as well as conditionally heteroscedastic versions of the RST model. We also noted seasonal and flow-dependent changes in the strength of the diurnal component of wind speed, and this suggests non-diurnal as well as diurnal versions of the predictive model. To summarize, we distinguish four variants of the RST model. The **RST-N** and **RST-N-CH** techniques do **not** attempt to model the diurnal component, and they use variance structures that are **h**omoscedastic and **c**onditionally **h**eteroscedastic, respectively. The **RST-D** and **RST-D-CH** techniques fit a **d**iurnal trend component, but do so only in the westerly regime. The variance structures are homoscedastic and conditionally heteroscedastic, respectively. We generally do not model the diurnal component in the easterly regime; we considered approaches of this type, and they did not result in improved predictive performance.

### 4.2 The RST-N and RST-N-CH techniques

We now describe the RST-N and RST-N-CH techniques for 2-hour forecasts of hourly average wind speed at Vansycle. Both methods employ a cut-off Gaussian predictive distribution with parameters  $\mu_{t+2}$  and  $\sigma_{t+2} > 0$ , in symbols,

$$\mathcal{N}^0\left(\mu_{t+2}, \sigma_{t+2}^2\right). \quad (7)$$

We recall from (3) and (4) that the true predictive median,  $\mu_{t+2}^+$ , is given by  $\mu_{t+2}^+ = \max(\mu_{t+2}, 0)$ , and that the predictive mean,  $\mu_{t+2}^0$ , is found as

$$\mu_{t+2}^0 = \mu_{t+2} \Phi\left(\frac{\mu_{t+2}}{\sigma_{t+2}}\right) + \sigma_{t+2} \phi\left(\frac{\mu_{t+2}}{\sigma_{t+2}}\right). \quad (8)$$

During our evaluation period, May–November 2003, we obtained nonnegative estimates of  $\mu_{t+2}$  for 5,134 of the 5,136 predictive distributions. Clearly, the true predictive median,  $\mu_{t+2}^+$ , and the predictive mean,  $\mu_{t+2}^0 \geq \mu_{t+2}^+$ , are always nonnegative. The point mass at zero was mostly but not always very small, and the predictive median and predictive mean were mostly but not always nearly indistinguishable. Hereinafter, we refer to the parameters  $\mu_{t+2}$  and  $\sigma_{t+2}$  of the cut-off normal distribution (7) as the predictive median and the predictive spread, respectively, with the tacit assumption that  $\mu_{t+2}$  is nonnegative.

We model  $\mu_{t+2}$  as a linear function of past and present observations of wind speed at the meteorological towers. In the westerly forecast regime, we put

$$\mu_{t+2} = a_0 + a_1V_t + a_2V_{t-1} + a_3K_t + a_4K_{t-1} + a_5G_t, \quad (9)$$

where  $V_t$ ,  $K_t$ , and  $G_t$  denote the wind speed at Vansycle, Kennewick and Goodnoe Hills at time  $t$ , respectively. This is akin to a vector autoregressive time series model (Brockwell and Davis 1991, Chapter 11; de Luna and Genton 2003) or an autoregressive distributed lag scheme (Zivot and Wang 2003, Section 6.4). However, the predictive distribution is cut-off Gaussian, and the model applies in the westerly regime only; so, conventional time series techniques do not apply. In the easterly forecast regime, we exclude information from the tower at Goodnoe Hills, which is downwind in this regime, and model the predictive median as

$$\mu_{t+2} = a_0 + a_1V_t + a_2V_{t-1} + a_3K_t. \quad (10)$$

As noted above, we distinguish the forecast regimes by the current, westerly or easterly, wind direction at Goodnoe Hills. The model selection process for the conditional linear models in (9) and (10) was guided by the regime characteristics, by the parsimony principle and by an exploratory analysis of the wind speed series in August–December 2002. For both forecast regimes, we considered multiple linear regression models for  $V_{t+2}$  on the predictor variables in Table 4. We started from the simplest model and added predictor variables until the decrease in the root-mean-square error (RMSE) became negligible, when compared to the persistence RMSE. Clearly, there are more formal, automated approaches to model selection that may or may not be appropriate here. We return to this point in Section 6 below.

It remains to model the predictive spread,  $\sigma_{t+2}$ . The homoscedastic RST-N technique assumes that the predictive spread is constant over time. The RST-N-CH method allows for conditional heteroscedasticity, by modeling

$$\sigma_{t+2} = b_0 + b_1v_t \quad (11)$$

as a linear function of the volatility value,  $v_t$ , with coefficients that are constrained to be nonnegative. The volatility value,

$$v_t = \left( \frac{1}{6} \sum_{i=0}^1 \left( (V_{t-i} - V_{t-i-1})^2 + (K_{t-i} - K_{t-i-1})^2 + (G_{t-i} - G_{t-i-1})^2 \right) \right)^{1/2}, \quad (12)$$

reflects the magnitudes of recent changes in wind speed at the three sites. Note that we use (12) to define volatility both in the westerly and in the easterly regime. It may surprise that we include information from Goodnoe Hills in the volatility values for the easterly regime. However, while volatility is flow-dependent it is also a diurnal effect. We attempted to model the diurnal variation of predictability directly, as suggested by Campbell and Diebold (2003), but this did not result in improved predictive performance.

We turn to the estimation of the conditional statistical models. The literature argues that maximum likelihood plug-in estimates may be suboptimal when the goal is prediction (Copas 1983; Friedman 1989). Gneiting, Westveld, Raftery and Goldman (2004) proposed the novel technique of minimum continuous ranked probability score (CRPS) estimation for estimating predictive distributions. In minimum CRPS estimation, we express the continuous ranked probability score for the training data as a function of the model parameters, and minimize that function with respect to the parameter values. This technique is tailored to probabilistic forecasting and can be interpreted within the framework of robust M-estimation (Gneiting and Raftery 2004). Here the predictive distributions are cut-off normal and we minimize the CRPS value (2), where each term is computed from (5). For RST-D-CH forecasts in the easterly regime, for instance, we find the minimum of the CRPS value as a function of the parameters in (10) and (11). This needs to be done numerically, and we use the Broyden-Fletcher-Goldfarb-Shanno algorithm (Press et al. 1992, Section 10.7) as implemented in the R language and environment ([www.cran.r-project.org](http://www.cran.r-project.org)). The algorithm is iterative, and starting values based on past experience usually give good solutions. Convergence to a global maximum cannot be guaranteed, and the solution reached can be sensitive to the initial values.

### 4.3 Modeling the diurnal component: The RST-D and RST-D-CH techniques

The RST-D and RST-D-CH techniques fit diurnal trend components to the wind speed series at Vansycle, Kennewick and Goodnoe Hills. This is done in the westerly regime only; in the easterly forecast regime, the RST-D and RST-D-CH techniques do not differ from the RST-N and RST-N-CH approaches, respectively.

Hence, suppose that the current regime is westerly. We employ the cut-off normal predictive distribution (7) with predictive median,  $\mu_{t+2}$ , and predictive spread,  $\sigma_{t+2}$ , to be specified below. The predictive mean,  $\mu_{t+2}^0$ , is found from (8). At each of the three sites, we fit the trigonometric function

$$D_t = d_0 + d_1 \sin\left(\frac{2\pi t}{24}\right) + d_2 \cos\left(\frac{2\pi t}{24}\right) + d_3 \sin\left(\frac{4\pi t}{24}\right) + d_4 \cos\left(\frac{4\pi t}{24}\right) \quad (13)$$

that uses two pairs of harmonics to regress wind speed,  $D_t$ , on the hour of the day, conditionally on the wind direction at Goodnoe Hills being westerly at time  $t - 2$ . We remove the respective ordinary least squares fit from the wind speed series at Vansycle, Kennewick and Goodnoe Hills, resulting in residual series that we denote by  $V_t^r$ ,  $K_t^r$  and  $G_t^r$ , respectively. The predictive median,  $\mu_{t+2}$ , is then modeled as

$$\mu_{t+2} = D_{t+2} + \mu_{t+2}^r,$$

where  $D_{t+2}$  is the ordinary least squares fit for the diurnal component at Vansycle, and where

$$\mu_{t+2}^r = a_0 + a_1 V_t^r + a_2 V_{t-1}^r + a_3 K_t^r + a_4 K_{t-1}^r + a_5 G_t^r \quad (14)$$

is a linear function of present and past values of the residual series at the three sites. The RST-D technique is homoscedastic and assumes that the predictive spread,  $\sigma_{t+2}$ , is constant. The RST-D-CH approach allows for conditional heteroscedasticity, by modeling the predictive spread,  $\sigma_{t+2} = b_0 + b_1 v_t^r$ , as a linear function of the volatility value,

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

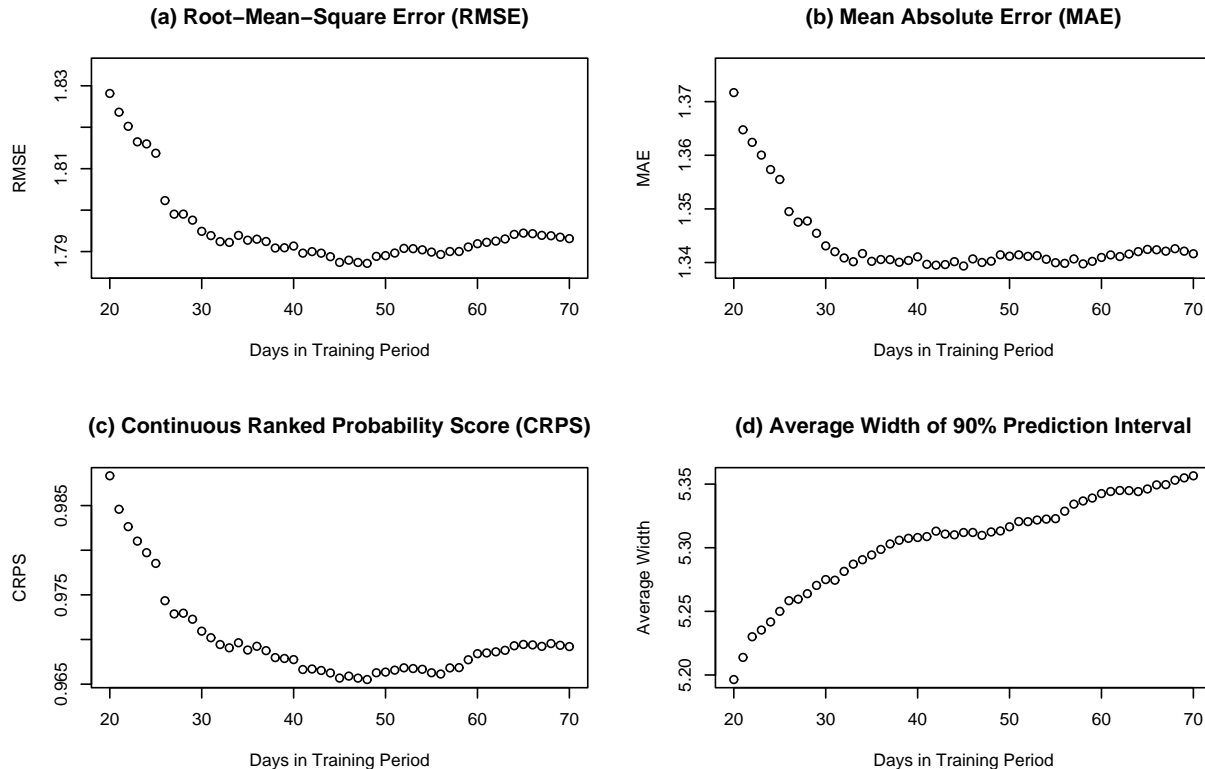


Figure 6: Measures of predictive performance for the RST-D-CH forecasts in dependence on the length of the sliding training period for a test period between 8 May 2003 and 30 November 2003, in  $\text{m}\cdot\text{s}^{-1}$ : (a) Root-mean-square error. (b) Mean absolute error. (c) Continuous ranked probability score. (d) Average width of 90% prediction interval.

with coefficients  $b_0$  and  $b_1$  that are constrained to be nonnegative. We use the minimum CRPS technique to estimate the conditional linear model (14) and the variance parameters, as described above.

#### 4.4 Choice of training period

What training period should be used to estimate the conditional predictive models? The lack of a historic meteorological record at Vansycle, which is typical for wind energy sites, suggests the use of the sliding window technique in which the training period consists of the recent past. Clearly, there is a trade-off in selecting the length of the sliding training period. Shorter training periods can adapt rapidly to seasonally varying atmospheric conditions, and thereby take account of seasonal non-stationarities. On the other hand, longer training periods reduce the statistical variability in the estimation. Intuitively, the trade-off may suggest training periods between 30 and 60 days.

There is no automatic way of finding the optimal length, and we studied the effect of the different training periods empirically, similarly to the experiments reported by Raftery, Balabdaoui, Gneiting and Polakowski (2003) and Gneiting et al. (2004). As noted above, the longest complete segment of data at Vansycle, Kennewick and Goodnoe Hills consists of the 279-day period between 25 February 2003 and 30 November 2003. We considered sliding training periods consisting of the  $m$

most recent 24-hour periods, where  $m = 20, 21, \dots, 70$ . In other words, the 2-hour forecast for the hourly average wind speed at time  $t + 2$  was trained on the patterns  $\mathbf{x}_s$ , defined in (6) above, at time  $s = t - 2, t - 3, \dots, t - 24m - 1$ . For comparability, we used the same test set in assessing all the training periods, that is, the 208-day period between 8 May 2003 and 30 November 2003. For reasons of computational tractability, we used a two-stage estimator that employs the ordinary least squares technique to fit the regression coefficients in (9), (10) or (14), respectively, and then finds the spread parameters using minimum CRPS estimation, with the regression coefficients fixed. The two-stage procedure is much faster than full minimum CRPS estimation, which optimizes over the regression coefficients and variance parameters simultaneously. The associated slight loss in predictive performance is immaterial for evaluating the training periods.

Figure 6 summarizes the results of the experiment for the RST-D-CH approach. The results for the RST-N, RST-N-CH and RST-D forecasts were similar. Figures 6(a) and 6(b) show the root-mean-square error (RMSE) and the mean absolute error (MAE) of the deterministic-style forecasts, respectively. These decrease sharply for training periods less than 30 days, stay about constant for training periods between 30 and 60 days, and tend to increase thereafter. Figure 6(c) shows the continuous ranked probability score (CRPS). The pattern is similar to those for the RMSE and the MAE. The average width of the 90% prediction interval is shown in Figure 6(d). The width increases with the length of the training period, but is about constant for training periods between 40 and 50 days. To summarize these results, there appear to be substantial gains in increasing the training period beyond 30 days, and a 45-day training period appeared about best. Hence, we consider RST forecasts that use a 45-day training period and full minimum CRPS estimation. However, the predictive performance was insensitive to the length of the training period, and training periods between 30 and 60 days generally seemed adequate.

## 5 Predictive performance

We now assess the predictive performance of the RST forecasts. First we describe more conventional prediction techniques that we use as benchmarks. We evaluate the deterministic-style forecasts and the probabilistic forecasts, and we give an explicit example of the RST forecasts for the 3-week period beginning 21 June 2003.

### 5.1 Reference forecasts

This section describes the reference forecasts: the persistence forecast, the new reference forecast of Nielsen et al. (1998), forecasts based on univariate autoregressive time series models, and forecasts using vector autoregressive techniques. We denote wind speed at Vansycle, Kennewick and Goodnoe Hills at time  $t$  by  $V_t$ ,  $K_t$  and  $G_t$ , respectively; and we write  $\hat{V}_{t+2}$  for a deterministic-style 2-hour forecast of hourly average wind speed at Vansycle.

The classical reference forecast in the literature is the **persistence** forecast,  $\hat{V}_{t+2} = V_t$ , that utilizes the most recent observation at hand. Clearly, the shorter the forecast lead time, the more competitive the persistence forecast. Nielsen et al. (1998) proposed a **new reference** forecast for wind energy applications that takes the form  $\hat{V}_{t+2} = \rho V_t + (1 - \rho)\bar{V}$ , where  $\rho$  and  $\bar{V}$  denote the 2-hour lagged correlation and the average wind speed in the historic record, respectively. The new reference forecast can be interpreted as a forecast based on an autoregressive time series model. The persistence and the new reference methods provide deterministic-style forecasts only.

Brown et al. (1984) proposed the use of autoregressive (AR) time series models for wind speed and wind power forecasts. They applied the square root transform to the series of hourly average wind speed at Goodnoe Hills, fit and extracted a diurnal trend component, and modeled the residual component as an AR process. This approach has found widespread use (Giebel 2003). For the forecasts at Vansycle, we experimented with the square root transform of Brown et al. (1984) and the more general transform proposed by Allcroft and Glasbey (2003), but found forecasts based on time series models for the non-transformed wind speed series to perform best. The homoscedastic **AR-N** technique utilizes the S-PLUS function `ar.yw` to fit an AR model of order at most 4. The conditionally heteroscedastic **AR-N-CH** technique, furthermore, uses the S+FINMETRICS function `garch` to fit a GARCH(1,1) model for the conditional variance of the innovations (Bollerslev 1986; Zivot and Wang 2003, Chapter 7). The **AR-D** and **AR-D-CH** methods estimate and extract a diurnal trend component of the form (13) and then proceeds as above. To obtain Gaussian predictive distribution functions from the time series models, we follow Brown et al. (1984) and Brockwell and Davis (1991, Section 5.4). In our experiments, a sliding 40-day training period appeared best for the AR forecasts. The results were insensitive to changes in the training period, and training periods between 30 and 50 days seemed adequate.

In contrast to the univariate time series techniques that rely on on-site predictors, the RST method utilizes observations from the meteorological towers at Kennewick and Goodnoe Hills, respectively. Multivariate time series techniques (Brockwell and Davis 1991, Chapter 11) offer a more traditional way of incorporating information from Kennewick and Goodnoe Hills into the predictive model. De Luna and Genton (2003) proposed a vector autoregressive (VAR) specification that is geared to provide time-forward predictions in environmental applications, where there are as many time series as stations. They introduced a model selection strategy that takes the spatio-temporal dependence structure into account and borrows its strength from the Box and Jenkins (1976) strategy for univariate time series. S-PLUS and R code for this technique is available; however, the current implementation does not provide prediction bounds. We implemented the de Luna and Genton (2003) approach in two variants, using a maximal order of 6 for the vector autoregression and a sliding 45-day training period. The **VAR-N** technique fits the vector AR model directly to the wind speed series at Vansycle, Kennewick and Goodnoe Hills. The **VAR-D** technique fits a diurnal trend component of the form (13) at each of the three sites, and applies the vector autoregression to the residual series.

Vector AR techniques lead to predictive models that resemble the schemes in (9), (10) and (14), respectively. That said, we note two key differences to the RST method. In contrast to the RST method, vector AR techniques do not identify and distinguish forecast regimes. However, the method of de Luna and Genton (2003) applies a model selection strategy for each forecast; hence, the predictor variables may vary from forecast to forecast. For the RST method, the parameter estimates vary with the sliding training period, but the predictive models themselves are fixed.

## 5.2 Assessment of deterministic-style forecasts

We now assess the predictive performance of the various forecasts. As noted before, the longest continuous data record at Vansycle, Kennewick and Goodnoe Hills ranges from 25 February 2003 through 30 November 2003. The forecasts based on the univariate AR models used a 40-day training period, and the VAR and RST forecasts were trained on a 45-day period. Our test period ranges from 1 May 2003 through 30 November 2003, and we report the results month by month. This allows us to study the consistency of the results and to assess seasonal effects.

Table 5: Root-mean-square error (RMSE) for deterministic-style 2-hour forecasts of hourly average wind speed at Vansycle in May–November 2003, in  $\text{m}\cdot\text{s}^{-1}$ . The lowest value in each column is given in bold fonts.

Forecast	May	Jun	Jul	Aug	Sep	Oct	Nov
Persistence	2.14	1.97	2.37	2.27	2.17	2.38	2.11
New reference	2.06	1.93	2.27	2.18	2.12	2.31	2.12
AR-N	2.04	1.92	2.19	2.13	2.10	2.31	2.08
AR-D	2.01	1.85	2.00	2.03	2.03	2.30	2.08
VAR-N	1.95	1.70	1.87	1.90	1.95	2.27	2.03
VAR-D	1.79	1.61	1.72	1.80	1.80	2.13	1.90
RST-N	1.78	1.58	1.80	1.84	1.81	2.08	<b>1.87</b>
RST-D	<b>1.75</b>	<b>1.56</b>	<b>1.70</b>	<b>1.78</b>	<b>1.77</b>	<b>2.07</b>	1.88

Table 6: Mean absolute error (MAE) for deterministic-style 2-hour forecasts of hourly average wind speed at Vansycle in May–November 2003, in  $\text{m}\cdot\text{s}^{-1}$ . The lowest value in each column is given in bold fonts.

Forecast	May	Jun	Jul	Aug	Sep	Oct	Nov
Persistence	1.60	1.45	1.74	1.68	1.59	1.68	1.51
New reference	1.57	1.44	1.68	1.62	1.59	1.66	1.55
AR-N	1.54	1.42	1.63	1.61	1.56	1.66	1.51
AR-D	1.54	1.38	1.50	1.54	1.53	1.67	1.53
VAR-N	1.52	1.31	1.47	1.44	1.56	1.71	1.50
VAR-D	1.38	1.24	1.36	1.35	1.40	1.57	1.44
RST-N	1.35	1.20	1.40	1.37	1.39	1.50	1.38
RST-D	<b>1.32</b>	<b>1.18</b>	<b>1.33</b>	<b>1.31</b>	<b>1.36</b>	<b>1.48</b>	<b>1.37</b>

In evaluating deterministic-style forecast skill, we ignore the distinction between the homoscedastic and the conditionally heteroscedastic versions of the AR-N, AR-D, RST-N and RST-D forecasts, respectively. We also avoid a comparison of forecasts using the predictive median and predictive mean. Typically, the forecasts based on the conditionally heteroscedastic models showed slightly smaller root-mean-square error (RMSE) than the respective forecasts based on the homoscedastic models; and forecasts using the predictive mean generally had RMSE slightly lower than forecasts using the predictive median. However, the differences were mostly negligibly small, and in almost all cases the respective RMSE or mean absolute error (MAE) values were identical when rounded to three decimal places.

Table 5 shows the RMSE for the various methods. The RST forecasts had consistently lower RMSE than the forecasts based on the vector AR techniques; the latter outperformed the forecasts based on the univariate AR models; and the forecasts based on the univariate AR models had consistently lower RMSE than the persistence and new reference forecasts. In May through October 2003, the RST-D forecasts outperformed all others and had RMSE 18.3%, 20.9%, 28.6%, 21.6%, 18.7% and 13.2% less than the persistence forecasts, respectively. In November 2003, the RST-N

Table 7: Continuous ranked probability score (CRPS) for probabilistic 2-hour forecasts of hourly average wind speed at Vansycle in May–November 2003, in  $\text{m}\cdot\text{s}^{-1}$ . The lowest value in each column is given in bold fonts.

Forecast	May	Jun	Jul	Aug	Sep	Oct	Nov
AR-N	1.13	1.05	1.20	1.17	1.14	1.22	1.12
AR-N-CH	1.12	1.04	1.19	1.16	1.13	1.22	1.10
AR-D	1.12	1.02	1.10	1.11	1.11	1.22	1.13
AR-D-CH	1.11	1.01	1.10	1.11	1.10	1.22	1.10
RST-N	0.99	0.86	1.00	1.00	1.00	1.10	1.01
RST-N-CH	0.98	0.86	1.00	0.99	0.99	1.09	<b>1.00</b>
RST-D	0.97	<b>0.85</b>	<b>0.95</b>	0.96	0.98	1.10	1.01
RST-D-CH	<b>0.96</b>	<b>0.85</b>	<b>0.95</b>	<b>0.95</b>	<b>0.97</b>	<b>1.08</b>	<b>1.00</b>

technique performed best, with an RMSE that was 11.4% lower than the persistence RMSE. The improved performance of the RST-N forecasts toward the end of the test period is not surprising, given the lack of a pronounced diurnal trend component during the cool season. The diurnal vector AR technique, VAR-D, was the closest competitor to the RST forecasts. However, the RST forecasts had consistently lower RMSE. The results in terms of the mean absolute error (MAE) are shown in Table 6; they were similar, except that the relative improvement of the RST-D over the VAR-D forecasts was more pronounced than in terms of the RMSE. The RST-D forecasts had consistently the lowest MAE.

### 5.3 Assessment of probabilistic forecasts

To evaluate the probabilistic forecast skill of the univariate AR and the RST techniques, we use the continuous ranked probability score (CRPS) and the diagnostic tools described in Section 2.2.

Table 7 shows the CRPS for the various techniques. The forecasts that allowed for conditional heteroscedasticity generally had lower CRPS than their homoscedastic counterparts; and the techniques that fit a diurnal trend component mostly performed better than the respective non-diurnal methods. The RST predictive CDFs showed substantially lower CRPS than the predictive CDFs using the AR techniques. For each month in the test period, the RST-D-CH forecasts had the lowest CRPS. In June 2003 and July 2003 the top performance was tied with the RST-D forecasts, and in November 2003 it was shared with the RST-N-CH forecasts.

The probability integral transform (PIT) histograms for the predictive CDFs based on the univariate AR and the RST methods are shown in Figures 7 and 8, respectively. The PIT histograms for the AR-D and AR-D-CH forecasts were inverse U-shaped, thereby indicating overdispersed predictive distributions. The effect was slightly less pronounced for the conditionally heteroscedastic AR-D-CH forecasts, but still was prominent. Gneiting et al. (2004) argue that the use of maximum likelihood plug-in estimates tends to result in overdispersed predictive distributions; and a similar argument may apply here, too. The RST techniques use minimum CRPS estimation, and the PIT histograms were nearly uniform, except for a small spike at the right-most bin. The PIT histogram for the RST-D-CH forecasts is much more uniform than the histograms typically seen in the literature and implies close to nominal coverage of the prediction intervals at all levels.

Table 8: Average width of the 90% prediction intervals for 2-hour forecasts of hourly average wind speed at Vansycle in May–November 2003, in  $\text{m}\cdot\text{s}^{-1}$ . The lowest value in each column is given in bold fonts.

Forecast	May	Jun	Jul	Aug	Sep	Oct	Nov
AR-N	7.22	6.41	6.62	6.96	6.69	6.58	6.85
AR-N-CH	6.91	6.16	6.47	6.79	6.53	6.89	6.74
AR-D	6.98	6.22	6.21	6.38	6.37	6.40	6.78
AR-D-CH	6.68	5.95	6.04	6.29	6.24	6.66	6.67
RST-N	6.04	5.02	5.31	5.44	5.28	5.35	5.43
RST-N-CH	5.96	4.92	5.32	5.50	5.34	5.64	5.54
RST-D	5.96	4.91	5.15	<b>5.14</b>	<b>5.07</b>	<b>5.15</b>	<b>5.35</b>
RST-D-CH	<b>5.93</b>	<b>4.83</b>	<b>5.14</b>	5.22	5.15	5.45	5.46

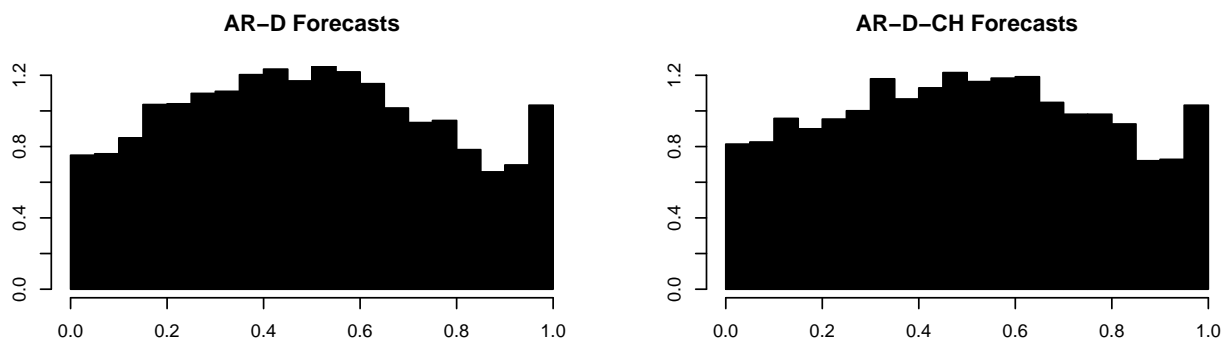


Figure 7: Probability integral transform (PIT) histograms for AR-D (left) and AR-D-CH (right) predictive CDFs of hourly average wind speed at Vansycle in May–November 2003.

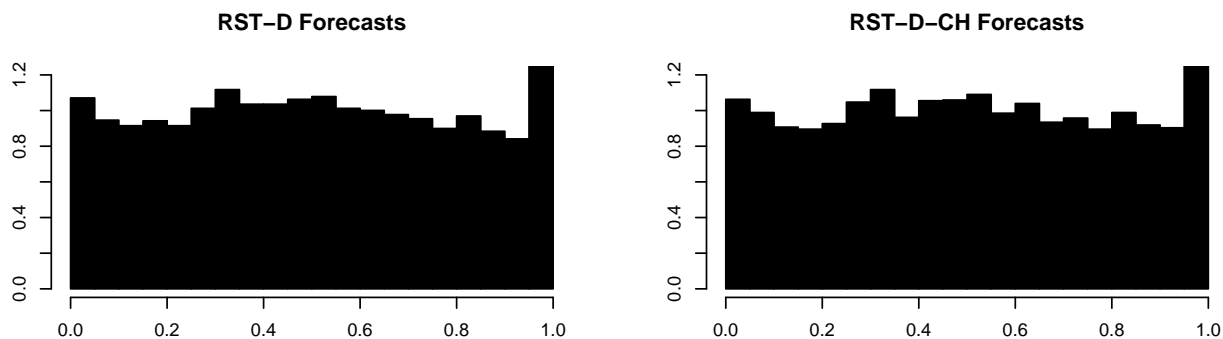


Figure 8: Probability integral transform (PIT) histograms for RST-D (left) and RST-D-CH (right) predictive CDFs of hourly average wind speed at Vansycle in May–November 2003.

Table 8 shows the average width of the 90% central prediction interval. The prediction intervals for the diurnal techniques were consistently shorter than the intervals for the respective non-diurnal approaches. The effect of modeling conditional heteroscedasticity was less clear-cut. The RST predictive CDFs were substantially sharper than the predictive CDFs derived from the univariate AR models. On average, the RST-D-CH prediction intervals were about 18% shorter than the AR-D-CH intervals, while maintaining adequate coverage, thereby achieving significant reductions in forecast uncertainty.

## 5.4 Example

To summarize the above results, the RST-D-CH technique outperformed all the others during the warm season. Our test period did not include the winter months, but the results for November 2003 suggest that during the cool season the RST-N-CH forecasts may perform best.

Figure 9 shows 2-hour forecasts of hourly average wind speed at Vansycle for the 3-week period beginning on 21 June 2003, using the RST-D-CH approach. The green line shows the mean of the cut-off normal predictive distribution, and the broken red lines correspond to the lower and upper end points of the 90% central prediction interval, respectively. The observed values of the hourly average wind speed are shown as black circles and, indeed, the empirical coverage was close to nominal. The forecasts issued in the westerly regime are indicated by the blue marks at top. The westerly regime dominated and showed a pronounced diurnal component, with wind speeds that peaked at night. In the easterly regime, the wind speeds were generally lower, there was little evidence of a diurnal trend component, and the prediction intervals were considerably shorter than in the westerly regime.

## 6 Discussion

We introduced the regime-switching space-time (RST) method that merges meteorological and statistical expertise to obtain accurate and calibrated, fully probabilistic forecasts of wind resources. The model formulation is parsimonious yet takes account of the salient features of wind speed: alternating atmospheric regimes, temporal and spatial correlation, diurnal and seasonal non-stationarity, range restrictions, and conditional heteroscedasticity. The RST method identifies forecast regimes at the wind energy site and utilizes geographically dispersed meteorological observations in the vicinity of the wind farm as off-site predictors.

We applied the RST technique to 2-hour forecasts of hourly average wind speed at the Vansycle ridge in the US Pacific Northwest. In July 2003, for instance, the RST forecasts had root-mean-square error (RMSE) 28.6% lower than the persistence forecasts. The RST method provides probabilistic forecasts in the form of predictive distributions, and those were well calibrated and sharp. The RST prediction intervals were substantially shorter on average than prediction intervals derived from univariate time series techniques, and still showed adequate coverage, as reflected by a nearly uniform probability integral transform (PIT) histogram. During the warm season, the RST-D-CH variant of the RST technique performed best. This method models the diurnal component of wind speed and takes conditional heteroscedasticity into account. For forecasts in the cool season, we recommend the use of the RST-N-CH technique that does not model the diurnal component.

We proceed with a discussion of possible extensions as well as limitations of the RST approach. In its current implementation, the RST technique uses empirical results to select fixed sets of

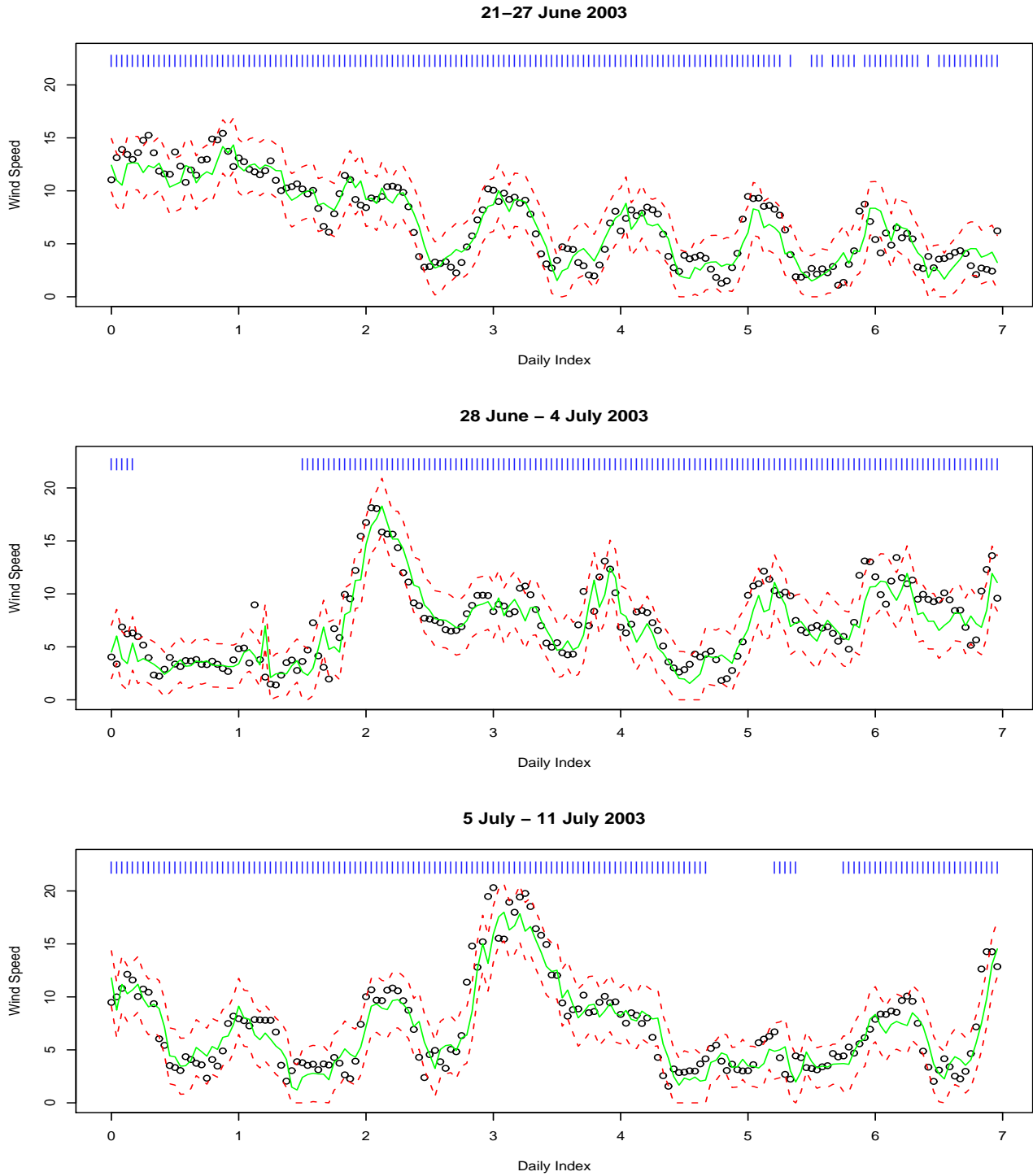


Figure 9: 2-Hour RST-D-CH forecasts of hourly average wind speed at Vansycle for the 3-week period beginning on 21 June 2003, in  $\text{m}\cdot\text{s}^{-1}$ . The mean of the predictive distribution is shown by the green line, along with the 90% central prediction interval that is bordered by the broken red lines. The observed wind speeds are shown as black circles, and forecasts issued in the westerly regime are identified by the blue marks at top.

predictor variables in the conditional linear models (9), (10) and (14), respectively. In our case study, the choice of the predictor variables was comparably straightforward, because there were only three stations and two regimes, and because the local climatology was well studied. In other cases, there could be more stations and more regimes, and the model selection problem could be more difficult. The closest competitor for the RST method, the vector autoregressive (VAR) technique in the implementation of de Luna and Genton (2003), utilizes off-site observations, too, but does not distinguish forecast regimes, and does not provide predictive distributions. However, the VAR technique uses a model selection strategy that is geared to provide time-forward predictions in environmental applications and takes the spatio-temporal dependence structure into account. It may be possible to combine the strengths of the two approaches, by applying similar automated model selection strategies to the conditional RST models. The resulting improvements in forecast skill, if any, are likely to be incremental. That said, even small improvements in forecast accuracy may result in substantial financial gains.

The conditional linear RST models allow for additional predictor variables. For instance, the pressure gradient between The Dalles, Oregon in the Columbia Gorge and Portland, Oregon is commonly used by operational forecasters to predict the strength of the Columbia Gorge gap flow (Sharp and Mass 200x). Other meteorological variables or numerical weather prediction forecasts could be considered. The benefits are likely to be small, unless meteorological observations from stations east of Vansycle were to become available.

As noted in the introduction, numerical weather prediction (NWP) forecasts are not competitive at the forecast lead time that we consider here. However, NWP forecasts and mesoscale ensemble prediction systems (Grimt and Mass 2002) may provide independent information that could be used in the form of additional predictor variables for the conditional linear models, in identifying the forecast regimes, or in assessing and modeling conditional heteroscedasticity. These are topics for future research.

To conclude, we anticipate that the RST approach can be successfully applied at wind energy sites all over the world. The approach borrows its strength from collaborative efforts between atmospheric scientists and statisticians and yields physically interpretable predictive models. Atmospheric regimes can be identified globally, and in many parts of the world quality off-site observations may already be available in real time, although the placement of the sites may or may not be ideal for wind forecasting applications. The resulting improvements in deterministic-style and probabilistic forecast skill can contribute to improved wind energy integration, a more reliable power grid, improved risk management, economic savings, and ultimately a greater acceptance of wind power and a lesser reliance on non-renewable energy sources throughout the world.

## 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.
- Allcroft, D. J. and Glasbey, C. A. (2003). A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation. *Applied Statistics*, **52**, 487–498.
- Bardossy, A. and Plate, E. J. (1992). Space-time models for daily rainfall using atmospheric circulation patterns. *Water Resources Research*, **28**, 1247–1259.

- Bernardo, J. M. (1979). Expected information as expected utility. *Annals of Statistics*, **7**, 686–690.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *Journal of Econometrics*, **31**, 307–327.
- Box, G. E. P. and Jenkins, G. M. (1976). *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco.
- Bremnes, J. B. (2004). Probabilistic wind power forecasts using local quantile regression. *Wind Energy*, **7**, 47–54.
- Brockwell, P. and Davis, R. A. (1991). *Time Series: Theory and Methods*, second edition. Springer, New York.
- 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.
- Campbell, S. D. and Diebold, F. X. (2003). Weather forecasting for weather derivatives. Working Paper no. w10141, National Bureau of Economic Research. Available online at [papers.nber.org/papers/W10141](http://papers.nber.org/papers/W10141).
- Carroll, R. J., Chen, R., George, E. I., Li, T. H., Newton, H. J., Schmiediche, H. and Wang, N. (1997). Ozone exposure and population density in Harris County, Texas (with discussion). *Journal of the American Statistical Association*, **92**, 392–415.
- Copas, J. B. (1983). Regression, prediction and shrinkage. *Journal of the Royal Statistical Society, Ser. B*, **45**, 311–354.
- Cripps, E. and Dunsmuir, W. T. M. (2003). Modeling the variability of Sydney Harbor wind measurements. *Journal of Applied Meteorology*, **42**, 1131–1138.
- Dawid, A. P. (1984). Statistical theory: The prequential approach. *Journal of the Royal Statistical Society, Ser. A*, **147**, 278–292.
- de Luna, X. and Genton, M. G. (2003). Predictive spatio-temporal models for spatially sparse environmental data. Institute of Statistics Mimeo Series #2546. Revised for *Statistica Sinica*. Available online at [www4.stat.ncsu.edu/~mggenton/stpredict.ps](http://www4.stat.ncsu.edu/~mggenton/stpredict.ps).
- Diebold, F. X., Gunther, T. A. and Tay, A. S. (1998). Evaluating density forecasts with applications to financial risk management. *International Economic Review*, **39**, 863–883.
- Fiebrich, C. A. and Crawford, K. C. (2001). The impact of unique meteorological phenomena detected by the Oklahoma mesonet on automated quality control. *Bulletin of the American Meteorological Society*, **82**, 2173–2187.
- Friedman, J. H. (1989). Regularized discriminant analysis. *Journal of the American Statistical Association*, **84**, 165–175.
- Garratt, A., Lee, K., Pesaran, M. H. and Shin, Y. (2003). Forecast uncertainties in macroeconomic modelling: An application to the UK economy. *Journal of the American Statistical Association*, **98**, 829–838.
- Gel, Y., Raftery, A. E. and Gneiting, T. (2004). Calibrated probabilistic mesoscale weather field forecasting: The geostatistical output perturbation (GOP) method (with discussion). *Journal of the American Statistical Association*, **99**, in press.
- Giebel, G. (2003). The state-of-the-art in short-term prediction of wind power. Deliverable Report D1.1, Project Anemos. Available online at [anemos.cma.fr/modules.php?name=Downloads&d\\_op=viewdownload&cid=3](http://anemos.cma.fr/modules.php?name=Downloads&d_op=viewdownload&cid=3).

- Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association*, **97**, 590–600.
- Gneiting, T. and Raftery, A. E. (2004). Strictly proper scoring rules, prediction, and estimation. Technical Report no. 463, Department of Statistics, University of Washington. Available online at [www.stat.washington.edu/tech.reports](http://www.stat.washington.edu/tech.reports).
- Gneiting, T., Raftery, A. E., Balabdaoui, F. and Westveld, A. (2003). Verifying probabilistic forecasts: Calibration and sharpness. In *Proceedings of the Workshop on Ensemble Forecasting, Val-Morin, Québec*. Available online at [www.cdc.noaa.gov/~hamill/ef\\_workshop\\_2003.html](http://www.cdc.noaa.gov/~hamill/ef_workshop_2003.html).
- Gneiting, T., Westveld, A. H., Raftery, A. E. and Goldman, T. (2004). Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Technical Report no. 449, Department of Statistics, University of Washington. Revised for *Monthly Weather Review*. Available online at [www.stat.washington.edu/tilmann/#techreports](http://www.stat.washington.edu/tilmann/#techreports).
- Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society, Ser. B*, **14**, 107–114.
- Grimit, E. P. and Mass, C. F. (2002). Initial results of a mesoscale short-range forecasting system over the Pacific Northwest. *Weather and Forecasting*, **17**, 192–205.
- Haslett, J. and Raftery, A. E. (1989). Space-time modelling with long-memory dependence: Assessing Ireland’s wind-power resource (with discussion). *Applied Statistics*, **38**, 1–50.
- Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, **15**, 559–570.
- Huang, H.-C. and Hsu, N.-J. (2004). Modeling transport effects on ground-level ozone using a non-stationary space-time model. *Environmetrics*, **15**, 151–268.
- Hughes, J. P. and Guttorp, P. (1994). A class of stochastic models for relating synoptic atmospheric patterns to regional hydrologic phenomena. *Water Resources Research*, **30**, 1535–1546.
- Hughes, J. P., Guttorp, P. and Charles, S. P. (1999). A non-homogeneous hidden Markov model for precipitation occurrence. *Applied Statistics*, **48**, 15–30.
- Kretzschmar, R., Eckert, P., Cattani, D. and Eggimann, F. (2004). Neural network classifiers for local wind prediction. *Journal of Applied Meteorology*, **43**, 727–738.
- Larson, K. and Gneiting, T. (2004). Improving wind energy forecasts. *Bulletin of the American Meteorological Society*, **85**, 343–344.
- Nielsen, T. S., Joensen, A., Madsen, H., Landberg, L. and Giebel, G. (1998). A new reference for wind power forecasting. *Wind Energy*, **1**, 29–34.
- Palmer, T. N. (2002). The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. *Quarterly Journal of the Royal Meteorological Society*, **128**, 747–774.
- Pinson, P. and Kariniotakis, G. (2004). On-line assessment of prediction risk for wind power production forecasts. *Wind Energy*, **7**, 119–132.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. (1992). *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, second edition. Cambridge University Press, Cambridge.
- Raftery, A. E., Balabdaoui, F., Gneiting, T. and Polakowski, M. (2003). Using Bayesian model averaging to calibrate forecast ensembles. Technical Report no. 440, Department of Statistics, University of Washington. Available online at [www.stat.washington.edu/tech.reports](http://www.stat.washington.edu/tech.reports).

- Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics*, **23**, 470–472.
- Roulston, M. S. and Smith, L. A. (2002). Evaluating probabilistic forecasts using information theory. *Monthly Weather Review*, **130**, 1653–1660.
- Roulston, M. S., Kaplan, D. T., Hardenberg, J. and Smith, L. A. (2003). Using medium-range weather forecasts to improve the value of wind energy production. *Renewable Energy*, **28**, 585–602.
- Sansó, B. and Guenni, L. (2000). A nonstationary multisite model for rainfall. *Journal of the American Statistical Association*, **95**, 1089–1100.
- Shafer, M. A., Fiebrich, C. A., Arndt, D. S., Fredrickson, S. E. and Hughes, T. W. (2000). Quality assurance procedures in the Oklahoma mesonet. *Journal of Atmospheric and Oceanic Technology*, **17**, 474–494.
- Sharp, J. M. and Mass, C. (2002). Columbia Gorge gap flow – Insights from observational analysis and ultra-high-resolution simulation. *Bulletin of the American Meteorological Society*, **83**, 1757–1762.
- Sharp, J. M. and Mass, C. (200x). Columbia Gorge gap winds: Their climatological influence and synoptic evolution. *Weather and Forecasting*, in press. Available online at [www.atmos.washington.edu/~justin](http://www.atmos.washington.edu/~justin).
- Staley, D. O. (1959). Some observations of surface-wind oscillations in a heated basin. *Journal of Meteorology*, **16**, 364–370.
- Stein, M. L. (200x). Space-time covariance functions. *Journal of the American Statistical Association*, in press. Available online at [galton.uchicago.edu/~cises/research/tr.html](http://galton.uchicago.edu/~cises/research/tr.html).
- Székely, G. J. (2003).  $\mathcal{E}$ -Statistics: The energy of statistical samples. Technical Report no. 2003–16, Department of Mathematics and Statistics, Bowling Green State University.
- Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica*, **26**, 24–36.
- Wan, Y., Milligan, M. and Parsons, B. (2003). Output power correlation between adjacent wind power plants. *Journal of Solar Energy Engineering*, **125**, 551–555.
- Weigend, A. S., Mangeas, M. and Srivastava, A. N. (1995). Nonlinear gated experts for time series: Discovering regimes and avoiding overfitting. *International Journal of Neural Systems*, **6**, 373–399.
- Wikle, C. K. and Cressie, N. (1999). A dimension reduction approach to space-time Kalman filtering. *Biometrika*, **86**, 815–829.
- Zivot, E. and Wang, J. (2003). *Modeling Financial Time Series with S-Plus*, second edition. Springer, New York.