Calibrated Probabilistic Forecasting at the Stateline Wind Energy Center: The Regime-Switching Space–Time Method

Tilmann GNEITING, Kristin LARSON, Kenneth WESTRICK, Marc G. GENTON, and Eric ALDRICH

With the global proliferation of wind power, the need for accurate short-term forecasts of wind resources at wind energy sites is 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 into account all of the salient features of wind speed: alternating atmospheric regimes, temporal and spatial correlation, diurnal and seasonal nonstationarity, conditional heteroscedasticity, and non-Gaussianity. The RST method identifies forecast regimes at a 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 near the Stateline wind energy center in the U.S. Pacific Northwest. The RST point forecasts and distributional forecasts were accurate, calibrated, and sharp, and they compared favorably with predictions based on state-of-the-art time series techniques. This suggests that quality meteorological data from sites upwind of wind farms can be efficiently used to improve short-term forecasts of wind resources.

KEY WORDS: Continuous ranked probability score; Minimum continuous ranked probability score estimation; Predictive distribution; Spatiotemporal; Truncated normal; Weather prediction.

1. INTRODUCTION

Wind power is the fastest-growing energy source today. Globally, the annual average growth rate of wind energy has exceeded 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 attributed chiefly to advances in wind turbine design—which 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. Up-to-date information on the status of wind energy is available at the American Wind Energy Association website (*www.awea.org*). Nonetheless, 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 can help address these concerns. One requirement is for the development of improved prediction techniques on the 2-hour forecast horizon, a typical lead time for transmission scheduling, resource allocation, and generation dispatch.

Traditionally, short-term forecasts have used 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, Brownsword, and Kariniotakis (2003) surveyed 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. Consequently, we introduce the *regime-switching space–time* (RST) method, which 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 into account all of the salient features of wind speed: alternating atmospheric regimes, temporal and spatial correlation, diurnal and seasonal nonstationarity, conditional heteroscedasticity, and the nonnegativity of the predictand. The RST method identifies atmospheric regimes at the wind energy site and fits conditional predictive models 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 regime-dependent autoregressive time series model of Zwiers and von Storch (1990); to the space–time model for daily rainfall developed by Bardossy and Plate (1992), which uses multivariate time series models with parameters depending on the atmospheric circulation pattern; and to the hidden-variable models of Hughes, Guttorp, and Charles (1999) and Berliner, Wikle, and Cressie (2000), which apply stochastic regime switching to atmospheric science problems.

Whenever appropriate and feasible, the RST method uses geographically dispersed meteorological observations in the

Tilmann Gneiting is Associate Professor, Department of Statistics, University of Washington, Seattle, WA 98195 (E-mail: *tilmann@stat.washington.edu*). Kristin Larson is Senior Research Meteorologist (E-mail: *klarson@3tiergroup.com*) and Kenneth Westrick is Founder and CEO (E-mail: *kwestrick@3tiergroup.com*), 3 Tier Environmental Forecast Group, Inc., Seattle, WA 98121. Marc G. Genton is Associate Professor, Department of Statistics, Texas A&M University, College Station, TX 77843 (E-mail: *genton@stat.tamu.edu*). Eric Aldrich is Ph.D. Student, Department of Economics, Duke University, Durham, NC 27708 (E-mail: *ealdrich@gmail.com*). Part of this work was done while Gneiting was on sabbatical leave at the Soil Physics Group, Universität Bayreuth, Bayreuth, Germany. The authors are grateful to the Washington Technology Center (WTC) for supporting this work through the WTC Research & Technology Development program. Gneiting also acknowledges support by the National Science Foundation (NSF) under award 0134264 and by the DoD Multidisciplinary University Research Initiative program administered by the Office of Naval Research under grant N00014-01-10745. 3 Tier acknowledges the support of PPM Energy, Inc. The work of Genton was supported in part by NSF grant DMS-05-04896. 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 the editors, the referees, Jeff Baars, Barbara G. Brown, Vito Giarrusso, Ralf Kretzschmar, Clifford F. Mass, Adrian E. Raftery, Justin Sharp, Pascal Storck, Werner Stuetzle, and Toby White for helpful comments.

Journal of the American Statistical Association
September 2006, Vol. 101, No. 475, Applications and Case Studies
DOI 10.1198/016214506000000456

vicinity of the wind farm. Because changes in wind often propagate with the wind, it is possible to use upwind observations to detect precursors to changes in wind speed at the wind energy site. The RST technique introduces a new generation of forecast algorithms in that 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 (cdf's). Alexiadis, Dokopoulos, and Sahsamanoglou (1999) considered forecasts of wind speed and wind power at Thessaloniki Bay, Greece and showed that the use of off-site predictors can improve forecast accuracy; however, 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 regime and thereby addresses these concerns. Our focus on predictive distributions is in line with the recent surge of interest in probabilistic forecasts of wind resources (Roulston, Kaplan, Hardenberg, and Smith 2003; Bremnes 2004; Pinson and Kariniotakis 2004).

The remainder of the article is organized as follows. In Section 2 we introduce 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 U.S. Pacific Northwest. The Vansycle tower is adjacent to the Stateline wind energy center at the eastern end of the Columbia River Gorge, on the border between Oregon and Washington. Stateline is currently the world's largest single wind energy project. In Section 3 we describe various versions of the RST model for 2-hour forecasts of hourly average wind speed at Vansycle. We also discuss strategies for selecting predictor variables, estimating predictive models, and choosing adequate training periods. 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). In Section 4 we assess the predictive performance of the RST method compared with more conventional forecast techniques that we use as benchmarks, namely the persistence (no-change) forecast, 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 of hourly average wind speed at Vansycle had a root mean squared prediction error (RMSE) 28.8% lower than that of the persistence forecasts. Forecasts based on univariate time series techniques used on-site information only and reduced the persistence RMSE by up to 16.0%. The RST method entertains probabilistic forecasts in the form of truncated normal predictive distributions, which were well calibrated and sharp, with prediction intervals that were substantially shorter on average than prediction intervals based on univariate time series techniques. We close the article with a discussion in Section 5.

2. WIND DATA

2.1 Wind Patterns at Vansycle, Kennewick, and Goodnoe Hills

The meteorological data used herein were collected by Oregon State University at three meteorological towers in the U.S. 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. Table 1 provides site information, and Figure 1 shows the locations of the three stations along the Columbia River Gorge and the Oregon–Washington border. Goodnoe Hills lies 146 km west of Vansycle, and the tower at Kennewick is situated 39 km northwest of the tower at Vansycle.

The raw data record at the three sites is largely but not entirely complete. We imputed a minimal amount of isolated missing data ($<.03\%$ for calendar year 2003) by linear interpolation; adopted the quality assurance procedures proposed by Fiebrich and Crawford (2001), which led us to discard various stretches of data; and aggregated the original 10-minute averages to obtain time series of hourly average wind speed at Vansycle, Kennewick, and Goodnoe Hills. To obtain hourly time series of wind direction, we took the observations for the final 10-minute interval. The longest continuous hourly records at all three sites jointly comprise 55 and 279 days and range from September 4 to October 28, 2002 and from February 25 to November 30, 2003.

Figure 2 shows the monthly average wind speed and the relative frequency of westerly winds at the three sites in 2003. At Vansycle and Kennewick, wind speeds tended to peak in the cool season; at Goodnoe Hills, winds tended to be stronger in summer, much in line with the historical seasonal averages at Kennewick and Goodnoe Hills (*me.oregonstate.edu/ERRL/bpa_hist.html*). At all three stations and throughout the year, westerly winds were more frequent than easterly winds. Figure 3 shows the wind speeds at Vansycle, Kennewick, and Goodnoe Hills for the 7-day period beginning on August 3, 2002, when observations at Vansycle became available. Temporal and spatial correlation as well as a diurnal trend component can be observed. The autocorrelation and cross-correlation functions of wind speed at the three sites in August–December 2002 are shown in Figure 4. The autocorrelation functions demonstrate a steady decline with the temporal lag, except for

Table 1. Site Information for the Meteorological Towers at Vansycle, Kennewick, and Goodnoe Hills

Site	*Latitude*	*Longitude*	*Elevation*	*Anemometer height*	*Data record starting*
Vansycle	45°57′ 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

NOTE: Latitude and longitude are given in degrees and minutes; elevation and anemometer height are given in meters. Further information is available at *http://me.oregonstate.edu/ERRL/bpa_info.html*.

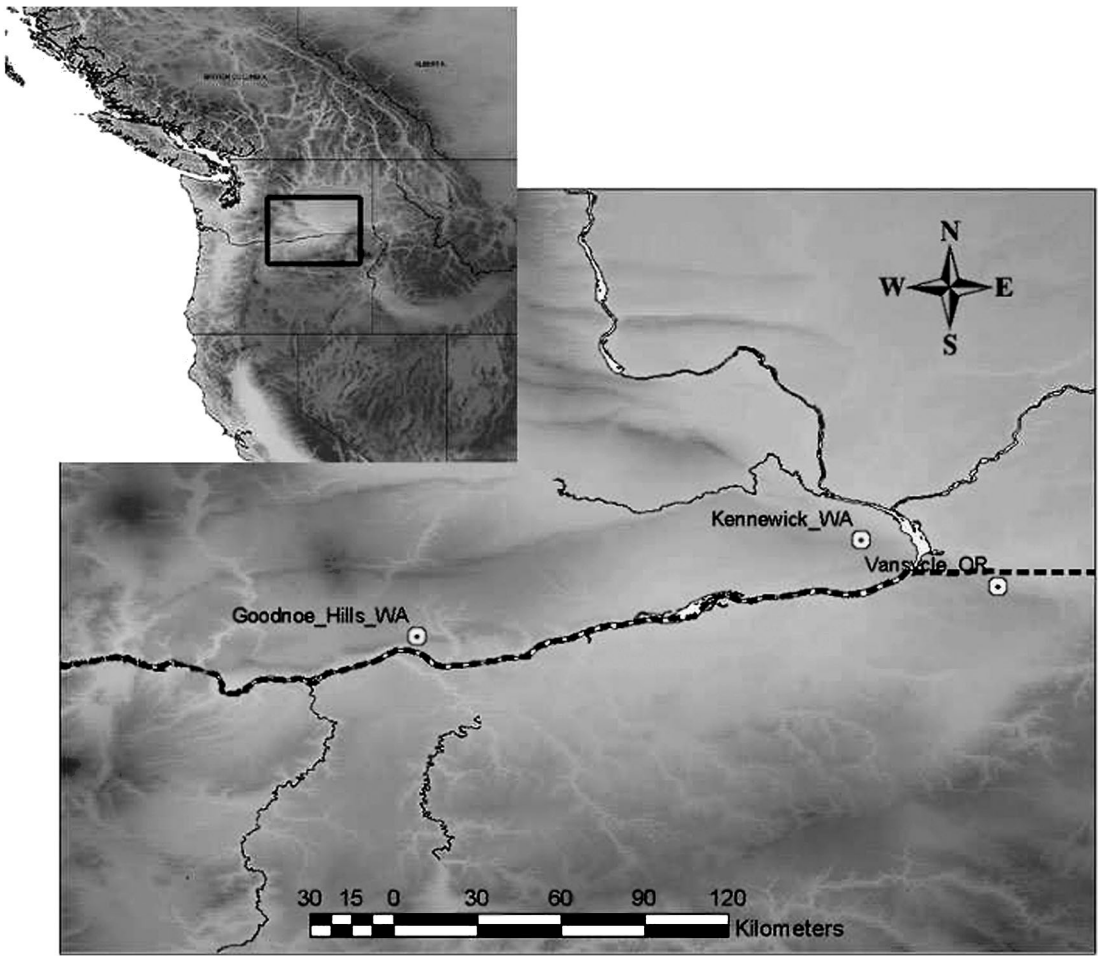

Figure 1. Locations of the Meteorological Towers at Vansycle, Kennewick, and Goodnoe Hills Along the Columbia River Gorge and the Oregon–Washington Border in the U.S. Pacific Northwest. The topography is indicated by different shading; light shading indicates low altitude, and dark shading indicates high altitude.

local maxima at lags that are multiples of 24. The local maxima correspond to the diurnal pattern, which is more pronounced at Vansycle and Goodnoe Hills than at Kennewick. A particularly noteworthy feature is the asymmetry of the cross-correlation function between Vansycle and Goodnoe Hills, which peaks at lags of 2 and 3 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; therefore, it seems plausible that wind speeds at Goodnoe Hills tend to lead those at Vansycle, with lead times of a few hours. Gneiting (2002), de Luna and Genton (2005), and Stein (2005) 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.

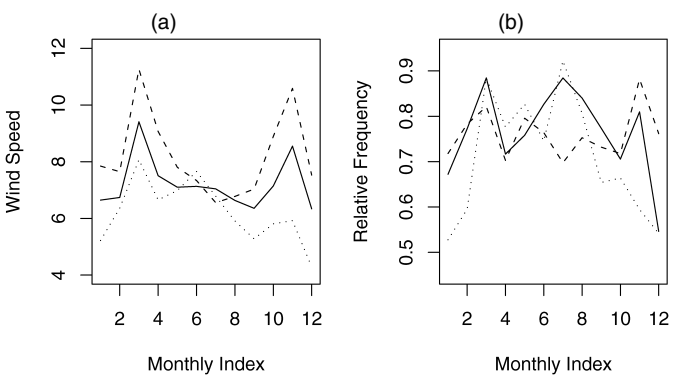

Figure 2. Seasonal Component of Wind Speed and Wind Direction at Vansycle (——), Kennewick (- - - -), and Goodnoe Hills (·····) in 2003. (a) Mean wind speed in m s^{-1}. (b) Relative frequency of westerly winds.

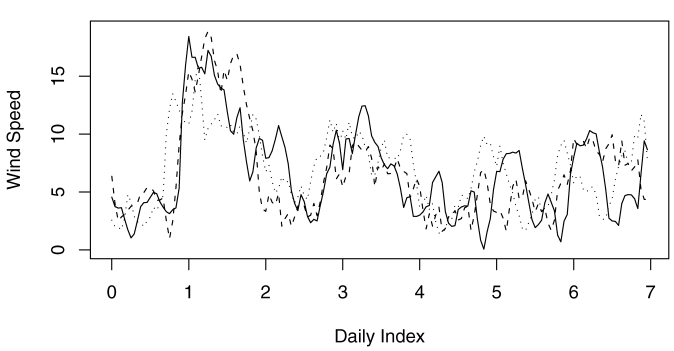

Figure 3. Wind Speed at Vansycle (——), Kennewick (- - - -), and Goodnoe Hills (·····) for the 7-Day Period Beginning on August 3, 2002, in m s^{-1}.

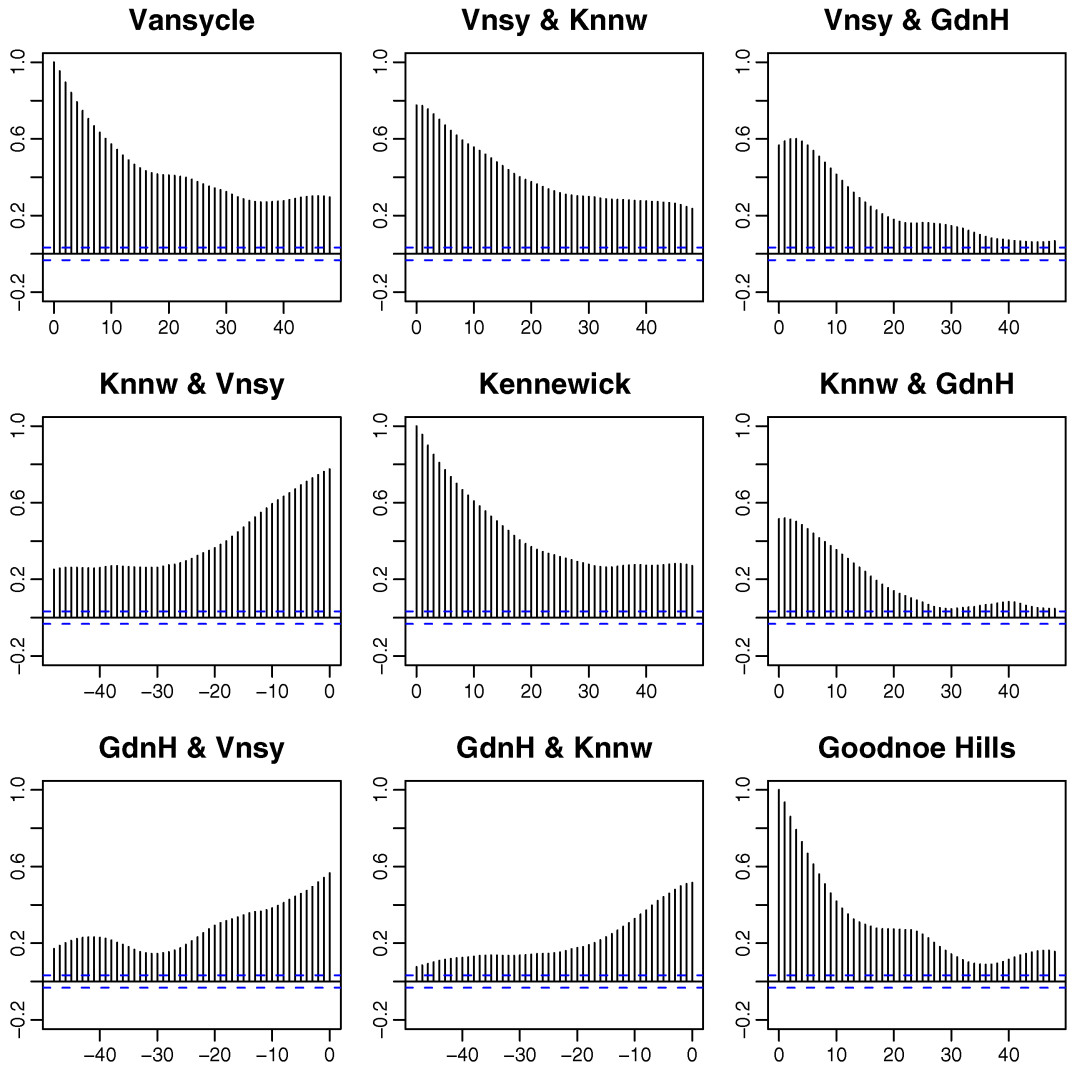

Figure 4. Autocorrelation and Cross-Correlation Functions of Hourly Average Wind Speed at Vansycle, Kennewick, and Goodnoe Hills in August–December 2002. Goodnoe Hills lies west of Kennewick, and Kennewick is located west of Vansycle. Positive lags indicate observations at the westerly station leading those at the easterly site.

2.2 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 (*http://me.oregonstate.edu/ERRL/goodnoe_hills.html*). 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. Because of the high terrain north and south of the Columbia Gorge, airflow through the gap is constrained to be parallel to the channel walls; hence winds are either generally westerly or generally easterly, depending on the pressure gradient across the Cascade Mountains. When the pressure is higher west of the Cascades, the wind is westerly; when the pressure is higher east of the Cascades, the wind is easterly. In the summertime, pressure is typically high over the Pacific Ocean, whereas strong heating in the arid zone east of the Cascades lowers the surface pressure east of the Gorge. Thus the prevailing wind is westerly, with a strong diurnal component (Sharp and Mass 2002). In the cool season, temperature inversions in the Columbia Basin east of the Cascades lead to frequent easterly flow through the Gorge, punctuated by strong westerly winds due to the passage of continental-scale weather systems (Sharp and Mass 2002, 2004).

The alternation of westerly and easterly 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 prevailing westerly wind suggests that the wind direction at the most westerly station, Goodnoe Hills, is a more useful indicator of the forecast regime than the wind direction at Vansycle. Figure 5 provides strong evidence in favor of this hypothesis. The boxplots in Figure 5(a) compare wind speeds at Vansycle 2 hours after observing westerly and easterly flow at Goodnoe Hills. 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 Figure 5(b). Hence we distinguish the westerly and the easterly forecast regime based on the current (westerly or easterly) wind direction at Goodnoe Hills.

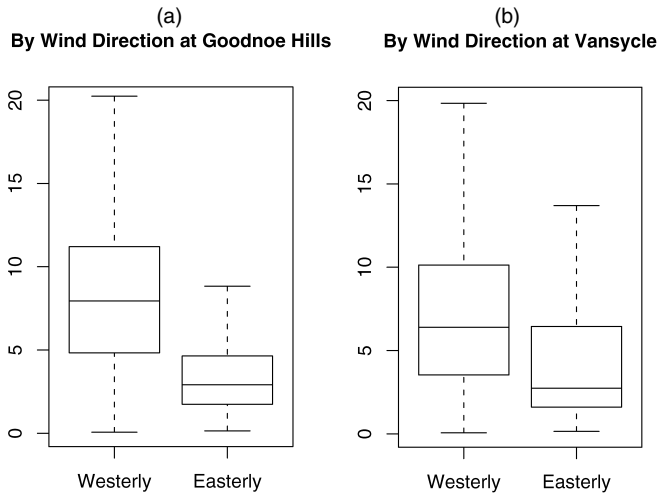

Figure 5. Boxplots of Hourly Average Wind Speed at Vansycle in August–December 2002 in Dependence on the Wind Direction at Goodnoe Hills (a) and Vansycle (b) 2 Hours Before, in m s^{-1}.

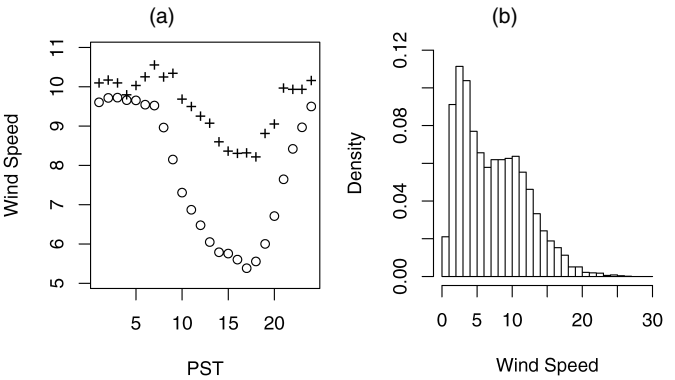

Figure 6. Wind Speed at Vansycle in 2003, in m s^{-1}. (a) Diurnal component in the warm season (April–September; ∘) and in the cool season (January–March and October–December; +), conditional on the wind direction at Goodnoe Hills being westerly 2 hours before. (b) Unconditional histogram of hourly average wind speed.

Hereinafter, we denote hourly average wind speed at Vansycle, Kennewick, and Goodnoe Hills at time (hour) t by V_t, K_t, and G_t. Table 2 shows the correlation between wind speed at Vansycle 2 hours ahead, V_{t+2}, and present and past wind speeds at the three sites, conditional on the forecast regime. The correlations for the westerly regime were computed from the data vectors $(V_{t+2}, V_t, V_{t-1}, V_{t-2}, K_t, K_{t-1}, K_{t-2}, G_t, G_{t-1}, G_{t-2})$ at the times t with westerly flow at Goodnoe Hills, and similarly for the easterly regime. The difference in the regime-specific spatiotemporal dependence structures is stark, in that the westerly regime shows substantially higher correlations between future wind speed at Vansycle and present and past wind speeds at Goodnoe Hills.

The wind record shows a diurnal cycle, particularly in the westerly regime and during the warm season. Figure 6(a) shows the average wind speed at Vansycle in dependence on Pacific Standard Time (PST), conditional on the wind direction at Goodnoe Hills being westerly 2 hours before. There was a pronounced diurnal component in summer 2003, with wind speeds peaking at night. Based on real-time mesoscale weather simulations conducted by Kenneth Westrick, 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. The unconditional histogram of wind speed at Vansycle is shown in Figure 6(b). The histogram is bimodal and can be approximated by a mixture density, with mixture components that can be associated with the forecast regimes. The predictive distributions that we use later are truncated Gaussian; this works well, because the conditional forecast errors are approximately normally distributed, as opposed to the wind speed values themselves.

3. THE REGIME–SWITCHING SPACE–TIME METHOD

RST models merge meteorological and statistical expertise to obtain calibrated 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 spatiotemporal 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), which allows for flow-dependent, nonstationary covariance structures. The computational requirements of the RST method are modest, and the technique can be readily implemented in real time. The model formulation is parsimonious and physically interpretable.

3.1 Overview

We give an overview of the RST approach to 2-hour forecasts of hourly average wind speed at Vansycle. In Section 2.2 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, because the regional climatology in the Pacific Northwest is well studied, and the literature distinguishes westerly and easterly flow regimes in the Columbia Gorge. At other wind energy sites, such as the Swiss Alps example of Kretzschmar et al. (2004), the identification of forecast regimes may be less straightforward and may involve substantial amounts of exploratory data analysis as well as local meteorological expertise.

Predictive distributions are frequently taken to be Gaussian even though the predictand is a nonnegative quantity (e.g., wind

Table 2. Correlation Between Hourly Average Wind Speed at Vansycle 2 Hours Ahead and Present and Past Values of Hourly Average Wind Speed at Vansycle, Kennewick, and Goodnoe Hills in August–December 2002, in the Westerly Regime and in the Easterly Regime

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}
Westerly regime	.86	.79	.72	.72	.69	.66	.61	.62	.60
Easterly regime	.91	.86	.82	.74	.70	.67	.30	.28	.27

speed or pollutant concentrations) or a nonnegative transformation thereof. To address the nonnegativity of wind speed, we use truncated normal predictive distributions with a cutoff at 0. Specifically, the truncated normal distribution $\mathcal{N}^+(\mu, \sigma^2)$ with location parameter μ and scale parameter $\sigma > 0$ has cdf $F_{\mu,\sigma}(x) = 0$ for $x < 0$ and

$$F_{\mu,\sigma}(x) = \frac{\Phi(\frac{x-\mu}{\sigma}) - \Phi(-\frac{\mu}{\sigma})}{1 - \Phi(-\frac{\mu}{\sigma})} \tag{1}$$

for $x \geq 0$. Compared with the respective Gaussian distribution, the truncated normal distribution is concentrated on the positive half-axis, with a probability density function (pdf) proportional to the Gaussian pdf. Quantiles are readily computed from (1).

Recent work by Cripps and Dunsmuir (2003) and Campbell and Diebold (2005) suggests the presence of conditional heteroscedasticity (i.e., high frequency changes of predictability) in meteorological time series. To investigate this, we entertain homoscedastic as well as conditionally heteroscedastic versions of the RST model. Section 2 noted seasonal changes in the strength of the diurnal component of wind speed, suggesting nondiurnal as well as diurnal versions of the predictive model. To summarize, we distinguish four variants of the RST model, all using truncated normal predictive distributions but with differing conditional means and conditional variance dynamics. The RST–N–H and RST–N–CH techniques do not attempt to model the diurnal component, and they use variance structures that are homoscedastic and conditionally heteroscedastic, respectively. The homoscedastic RST–D–H method and the conditionally heteroscedastic RST–D–CH technique fit a diurnal trend component in the westerly regime. We generally do not model the diurnal component in the easterly regime; we considered approaches of this type but found that these did not result in improved predictive performance.

3.2 The RST–N Technique

We now describe the RST–N–H and RST–N–CH techniques for 2-hour-ahead probabilistic forecasts of hourly average wind speed at Vansycle, V_{t+2}. Both methods use a truncated normal predictive distribution,

$$\mathcal{N}^+(\mu_{t+2}, \sigma^2_{t+2}), \tag{2}$$

with parameters μ_{t+2} and $\sigma_{t+2} > 0$, which we call the predictive center and predictive spread. The associated deterministic-style or point forecast for V_{t+2} is the mean, μ^+_{t+2}, of the predictive distribution, given by

$$\mu^+_{t+2} = \mu_{t+2} + \sigma_{t+2}\varphi\left(\frac{\mu_{t+2}}{\sigma_{t+2}}\right) \Big/ \Phi\left(\frac{\mu_{t+2}}{\sigma_{t+2}}\right), \tag{3}$$

where φ and Φ denote the pdf and the cdf of the standard normal distribution. We model the predictive center, μ_{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

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

where V_t, K_t, and G_t denote the wind speed at Vansycle, Kennewick, and Goodnoe Hills at time t. This is akin to a vector autoregressive (VAR) time series model (Brockwell and Davis 1991, chap. 11; de Luna and Genton 2005) or an autoregressive (AR)-distributed lag scheme (Zivot and Wang 2003, sec. 6.4). However, the predictive distribution is truncated 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 μ_{t+2} as

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

The model selection process that led to the conditional linear models in (4) and (5) [and also in (9)] was guided by an analysis of the wind data in August–December 2002, before our evaluation period. For each forecast regime, we considered multiple linear regression models for V_{t+2} on the predictor variables in Table 2. We started from the simplest model and added predictor variables in a stepwise forward search until no further improvement in the Bayesian information criterion (BIC) was obtained. The simplicity of the linear model (5) for the conditional mean dynamics in the easterly regime is striking, although in line with the parsimony principle for time series forecasting.

It remains to model the predictive spread, σ_{t+2}. The homoscedastic RST–N–H technique assumes that the predictive spread is constant. The RST–N–CH method allows for conditional heteroscedasticity by modeling

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

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

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

reflects the magnitude of recent changes in wind speed at the three sites. In both regimes, the terms for Goodnoe Hills can be omitted from (7) with no notable degradation in the predictive performance of the RST–N–CH technique.

3.3 Modeling the Diurnal Component: The RST–D Technique

The RST–D technique fits 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 regime, the RST–D technique does not differ from the RST–N approach.

Hence, suppose that the current regime is westerly. We use the truncated normal predictive distribution (2) with predictive center μ_{t+2} and predictive spread σ_{t+2}, specified below. The predictive mean, μ^+_{t+2}, is found from (3). At each of the three sites, we fit the trigonometric function,

$$D_t = d_0 + d_1 \sin\left(\frac{2\pi t}{24}\right) + d_2 \cos\left(\frac{2\pi t}{24}\right) + d_3 \sin\left(\frac{4\pi t}{24}\right) + d_4 \cos\left(\frac{4\pi t}{24}\right), \tag{8}$$

which 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. The predictive center, μ_{t+2}, is then modeled as

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

where D_{t+2} is the fitted diurnal component at Vansycle and

$$\mu_{t+2}^r = a_0 + a_1 V_t^r + a_2 V_{t-1}^r + a_3 K_t^r + a_4 K_{t-1}^r + a_5 G_t^r \quad (9)$$

is a linear function of present and past values of the residual series at the three sites. The RST–D–H technique is homoscedastic and assumes that the predictive spread 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,

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

with coefficients b_0 and b_1 that are constrained to be nonnegative. Again, the terms for Goodnoe Hills can be omitted from the volatility value with no notable degradation in predictive performance.

3.4 Minimum Continuous Ranked Probability Score Estimation and Choice of Training Period

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, Raftery, Westveld, and Goldman (2005b) proposed the novel technique of minimum continuous ranked probability score (CRPS) estimation for estimating predictive distributions, which we use here.

To describe minimum CRPS estimation, we require a number of basic facts on scoring rules. Scoring rules assign numerical scores to probabilistic forecasts, based on the predictive distribution F and the value x that materializes. A well-known scoring rule is the logarithmic score (Good 1952; Roulston and Smith 2002), that is, the negative of the logarithm of the predictive density evaluated at the observation. The CRPS provides a more robust alternative. Specifically, if F is the predictive cdf and x materializes, then the CRPS is defined as

$$\mathrm{crps}(F, x) = \int_{-\infty}^{\infty} \big(F(y) - \mathbb{1}(y \geq x)\big)^2 \, \mathrm{d}y, \quad (10)$$

where $\mathbb{1}(y \geq x)$ attains the value 1 when $y \geq x$ and 0 otherwise. Gneiting and Raftery (2004) discussed alternative representations and showed that the CRPS can be reported in the same unit as the observations. Furthermore, the CRPS is proper, in the sense that the forecaster minimizes the expected score for an observation drawn from F if he or she issues the probabilistic forecast F rather than $G \neq F$. Probabilistic forecasts can be assessed and ranked by comparing the average value,

$$CRPS = \frac{1}{n} \sum_{i=1}^{n} \mathrm{crps}(F_i, x_i), \quad (11)$$

of the CRPS over an evaluation set. If each F_i is a point forecast, then the CRPS value reduces to the mean absolute prediction error (MAPE), and the CRPS and the MAPE can be directly compared. Of course, the smaller the CRPS, the better.

The classical maximum likelihood technique can be interpreted as minimizing the logarithmic score for the training data, when expressed as a function of the parameter values. In minimum CRPS estimation, we express the CRPS value (11) 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 truncated normal, and a tedious but straightforward calculation shows that if $x \geq 0$, then

$$\begin{aligned} &\mathrm{crps}(\mathcal{N}^+(\mu, \sigma^2), x) \\ &\quad = \sigma \Phi\left(\frac{\mu}{\sigma}\right)^{-2} \\ &\qquad \times \left(\frac{x-\mu}{\sigma} \Phi\left(\frac{\mu}{\sigma}\right) \left(2\Phi\left(\frac{x-\mu}{\sigma}\right) + \Phi\left(\frac{\mu}{\sigma}\right) - 2 \right) \right. \\ &\qquad \left. + 2\varphi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\frac{\mu}{\sigma}\right) - \frac{1}{\sqrt{\pi}} \Phi\left(\sqrt{2}\frac{\mu}{\sigma}\right) \right). \end{aligned} \quad (12)$$

For RST–D–CH forecasts in the westerly regime, for instance, we find the minimum of the CRPS value (11) for the training data as a function of the parameters in (4) and (6). This needs to be done numerically, and we use the Broyden–Fletcher–Goldfarb–Shanno algorithm (Press, Teukolsky, Vetterling, and Flannery 1992, sec. 10.7) as implemented in the R language and environment (*www.cran.r-project.org*). The algorithm is iterative, and starting values based on past experience usually give good solutions. In the sliding window implementation described later, we use the most recent estimate as initial value; to start off, we find initial values using standard linear regression. Convergence to a global minimum cannot be guaranteed, and the solution reached can be sensitive to the initial value. We investigated this empirically, and found that in numerical experiments with randomly generated initial values, the solutions were essentially independent of the starting values, except for cases in which the initial value was outrageously inadequate (results not shown here, to conserve space). However, the right side of (12) becomes numerically unstable as $\mu/\sigma \to -\infty$, which can misinform the optimization algorithm. The problem disappears when the right side of (12) is set to a large positive value if μ/σ is less than a threshold (here -4) that corresponds to physically unreasonable predictive distributions.

What training period should be used to estimate the conditional predictive models? The lack of a multi-year meteorological record at Vansycle, which is typical for wind energy sites, suggests using 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, in that shorter periods can adapt rapidly to seasonally varying atmospheric conditions, whereas longer training periods reduce the statistical variability in the estimation. There is no automatic way to find the optimal length, and we studied the effect of the different training periods empirically, similarly to experiments

reported by Gneiting et al. (2005b) and Raftery, Gneiting, Balabdaoui, and Polakowski (2005). We observed that summary measures of predictive performance, such as the RMSE of the RST point forecasts and the CRPS value for the RST forecast distributions, decrease sharply for training periods shorter than 30 days, stay about constant for training periods of 30–60 days, and tend to increase thereafter (results not shown, to conserve space). The average width of the 90% prediction interval increases with the length of the training period but is about constant for training periods of 40–50 days. Hence, there appear to be substantial gains in increasing the training period beyond 30 days, and training periods of 30–60 days generally seem adequate. The results reported hereinafter use a sliding 45-day training period to estimate the conditional predictive models. We intend to revisit these issues as multiyear records of wind speed at Vansycle become available, thereby allowing for the inclusion of training data from previous years, to better address seasonal effects.

4. PREDICTIVE PERFORMANCE

Here we assess the predictive performance of the RST method. We describe more conventional prediction techniques that we use as benchmarks, evaluate the point forecasts and the probabilistic forecasts, and give an explicit example of the RST forecasts for the 7-day period beginning on June 28, 2003.

4.1 Reference Forecasts

The classical reference forecast in the meteorological literature is the *persistence* (no-change) forecast, $\widehat{V}_{t+2} = V_t$, which uses the most recent observation at hand as a point forecast. Clearly, the shorter the forecast lead time, the more competitive the persistence forecast.

Brown et al. (1984) proposed using AR time series models in wind speed and wind power forecasting. They applied the square root transform to the series of hourly average wind speed at Goodnoe Hills, fitted and extracted a diurnal trend component, and modeled the residual component as an AR process. This approach has found widespread use since then (Giebel et al. 2003). For the forecasts at Vansycle, we experimented with the square root transform and a more general transform proposed by Allcroft and Glasbey (2003), but found that forecasts based on time series models for the nontransformed wind speed series performed best. The homoscedastic AR–N–H technique uses 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, chap. 7). The AR–D–H and AR–D–CH methods estimate and extract a diurnal trend component of the form (8) and proceed as outlined earlier. To obtain Gaussian predictive distributions from the time series models, we follow Brown et al. (1984) and Brockwell and Davis (1991, sec. 5.4). 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 of 30–50 days generally seem adequate.

In contrast to the univariate time series techniques that rely on on-site predictors, the RST method uses observations from the meteorological towers at Kennewick and Goodnoe Hills. Multivariate time series techniques (Brockwell and Davis 1991, chap. 11) offer a more traditional way of incorporating information from Kennewick and Goodnoe Hills into the predictive model. Recently, de Luna and Genton (2005) proposed a VAR specification 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 into account the spatiotemporal dependence structure and borrows its strength from the Box and Jenkins (1976) strategy for univariate time series. We implemented the de Luna and Genton (2005) approach in two variants, using a maximal order of 6 for the VAR 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 (8) at each of the three sites and applies the VAR to the residual series. Currently available code for this technique provides point forecasts only.

VAR techniques lead to predictive models that resemble the schemes in (4), (5), and (9), but they do not identify and differentiate forecast regimes. However, the method of de Luna and Genton (2005) applies a model selection strategy to the successive fits in the sliding window implementation; hence the predictor variables may vary from hour to hour. For the RST technique, the parameter estimates vary from hour to hour, but the predictive models themselves are fixed.

4.2 Assessment of Point Forecasts

We now assess the predictive performance of the forecast techniques. As noted earlier, the longest continuous data record at Vansycle, Kennewick, and Goodnoe Hills simultaneously extends from February 25 through November 30, 2003. In view of the sliding 45-day training period for the RST forecasts, our evaluation period extends from May 1 through November 30, 2003, and we report the various scores month by month. This allows us to study the consistency of the results and to assess seasonal effects.

In evaluating point forecast performance, we ignore the distinction between the homoscedastic and conditionally heteroscedastic versions of the AR–N, AR–D, RST–N, and RST–D forecasts and report the results only for the conditionally heteroscedastic schemes. Typically, the forecasts based on the conditionally heteroscedastic model showed slightly smaller RMSE than the respective forecasts based on the homoscedastic model, but the differences were very small. For example, the RMSE for point forecasts using the predictive mean (3) of the RST–D–CH forecast distributions in (all of) May–November 2003 was 1.787, and the respective number for the RST–D–H forecasts was 1.789. The maximal difference in the monthly scores was .007.

Table 3 gives the RMSE values for the various methods. The RST forecasts consistently had lower RMSE than the VAR forecasts; the latter outperformed the forecasts based on the univariate AR models; and the forecasts based on the univariate AR models consistently had lower RMSE than the persistence forecast. In May–October 2003, the RST–D forecast outperformed all others and had RMSE 19.2%, 21.0%, 28.8%, 21.7%, 18.8%, and 13.2% less than the persistence forecast. In November 2003, the top rank was tied with the RST–N technique. The

Table 3. RMSE for 2-Hour Point Forecasts of Hourly Average Wind Speed at Vansycle in May–November 2003, in m s^{-1}

	Forecast						
	May	*June*	*July*	*Aug.*	*Sep.*	*Oct.*	*Nov.*
Persistence	2.14	1.97	2.37	2.27	2.17	2.38	2.11
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	1.76	1.58	1.78	1.83	1.81	2.08	**1.87**
RST–D	**1.73**	**1.56**	**1.69**	**1.78**	**1.77**	**2.07**	**1.87**

improved performance of the RST–N technique toward the end of the evaluation period is not surprising, given the lack of a pronounced diurnal component during the cool season. The diurnal VAR technique, VAR–D, was the closest competitor to the RST forecasts, but the RST–D method consistently had lower RMSE.

Table 4 compares RST–N–CH point forecasts using competing sets of predictor variables in the nondiurnal westerly conditional linear model for the predictive center. The first entry shows the RMSE value in (all of) May–November 2003 for the model (4) chosen by the variable selection algorithm described in Section 3.4. This value, 1.886, ignores the forecasts in the easterly regime and hence is not directly comparable to the values in Table 3. The subsequent entries suggest that the addition or (to a lesser extent) the omission of a predictor variable has little effect on the RMSE, unless information from a site is completely excluded, which results in major forecast deterioration. Similar experiments for the diurnal westerly and easterly cases (results not tabulated, to conserve space) generally support this interpretation.

4.3 Assessment of Probabilistic Forecasts

In assessing probabilistic forecasts, we are guided by the paradigm that probabilistic forecasting strives to maximize the sharpness of the predictive distributions under the constraint of calibration (Gneiting, Balabdaoui, and Raftery 2005a). Calibration, the statistical consistency between the predictive distributions and the observations, is a joint property of the forecasts and the values or events that materialize. Sharpness, the spread of the predictive distributions, is a property of the forecasts only. The more concentrated the predictive distributions, the sharper the forecasts, and the sharper, the better, subject to calibration.

Table 4. RMSE for 2-Hour RST–N–CH Point Forecasts of Hourly Average Wind Speed at Vansycle in the Westerly Regime in May–November 2003 Using Competing Sets of Predictor Variables, in m s^{-1}

Predictor variables			*RMSE*
V01	K01	G0	1.886
V012	K01	G0	1.888
V01	K012	G0	1.880
V01	K01	G01	1.885
V0	K01	G0	1.918
V01	K0	G0	1.913
V01	K01		2.042

NOTE: The predictor variables are listed symbolically; for instance, the first entry refers to the conditional linear model (4) chosen by our variable selection algorithm.

Table 5. CRPS for Probabilistic 2-Hour Forecasts of Hourly Average Wind Speed at Vansycle in May–November 2003, in m s^{-1}

	Forecast						
	May	*June*	*July*	*Aug.*	*Sep.*	*Oct.*	*Nov.*
AR–N–H	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–H	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–H	.97	.86	.99	.99	.99	1.10	1.01
RST–N–CH	.97	.86	.99	.99	.99	**1.08**	**1.00**
RST–D–H	.96	**.85**	**.94**	.96	.97	1.09	1.01
RST–D–CH	**.95**	**.85**	**.94**	**.95**	**.96**	**1.08**	**1.00**

The CRPS value (11) provides a summary measure of probabilistic forecast performance, in that it addresses calibration as well as sharpness. The respective results are given in Table 5. 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 nondiurnal methods. The RST predictive cdf's showed substantially lower CRPS than the predictive cdf's using univariate AR techniques. For each month in the evaluation period, the RST–D–CH forecast had the lowest CRPS, with the top rank tied in June, July, October, and November.

The key tool in assessing calibration is the histogram of the probability integral transform (PIT) (Dawid 1984; Diebold, Gunther, and Tay 1998). The PIT is simply the value that the predictive cdf attains at the observation, that is, a number between 0 and 1. 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 consider the average width of the 90% central prediction interval.

PIT histograms for the predictive cdf's based on the AR–D and RST–D techniques are shown in Figures 7 and 8. The histograms for the AR forecasts were hump shaped, thereby indicating overdispersed predictive distributions, with prediction intervals that are unnecessarily wide. The effect was slightly less pronounced for the conditionally heteroscedastic AR forecast but still prominent. Gneiting et al. (2005b) argued that using maximum likelihood plug-in estimates tends to result in

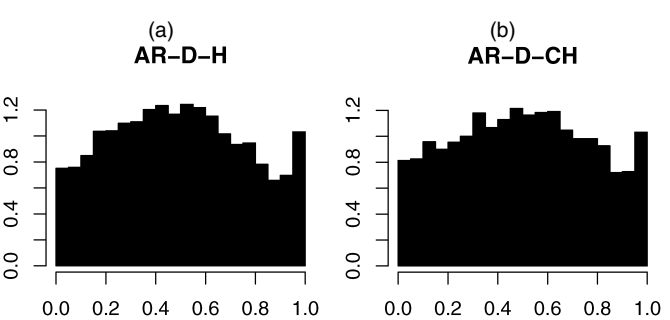

Figure 7. PIT Histograms for (a) AR–D–H and (b) AR–D–CH Predictive cdf's of Hourly Average Wind Speed at Vansycle in May–November 2003.

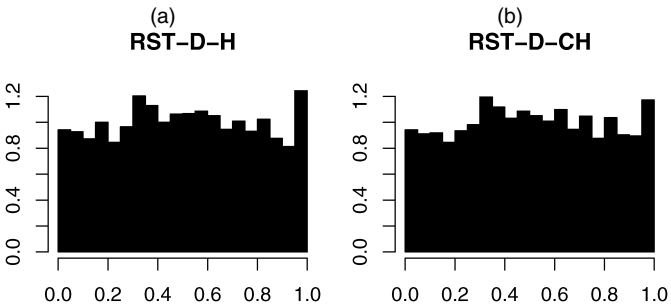

Figure 8. PIT Histograms for (a) RST–D–H and (b) RST–D–CH Predictive cdf's of Hourly Average Wind Speed at Vansycle in May–November 2003.

overdispersed predictive distributions, and a similar effect may apply here. The RST technique uses minimum CRPS estimation, and the respective PIT histograms were nearly uniform. The average width of the 90% central prediction interval was 5.36 m s^{-1} for the RST–D–H forecast and 5.44 m s^{-1} for the RST–D–CH forecast. Among the AR forecasts, the AR–D–CH technique was sharpest, with an average width of 6.36 m s^{-1}. Hence the RST–D–CH prediction intervals were at least 14.5% shorter on average than the AR intervals while maintaining adequate coverage, thereby achieving significant reductions in forecast uncertainty.

4.4 Example

To summarize the foregoing results, the RST–D–CH technique outperformed all of the others during the warm season. Our test period did not include the winter months, but the results for October and November suggest that in the cool season the RST–N–CH forecast may perform best.

Figure 9 shows 2-hour forecasts of hourly average wind speed at Vansycle for the 7-day period beginning on June 28, 2003, using the RST–D–CH approach. The solid line shows the point forecast, that is, the mean (3) of the truncated normal predictive distribution; the broken lines correspond to the lower and upper end points of the 90% central prediction interval. The observed values of hourly average wind speed are shown as black circles; forecasts issued in the westerly regime are indicated by the marks at the top. The westerly regime was prevalent and showed a pronounced diurnal component, with wind speeds peaking at night. In the easterly regime, wind speeds were generally lower, there was less evidence of a diurnal trend component, and the prediction intervals were considerably shorter than in the westerly regime.

5. DISCUSSION

We have introduced the RST method, which merges meteorological and statistical expertise to obtain accurate and calibrated, fully probabilistic forecasts of wind resources. The model formulation is parsimonious yet takes into account all of the salient features of wind speed: alternating atmospheric regimes, temporal and spatial correlation, diurnal and seasonal nonstationarity, conditional heteroscedasticity, and the nonnegativity of the predictand. The RST method identifies forecast regimes at the wind energy site and uses geographically dispersed meteorological observations in the vicinity of the wind farm as off-site predictors. The technique is tailored to single-site meteorological prediction problems, which are common in short-range wind energy forecasting and in the context of weather derivatives (Campbell and Diebold 2005). It does not address probabilistic forecasts on spatial grids that are essential in other types of applications, such as highway salting, aircraft and ship routing, and sailing course design (Gel, Raftery, and Gneiting 2004; Cripps, Nott, Dunsmuir, and Wikle 2005).

We applied the RST technique to 2-hour forecasts of hourly average wind speed at the Vansycle ridge in the U.S. Pacific Northwest, in the immediate vicinity of the Stateline wind energy center. In July 2003, for instance, the RST forecast had an RMSE 28.8% lower than that of the persistence forecast. The RST method provides probabilistic forecasts in the form of truncated normal predictive distributions, which were well calibrated and sharp. The RST prediction intervals were substantially shorter on average than the prediction intervals derived from univariate time series techniques, and still showed adequate coverage, as reflected by a nearly uniform PIT histogram. During the warm season, the RST–D–CH variant of the

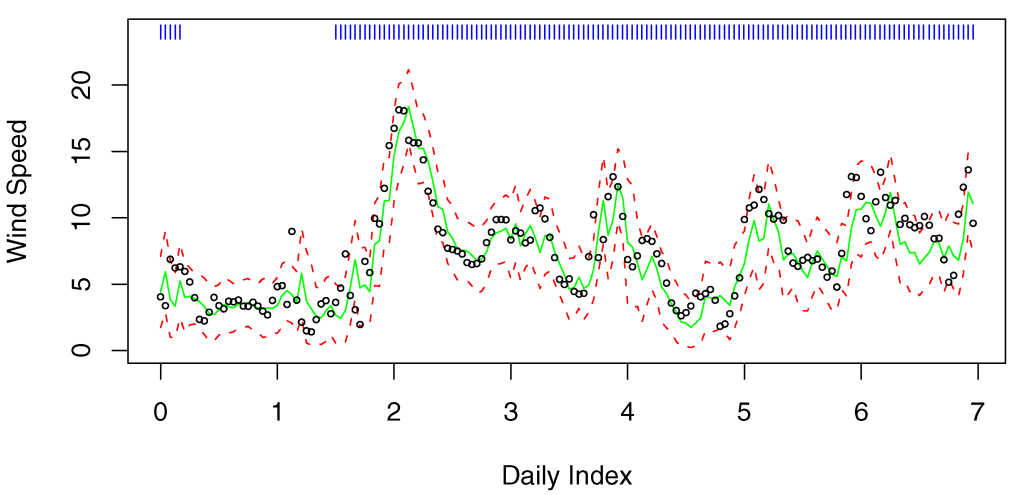

Figure 9. 2-Hour RST–D–CH Forecasts of Hourly Average Wind Speed at Vansycle for the 7-Day Period Beginning on June 28, 2003, in m s^{-1}. The point forecast (3) is shown by the solid line, along with the 90% central prediction interval bordered by the broken lines. The observed wind speeds are shown as circles, and forecasts issued in the westerly regime are identified by the marks at the top.

RST technique performed best. This method models the diurnal component of wind speed and takes into account conditional heteroscedasticity. For forecasts in the cool season, we recommend using the RST–N–CH technique, which 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 historical wind records and the BIC to select fixed sets of predictor variables for the regime-specific linear models (4), (5), and (9) that describe the conditional mean dynamics. Similar to the de Luna and Genton (2005) implementation of the VAR technique, it may be possible to apply variable selection algorithms within a sliding window implementation, thereby eliminating the need for an earlier wind record and allowing for structural change in the conditional linear models. However, the resulting improvements in forecast performance, if any, are likely to be incremental.

The conditional linear models allow for other types of predictor variables, in addition to current and past values of wind speed. 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 2004). Other meteorological variables or numerical weather prediction (NWP) model output could be considered as well. The benefits are likely to be small, unless meteorological observations from stations east of Vansycle were to become available.

In our record of hourly average wind speed at Vansycle, values that were exactly 0 did not occur. In other types of situations, vanishing wind speeds may occasionally be observed, and the predictive distributions may require a point mass at 0. A convenient and elegant way of entertaining this is through the use of cutoff normal (rather than truncated normal) predictive distributions, as done by Sansó and Guenni (2000) and Allcroft and Glasbey (2003), and an earlier implementation of the RST technique (Gneiting, Larson, Westrick, Genton, and Aldrich 2004). At Vansycle, forecasts using truncated normal predictive distributions are consistently superior in terms of RMSE and CRPS values.

As noted in Section 1, NWP forecasts are not competitive at the forecast lead time that we consider here. However, NWP forecasts and mesoscale ensemble prediction systems (Grimit 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.

In conclusion, 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, even though 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 greater acceptance of wind power and less reliance on nonrenewable energy sources throughout the world.

Note Added in Proof

When checking proofs, we realized that a note on sensor problems for the Goodnoe Hills wind speed data in July–December 2003 had been posted at *http://me.oregonstate.edu/ERRL/previous_data.html*. A corrected data set was generated. This affects the graph for Goodnoe Hills in Figure 2(a), in which the July–December averages ought to be slightly higher (by less than 1 m s^{-1}) than shown. The RST forecasts use wind speed at Goodnoe Hills as a predictor variable in the westerly regime only. Here, the corrected data set is immaterial to the prediction problem, since real-time forecasts are necessarily based on the original, uncorrected predictor variables that we used in the case study.

[Received September 2004. Revised July 2005.]

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.

Berliner, L. M., Wikle, C. K., and Cressie, N. (2000), "Long-Lead Prediction of Pacific SSTs via Bayesian Dynamic Modeling," *Journal of Climate*, 13, 3953–3968.

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*, San Francisco: Holden-Day.

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* (2nd ed.), New York: Springer-Verlag.

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. (2005), "Weather Forecasting for Weather Derivatives," *Journal of the American Statistical Association*, 100, 6–16.

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.

Cripps, E., Nott, D., Dunsmuir, W. T. M., and Wikle, C. (2005), "Space–Time Modelling of Sydney Harbour Winds," *Australian & New Zealand Journal of Statistics*, 47, 3–17.

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. (2005), "Predictive Spatio-Temporal Models for Spatially Sparse Environmental Data," *Statistica Sinica*, 15, 547–568.

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.

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, 575–590.

Giebel, G., Brownsword, R., and Kariniotakis, G. (2003), "The State-of-the-Art in Short-Term Prediction of Wind Power," Deliverable Report D1.1, Project Anemos, available at *http://anemos.cma.fr/download/ANEMOS_D1.1_StateOfTheArt_v1.1.pdf*.

Gneiting, T. (2002), "Nonseparable, Stationary Covariance Functions for Space–Time Data," *Journal of the American Statistical Association*, 97, 590–600.

Gneiting, T., Larson, K., Westrick, K., Genton, M. G., and Aldrich, E. (2004), "Calibrated Probabilistic Forecasting at the Stateline Wind Energy Center: The Regime-Switching Space–Time (RST) Method," Technical Report 464, University of Washington, Dept. of Statistics.

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2005a), "Probabilistic Forecasts, Calibration and Sharpness," Technical Report 483, University of Washington, Dept. of Statistics.

Gneiting, T., and Raftery, A. E. (2004), "Strictly Proper Scoring Rules, Prediction, and Estimation," Technical Report 463, University of Washington, Dept. of Statistics.

Gneiting, T., Raftery, A. E., Westveld, A. H., and Goldman, T. (2005b), "Calibrated Probabilistic Forecasting Using Ensemble Model Output Statistics and Minimum CRPS Estimation," *Monthly Weather Review*, 133, 1098–1118.

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 Result of a Mesoscale Short-Range Ensemble 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.

Huang, H.-C., and Hsu, N.-J. (2004), "Modeling Transport Effects on Ground-Level Ozone Using a Non-Stationary Space–Time Model," *Environmetrics*, 15, 251–268.

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.

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* (2nd ed.), Cambridge, U.K.: Cambridge University Press.

Raftery, A. E., Gneiting, T., Balabdaoui, F., and Polakowski, M. (2005), "Using Bayesian Model Averaging to Calibrate Forecast Ensembles," *Monthly Weather Review*, 133, 1155–1174.

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.

Roulston, M. S., and Smith, L. A. (2002), "Evaluating Probabilistic Forecasts Using Information Theory," *Monthly Weather Review*, 130, 1653–1660.

Sansó, B., and Guenni, L. (2000), "A Nonstationary Multisite Model for Rainfall," *Journal of the American Statistical Association*, 95, 1089–1100.

Sharp, J., 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.

——— (2004), "Columbia Gorge Gap Winds: Their Climatological Influence and Synoptic Evolution," *Weather and Forecasting*, 19, 970–992.

Stein, M. L. (2005), "Space–Time Covariance Functions," *Journal of the American Statistical Association*, 100, 310–321.

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*, New York: Springer-Verlag.

Zwiers, F., and von Storch, H. (1990), "Regime-Dependent Autoregressive Time Series Modeling of the Southern Oscillation," *Journal of Climate*, 3, 1347–1363.