LOCATION ROBUST ESTIMATION OF PREDICTIVE WEIBULL PARAMETERS IN SHORT-TERM WIND SPEED FORECASTING

Matthew J. Holland   *Kazushi Ikeda*

Graduate School of Information Science
Nara Institute of Science and Technology, Ikoma, Japan
matthew-h@is.naist.jp, kazushi@is.naist.jp

ABSTRACT

From turbine control systems at wind farms to extreme weather early-warning systems, short-term probabilistic wind speed forecasts are seeing widespread use in industrial applications. Successful modern forecast methods, often Weibull-based, have been shown to be extremely sensitive to even minor changes in location. We contend that this lack of robustness stems not from model selection, but rather the parameter estimation methods used, and propose a new proper scoring rule to be dynamically minimized. Tested on a weather array spanning the islands of Japan, we verify both superior short-term forecasting performance and model fit of the proposed method over all standard references, and empirically confirm the desired location robustness.

***Index Terms*—** Robust wind speed forecasting, density forecasting, proper scoring rule

1. INTRODUCTION

Wind speeds near the surface of the earth are inherently intermittent. Demand for reliable forecasts and stochastic wind information exists in numerous applications, including extreme weather early-warning systems [1, 2], supply planning for power grids using wind energy [3], and predictive control systems for optimizing turbine power output [4] or reducing turbine vibration [5]. Depending on the task, the required forecast "horizon" length may differ; for longer-term forecasts between 12–72 hours, numerical models capturing the dynamics of mesoscale wind flows have proven successful [6], using statistical post-processing to attempt spatial localization [7]. At shorter time scales, especially shorter than 6 hours, data requirements and computational cost become prohibitive, and statistical models not constrained to physical processes are the standard approach for spatially local short-term forecasts [8].

The research problem in this study concerns the robustness of statistical wind speed models (cf. [9]) to forecast location in the 1–5 hour case. Robustness to location naturally implies robustness to anemometer height and altitude, presence of nearby buildings and obstructions, as well as gradient and roughness of surrounding terrain. If such information is available, it can be effectively reflected in probabilistic wind speed models [10], though this is clearly a site-specific fine-tuning. Such models naturally do not generalize well off-site, but even in the case of very simple models without such fine-tuning, strong performance at one location by no means guarantees the performance at nearby sites [11]. The *ad hoc* manner in which methods in the literature have been proposed leads to two major problems. First, in general sufficient historical data and detailed terrain information cannot be assumed available, rendering key assumptions in past models invalid. Second, since successful models end up differing fundamentally site-to-site, a cross-location comparison of model performance becomes difficult, and insights into the stochastic character of wind speed are drastically limited.

In our study, we contend that a key factor limiting forecaster robustness is the method of parameter estimation used to generate each forecast, rather than the specifics of the model used. Here we propose a novel estimation method, using a new parsimonious metric to be dynamically minimized, which requires only the first and second moments of the predictive distribution, and has desirable theoretical properties as a cost function (section 3). Using observational data from the Japan Meteorological Agency (section 2), we verify both the location robustness of our method, and the location sensitivity of standard alternative methods, over sites nationwide capturing a variety of geographic and climatic conditions. The results are applicable for the highly general case of a Weibull-based forecaster with no fine-tuning. Key experimental results and subsequent analysis are given in section 4, and discussion with a look ahead in section 5 concludes the paper.

2. DATA AND EXPERIMENT INFORMATION

The data to be used in this research comes from the Japan Meteorological Agency (JMA) observation network AMeDAS, whose historical data is made public. Having aggregated the observations from over 1100 sites on the network, for this study we limit the focus to "target" sites (i.e., at locations at which to forecast) with both reasonably high average wind speeds as well as sufficiently long sets of contiguous data without missing or uncertain values over calendar year 2012. The minimum average annual wind speed allowed is 2.5m/s, and of the 52704 10-minute observations, we set a raw maximum missing value rate at 0.1%, and in addition require contiguous data free of missing values no shorter than 250 days worth of measurements.

The resulting 44 high-quality data sites (Fig. 1) span a large geographical and climatic range, from Okinawa and Kyushu through central Honshu, and up to Hokkaido. Half of these sites had contiguous sequences of length 52704 (zero missing values), and the average length was 46666, or roughly 325 days. In this study we do not explicitly look at robustness to seasonal conditions, and as the time-series are of sufficient length and forecasts are carried out independently for any given site, every site need not cover precisely the same time period.

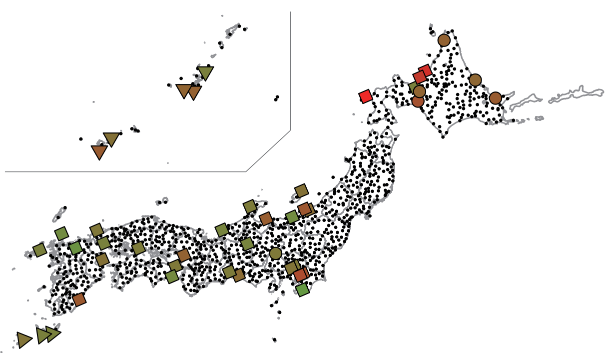

Fig. 1. AMeDAS network with filtered targets denoted by large symbols. Shapes denote annual average temperature (C): circles are < 8, diamonds are [8,18), triangles are ≥ 18 over 2012–2013. The green-red colour gradient represents normalized standard deviation of wind speed (m/s), with minimum of 1.5 and maximum 3.8.

3. PREDICTIVE PARAMETER ESTIMATION METHOD

We now elucidate the details of the problem and our proposed solution. At present time t, our task is to forecast wind speed x_{t+k} for forecast horizon $k > 0$. Assuming observations $\mathbf{X}_t = (\mathbf{x}_{t,1}, \ldots, \mathbf{x}_{t,m})$, where $m \geq 1$ and the index m includes both wind speeds at remote sites and other local variables, we model $x_{t+k} \sim F(\boldsymbol{\theta}(\mathbf{X}_t; \mathbf{w}))$. Conditional on data $\mathbf{X}_t$, determining $\mathbf{w}$ specifies the parametric model F. Previous studies tend to focus on finding a particularly good F which may take the form of a truncated Normal, log-Normal, Gamma, Rayleigh, or most commonly a (uni/bi-modal) Weibull distribution [12], and parameter model $\boldsymbol{\theta}(\cdot; \mathbf{w})$. In these works, the method of estimating $\mathbf{w}$ is in most cases determined by computational convenience dependent on the model. If site-specific assumptions (logit-Normal, etc.) allow it, recursive methods typically minimizing squared forecast error (e.g., Kalman filters) [8] may be utilized, though in the case of the task for general locations, log likelihood [13] and continuous ranked probability score (CRPS) [2, 14] can be considered wind industry standards.

We take the approach of starting with the task of finding an appropriate metric to minimize, and then evaluating it for important model classes. Consider a measurable space $(\Omega, \mathcal{A})$ where $\mathcal{A}$ is a σ-algebra of subsets of Ω. Let $\mu : \mathcal{A} \to \mathbb{R}$ be a measure and $\mathbb{P} = \{\mu : \mu(\Omega) = 1\}$. Then any non-empty $\mathcal{P} \subset \mathbb{P}$ is a model, and for any $f : \Omega \to \mathbb{R}$, and $f^{\pm} := (|f| \pm f)/2$, if we simply require that one or more of f^{+}, f^{-} be integrable with respect to every $P \in \mathcal{P}$, then any such f is called $\mathcal{P}$-quasi-integrable [15]. Given a model $\mathcal{P}$ then, we call a function $\mathcal{L} : \mathcal{P} \times \Omega \to \mathbb{R}$ a cost function if it is $\mathcal{P}$-quasi-integrable. We seek a metric which is compatible with a large class $\mathcal{P}$ and has nice theoretical qualities as a cost function. We say that if

$$\mathbb{E}_Q[\mathcal{L}(Q,\omega)] \leq \mathbb{E}_Q[\mathcal{L}(P,\omega)], \quad \forall\, P, Q \in \mathcal{P}, \tag{1}$$

where of course $\omega \sim Q$, then $\mathcal{L}$ is *proper* with respect to $\mathcal{P}$. *Strict* propriety holds if equality in (1) holds $\iff P = Q$. For probability measures on $\Omega = \mathbb{R}^n$, let $\boldsymbol{\mu}_P$ denote the mean vector, and $\mathbf{V}_P$ the covariance (or dispersion) matrix of some $P \in \mathcal{P}$. Our proposed metric is for the $n = 1$ case, and is given as

$$\mathcal{J}(P,\omega) = \frac{(\omega - \mu_P)^2}{V_P} + \log(V_P). \tag{2}$$

Clearly we only require from P that its first and second moments are finite. Next we show the propriety of this metric following an argument in the spirit of [16].

We consider the case of general $n \geq 1$ here. Let $\mathbb{S}_n^{+}$ be the set of positive definite $n \times n$ symmetric matrices on $\mathbb{R}$. If $\mathcal{P}_n = \{P : \mathbf{V}_P \in \mathbb{S}_n^{+}\}$, for $\lambda \in [0,1]$, standard integration properties give us for $R = \lambda P + (1-\lambda)Q$ where $P, Q \in \mathcal{P}_n$, that $\mathrm{Cov}_R[i,j] = \lambda \mathrm{Cov}_P[i,j] + (1-\lambda)\mathrm{Cov}_Q[i,j]$ that is $V_R = \lambda V_P + (1-\lambda)V_Q$. Positive definiteness of $\mathbf{V}_R$ follows immediately from $\mathbf{V}_P, \mathbf{V}_Q \in \mathbb{S}_n^{+}$, and thus $\mathcal{P}_n$ is convex.

Define $K : \mathbb{S}_n^{+} \to \mathbb{R}$ by $K(\mathbf{V}) := \log(\det \mathbf{V})$. Basic results for convex functions (cf. [17]) are used as follows. Note the epigraph of $-K$ is clearly non-empty and by positive definiteness the argument passed to the $\log$ will be positive, and as such $-K$ is a proper convex function. By theorem 6.4 of [17], one can easily confirm any $\alpha > 0$ is in the relative interior of $\mathrm{dom}(\log) = \mathbb{R}_{+}$. Thus using the natural extension of theorem 23.4 [17] from vectors to square matrices, and the differentiability of K, we have for $\mathbf{W} \in \mathbb{S}_n^{+}$

$$K(\mathbf{V}) \leq K(\mathbf{W}) + \mathrm{tr}\left(\partial_{\mathbf{W}} K(\mathbf{W})(\mathbf{V} - \mathbf{W})\right), \tag{3}$$

for all $\mathbf{V} \in \mathbb{S}_n^{+}$. This follows from using the subgradient inequality for convex $-K$, and homogeneity of the derivative to switch the inequality direction into the supergradient case. For some $P \in \mathcal{P}_n$, if we define a metric

$$\mathcal{M}(P,\boldsymbol{\omega}) = (\boldsymbol{\omega} - \boldsymbol{\mu}_P)^T \mathbf{V}_P^{-1}(\boldsymbol{\omega} - \boldsymbol{\mu}_P) + K(\mathbf{V}_P) - \mathrm{tr}(\partial_{\mathbf{W}} K(\mathbf{W})\mathbf{V}_P)$$

it is straightforward to confirm for $Q \in \mathcal{P}$

$$\mathbb{E}_Q[\mathcal{M}(P,\omega)] = (\boldsymbol{\mu}_Q - \boldsymbol{\mu}_P)^T \mathbf{V}_P^{-1}(\boldsymbol{\mu}_Q - \boldsymbol{\mu}_P) + K(\mathbf{V}_P) + \mathrm{tr}(\mathbf{V}_P^{-1}(\mathbf{V}_Q - \mathbf{V}_P)),$$

as $\partial_{\mathbf{W}} K(\mathbf{W}) = (\mathbf{W}^{-1})^T = \mathbf{W}^{-1}$ by symmetry (cf. [18]). Comparison of the eigenvalues of $\mathbf{V}_P$ with its inverse give positive semi-definiteness of $\mathbf{V}_P^{-1}$, and thus using the inequality (3) it follows that

$$\mathbb{E}_Q[\mathcal{M}(Q,\omega)] = K(\mathbf{V}_Q) \leq \mathbb{E}_Q[\mathcal{M}(P,\omega)].$$

This is precisely the propriety of metric $\mathcal{M}$, and since $\mathrm{tr}(\mathbf{V}_P^{-1}\mathbf{V}_P) = n$, this clearly implies that (1) holds for $\mathcal{J}(P,\omega)$ on $\mathcal{P}_n$, namely as the special case of this more general result where $n = 1$. We thus have propriety of $\mathcal{J}$ for all distributions with first and second moments defined, and strict propriety may be shown for distributions characterized by their first two moments. The utility of (1) is clear, given observations $\mathbf{X}_t$, as by the LLN for N large enough we need only minimize $\sum_{i=0}^{N-1} \mathcal{J}(F(\boldsymbol{\theta}_{t-i}), x_{t+k-i})/N$.

As noted above, applicability of the metric to the Weibull distribution is an important requirement. For a Weibull with shape $\kappa(\mathbf{w}_\kappa)$ and scale $\lambda(\mathbf{w}_\lambda)$ denoted $W(\lambda,\kappa)$, we see

$$\mathcal{J}(W,\omega) = \frac{(\omega - \lambda\Gamma_1)^2}{\lambda^2(\Gamma_2 - \Gamma_1^2)} + 2\log(\lambda) + \log(\Gamma_2 - \Gamma_1^2),$$

where $\Gamma_i := \Gamma(1 + i/\kappa)$. The gradient is readily obtained, and for parameter initialization in the case of $\kappa(\mathbf{w}_\kappa), \lambda(\mathbf{w}_\lambda)$ being linear combinations of input features, we initialize the two intercept terms by a basic moment method, in which we seek the root of

$$f(\kappa) = \frac{\sqrt{\Gamma_2 - \Gamma_1^2}}{\Gamma_1} - \frac{\widehat{\sigma}_N}{\bar{x}_N}$$

and use $\hat{\kappa}$ to solve $\lambda = \bar{x}_N/\Gamma_1$. The results motivating this initializer come from [19]. The remaining parameters are uniformly set to small positive values. We may then use fast quasi-Newton optimizers such as the BFGS method as implemented in R [20].

4. EXPERIMENTAL RESULTS AND ANALYSIS

Our goal here is to make a direct comparison of competing parameter estimation methods, leaving the model fixed for each forecasting task, and to evaluate their robustness to location-dependent factors, across all time horizons of interest (1–5h). We fix the models at time t to $W(\mathbf{w}_\lambda^T \mathbf{x}_{t,l}, \kappa_t)$,

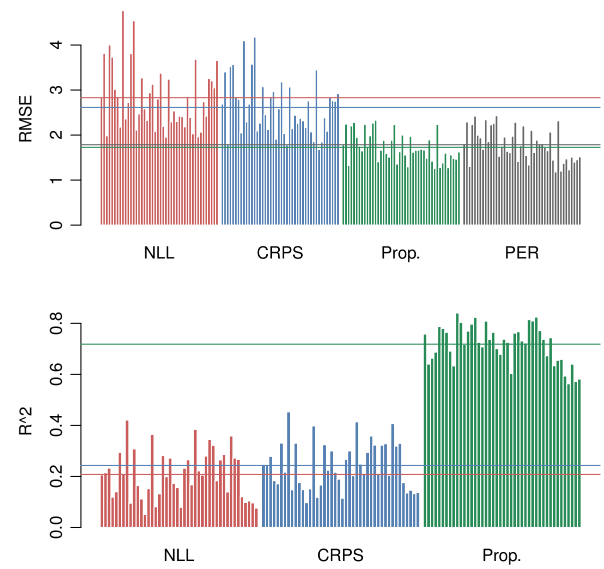

Fig. 2. RMSE (top) and R^2 (bottom) at each site for 5h forecast. Lines denote averages taken over sites per method.

where $\mathbf{x}_{t,l} = (1, x_t, \ldots, x_{t-l})$. The order of the AR model is set to $l = k + 3$, where $k = 1, \ldots, 5$. Reference probabilistic methods use this model and negative log likelihood (NLL) or CRPS (cf. [21]) for parameter estimation. Persistence (PER), a deterministic reference, is simply a random walk forecast, $\widehat{x}_{t+k} = x_t$. Spatial models capturing observations at surrounding sites have been shown to be effective for potentially fine-tuning more basic models [21], though inclusion of such site-dependent information will dilute any conclusions regarding location robustness of estimation methods, and thus we use the benchmark model as stated. Several parameter initializers were used, including the Newby moment approach [19], fixed values (at 1), random values (1+$|\mathcal{N}(0, 0.25)|$), past values (at t use estimates from $t-1$), and ML estimates for the intercept terms, and the best-performing results (by forecast error) were used as references.

The task is carried out at each site (Fig. 1) and each horizon 1–5h separately, with no cross-referencing of forecasts. Window length N is fixed at 15 days worth of observations. Standard model selection algorithms (AIC/BIC, KL-IC, etc.) can be used to automate this by-site. Beginning with Fig. 2, we see forecast accuracy and model fit R^2 (cf. [12]) over all the sites at the 5 hour horizon, though as we will see, the same trends hold at all horizons. Clearly models using the proposed estimation method performed far better; in fact, at all horizons the output is as good or better than persistence on average, even with such simple models free of fine-tuning. Strong calibration is further reinforced by the PIT histograms given in Fig. 3, where the proposed method is far more uniform than the under-dispersed rivals.

Next we look more explicitly at the location robustness.

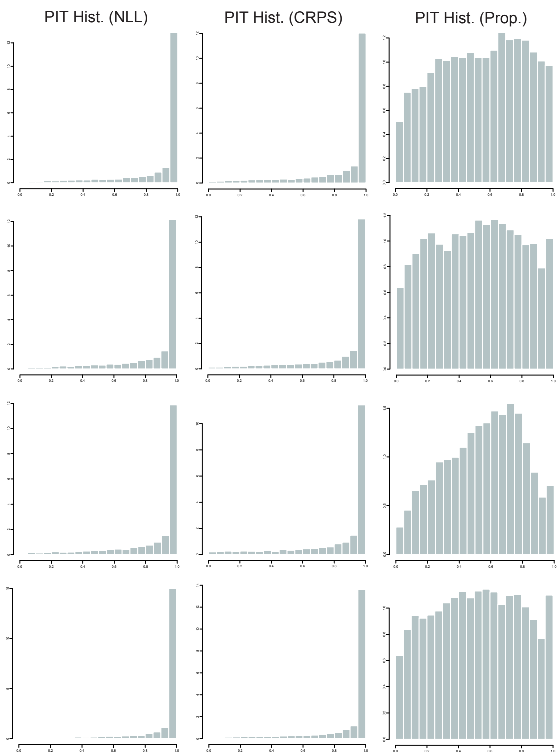

Fig. 3. Histogram of $F(x_{t+k}; \widehat{\boldsymbol{\theta}}_t)$ at four randomly selected sites for 5h task; each column denotes a method.

In Table 1, we have the standard deviation of several metrics, taken over all the test sites, at the shortest and longest horizons. Volatility difference is simply the difference in sample variance between the generated point forecasts and the true observations. The proposed estimation method results in notably smaller variance over sites, particularly in accuracy metrics, providing perhaps the clearest indicator of superior location robustness. Related measurements in Fig. 4 look at RMSE as a function of site wind speed and temperature. First, we note that the reference performance drops at windier sites, while under the proposed method the change is minimal. As well, the rate of increase in error as the horizon length increases is slower. We see that the references tend to be particularly sensitive to colder sites, whereas the proposed approach sees minimal variability. Such trends were observed for error as a function of anemometer height, and across the board for MAE and R^2 as well.

5. DISCUSSION AND CONCLUSIONS

In this study we proposed a new method for parameter estimation in short-term wind speed forecasting, and verified the desired robustness to location empirically over a large observation array. This robustness applies to both forecast accuracy and model fit, and was found to outperform strong standard references. The experimental results and their interpretation are clear, but why does the proposed method succeed? The first numerator is a proxy for forecast error, and variance in the denominator prevents "overconfident" forecasts. On the other hand, the log variance term ensures that density is not over-dispersed, and in practice works to balance point forecast accuracy and long-run model fit. MLE fits a joint distribution under independence assumptions, and CRPS a distribution function; depending on the model both may not optimize accuracy, or be without a mechanism to stop overfitting on the sample window.

Future lines of work will involve verification of the performance seen here over remaining distributions of interest, and subsequently to investigate whether analogous results can be found for site-adaptive model selection algorithms. As well, quantitative investigation of the scalability of candidate estimation methods to high model complexity or very short horizons (< 1min), may lead to beneficial new insights.

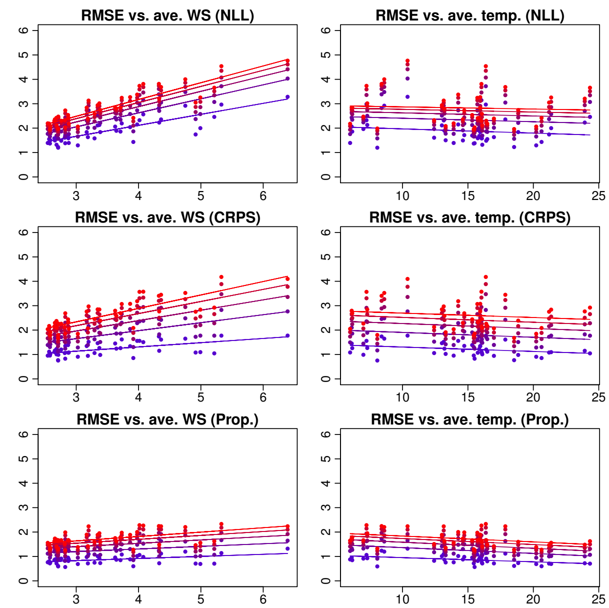

Fig. 4. Forecast accuracy as a function of average annual wind speed. Blue-red gradient indicates horizons 1–5h.

	RMSE	*MAE*	*Vol diff.*	R^2
NNL	0.502	0.415	0.289	0.085
CRPS	0.280	0.219	0.172	0.051
Prop.	**0.168**	**0.126**	**0.109**	**0.045**
NNL	0.702	0.617	0.369	0.094
CRPS	0.619	0.529	0.330	0.092
Prop.	**0.319**	**0.243**	**0.221**	**0.074**

Table 1. Standard deviation of each metric, taken over all sites, by method. First three rows for 1h, latter for 5h task.

6. REFERENCES

[1] François Lalaurette, "Early detection of abnormal weather conditions using a probabilistic extreme forecast index," *Quarterly Journal of the Royal Meteorological Society*, vol. 129, no. 594, pp. 3037–3057, 2003.

[2] S Lerch and T L Thorarinsdottir, "Comparison of non-homogeneous regression models for probabilistic wind speed forecasting," *Tellus A*, vol. 65, pp. 21206, 2013.

[3] Gregor Giebel, Richard Brownsword, George Kariniotakis, Michael Denhard, and Caroline Draxl, "The state-of-the-art in short-term prediction of wind power: A literature overview," Tech. Rep., ANEMOS.plus, 2011.

[4] Boubekeur Boukhezzar and Houria Siguerdidjane, "Nonlinear control with wind estimation of a DFIG variable speed wind turbine for power capture optimization," *Energy Conversion and Management*, vol. 50, no. 4, pp. 885–892, 2009.

[5] Andrew Kusiak, Zhe Song, and Haiyang Zheng, "Anticipatory control of wind turbines with data-driven predictive models," *Energy Conversion, IEEE Transactions on*, vol. 24, no. 3, pp. 766–774, 2009.

[6] Peter Lynch, "The origins of computer weather prediction and climate modeling," *Journal of Computational Physics*, vol. 227, no. 7, pp. 3431–3444, 2008.

[7] Adrian E Raftery, Tilmann Gneiting, Fadoua Balabdaoui, and Michael Polakowski, "Using Bayesian model averaging to calibrate forecast ensembles," *Monthly Weather Review*, vol. 133, no. 5, pp. 1155–1174, 2005.

[8] Pierre Pinson, "Very-short-term probabilistic forecasting of wind power with generalized logit–normal distributions," *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, vol. 61, no. 4, pp. 555–576, 2012.

[9] Tilmann Gneiting, Adrian E Raftery, Anton H Westveld III, and Tom Goldman, "Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation," *Monthly Weather Review*, vol. 133, no. 5, pp. 1098–1118, 2005.

[10] Takeshi Ishihara and Atsushi Yamaguchi, "Prediction of the extreme wind speed in the mixed climate region by using Monte Carlo simulation and measure-correlate-predict method," *Wind Energy*, 2014.

[11] Amanda S Hering and Marc G Genton, "Powering up with space-time wind forecasting," *Journal of the American Statistical Association*, vol. 105, no. 489, pp. 92–104, 2010.

[12] Eugene C Morgan, Matthew Lackner, Richard M Vogel, and Laurie G Baise, "Probability distributions for offshore wind speeds," *Energy Conversion and Management*, vol. 52, no. 1, pp. 15–26, 2011.

[13] JV Seguro and TW Lambert, "Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 85, no. 1, pp. 75–84, 2000.

[14] Tilmann Gneiting, Kristin Larson, Kenneth Westrick, Marc G Genton, and Eric Aldrich, "Calibrated probabilistic forecasting at the stateline wind energy center: The regime-switching space-time method," *Journal of the American Statistical Association*, vol. 101, no. 475, pp. 968–979, 2006.

[15] Tilmann Gneiting and Adrian E Raftery, "Strictly proper scoring rules, prediction, and estimation," *Journal of the American Statistical Association*, vol. 102, no. 477, pp. 359–378, 2007.

[16] A Philip Dawid and Paola Sebastiani, "Coherent dispersion criteria for optimal experimental design," *Annals of Statistics*, pp. 65–81, 1999.

[17] R Tyrrell Rockafellar, *Convex Analysis*, Number 28. Princeton University Press, 1997.

[18] Jan R Magnus and Heinz Neudecker, "Matrix differential calculus with applications in statistics and econometrics," 1988.

[19] MJ Newby, "The properties of moment estimators for the Weibull distribution based on the sample coefficient of variation," *Technometrics*, vol. 22, no. 2, pp. 187–194, 1980.

[20] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013.

[21] Matthew J. Holland and Kazushi Ikeda, "Forecasting in wind energy applications with site-adaptive Weibull estimation," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2014)*, Florence, Italy, 2014.