

# FORECASTING IN WIND ENERGY APPLICATIONS WITH SITE-ADAPTIVE WEIBULL ESTIMATION

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 optimal supply decisions to anticipatory control systems, wind-based energy applications rely heavily upon accurate, local, short-term forecasts of future wind speed. Recent studies have shown continuous ranked probability score (CRPS) minimizing models with Gaussian assumptions to be effective for well-researched sites where those assumptions are appropriate. We consider the more general case where Gaussianity is not assumed and access to historical data may be constrained. Deriving a CRPS expression for a minimum Extreme Value distribution, we use it to propose a site-adaptive Weibull-based CRPS-minimizing model, which is tested and shown to perform better than both deterministic and probabilistic reference models on a ground-based array of weather observation sites in northern Japan.

*Index Terms*— Wind energy, wind power forecasting, continuous ranked probability score, wind turbine control

## 1. INTRODUCTION

Wind-based power projects are subject to significant uncertainty due to the intermittent nature of the underlying natural resource. This uncertainty impacts both economic and engineering decisions such as turbine generator torque and blade pitch control [1], turbine maintenance frequency [2], production planning and supply contract quantities in power markets [3]. Predictions of the future wind state play a vital role in all such applications, and depending on the task the forecast requirement ranges from a few minutes in advance [4], to several hours or even days ahead [5]. Many anticipatory turbine control tasks make use of wind forecasts on the order of 1–6 hours, and here we elect to focus on this short-term time scale.

The first of two major forecasting paradigms is based on physical models of atmospheric flows. Known as numerical weather prediction (NWP) and extremely computationally intensive, the efficacy of NWP on 36- and 72-hour time scales has improved significantly with increasing computational capabilities [6], though spatial and temporal resolution

is limited, and both data-related and computational costs may be a barrier to independent power providers. The second paradigm is statistically-oriented and is feasible for making local, short-term forecasts [7]. Such methods take a more mechanical approach to the problem with varying degrees of expert knowledge present. Past work on short-term forecasting ranges from autoregressive time-series [8] and Kalman filters [9] to more recently Bayesian structural break models [4] and other probabilistic methods which capture spatio-temporal information in intuitive ways [5].

In particular, the regime-switching space-time (RST) model from Gneiting *et al.* (2006) [5] and trigonometric direction diurnal (TDD) model from Hering and Genton (2010) [10] have been shown to outperform persistence, vector AR, and other references on the multi-hour timescale in a well-researched region of the American Pacific northwest, and provide the context for our work. In real wind energy applications, constraints on historical data access will often limit the expert knowledge available for a given forecast site, and thus we propose a new CRPS-minimizing Weibull density forecast model which aims to be site-adaptive, requiring minimal background knowledge of a particular observation site or its neighbours. In this study we utilize a ground-based weather observation network in Hokkaido, Japan.

In section 2 we discuss the prediction problem formulated in this study, and review experimental details. In section 3 we derive a closed-form CRPS expression for the Generalized Extreme Value distribution which lets us model a predictive Weibull, and introduce the model-building and learning algorithms. An evaluation of the results of our experiment using both deterministic and probabilistic metrics and a subsequent discussion are the focus of section 4. We end the paper in section 5 with conclusions drawn from this study and discuss the outlook for future lines of work.

## 2. DATA AND EXPERIMENT DETAILS

The Japan Meteorological Agency (JMA) operates a network of over 1300 land-based sites capable of measuring and

recording meteorological variables, called *Automated Meteorological Data Acquisition System* (AMeDAS), of which approximately 840 sites (one per 21km<sup>2</sup> on average) are equipped with the ability to measure wind speed and direction. Historical AMeDAS data is made public by the JMA online (embedded in HTML tables), which is easily acquired and processed using an algorithmic approach. In this study we utilize a subset of the full dataset, namely 10-minute and 1-hour wind speed and direction observations for a cluster of seven sites in northwestern Hokkaido.

As a general problem formulation, at present time  $t$ , assuming constant sampling period the basic task is to generate a forecast of wind speed  $y_{t+k}$ , where  $k > 0$  is the forecast lag, under the following data constraints: we have no topographical information, only a small set of past wind velocity/direction data ( $< 1$  year), no prior information about the stochastic wind profile of any forecast sites, but at each time step we will collect new observations. In this study we sample at 10-minute intervals, pre-process hourly averages for direct input into the model, and design and test a model which forecasts  $k$ -step ahead hourly averages for any specified site in the cluster.

Our model uses a fixed size training set of  $M > 0$  time-steps worth of observations, which is updated and used to re-train the model after each time step. If we let  $\mathcal{S} = \{1, 2, \dots, S\}$  be an index of discrete forecast sites in our region of interest, then at time  $t$  at each  $s \in \mathcal{S}$  we have  $\mathcal{D}_{s,t}(t) = (\mathbf{x}_{s,T_s}, \dots, \mathbf{x}_{s,t})$ , and our set of past observations as of  $t$  will be simply  $\mathcal{D}_t = \bigcup_{s \in \mathcal{S}} \mathcal{D}_{s,t}$ . In advance, a fixed “window” of most recent  $M$  time-steps per site is specified; thus to start with, when data is scarce this  $M$  may have a small upper bound. Given an  $M$  value, in this experiment the model is built over the first 100 days’ worth of forecasts in 2006, and results are evaluated over the subsequent 100 days.

### 3. WIND FORECASTING MODEL

Both unary point-forecasts as well as density forecasts which specify a predictive distribution for future wind speed are feasible solutions to our forecasting problem, and while for many end-users, a single numerical forecast may be easier to understand and interpret, valuable information about the probabilistic profile of the wind can only be captured by a density forecast [11]. If wind speed is denoted simply  $x$ , at time  $t$  we specify  $\boldsymbol{\theta}_t$  and model a  $k$ -step ahead parametrized distribution  $F(x_{t+k} | \boldsymbol{\theta}_t)$ .

The *continuous ranked probability score* (CRPS) is defined for predictive distribution  $F(\boldsymbol{\theta})$  and realized observation  $x \in \mathbb{R}$  as

$$\text{CRPS}(F_{\boldsymbol{\theta}}, x) = \int_{-\infty}^{\infty} (F(t|\boldsymbol{\theta}) - \mathbf{1}[t \geq x])^2 dt,$$

and is a proper scoring rule which has been shown to generally perform better and to be more robust in meteorological

applications than maximum likelihood inference [12]. Its use as a loss function for parameter estimation has been advocated by Gneiting *et al.* (2005) [13], and CRPS-minimization inference was used to optimize truncated Normal predictive distributions of the RST and TDD models. Prior analysis led these authors to the conclusion that a truncated Normal was a sufficient approximation. In general however, non-Gaussianity is observed frequently [8], and the issue faced with CRPS inference is a distinct lack of closed-form expressions for many distributions. A large body of work in regions around the world indicates that the Weibull distribution tends to describe the stochastic character of wind velocity well [14, 15, 16], and here we seek a predictive Weibull using CRPS minimization inference. The Weibull is defined with CDF

$$\mathcal{W}(x|\kappa, \lambda) = 1 - \exp\left(-\left(\frac{x}{\lambda}\right)^\kappa\right),$$

with shape and scale parameters  $\kappa > 0$  and  $\lambda > 0$ . We take an indirect approach to a CRPS expression for a predictive Weibull  $\mathcal{W}(\kappa, \lambda)$  by detouring through the minimum Generalized Extreme Value (GEV) type-I distribution  $\mathcal{G}(\mu, \sigma)$  [17], which is defined by CDF

$$\mathcal{G}(y|\mu, \sigma) = 1 - \exp\left(-\exp\left(\frac{y - \mu}{\sigma}\right)\right).$$

One can easily confirm that if  $X \sim \mathcal{W}(\kappa, \lambda)$ , and we define some  $Y = \log(X)$ , then  $Y \sim \mathcal{G}(\mu, \sigma)$ , with  $\mu = \log(\lambda)$ ,  $\sigma = 1/\kappa$ . Laio and Tamea (2007) [18] showed the CRPS to be equivalent to the area under the curve of a particular single-parameter loss function, here equivalently denoted

$$\begin{aligned} \text{CRPS}(F, x) &= 2 \int_0^1 \tau(x - F^{-1}(\tau)) \\ &\quad - \mathbf{1}[F(x) < \tau](x - F^{-1}(\tau)) d\tau. \end{aligned}$$

Once a distribution is determined, several difficult integrals often remain, and techniques for approaching them in the case of *maximum* GEV distributions are elucidated by Friederichs and Thorarinsdottir (2012) [19]. Here however, we explore the case of a *minimum* GEV distribution, so for our purposes we expand the right-hand side of the CRPS identity when  $F = \mathcal{G}(\mu, \sigma)$ , and take advantage of key properties of the exponential integral  $Ei(\cdot)$  and its siblings [20]. Namely, we consider for  $z \in \mathbb{C}$  and integer  $n > 0$  the integral  $E_n(z) = \int_1^\infty e^{-zt}/t^n dt$ , for which

$$E_1(z) = \int_z^\infty \frac{e^{-t}}{t} dt$$

holds for  $|\varphi| < \pi$ , recalling  $z = |z|e^{i\varphi}$ . This clearly holds if  $z \in \mathbb{R}$  and  $z > 0$ , and we can confirm that  $Ei(z) = -E_1(-z)$ . The derivative of  $E_n(z)$  is defined recursively as  $E'_n(z) = -E_{n-1}(z)$  and the base case at  $n = 0$  is a simple integral to evaluate. This allows us to verify that

$$\frac{d}{d\tau} Ei(\log(1 - \tau)) = \frac{1}{\log(1 - \tau)}$$

indeed holds, and here using the series expansion

$$Ei(z) = \gamma + \log(|z|) + \sum_{n=1}^{\infty} \frac{z^n}{(n!)n}$$

and with careful analysis of the limiting behaviour of the expanded terms, we can derive the following expression for the minimum GEV type-I CRPS

$$\text{CRPS}(\mathcal{G}, z) = z - \mu + \sigma(\gamma - \log(2) - 2Ei(\mathcal{L}(z)))$$

where  $\gamma = 0.57721566\dots$  is the Euler-Mascheroni constant, and  $\mathcal{L}(z) = \log(1 - \mathcal{G}(z))$ . We can thus model  $\kappa$  and  $\lambda$  in the wind speed domain, or equivalently model  $\mu$  and  $\sigma$  in the log-transformed domain.

Any wind speed  $x$  will be a non-negative real value, but  $y = \log(x)$  clearly may be negative irrespective of units, which would imply  $\arg(y) = \pi$ . As a practical issue, so long as we deal with rare zero values appropriately, then shifting the data by 1 and using a third (shift) parameter to slide back the resulting predictive Weibull by 1, output almost uniformly improves. Let  $\phi \in \{1\} \times \mathbb{R}^d$  be the input feature vector; specific features to be included will be decided by an endogenous model-building algorithm described below. We elect to work in the wind speed domain, modelling the scale parameter using a simple linear combination  $\lambda(\alpha) = \alpha^T \phi$ , and the shape as  $\kappa(\beta) = \beta_0 + \beta_1 v + \beta_2 g$ , where

$$v = \frac{1}{k|\mathcal{S}^*|} \sum_{i=0}^{k-1} \left( \sum_{s \in \mathcal{S}^*} \Delta x_{s,t-i}^2 \right)^{\frac{1}{2}},$$

$$g = \frac{\frac{1}{M} \sum_{i=0}^{M-1} (x_{s',t-i} - \bar{x}_{s'})^3}{\left( \frac{1}{M} \sum_{i=0}^{M-1} (x_{s',t-i} - \bar{x}_{s'})^2 \right)^{3/2}},$$

that is, a linear combination of sample volatility and skewness terms. Note  $\mathcal{S}^*$  is the index set of sites with a non-zero number of factors included in the model for a given site,  $s'$  is the index for the forecast site,  $\Delta x_t = x_t - x_{t-1}$ , and  $\bar{x}_{s'}$  is the sample mean of the re-training set of wind speeds for site  $s'$ . Following Gneiting *et al.* (2005) [13], at present time  $t$ , to output a predictive distribution  $\mathcal{W}(\alpha^*, \beta^*)$  for time  $t + k$ , our learning algorithm minimizes the arithmetic mean of the CRPS over the sliding training set. That is, since the GEV depends on the same controllable parameters,

$$\arg \min_{\alpha, \beta} \frac{1}{M} \sum_{i=0}^{M-1} \text{CRPS}(\mathcal{G}_{t-i}(\alpha, \beta), y_{t-i})$$

determines parameters and thus the predictive distribution  $\mathcal{W}$  at  $t$ . If data is scarce,  $M$  in these equations may have to be reduced for the model-building stage. Optimization may require linear inequality constraints on the parameters, which can be readily implemented in the R language and environment [21] using logarithmic barrier functions.

---

#### Algorithm 1 Pseudocode for model-building subroutine

---

```

1:  $\epsilon \geq 0, \phi(t) \leftarrow 1, h_0 \leftarrow \text{BIC}(\text{MinCRPS}(\phi)) \triangleright$  Initialize
2: for  $s \in \mathcal{S}$  do
3:   for  $x \in \{x_{s,t}, x_{s,t-1}, \dots\}$  do
4:      $\phi(t) \leftarrow \text{Append}(\phi(t), x - \text{OLS}(x))$ 
5:      $h \leftarrow \text{BIC}(\text{MinCRPS}(\phi))$ 
6:     if  $h_0 - h > \epsilon$  then
7:       break
8:     else
9:        $h_0 \leftarrow h$ 
10:    continue
11:   end if
12: end for
13: end for

```

---

Input features denoted  $\phi = (1, \phi_1, \dots, \phi_d)$  are selected on a disjoint data subset used exclusively for model-building, following the simple Algorithm 1. Having set threshold  $\epsilon$  and using the BIC as a proceed/site-switch criterion, like Hering and Genton (2010) [10] we create residual series, where the OLS operator regresses its argument's re-training time-series onto two trigonometric functions of the hour of day, and that is removed from the observed response.  $\text{MinCRPS}(\phi)$  executes the model which uses factors  $\phi$  and returns a set of forecasts to be plugged into the BIC subroutine.

## 4. RESULTS AND OBSERVATIONS

The experiment was carried out as discussed in section 2 for seven sites in the northwestern Hokkaido AMEDAS cluster, at forecast lag  $k = 2, 4, 6$ . That is, starting from the same fixed date for each  $k$ , the subsequent 2400 predictive distributions at  $k$ -steps ahead were computed. Naturally, different models are learned at different sites, but results were similar across sites, so avoiding redundancy here we discuss representative results from the Ishikari site. As deterministic reference models we used persistence (PER) and a correlation-weighted moving average (CMA) proposed by Nielsen *et al.* (1998) [22]. For a non-deterministic reference, we selected the truncated Normal TDD [10], the strongest model to come out of the line of work following the original RST model.

The results of deterministic evaluations using RMSE and MAE for the proposed (W-GEV) and reference models are contained in Table 1. We observe that in relative terms the gap in performance between persistence and the probabilistic models increases in with forecast lag; W-GEV is competitive with CMA over all forecast lags and is seen to be as good or better than the probabilistic reference over all  $k$  considered. In absolute terms error increases in  $k$ , an observation validated by posterior analysis of site wind speed cross-correlations, which start strong and grow weaker as lag  $k$  increases.

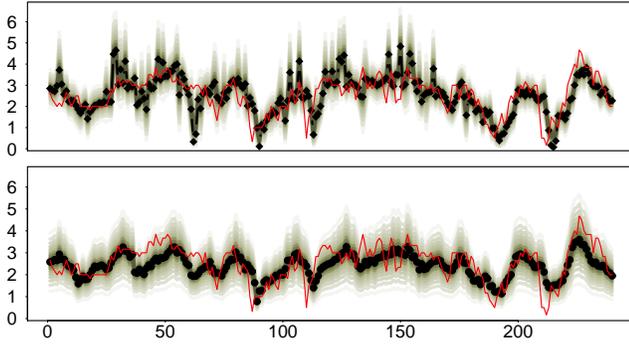
Next we consider a framework of *sharpness* subject to *calibration* [23] of density forecasts. First considering sharp-

	RMSE			MAE		
	$k = 2$	4	6	2	4	6
<b>PER</b>	0.59	0.82	0.98	0.42	0.61	0.72
<b>CMA</b>	0.55	0.72	<b>0.79</b>	<b>0.41</b>	0.54	<b>0.61</b>
<b>TDD</b>	0.56	0.71	0.81	0.44	0.54	0.63
<b>W-GEV</b>	<b>0.53</b>	<b>0.69</b>	0.81	<b>0.41</b>	<b>0.52</b>	<b>0.61</b>

**Table 1.** Deterministic forecast errors over the 2400-hour period at lag  $k = 2, 4, 6$ .

	90%	Ave. CRPS
<b>TDD</b>	<b>4.40</b>	0.33
<b>W-GEV</b>	5.27	<b>0.28</b>

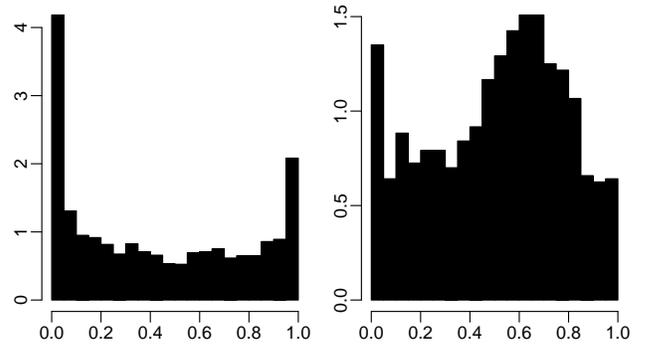
**Table 2.** Non-deterministic evaluation using average predictive intervals and CRPS score.



**Fig. 1.** Observations (red line), 10–90% prediction intervals (increasingly transparent grey lines), and point forecasts (points) for TDD (top) and W-GEV (bottom).

ness (for  $k = 2$ , similar results for other  $k$ ), the model 10–90% prediction intervals, point forecasts, and actual observations for the first 10 days of the test period are shown as an example in Fig. 1. The average 90% interval widths are contained in Table 2. TDD has the tighter interval but W-GEV has stronger deterministic output, suggesting the latter may be more robust towards lower-velocity sites which tend to have noisier, less well-defined wind profiles. In the same table, we see that W-GEV predictive distributions had a smaller average CRPS value. The deviation between the two metrics, and the lower overall CRPS values for W-GEV leads us to infer that over the test set W-GEV has better calibration than TDD. To validate this, we can make use of *probability integral transform* (PIT) histograms, where the more Uniformly distributed outputs suggest better calibration. The TDD looks to be “overconfident” (Fig. 2), in that its density is sharply concentrated, yet may not be well-calibrated; this observation generally matches the tighter prediction intervals seen for

TDD and smaller CRPS values seen for W-GEV.



**Fig. 2.** PIT histograms (relative frequency) for TDD (left) and W-GEV (right).

W-GEV results are in general robust to initial values, though appropriate ranges may require empirical investigation; the optimal sliding training set length  $M$  differs between models, and in this experiment given a built model at a given site, optimal  $M^*$  for each model was selected over  $M = 20, 21, \dots, 100$  according to the value that minimized RMSE at each time scale. TDD tended to perform better in the 25–35 day range, while W-GEV generally performed better in the 35–45 day range.

## 5. CONCLUDING REMARKS

On the basis of deterministic forecast evaluation metrics, the Weibull-GEV model proposed in this study was shown to outperform reference models for forecast lag beyond 3 hours, and to outperform the reference predictive distribution over a cluster of sites where lower-speed, noisier wind profiles may be present. Within the sharpness/calibration framework, for a 100-day test set we observed that the reference model was sharp but not necessarily well-calibrated, while subject to sufficient calibration, the W-GEV model had better sharpness, suggesting that Weibull models estimated using a CRPS-minimization algorithm may indeed be a valid choice for site-adaptive density forecasting applications with minimal prerequisites in terms of site knowledge or historical data.

The model introduced here was designed to be simple and intuitive, such that the merits of the derived Weibull-GEV CRPS might be explicitly examined, and compared with previous well-known models. As a result however, information about spatio-temporal correlations in the model was very limited. Future lines of work may pursue models which more effectively capture the rich correlations that may be present in such networked observations. As well, posterior analysis suggests that a multi-modal distribution may more accurately represent the data, and it can be thought that allowing for extension to mixture models may lead to further promising results.

## 6. REFERENCES

- [1] Andrew Kusiak and Zijun Zhang, “Control of wind turbine power and vibration with a data-driven approach,” *Renewable Energy*, vol. 43, pp. 73–82, 2012.
- [2] Christopher A Walford, *Wind turbine reliability: understanding and minimizing wind turbine operation and maintenance costs*, United States Department of Energy, 2006.
- [3] Pierre Pinson, Christophe Chevallier, Georges Kariniotakis, et al., “Optimizing benefits from wind power participation in electricity market using advanced tools for wind power forecasting and uncertainty assessment,” in *Proceedings of the European Wind Energy Conference (EWEC)*, 2004.
- [4] Yu Jiang, Zhe Song, and Andrew Kusiak, “Very short-term wind speed forecasting with Bayesian structural break model,” *Renewable Energy*, vol. 50, pp. 637–647, 2013.
- [5] 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.
- [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] Xinxin Zhu and Marc G Genton, “Short-term wind speed forecasting for power system operations,” *International Statistical Review*, vol. 80, no. 1, pp. 2–23, 2012.
- [8] Barbara G Brown, Richard W Katz, and Allan H Murphy, “Time series models to simulate and forecast wind speed and wind power,” *Journal of Climate and Applied Meteorology*, vol. 23, pp. 1184–1195, 1984.
- [9] Philippe Crochet, “Adaptive Kalman filtering of 2-metre temperature and 10-metre wind-speed forecasts in iceland,” *Meteorological Applications*, vol. 11, no. 2, pp. 173–187, 2004.
- [10] 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.
- [11] Pierre Pinson, *Estimation of the uncertainty in wind power forecasting*, Ph.D. thesis, École des Mines de Paris, 2006.
- [12] 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.
- [13] 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.
- [14] CG Justus, WR Hargraves, and Ali Yalcin, “Nationwide assessment of potential output from wind-powered generators,” *Journal of Applied Meteorology*, vol. 15, no. 7, pp. 673–678, 1976.
- [15] Erik Lundtang Petersen, Ib Troen, S Frandsen, and K Hedegaard, “Windatlas for Denmark,” *Roskilde, Risø*, vol. 229, 1981.
- [16] K Conradsen, LB Nielsen, and LP Prahm, “Review of Weibull statistics for estimation of wind speed distributions,” *Journal of Climate and Applied Meteorology*, vol. 23, no. 8, pp. 1173–1183, 1984.
- [17] Horst Rinne, *The Weibull distribution: a handbook*, CRC Press, 2010.
- [18] Francesco Laio and Stefania Tamea, “Verification tools for probabilistic forecasts of continuous hydrological variables,” *Hydrology and Earth System Sciences*, vol. 11, no. 4, pp. 1267–1277, 2007.
- [19] Petra Friederichs and Thordis L Thorarinsdottir, “Forecast verification for extreme value distributions with an application to probabilistic peak wind prediction,” *Environmetrics*, vol. 23, no. 7, pp. 579–594, 2012.
- [20] Milton Abramowitz and Irene A Stegun, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables*, vol. 55, 1964.
- [21] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2013.
- [22] Torben Skov Nielsen, Alfred Joensen, Henrik Madsen, Lars Landberg, and Gregor Giebel, “A new reference for wind power forecasting,” *Wind Energy*, vol. 1, no. 1, pp. 29–34, 1998.
- [23] Tilmann Gneiting, Fadoua Balabdaoui, and Adrian E Raftery, “Probabilistic forecasts, calibration and sharpness,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 2, pp. 243–268, 2007.