# Uncertainty Sets For Wind Power Generation

Yury Dvorkin, *Student Member, IEEE*, Miles Lubin, Scott Backhaus, Michael Chertkov, *Senior Member, IEEE*.

*Abstract*—As penetration of wind power generation increases, system operators must account for its stochastic nature in a reliable and cost-efficient manner. These conflicting objectives can be traded-off by accounting for the variability and uncertainty of wind power generation. This letter presents a new methodology to estimate uncertainty sets for parameters of probability distributions that capture wind generation uncertainty and variability.

*Index Terms*—Wind power uncertainty, wind power variability, power system operations.

## I. INTRODUCTION

WIND power generation (WPG) introduces variability and uncertainty in operational planning. As defined in [1], *variability* of WPG is the random fluctuation of wind speed caused by physical processes in the atmosphere while *uncertainty* of WPG results from wind forecast errors. To account for variability and uncertainty, approaches to robust unit commitment (RUC) [2], chance constrained optimal power flow (CC-OPF) [3] and distributionally robust CC-OPF (RCC-OPF) [3],[4] have been formulated and tested. However, the performance of these models depend on the accuracy of parameters of probability distributions that define the uncertainty sets used; therefore, the uncertainty in distribution parameters must be accounted for. This letter makes two contributions: i) we use a data-driven analysis to relate intra-hour wind speed variability to the hourly-average wind speed and ii) using this relationship we construct uncertainty sets that can be interpreted in terms of variability and uncertainty of WPG.

Bienstock et al [3] show that the CC-OPF formulation with Gaussian-distributed deviations of WPG and precisely known mean and variance can be extended to the RCC-OPF formulation, where the parameters of the Gaussian deviations (both mean and variance) fall within uncertainty sets. This RCC-OPF is implemented and tested on a large-scale system in [4]. In this letter, we interpret the set for the mean in terms of the wind power uncertainty, while the set for the variance is deemed as the wind power variability. This approach differs from the uncertainty sets on a random variable in [2] by introducing uncertainty sets on distribution parameters describing the random variable. We derive these sets with a data-driven approach such that the resulting RCC-OPF remains tractable [3], [4], yielding a practical approach for modeling WPG. Although the RCC-OPF model is nominally more conservative than the CC-OPF model, its solution incurs a lower operating cost when tested against actual realizations of WPG [4].

## II. METHODOLOGY

*1) Data:* We use historical wind speed measurements at 5-minute resolution from the Goodnoe meteorological station in the Bonneville Power Authority (BPA) system [5] and one-hour resolution wind speed forecasts produced by the NOAA Rapid Refresh numerical weather prediction model [6] for the same location. The historical measurements and forecasts are detrended by using data from the same calendar season (December–February) and range of hours (00:00–04:00 AM).

*2) Wind Speed Variability:* For any subinterval $\tau$ of hour-long interval $t$ wind speed $w_\tau$ can be written as $w_\tau = \mu_t + \epsilon_\tau$ [3], where $\mu_t$ and $\epsilon_\tau$ are the hourly-average wind speed and intra-hour, zero-mean variability around $\mu_t$ [1], respectively. To parametrize $w_\tau$, we calculate the average wind speed ($\mu_t$) for each hourly interval within the studied period using the 5-minute resolution Goodnoe data [5]. Next, we bin the hourly intervals based on their hourly-averaged wind speed $\mu_t$. This binning is in the range from 0 m/s to 25 m/s (the typical operating range of wind turbines) and has the width of 1 m/s. For each bin we obtain a histogram that represents an empirical intra-hour wind speed distribution $f^E(w_\tau|\mu_t)$ conditioned by the hourly-average wind speed ($\mu_t$). Each histogram is then fit to the Gaussian distribution, yielding estimated mean ($\mu^*$) and standard deviation values ($\sigma^*$) for every bin. Both the histogram and its Gaussian distribution are shown in Fig. 1a) for several values of $\mu_t$. Figure 1b) shows that $\sigma^*$ scales linearly with $\mu^*$ as $\sigma^*(\mu^*) = 0.231 + 0.197 \cdot \mu^*$ for $\mu^* \in [0, 25]$ m/s, which is consistent with velocity distributions for high Reynolds number atmospheric flows [7]. If there is no error in a given hourly-averaged wind speed forecast $\mu_t^f$, the wind speed variability is estimated using the linear mapping $\sigma^*(\mu^*)$ shown in Fig. 1b). The resulting distribution is then given as $N[w_\tau; \mu_t^f, (\sigma^*(\mu_t^f))^2]$, and can be used in [2] and [3] for uncertainty sets accounting for wind speed variability.

*3) Wind Speed Uncertainty and Variability:* Wind forecast errors cause $\mu_t \neq \mu_t^f$. The wind forecast error $e_t$ is calculated from the NOAA [6] data as $e_t(\Delta T) = \mu_t - \mu_t^f(\Delta T)$, where $\mu_t^f(\Delta T)$ is the forecast for hour $t$ made $\Delta T$ hours in advance. The empirical distribution $f^E(e_t; \Delta T)$ for $\Delta T = 1$ hour is shown in Fig. 2a). We propose to represent $f^E(\cdot)$ with a generalized Gaussian distribution, $f^G(\cdot)$:

$$f^G(e_t; \Delta T, \mu^-, \mu^+) = \frac{\int_{\mu^-}^{\mu^+} d\mu N[e_t; \mu, \sigma^*(\mu)]}{(\mu^+ - \mu^-)}, \quad (1)$$

where $\sigma^*(\mu)$ is the fit from Fig. 1b) and the dependence of $\mu^+$ and $\mu^-$ on $\Delta T$ is suppressed. The advantage of $f^G(\cdot)$ over a single Gaussian distribution is that it improves the
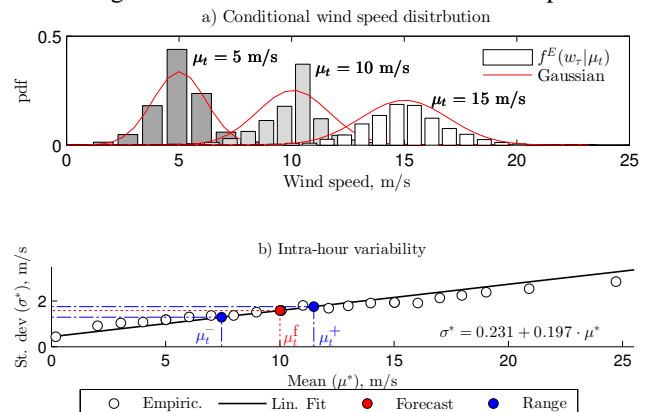


Fig. 1: a) Empirical distribution $f^E(w_t|\mu^*)$ (histograms) and their Gaussian best fits for a few $\mu^*$. b) Empirical relationship between $\sigma*$ and $\mu^*$ and its linear fit $\sigma^*(\mu^*)$.
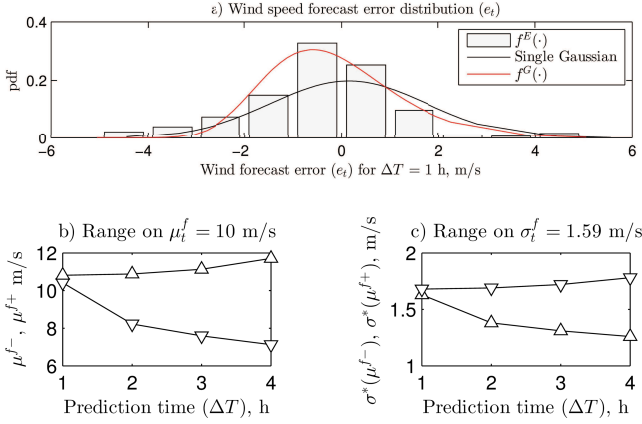
Fig. 2: a) Empirical distribution $f^E(e_t; \Delta T)$ (histogram), the best fit single Gaussian distribution (black) and the best fit distribution $f^G(\cdot)$ (red) from Eq. (1) for $\Delta T = 1$ h. b) $\mu_t^{f-}$ and $\mu_t^{f+}$ versus $\Delta T$ for $\mu_t^f = 10$ m/s. c) $\sigma^*(\mu_t^{f-})$ and $\sigma^*(\mu_t^{f+})$ for the data in b).
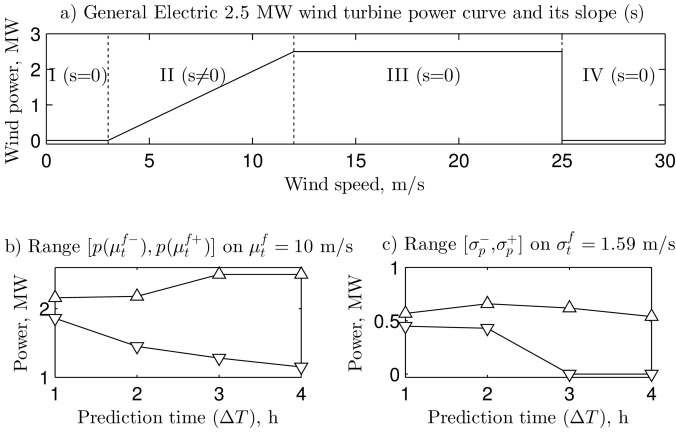


Fig. 3: a) Typical wind power turbine curve $p(\mu)$ with four nominal operating regions [9]. b) Ranges for the on the mean WPG (i.e. forecast error) computed using Eq. (3) for the data in Fig. 2c). c) Ranges on WPG standard deviation (variability) computed using Eqs. (4) and (5) for the data in b).

goodness-of-fit to empirical data [8] by encapsulating a linear superposition of Gaussian distributions over a range of means $[\mu^-, \mu^+]$ that represents the wind uncertainty. Furthermore, [8] explains that other single non-Gaussian distributions can be modelled using $f^G(\cdot)$. The best fit $\mu^-$ and $\mu^+$ are computed by solving the optimization problem:

$$\underset{\mu^-, \mu^+}{\arg\min} \int [f^G(e_t; \Delta T, \mu^-, \mu^+) - f^E(e_t; \Delta T)]^2 de_t, \quad (2)$$

which minimizes the mean square difference between $f^E(\cdot)$ and $f^G(\cdot)$ and ensures a better fit to the historical data than a single Gaussian distribution, as shown in Fig. 2a). We interpret the range $[\mu_t^{f-}, \mu_t^{f+}]$, where $\mu_t^{f-} = \mu_t^f + \mu^-$ and $\mu_t^{f+} = \mu_t^f + \mu^+$, as the bounds of the uncertainty set for the mean wind speed. Fig. 2b) displays $\mu_t^{f-}$ and $\mu_t^{f+}$ for $\mu_t^f = 10$ m/s for different $\Delta T$. Using $\sigma^*(\mu^*)$ from Fig. 1b), we compute the bounds on $\sigma^*$ as $\sigma^*(\mu_t^{f-})$ and $\sigma^*(\mu_t^{f+})$, which are shown in Fig. 1c). The ranges $[\mu_t^{f-}, \mu_t^{f+}]$ and $[\sigma^*(\mu_t^{f-}), \sigma^*(\mu_t^{f+})]$, if converted to wind generation as explained below, can be used in the RCC-OPF formulation from [3], [4].

*4) Conversion to Wind Power:* We illustrate the conversion to wind power using the single wind turbine power curve $p(\mu)$ shown in Fig. 3a), [9]. This procedure can be generalized to

multiple turbines by using an aggregated wind power curve [10]. The conversion of the range $[\mu_t^{f-}, \mu_t^{f+}]$ is given by:

$$[\mu_t^{f-}, \mu_t^{f+}] \rightarrow [p(\mu_t^{f-}), p(\mu_t^{f+})]. \quad (3)$$

Figure 3b) shows $p(\mu_t^{f-})$ and $p(\mu_t^{f+})$ corresponding to $\mu_t^{f-}$ and $\mu_t^{f+}$, respectively, from Fig. 2b). In Fig. 2b), the growth of $p(\mu_t^{f+})$ at larger $\Delta T$ is eventually clipped by $p(w)$ as $\mu_t^{f+}$ enters Region III of the turbine curve. If the entire range $[\mu_t^{f-}, \mu_t^{f+}]$ is in Region III, then the range $[p(\mu_t^{f-}), p(\mu_t^{f+})]$ collapses to zero width around the maximum output.

The conversion of the range $[\sigma^*(\mu_t^{f-}), \sigma^*(\mu_t^{f+})]$ is shaped by the RCC-OPF formulation in [4]. Note that $\sigma_p$ can be obtained $\forall p(\mu_t) \in [p(\mu_t^{f-}), p(\mu_t^{f+})]$ using the relation $\sigma^*(\mu)$ from Fig. 1b) and the slope $s$ of the turbine curve as $\sigma_p = s(\mu_t) \cdot \sigma^*(\mu_t)$. However, using the wind turbine power curve from Fig. 3a) results in difficult nonconvexity in distributionally robust formulations (e.g., Eq. (27)-(28) in [4]). To avoid this nonconvexity, we assume that the range on the mean value in (Eq. 3) and range on the standard deviation are independent. Therefore, the range on the standard deviations is given by:

$$\sigma_p^- = \min_{\mu \in [\mu_t^{f-}, \mu_t^{f+}]} s(\mu)\sigma^*(\mu) \quad (4)$$

$$\sigma_p^+ = \max_{\mu \in [\mu_t^{f-}, \mu_t^{f+}]} s(\mu)\sigma^*(\mu) \quad (5)$$

Figure 3c) shows the conversion to the robust interval on the standard deviation of WPG for the data in Fig. 2c).

## III. Conclusion

We have presented a data-driven method to develop robust intervals for distribution parameters, which preserves the physical relationship between instantaneous and hour-average wind speed and power. This method is suitable for uncertainty sets in the RUC [2] and RCC-OPF [3], [4], which can be interpreted in terms variability and uncertainty of WPG. The case study in [4] shows that the RCC-OPF model with these uncertainty sets outperforms several benchmarks in terms of several cost and reliability metrics.

## References

[1] E. Ela and M. O'Malley, "Studying the Variability and Uncertainty Impacts of Variable Generation at Multiple Timescales," *IEEE Trans. Power Syst.*, Vol. 27, No. 3, pp. 1324-1333, 2012.
[2] Y. Guan and J. Wang, "Uncertainty Sets for Robust Unit Commitment," *IEEE Trans. Power Syst.*, Vol. 29, No. 3, pp. 1439-1440, 2014.
[3] D. Bienstock *et al*, "Chance-Constrained Optimal Power Flow: Risk-Aware Network Control under Uncertainty," *SIAM Review*, Vol. 56, No. 3, pp. 461-495, 2014.
[4] M. Lubin, Y. Dvorkin, S. Backhaus, "A Robust Approach to Chance Constrained Optimal Power Flow with Renewable Generation," subm. to *IEEE Trans. Power Syst.* Available at: http://arxiv.org/abs/1504.06011.
[5] BPA Meteorological Data. [Online]. Available at: http://goo.gl/f1U3ys.
[6] NOAA Archive. [Online]. Available at: http://goo.gl/OPUqAm.
[7] J. Kaimal, "Turbulenece spectra, length scales and structure parameters in the stable surface layer," *Bound.-Layer Met.*, Vol. 4, pp. 289-309, 1973.
[8] G. J. Hahn and S. S. Shapiro, *Statistical Models in Engineering*. New York, NY, USA: John Wiley & Sons, 1967.
[9] GE 2.5MW Turbine Series. [Online]. Available at: http://goo.gl/4QUqE0.
[10] B. P. Hayes *et al*, "Equivalent power curve model of a wind farm based on field measurement data," in Proc. of *2011 IEEE Trondheim PowerTech*, 2011, pp. 1-7.