جامعــة خليفــة
Khalifa University

IEEE Canada
EPEC 2018

Evaluation of Parametric Statistical Models for Wind Speed Probability Density Estimation

Maisam Wahbah, *Omar Alhussein*, Tarek H.M. EL-Fouly, Bashar Zahawi, Sami Muhaidat

11 OCTOBER 2018

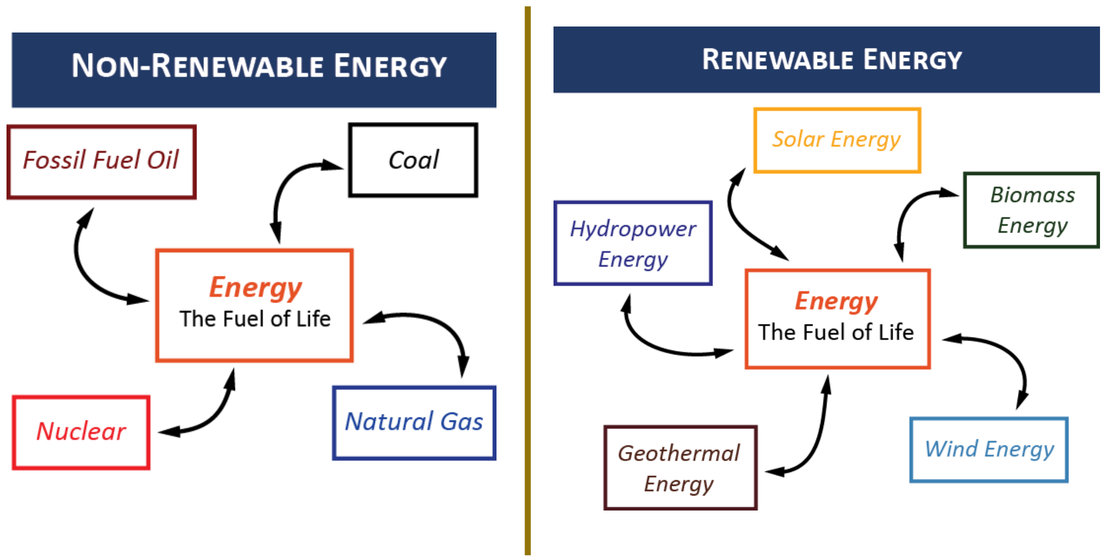
Session Wind Energy Systems
Presentation S11.2

Outline

- **Challenges**
- **Gaussian Mixture Model**
- **Expectation-Maximization**
- **Results and Analysis**
- **Summary and Conclusions**

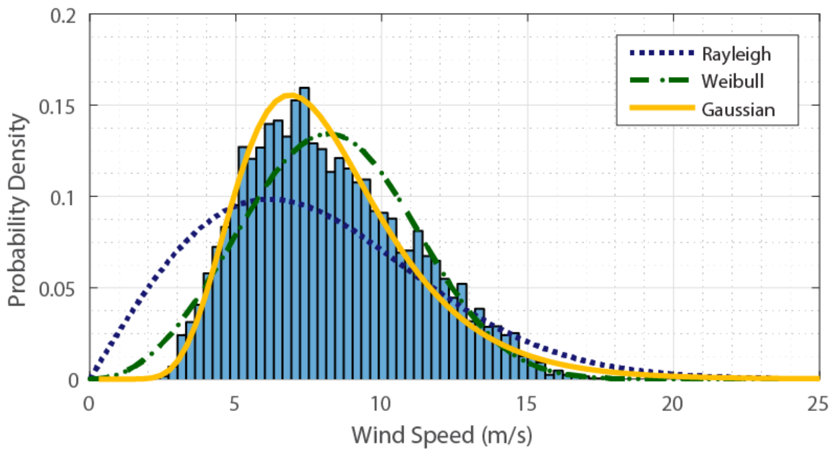
Challenges in Intermittent Renewables

- **Intermittent generation**
- **Frequency control**
- **System inertia**
- **Power quality**



Statistical Parametric Models

- **Parametric Families of Probability Distributions (PFPDs):**
  - Wind Speed
    - Rayleigh and Weibull distributions.
- **How are they used?**
  - Commonly used in power system and microgrid planning.



Limitations in Common Parametric Distribution Models

- Selecting the appropriate parametric distribution.
- Bimodal or multimodal probability density.
- May be appropriate for a site but may not be appropriate for another.

- Common parametric models are too restrictive and may very well end up in a model mis-specification, which in turn can lead to catastrophic estimates.

Gaussian Mixture Model (1/2)

- Widely used in:
  - Clustering, machine learning, pattern recognition.
  - Modeling non-Gaussian noise processes, and statistical modeling.

- Gaussian mixtures are often used due to the fact their individual densities are efficiently characterized by first two moments.

- Attributed to have “Universal-approximation” property as proven by Weiner’s approximation theorem.

- We are given a dataset $\boldsymbol{x} = \{x_1, \dots, x_N\}$, which we believe contains C classes. Then, express PDF as as finite convex sum of Gaussian densities.

$$f_X(x_j|\theta) = \sum_{i=1}^{C} \omega_i \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{\left(-\frac{(x-\mu_i)^2}{2\sigma_i^2}\right)}, \qquad j = \{1, \dots, n\}$$

Gaussian Mixture Model (2/2)

- **Goal:**
  - To determine parameters $(\theta = \{\omega_i, \mu_i, \sigma_i^2\}, \forall i = \{1, \dots, C\})$, such that likelihood function, $\Pr(\boldsymbol{x}|\theta, C)$, is maximized.

- **Cannot be solved directly:**
  - Use Expectation-Maximization (EM) algorithm to maximize the log-likelihood function instead.

- **Expectation-Maximization (EM)**
  - Coined by Dempster *et al.* in 1977.
  - Iterative Solution to Maximum Likelihood Estimation.
  - Used for Mixture Modeling or Clustering.

Expectation-Maximization (1/2)

- **Pros:**
  - Automated (Unsupervised Algorithm).
  - Can be Semi-supervised.
  - Guaranteed not to get worse, as it iterates by.
- **Cons:**
  - Takes time to converge.
  - Susceptible to over fitting.

- **Over-fitting**
  - Very large C maximizes log-likelihood but causes over-fitting.
  - Use BIC approach to penalize log-likelihood as C increases

$$BIC(C) = -2\ln \Pr\left(\boldsymbol{x} \middle| \hat{\theta}, C\right) + C\ln(n)$$

where $\hat{\theta} = \left\{\widehat{\omega_i}, \widehat{\mu_i}, \widehat{\sigma_i^2}\right\}, i = \{1, \ldots, C\}$ are the parameter such that log-likelihood is maximized.

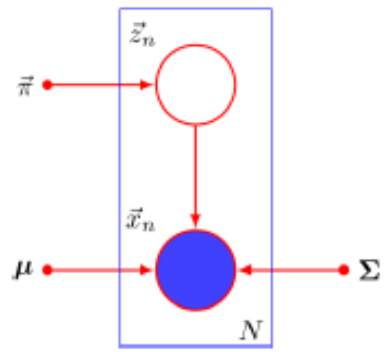
Expectation-Maximization (2/2)

- How to Run Algorithm:
  - Choose Initial values for $\mu_k, \sigma_k, \omega_k$.
  - Calculate responsibilities $\gamma(z_k)$ (closed-form equation exists)
  - Using $\gamma(z_k)$, update values of μ_k, σ_k, π_k (closed-form equation)
  - Repeat until Log-likelihood converges.

- BIC-assisted EM:
  - Do above procedure by varying C and running EM until BIC measure gets worse.

Expectation-Maximization*

- Suppose we are given a dataset $x = \{x_1, \dots, x_N\}$, which we believe contains C classes, each of which is independent and follow some exponential distribution.
  - Observed Variable:
    - Data, namely $\boldsymbol{x} = \{x_1, \dots, x_N\}$
  - Latent (Unobserved) Variable:
    - Which Data belongs to what? i.e.
    - $\Pr(z_k) = \pi_k$.
  - Joint Data, $Y = (\boldsymbol{x}, Z)$.



Expectation-Maximization*

- Consists of Two Main Steps:
  - Decide an initial C (number of components)
    - Expectation Step: Compute Posterior Probability
      - Given observed Data $\boldsymbol{x}$, calculate how likely it is that the complete data is y, i.e. to calculate $\Pr(y|D, \theta^m)$

$$\gamma(z_k) = \Pr(z_k = k|\boldsymbol{x}) \frac{\pi_k N(\boldsymbol{x}|\mu_k, \sigma_k^2)}{\sum_{l=1}^{C} \pi_l N(\boldsymbol{x}|\mu_l, \sigma_l^2)}, k = \{1, \ldots, C\}$$

    - Maximization Step:
      - Given the guess of θ^m, we want to maximize $E\{\log \Pr(x|\theta)\}$.

$$\vec{\mu}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk}) \vec{x}_n, \qquad \vec{\Sigma}_k = \frac{1}{N_k} \sum_{n=1}^{N} \gamma(z_{nk})(\vec{x}_n - \vec{\mu}_k)(\vec{x}_n - \vec{\mu}_k)^{\mathrm{T}},$$

    - Termination:
      - Keep iterating until Log-likelihood does not improve significantly.
- BIC-assisted EM:
  - Do above procedure by varying C and running EM until BIC measure gets worse.

Wind Speed Data

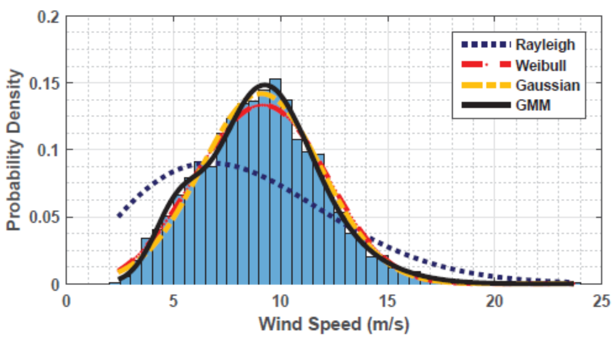
- Performance of the proposed GMM for estimating wind speed probability density is assessed via comparisons with conventional parametric probability density models.

- Hourly wind speed data from www.renewables.ninja

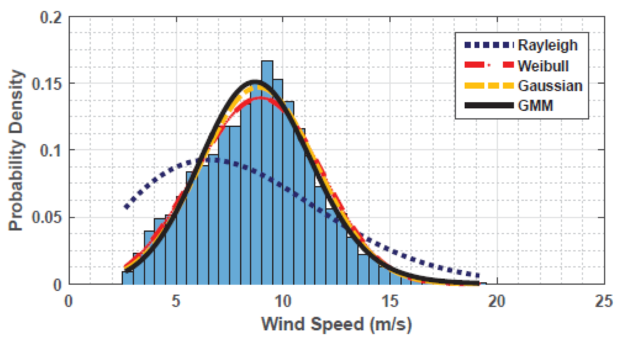
Site	Location	Geographical Coordinates (°)		Wind Speed (m/s)				
		Latitude	Longitude	Maximum	Minimum	Mean	Median	Standard Deviation
Site 1	Copenhagen, Denmark	55.69	12.57	23.69	2.41	9.10	9.09	2.82
Site 2	Kiel, Germany	54.46	10.20	19.17	2.63	8.87	8.90	2.73
Site 3	Black Law, Scotland	55.83	-3.82	19.10	1.85	7.40	7.01	2.79
Site 4	Whitelee, Scotland	55.71	-4.34	22.20	2.06	8.43	7.95	3.23
Site 5	Schaw, Scotland	55.46	-4.46	23.26	2.40	9.06	8.63	3.31
Site 6	Malin Head, Ireland	55.38	-7.40	24.32	2.37	10.05	9.53	3.80

Probability density plots and histogram of wind speed at Denmark, Germany and Ireland

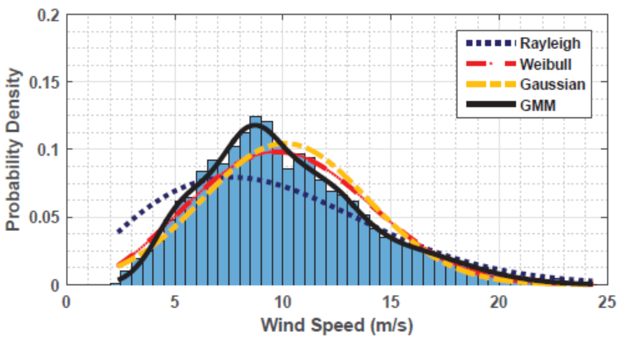
Copenhagen (Denmark)



Kiel (Germany)

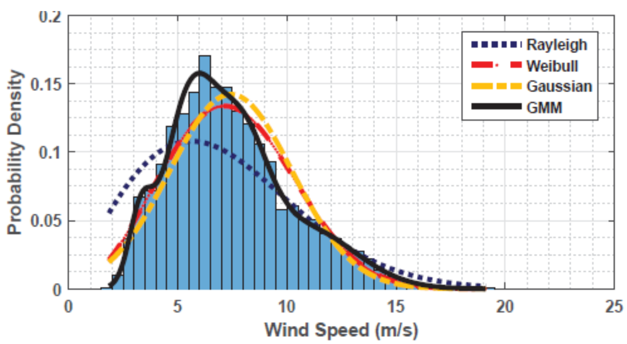


Malin Head (Ireland)

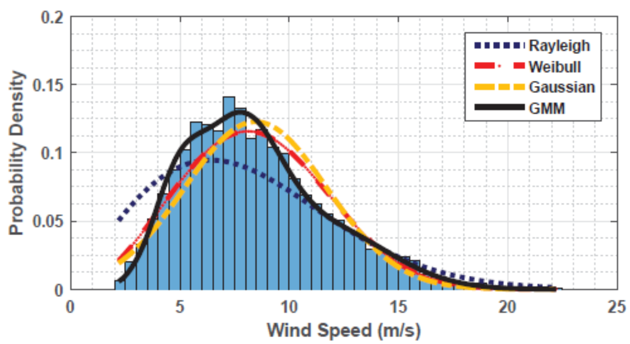


Probability density plots and histogram of wind speed at Three Sites in Scotland

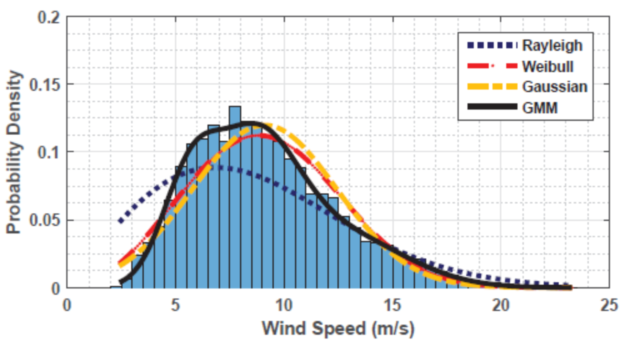
Black Law



Whitelee



Schaw



Parametric Probability Density Functions

Distribution	Probability Density Function	Domain	Parameters
Rayleigh	$f_X(x) = \frac{x}{b^2} e^{-(x^2/2b^2)}$	$x \in [0, \infty)$	Scale: $b \in (0, \infty)$
Weibull	$f_X(x) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-(x/\lambda)^k}$	$x \in [0, \infty)$	Scale: $\lambda \in (0, \infty)$ Shape: $k \in (0, \infty)$
Gaussian	$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	$x \in (-\infty, \infty)$	Mean: $\mu \in (-\infty, \infty)$ Variance: $\sigma^2 \in (0, \infty)$

- Estimated parameters completely define the distribution function for all values of x.

Parameter Estimation of PFPDs

- Estimated using Maximum Likelihood Estimation.

Site	Rayleigh	Weibull		Gaussian	
	b	λ	k	μ	σ^2
Site 1	6.73	10.09	3.50	9.09	7.89
Site 2	6.54	9.81	3.55	8.84	7.35
Site 3	5.62	8.35	2.83	7.43	7.86
Site 4	6.41	9.52	2.78	8.46	10.56
Site 5	6.85	10.21	2.91	9.10	11.09
Site 6	7.64	11.35	2.82	10.10	14.59

Quantitative Assessment

- Performance of the GMM is assessed against the three selected parametric distributions using:
  - Kolmogorov-Smirnov goodness-of-fit test.
  - Root Mean Square Error.
  - Mean Absolute Error.

- Quantitative way to assess the quality of a given statistical model.

Kolmogorov-Smirnov Goodness-of-Fit Test

- **Used to check whether wind speed data come from a particular statistical model.**

Can we ***reject*** the null hypothesis that the observed data are independently sampled from the six distribution models?

- **p-value**
  - A measure of the evidence against the null hypothesis, with low p-values corresponding to stronger evidence against the null hypothesis.
  - $p < 0.01$.

- M. DeGroot and M. Schervish, Probability and Statistics. London: Pearson Education, Inc., 4th ed., 2012.
- E. L. Lehmann and J. P. Romano, esting Statistical Hypotheses. New York: Springer, 3th ed., 2005.

p-values resulting from K-S Test

Site	Rayleigh	Weibull	Gaussian	GMM
Site 1	6.41×10^{-80}	**24.53×10^{-02}**	**58.67×10^{-02}**	**6.70×10^{-02}**
Site 2	2.00×10^{-90}	**5.31×10^{-02}**	**27.27×10^{-02}**	**5.57×10^{-02}**
Site 3	4.18×10^{-39}	6.85×10^{-10}	2.37×10^{-13}	**84.42×10^{-02}**
Site 4	9.28×10^{-38}	2.59×10^{-09}	5.15×10^{-15}	**49.65×10^{-02}**
Site 5	9.01×10^{-49}	2.66×10^{-12}	9.36×10^{-14}	**18.71×10^{-02}**
Site 6	8.88×10^{-34}	2.76×10^{-08}	7.91×10^{-13}	**23.18×10^{-02}**

Error Measures

Root Mean Square Error

$$RMSE = \sqrt{\frac{1}{t}\sum_{i=1}^{t}(y_i - \hat{y}_i)^2}$$

Mean Absolute Error

$$MAE = \frac{1}{t}\sum_{i=1}^{t}|y_i - \hat{y}_i|$$

Root Mean Square Error

- The proposed method gives the **lowest** RMSE value among all evaluated methods for all six sites.

Site	Rayleigh	Weibull	Gaussian	GMM
Site 1	1.24×10^{-2} (–)	3.07×10^{-3} (75.2%)	2.95×10^{-3} (76.3%)	2.83×10^{-3} **(77.2%)**
Site 2	1.19×10^{-2} (–)	4.20×10^{-3} (64.6%)	4.00×10^{-3} (66.3%)	3.81×10^{-3} **(67.9%)**
Site 3	8.85×10^{-3} (–)	5.76×10^{-3} (35.0%)	6.45×10^{-3} (27.1%)	3.29×10^{-3} **(62.8%)**
Site 4	8.64×10^{-3} (–)	5.88×10^{-3} (32.0%)	6.75×10^{-3} (21.9%)	4.19×10^{-3} **(51.5%)**
Site 5	9.05×10^{-3} (–)	5.77×10^{-3} (36.2%)	6.34×10^{-3} (29.9%)	3.89×10^{-3} **(56.9%)**
Site 6	8.22×10^{-3} (–)	5.33×10^{-3} (35.2%)	6.09×10^{-3} (25.9%)	3.98×10^{-3} **(51.6%)**

Mean Absolute Error

- The proposed method gives the **lowest** MAE value among all evaluated methods for all six sites.

Site	Rayleigh	Weibull	Gaussian	GMM
Site 1	9.02×10^{-3} (–)	2.26×10^{-3} (75.0%)	2.05×10^{-3} (77.3%)	2.00×10^{-3} (**77.9%**)
Site 2	8.78×10^{-3} (–)	2.91×10^{-3} (66.8%)	2.79×10^{-3} (68.2%)	2.42×10^{-3} (**72.4%**)
Site 3	5.81×10^{-3} (–)	3.94×10^{-3} (32.2%)	4.69×10^{-3} (19.3%)	2.34×10^{-3} (**59.7%**)
Site 4	5.82×10^{-3} (–)	4.01×10^{-3} (31.0%)	4.85×10^{-3} (16.6%)	2.66×10^{-3} (**54.2%**)
Site 5	6.57×10^{-3} (–)	4.07×10^{-3} (38.1%)	4.51×10^{-3} (31.3%)	2.86×10^{-3} (**56.5%**)
Site 6	5.56×10^{-3} (–)	3.83×10^{-3} (31.2%)	4.43×10^{-3} (20.3%)	2.93×10^{-3} (**47.3%**)

Summary and Conclusions

- Common Parametric Distribution Models
- Proposed Gaussian Mixture Model for KDE bandwidth selection
  - BIC-assisted Expectation Minimization Framework
- Probability density plots and histogram of wind speed
- Results of the Kolmogorov-Smirnov Goodness-of-Fit test
- The proposed method gives the lowest RMSE and MAE values among all evaluated methods for all six sites.

Thank You

maisam.wahbah@ku.ac.ae

oalhussein@uwaterloo.ca