A COMPARISON OF ECONOMIC DISPATCH SIMULATIONS WITH DIFFERING STOCHASTIC WIND POWER MODELS

A Report

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Engineering

by

Evan Kenneth Halloran

December 2019

ABSTRACT

With the threats of climate change, increased penetration of renewable energy on the electricity grid is a crucial method to reduce fossil fuel emissions. One of the main issues with renewable energy sources, such as wind, is that they produce intermittent power. There are many statistical distributions that are used to represent wind speed in a given location. These distributions can be used to develop stochastic models of wind power in a power flow simulation. This paper seeks to develop a framework for generating stochastic models from four wind speed distributions and begin to understand the effects of using a certain model in an electricity grid power flow simulation. Results demonstrate that the choice of stochastic wind model can have a noteworthy impact on the amount of additional traditional generation required in the system, and therefore great effort must be made to ensure that any model used is an accurate representation of the wind resources in the study area.

ACKNOWLEDGMENTS

Throughout my time in the Master of Engineering program I have received a great amount of support and assistance. I would first like to thank my primary project advisor, Professor Lindsay Anderson. Professor Anderson's expertise in power flow modeling and wind speed statistics were invaluable for this project, and her support with defining the scope of my project was incredibly helpful. Professor Anderson also served as my academic advisor and ensured that the courses I was taking fit my interests and would help me to develop as a professional.

I would also like to thank my research supervisor, Dr. Zongjie (Lisa) Wang, a Postdoctoral Associate in Professor Anderson's lab. Her insights were very useful for framing my project plans and understanding what needed to be achieved. Lisa's technical acumen was also of great assistance when I needed help, and she was always very eager to do so.

I also wish to acknowledge the great strides made by Trisha Ray, a 2019 graduate of Cornell University and former member of Professor Anderson's lab, which made this project possible. Her framework on power flow modeling allowed me to get the results I needed in a one-year program, which would not have been possible otherwise.

Lastly, I would like to thank those who have supported me throughout my life and during my master's program, from friends and family to fellow students. Their encouragement has certainly helped me through some busy times.

TABLE OF CONTENTS

I. INTRODUCTION

The threat of climate change is more pressing now than ever. Global temperatures are rising, extreme weather events are increasing in frequency, and polar ice caps are melting. Unless serious action is taken soon, the future may be bleak for humans and for the millions of other species on this planet. One effort to curtail climate change is to reduce emissions by replacing electricity generated through burning fossil fuels with clean, renewable sources. A form of renewable generation that has shown significant promise is wind energy. In terms of potential alone, there is more kinetic energy in the atmospheric boundary layer from wind than the total global energy demand.[1] Wind turbines provide a method to capture this energy and convert it into electricity that can be used in the same way as any other form of generation feeding into the electric grid. Using an aggressive wind turbine deployment strategy could reduce emissions to such an extent that the two-degree Celsius warming threshold for dangerous climate change, set by the Intergovernmental Panel on Climate Change (IPCC), could be avoided entirely.[1] One of the issues with large-scale deployment of wind energy is its intermittency, as fluctuations in wind speed result in different, largely unpredictable outputs from wind turbines. In order to have effective electricity grids with a large percentage of energy derived from wind, we must understand the variation of wind speed so that electricity can be properly dispatched to minimize cost. There are several statistical models used to characterize the wind speed, and although much previous work compares the fit of these models to the observed wind speed data, the models are not often compared side by side in a power flow simulation. This project will compare using four wind speed distributions to generate different stochastic wind power models that will be used in an economic dispatch model to begin to understand the differences between the models in a power flow.

II. WIND SPEED DISTRIBUTIONS

The most common wind speed model is the Weibull Distribution. This two-parameter model was developed by Waloddi Weibull in 1951 to be applied to a variety of fields, including strength of materials.[2] The Weibull Distribution is defined by one shape parameter (k) and one

scale parameter (c). The probability density function (pdf) and cumulative density function (cdf) are given by

$$f_{Weibull}(v) = \frac{k}{c}\left(\frac{v}{c}\right)^{k-1} exp\left(-\left(\frac{v}{c}\right)^{k}\right)$$

$$F_{Weibull}(v) = 1 - exp\left(-\left(\frac{v}{c}\right)^{k}\right)$$

where v is the wind speed, as will be the case for all distributions introduced in this section.[3] There are several methods for determining the shape and scale parameter from wind speed data, but the simplest involves two formulas given as

$$k = (\sigma/\bar{v})^{-1.086}$$

$$c = \frac{\bar{v}}{\Gamma(1+\frac{1}{k})}$$

where $\bar{v}$ is the average wind speed and σ is the standard deviation of the data.[3] The Weibull Distribution tends to fit typical wind speed distributions very well, as is seen through Figure 1.[4]

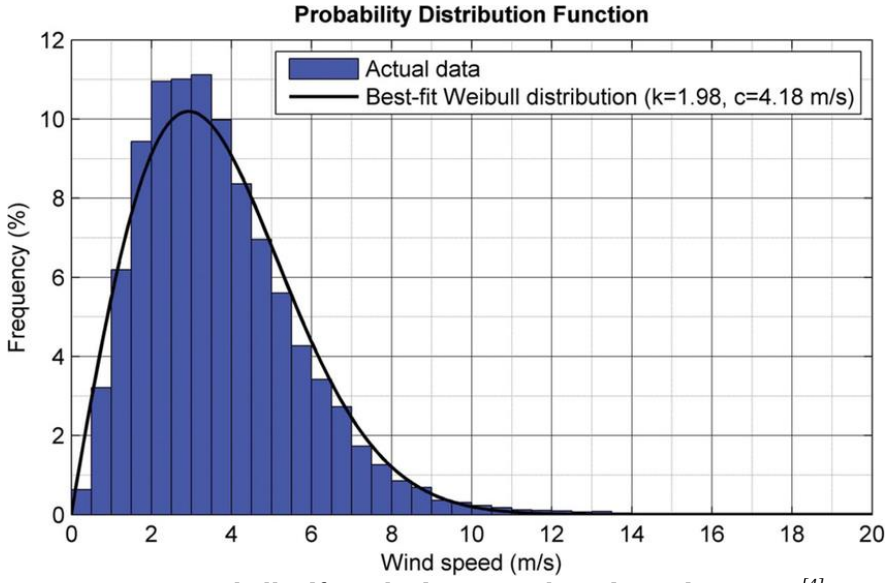

Figure 1: Weibull pdf overlaid on typical wind speed statistics[4]

The Rayleigh Distribution is a simplified version of the Weibull Distribution, in the case where $k = 2$ and $c = \sigma\sqrt{2}$. This results in a simple distribution, which can be solved solely using the average wind speed. The pdf and cdf for the Rayleigh Distribution are given by[5]

$$f(v) = \frac{\pi}{2}\left(\frac{v}{\bar{v}^2}\right) exp\left[-\frac{\pi}{4}\left(\frac{v}{\bar{v}}\right)^2\right]$$

$$F(v) = 1 - exp\left[-\frac{\pi}{4}\left(\frac{v}{\bar{v}}\right)^2\right].$$

Although the Weibull Distribution has become the standard for modeling wind speeds, this does not necessarily mean that it is the most accurate. There are other common two-parameter distributions, as well as additional distributions with a greater number of parameters used to characterize wind speeds.

The Lognormal Distribution, developed by Francis Galton in 1879, is another two-parameter statistical distribution that is used to characterize wind speed.[6] This distribution is defined by the mean (μ) and the standard deviation (σ). The Lognormal pdf and cdf are given by

$$f_{Lognormal}(v) = \frac{1}{\sigma v\sqrt{2\pi}} exp\left[\frac{-(\ln(v) - \mu)^2}{2\sigma^2}\right]$$

$$F_{Lognormal}(v) = \frac{1}{2} + \frac{1}{2} erf\left[\frac{\ln(v) - \mu}{\sigma\sqrt{2}}\right]$$

where $erf()$ is the error function from the normal distribution.[7] When applied to wind speed data, Lognormal distributions typically have better fits when method of moments estimators are employed for μ and σ as opposed to maximum likelihood estimators. This method will therefore be used in this research, with $\hat{\mu}$ and $\hat{\sigma}$ being calculated as

$$\hat{\mu} = ln\left(\frac{\bar{v}}{\sqrt{1 + \frac{S^2}{\bar{v}^2}}}\right)$$

$$\hat{\sigma} = \sqrt{\ln\left(1 + \frac{S^2}{\bar{\bar{v}}^2}\right)}$$

where S^2 is the variance of the data set.[7]

The final two-parameter distribution to be considered is the Beta Distribution, which was developed by Karl Pearson in 1904.[8] The convention for the Beta Distribution differs from the previous distributions as the pdf is not for the wind speed directly, but rather a dimensionless version of the wind speed. This dimensionless quantity, x, is defined as

$$x = \frac{\bar{v}}{v_{max}}$$

where $\bar{v}$ is the average wind speed and v_{max} is the maximum wind speed for the time interval under investigation. It then must be true that the value of x must always be between 0 and 1. There are two parameters, α and β that are used to define the Beta Distribution. The pdf is given in terms of x as

$$f_{Beta}(x) = x^{\alpha-1}(1-x)^{\xi-1}.$$

The parameters α and β are defined as

$$\alpha = \frac{1}{1+\eta}\left(\frac{\eta}{I} - 1\right)$$

$$\xi = \eta\alpha$$

where

$$\eta = \frac{v_{max} - \bar{v}}{\bar{v}}$$

$$I = \frac{\sigma^2}{\bar{v}^2}$$

and σ is the standard deviation of the wind speed data.[9]

The two-parameter statistical distributions typically tend to fit very well for most wind speed locations. There are instances, however, when the wind speed data follows an abnormal

distribution. One common such distribution occurs when there is a large spike of low wind speeds, and then a long tail trailing into the higher wind speeds.[10] Figure 2 displays this abnormal wind speed characteristic.

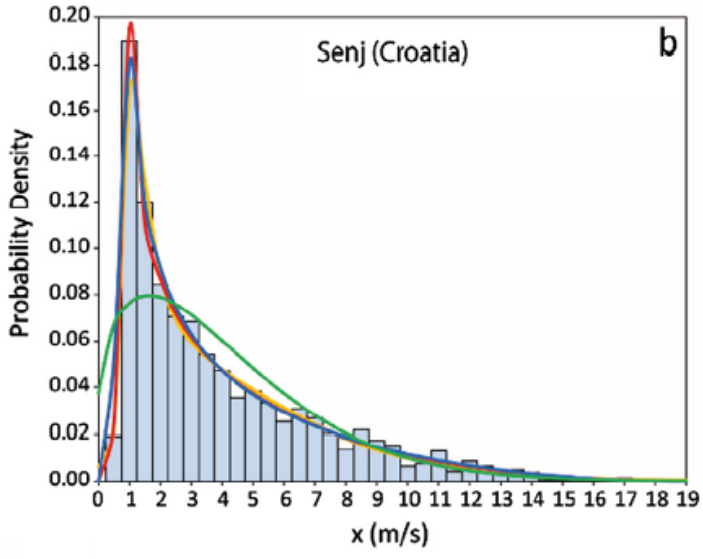

Figure 2: Long-tail wind speed data from a site in Croatia[10]

In circumstances such as these, a two-parameter statistical model may result in misleading distributions, and it is necessary to use a statistical model with a greater number of parameters.[10] The four-parameter Kappa Distribution is one such model. It was developed by J.R.M. Hosking in 1994 as a variation of the three-parameter Kappa Distribution and has since been applied to characterize wind speeds.[11] The distribution is characterized by the parameters ξ, α, k, and h. The pdf and cdf are given by[12]

$$f_{Kappa}(v) = \alpha^{-1}\left[\frac{1-k(v-\xi)}{\alpha}\right]^{\frac{1}{k-1}} F(v)^{1-h}$$

$$F_{Kappa}(v) = \left\{1 - h\left[1 - \frac{k(v-\xi)}{\alpha}\right]^{\frac{1}{k}}\right\}^{\frac{1}{h}}.$$

The four parameters are not solved for directly, but rather are estimated by solving relationships between the parameters, the sample L-moments (λ_1, λ_2), and the L-moment ratios (τ_1, τ_2). The following system of equations must be solved

$$\lambda_1 = \hat{\xi} + \frac{\hat{\alpha}(1 - g_1)}{\hat{k}}$$

$$\lambda_2 = \frac{\hat{\alpha}(g_1 - g_2)}{\hat{k}}$$

$$\tau_3 = \frac{-g_1 + 3g_2 - 2g_3}{g_1 - g_2}$$

$$\tau_4 = \frac{-g_1 + 6g_2 - 10g_3 + 5g_4}{g_1 - g_2}$$

where

$$g_r = \begin{cases} \dfrac{r\Gamma(1 + \hat{k})\Gamma(r/\hat{h})}{\hat{h}^{1+\hat{k}}\Gamma(1 + \hat{k} + r/\hat{h})}, & \hat{h} > 0 \\ \dfrac{r\Gamma(1 + \hat{k})\Gamma(-k - r/\hat{h})}{(-\hat{h})^{1+\hat{k}}\Gamma(1 - r/\hat{h})}, & \hat{h} < 0. \end{cases}$$

An iterative process is used to solve for the parameters that can then be input into the pdf to model the wind speed characteristics.[12]

Another model that can capture the long-tail wind speed behavior is the Wakeby Distribution. This five-parameter distribution was developed by John C. Houghton in 1978.[13] The Wakeby distribution is defined by the parameters $\xi, \alpha, \beta, \gamma$, and δ.[13] It is generally represented by the inverse cdf, which is given as[12]

$$F^{-1}_{Wakeby}\big(F(v)\big) = \xi + \frac{\alpha}{\beta}\Big[1 - \big(1 - F(v)\big)^{\beta}\Big] - \frac{\gamma}{\delta}\Big[1 - \big(1 - F(v)\big)^{-\delta}\Big].$$

As is the case with the Kappa distribution, the parameters in the Wakeby distribution cannot be solved for directly. The method to estimate the parameters depends on the sample L-moments. The four sample L-moments, $\lambda_1, \lambda_2, \lambda_3$, and λ_4, are given as

$$\lambda_1 = \hat{\xi} + \frac{\hat{\alpha}}{1+\hat{\beta}} + \frac{\hat{\gamma}}{1-\hat{\delta}}$$

$$\lambda_2 = \frac{\hat{\alpha}}{(1+\hat{\beta})(2+\hat{\beta})} + \frac{\hat{\gamma}}{(1-\hat{\delta})(2-\hat{\delta})}$$

$$\lambda_3 = \frac{\hat{\alpha}(1-\hat{\beta})}{(1+\hat{\beta})(2+\hat{\beta})(3+\hat{\beta})} + \frac{\hat{\gamma}(1+\hat{\delta})}{(1-\hat{\delta})(2-\hat{\delta})(3-\hat{\delta})}$$

$$\lambda_4 = \frac{\hat{\alpha}(1-\hat{\beta})(2-\hat{\beta})}{(1+\hat{\beta})(2+\hat{\beta})(3+\hat{\beta})(4+\hat{\beta})} + \frac{\hat{\gamma}(1+\hat{\delta})(2+\hat{\delta})}{(1-\hat{\delta})(2-\hat{\delta})(3-\hat{\delta})(4-\hat{\delta})}.$$

for $\delta > 0$.[14] The parameters are then estimated by following an algorithm, with the full details outlined in Hosking and Wallis.[15]

There is one uncommon wind speed distribution shape that even the four-parameter Kappa and five-parameter Wakeby cannot capture. This is the bimodal wind speed distribution, where there are two spikes around two separate wind speeds, as displayed in Figure 3.[10]

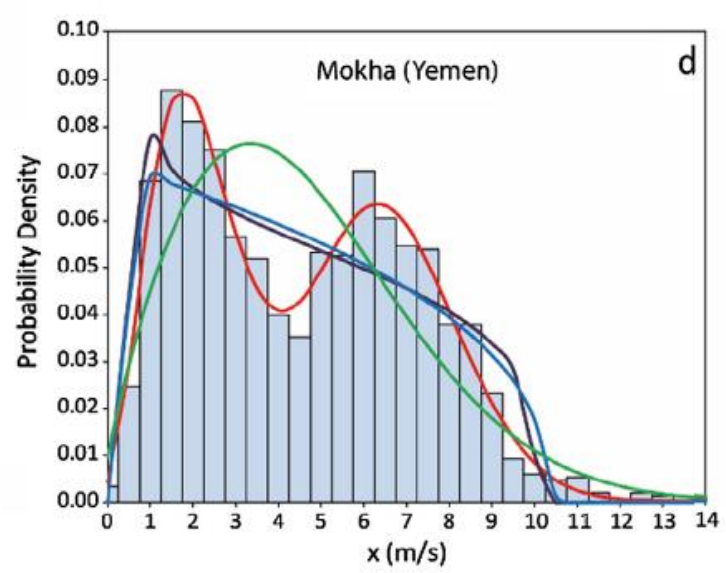

Figure 3: Bimodal wind speed data from a site in Yemen[10]

In this rare circumstance, a seven-parameter model must be used. The seven-parameter Burr-Generalized Extreme Value (B-GEV) mixture pdf can accurately depict the bimodal behavior

through a statistical distribution. This distribution implements the Burr function as a generalized extreme value mixture pdf and was developed by Christopher Jung in 2017 specifically for the purposes of characterizing bimodal wind speed distributions. The pdf and cdf are a combination of the Burr and Generalized Extreme Value (GEV) pdfs and cdfs. The full B-GEV distribution is characterized by the Burr parameters o, σ, and χ, the GEV parameters η, ι, and μ, and the B-GEV parameter ω. The Burr pdf and cdf are given as

$$f_{Burr}(v) = \frac{o\chi(\frac{v}{\sigma})^{o-1}}{\sigma\left[1+\left(\frac{v}{\sigma}\right)^{o}\right]^{\chi+1}}$$

$$F_{Burr}(v) = 1 - \left[1+\left(\frac{v}{\sigma}\right)^{o}\right]^{-\chi}.$$

and the GEV pdf and cdf are given as

$$f_{GEV}(v) = \frac{1}{\eta}\left[1-\frac{\iota}{\eta}(v-\mu)\right]^{\frac{1}{\iota}-1} exp\left\{-\left[1-\frac{\iota}{\eta}(v-\mu)\right]^{\frac{1}{\iota}}\right\}$$

$$F_{GEV}(v) = exp\left\{-\left[1+\iota\frac{(v-\mu)}{\eta}\right]^{-\frac{1}{\iota}}\right\}.$$

The full B-GEV pdf and cdf is then given as[10]

$$f_{B,GEV}(v) = \omega f_{Burr}(v) + (1-\omega) f_{GEV}(v)$$

$$F_{B,GEV}(v) = \omega F_{Burr}(v) + (1-\omega) F_{GEV}(v).$$

Estimation of the B-GEV parameters is achieved by the least-squares estimation method using the Levenberg-Marquadt algorithm. The full algorithm details are expanded upon in Marquadt's 1963 paper.[15]

This is by no means an exhaustive list of statistical distributions used to model wind speeds. For this research, the focus will be on the four distributions with one or two parameters, namely

the Weibull, Rayleigh, Lognormal and Beta distributions. Though the Kappa, Wakeby, and B-GEV distributions can prove useful for uncommon wind speed scenarios, the focus of this research is to compare the results of an economic dispatch model using various wind speed distributions for stochastic modeling, and therefore the addition of the more complex distributions does not add much value. Direct comparison of the four common distributions and how they affect a power flow system with all other conditions being the same can give insight as to how differences in statistical models with similar structures can affect the system.

III. WIND SPEED TO WIND POWER

While the statistical distributions give insight into the probability of observing particular wind speeds, they do not actually predict the power generated from a wind turbine generator. The power produced will vary depending on the wind turbine model, and the resultant power curve can be divided into three regions, displayed by Figure 4. The relationship between wind speed and wind power is

$$P = \begin{cases} 0, & v < v_{ci} \ or \ v > v_{co} \\ P_r\left(\dfrac{v - v_{ci}}{v_r - v_{ci}}\right), & v_r \le v \le v_r \\ P_r, & v_r \le v \le v_{co} \end{cases}$$

Where P is the power produced, P_r is the wind turbine rated power, v_r is the rated wind speed, v_{ci} is the cut-in wind speed, and v_{co} is the cut-out wind speed.[16]

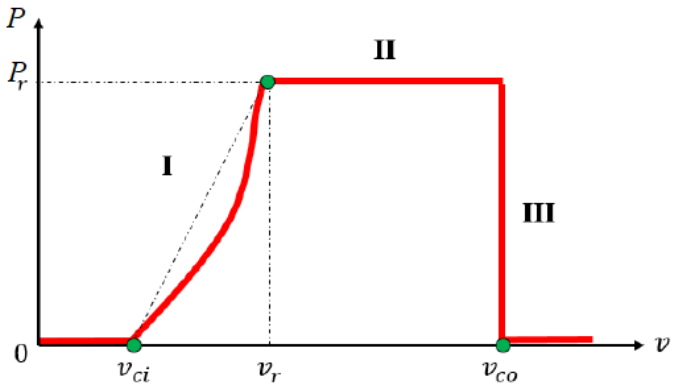

Figure 4: Wind power curve[16]

The wind speed distribution from any statistical model can work in tandem with the manufacturer's power curve for a given wind turbine generator to calculate the probability of producing a given power output.

IV. WIND SPEED DATA

Time series wind speed data is a vital component for carrying out a statistical analysis and developing a stochastic power flow model. The model for this research is based on the New England power grid, as will be explained in Section V, so wind speed data from New England is used. Selecting a data set that is representative of the wind patterns in the entire New England area is impossible, however, by selecting several distinct data sets in regions with good wind resources and in the general geographic areas where wind farms have been installed, wind patterns can be fairly represented. Wind speed data was gathered from the National Renewable Energy Laboratory's (NREL) Wind Prospector Tool. The data set contains the wind speed at 100 meters for 5-minute intervals for the years 2007-2012.[17] Figure 5 displays the wind resource from the NREL Wind Prospector and Figure 6 displays the installed wind turbines, both for the New England area. Four geographic locations were chosen for wind speed data collection. Each location has above average resources (with respect to the New England area), is in a general area where turbines have been installed, and each of the four locations are distinct in their resources. The locations are labeled in Figures 5 and 6 from one to four.

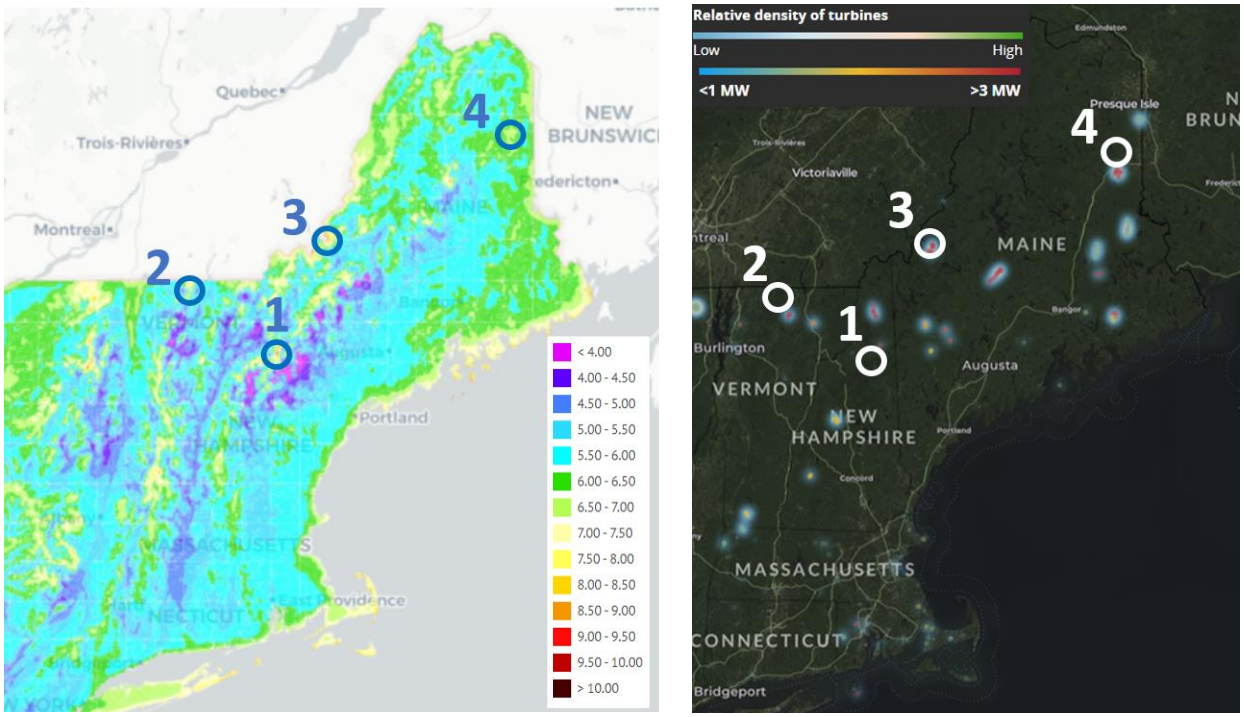

Figure 5 and Figure 6: Wind speed resources based on average wind speeds displayed in the colors based on the key in meters per second (left) and installed wind turbines with relative density and size of the turbine displayed on the key (right).[17],[18] *The location of the data used is shown in the blue circles on both figures.*

The main takeaway from observing these wind speed data sets is that wind speed is extremely variable. Therefore, it is desired to model wind power as stochastic. The variability of the data set can be seen in Figure 7, which displays the wind speed and wind power for an example 2-megawatt (MW) turbine for the beginning of the year 2012 using the aforementioned New England data from Location 1. Only Location 1 data is displayed in this figure as it represents the variability of wind in a given location.

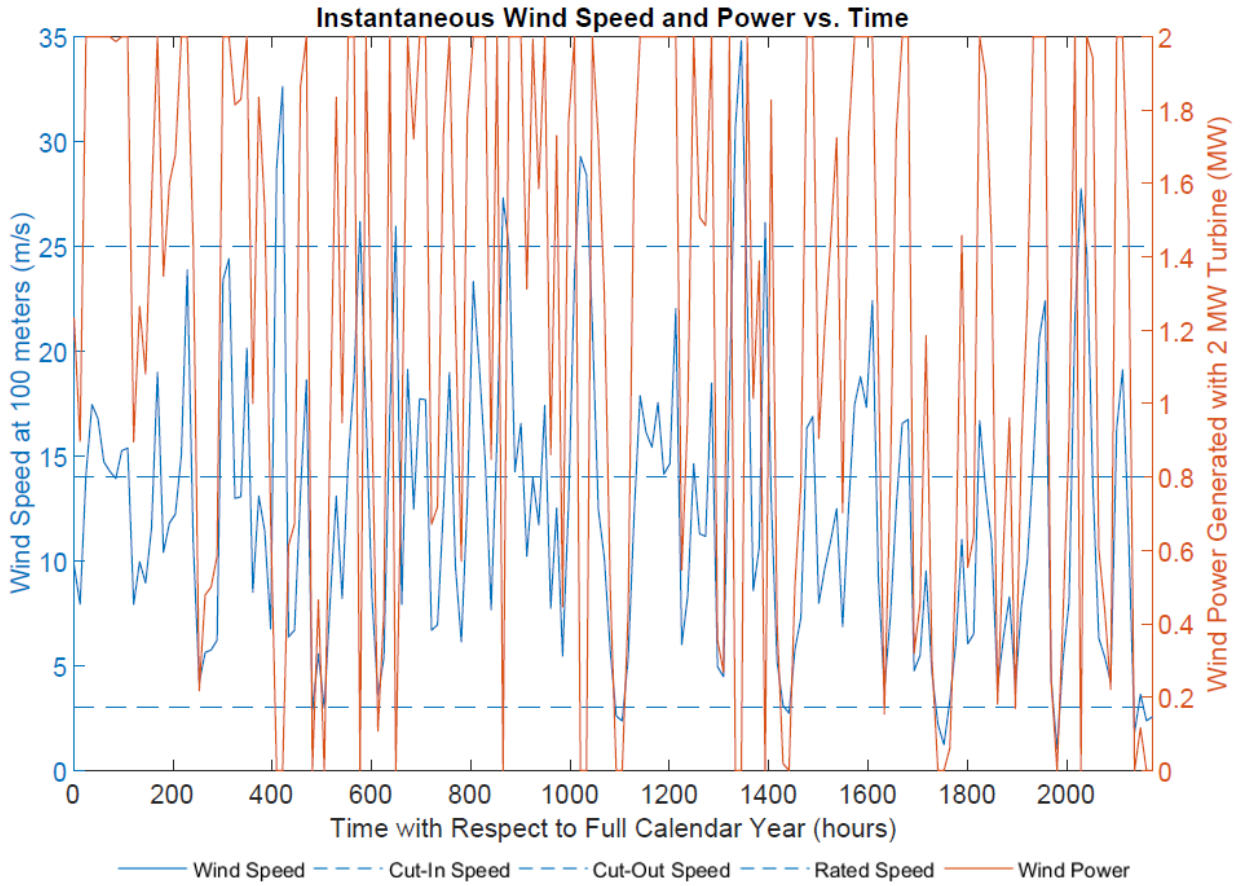

Figure 7: Instantaneous wind speed at 100 meters and power genereated by a 2 MW turbine with 12-hour averages for winter 2012. The left axis measures instantaneous wind speed and the right axis measures instantaneous wind power generated by a 2 MW turbine.

It is difficult to gain any insight from the instantaneous values for wind speed and wind power, but this data can be used to develop statistical distributions that can be used to help predict wind speed. Using all the locations' data sets, the probability density function is generated for the Weibull, Rayleigh, Lognormal, and Beta distributions. Figure 8 displays the four pdfs created from the full data set for comparison.

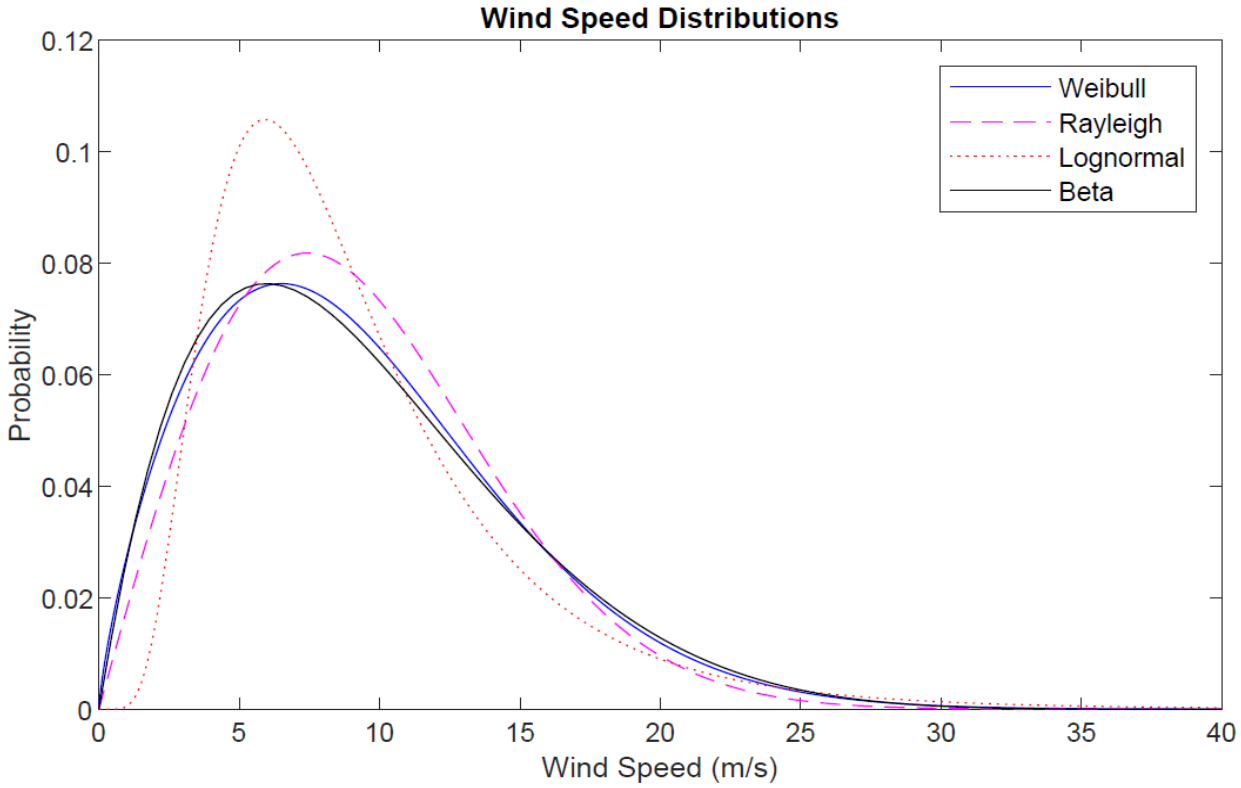

Figure 8: Probability density functions for Weibull, Rayleigh, Lognormal, and Beta distributions using Location 1 data

Figure 8 clearly displays the differences in the pdfs for each of these distributions, which can allow for insight into their behavior. The Weibull and Beta distributions appear to be the most similar of the group, with a leftward peak around five to eight meters per second wind speed and a long tail trailing off toward the higher wind speeds. The Beta distribution has slightly higher probabilities for the lower wind speeds, slightly lower probabilities for the average wind speeds, and slightly higher probabilities for the higher wind speeds when compared to the Weibull, however. The Rayleigh distribution, on the other hand, has lower probabilities for the lower wind speeds compared to the Weibull and Beta, and it also has a peak at a higher wind speed of nine to ten meters per second. The Rayleigh distribution has high probabilities for the average wind speeds, and then lower probabilities for the higher wind speeds. Lastly, the Lognormal distribution is the most visually dissimilar from the other three. It has a very sharp peak at roughly six meters per second wind speed, resulting in very high probabilities for the average wind speeds. As a result,

the Lognormal has quite low probabilities for lower and higher wind speeds compared to the other distributions. The exception is the Rayleigh distribution, which has even lower probabilities than the Lognormal for very high wind speeds above roughly 21 meters per second.

It is also important to understand the impacts of using these distributions to calculate on wind power generated. As a test case, the cumulative power distribution for all locations' data sets using each of the four distributions is displayed in Figure 9.

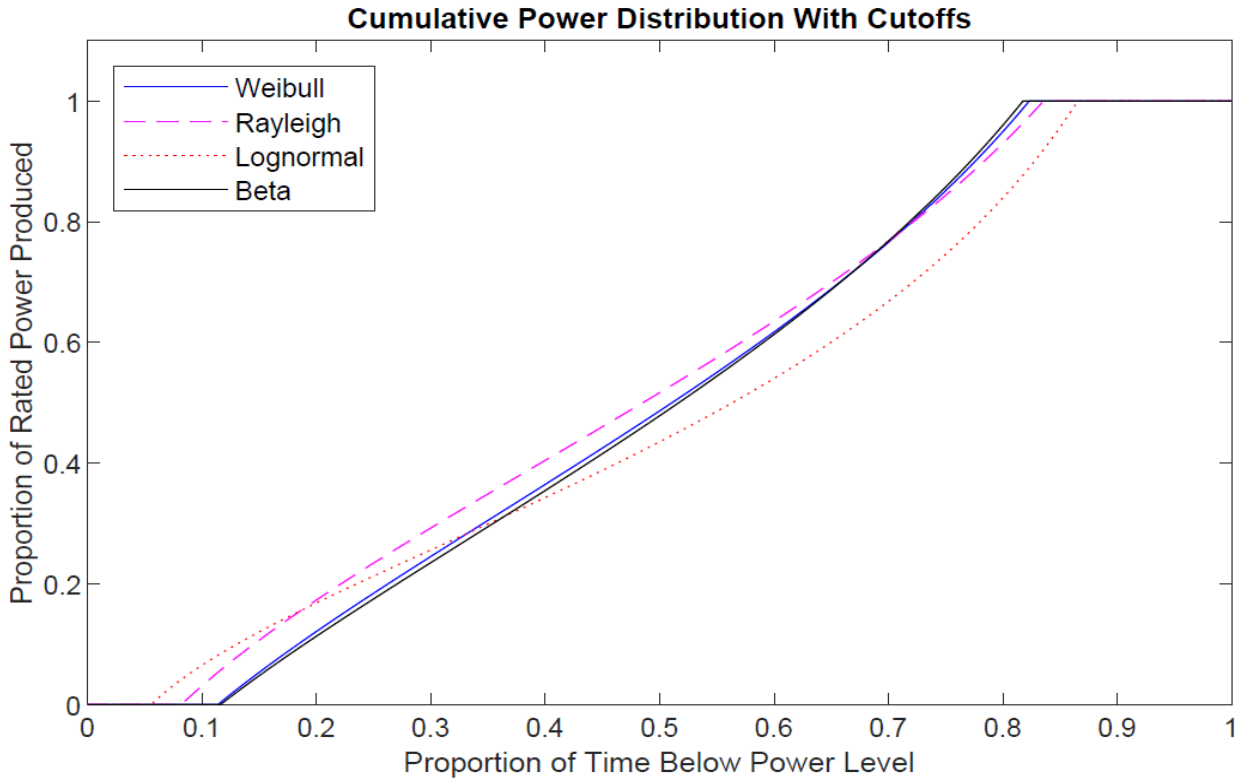

Figure 9: Cumulative power distribution for a 2 MW turbine using Weibull, Rayleigh, Lognormal, and Beta distributions. X-axis displays the proportion of time that the turbine will produce below the corresponding power output in terms of proportion of rated power on the Y-axis.

Figure 9 displays that the Rayleigh and Lognormal distribution have the lowest proportion of time producing zero power. This is a result of these distributions having lower probabilities for low wind speeds, and higher probabilities for the average speeds. The Rayleigh distribution tends to result in a higher power produced for a given proportion of time below that power level when compared to the Lognormal distribution, a result of the sharper peak around lower wind speeds for the Lognormal. The Rayleigh distribution also has a much higher proportion of time above the

rated power output than the Lognormal, with roughly 18% compared to only 12%. This is a result of the higher probabilities of the Lognormal for very high wind speeds above the cut-out speed, compared to the Rayleigh which has very low probabilities above cut-out. The Weibull and Beta distributions again show fairly similar results. They produce zero power approximately 12% of the time, and the rated power approximately 20% of the time. The Weibull results in higher power produced for a given proportion of time below that power level until the 0.68 proportion, when the Beta produces higher power for the remainder of the cumulative power distribution.

Through comparing the results from each of these four distributions when given the same data set, their behavior with respect to one another is better understood. It is also useful to dive slightly deeper into each individual distribution to study how it is affected by changing the parameters for each respective distribution. For each distribution, the actual pdf from the Location 1 data set will be plotted, along with three additional pdfs generated by altering one of the parameters and keeping all others constant. Figures 10 and 11 display the effects of changing the shape and scale parameters on the Weibull pdf.

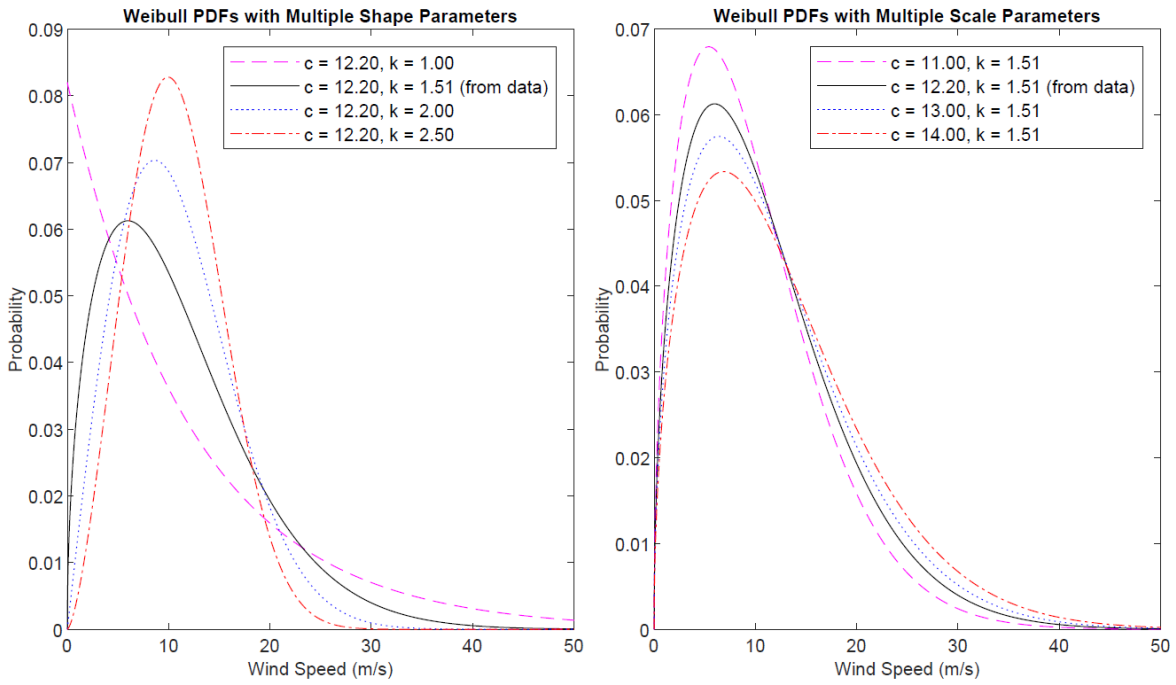

Figure 10 and Figure 11: Effect of changing the shape parameter on a Weibull pdf (left) and effect of changing the scale parameter on a Weibull pdf (right)

Figure 10 displays that as the shape parameter is increased, the peak of the pdf shifts to the right and corresponds to higher probabilities. This shifting of the peak results in lower probabilities for lower wind speeds and higher wind speeds, as more probability is focused around the average. The exception is when the shape parameter equals one and there is no peak. Figure 11 displays that as the scale parameter is increased, the entire pdf is "stretched" upward and slightly to the left. This results in higher probabilities around the average and lower wind speeds and lower probabilities around the higher wind speeds.

The Rayleigh distribution, as mentioned in Section II, only depends on the mean wind speed. Figure 12 displays the effects of altering the mean wind speed on the Rayleigh pdf.

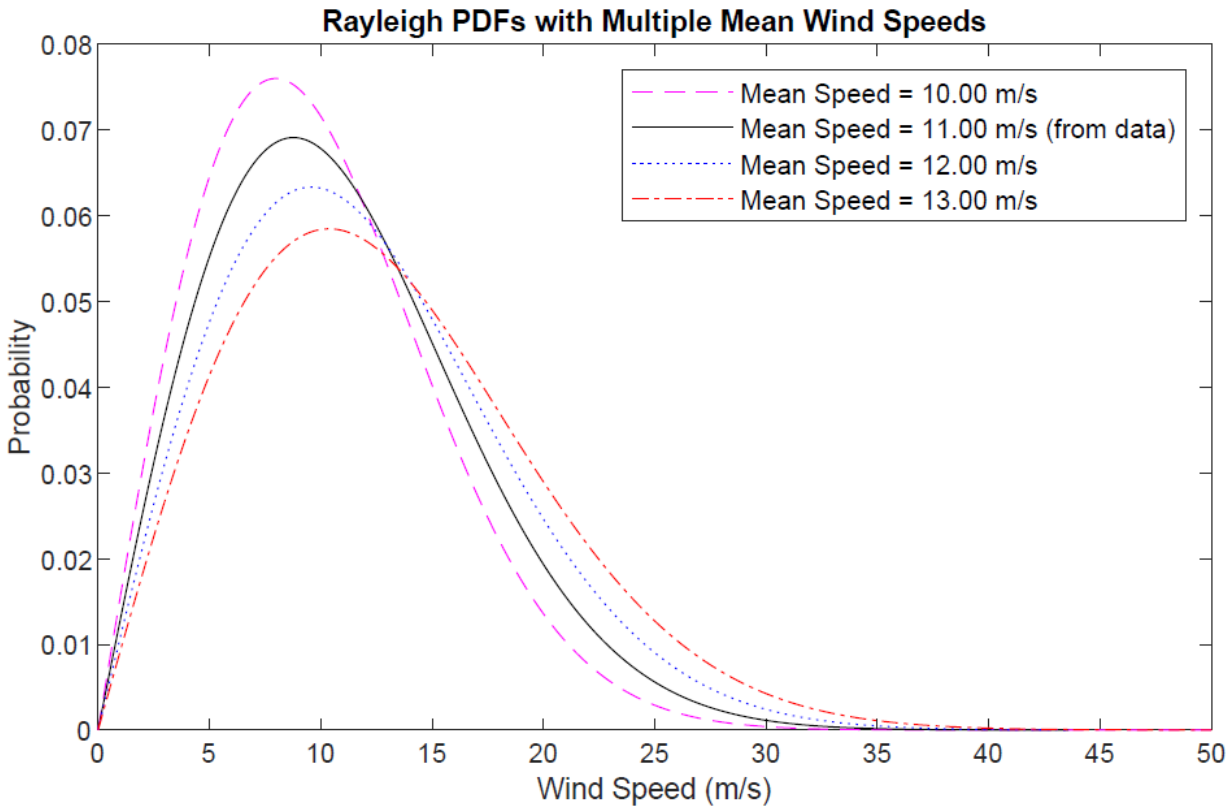

Figure 12: Effect of changing the mean wind speed on a Rayleigh pdf

Figure 12 shows that higher mean wind speeds result in peaks that are at lower probabilities and higher wind speeds. The result is that Rayleigh distributions with higher mean wind speeds have

a larger range of wind speeds, lower probabilities for lower wind speeds, and higher probabilities for higher wind speeds.

Figures 13 and 14 display the effects of modifying the $\hat{\mu}$ and $\hat{\sigma}$ estimators on the Lognormal distribution pdf.

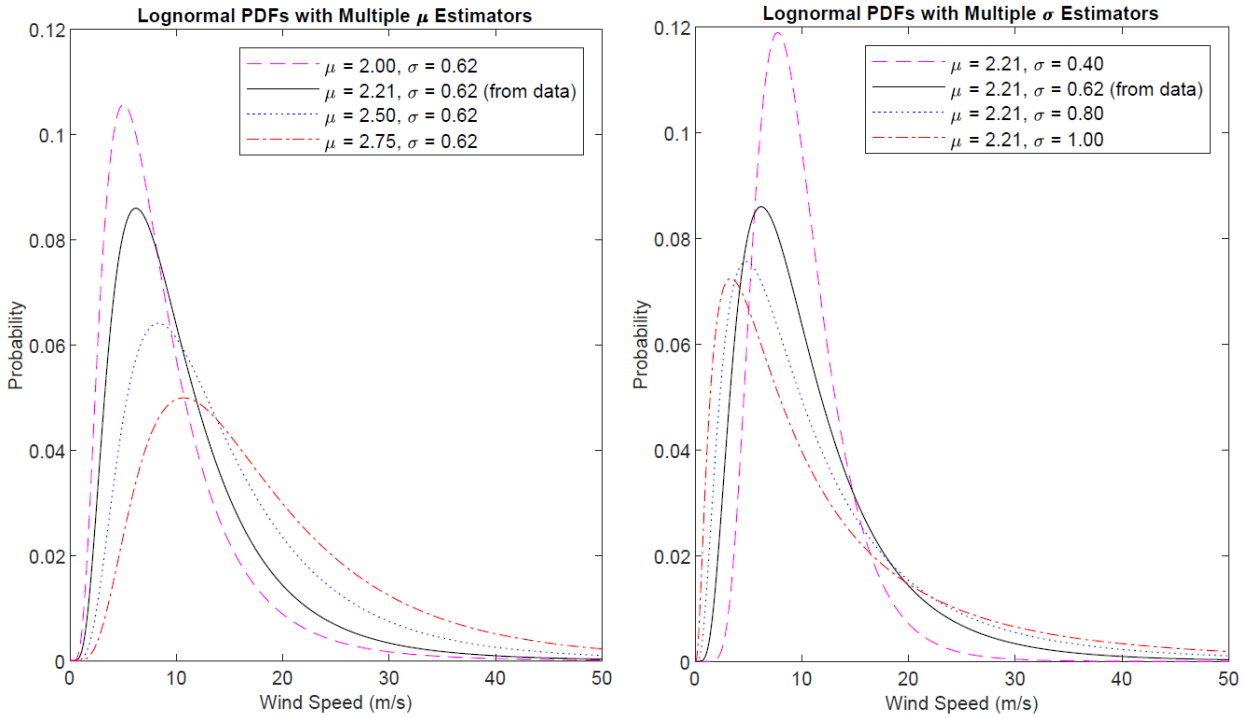

Figure 13 and Figure 14: Effect of changing the $\hat{\mu}$ estimator on a Lognormal pdf (left) and effect of changing the $\hat{\sigma}$ estimator on a Lognormal pdf (right)

Figure 13 displays that as the $\hat{\mu}$ estimator is increased, the peak of the pdf shifts to the right and down. This shifting results in less sharp peaks and corresponding higher probabilities for higher wind speeds and lower probabilities for lower wind speeds. Figure 14 displays that as the $\hat{\sigma}$ estimator is increased, the peak of the pdf moves down and to the right, resulting in higher probabilities for lower wind speeds. Increasing the $\hat{\sigma}$ estimator also results in a longer tail, generating higher probabilities for higher wind speeds as well. The decrease in probability is seen around the average wind speeds for an increase in the $\hat{\sigma}$ estimator.

Lastly, Figures 15 and 16 display the effects of modifying the α and ξ parameters on the Beta distribution pdf.

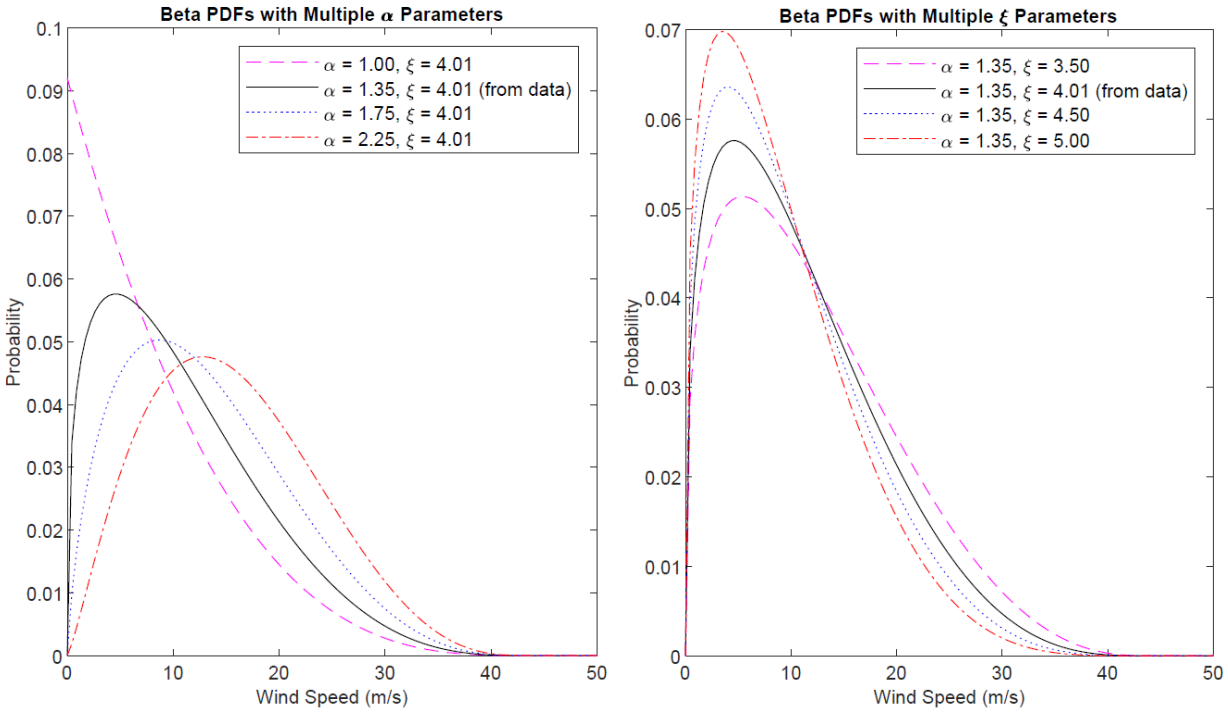

Figure 15 and Figure 16: Effect of changing the α parameter on a Beta pdf (left) and effect of changing the ξ parameter on a Beta pdf (right)

Figure 15 shows that as the α parameter is increased, the peak of the pdf becomes less sharp and is shifted down and to the left. This results in the distribution looking more like a normal distribution having higher probabilities for higher wind speeds and lower probabilities for lower wind speeds. The exception is when the α parameter equals one and there is no peak. Figure 16 shows that as the ξ parameter is increased, the pdf is shifted up and to the left. This results in lower probabilities for higher wind speeds and higher probabilities for lower wind speeds, as well as increased probabilities around the average wind speeds.

V. MODELING THE NEW ENGLAND POWER GRID

This research uses the New England power grid system as its test case. The IEEE 39 Bus Case System is commonly used as a representation of the New England grid with 10 generators and 46 lines. The grid system is broken up into eight different load zones which are geographically grouped together and represent an aggregate of load nodes used for wholesale electricity market pricing. Figure 17 displays the load breakdown.

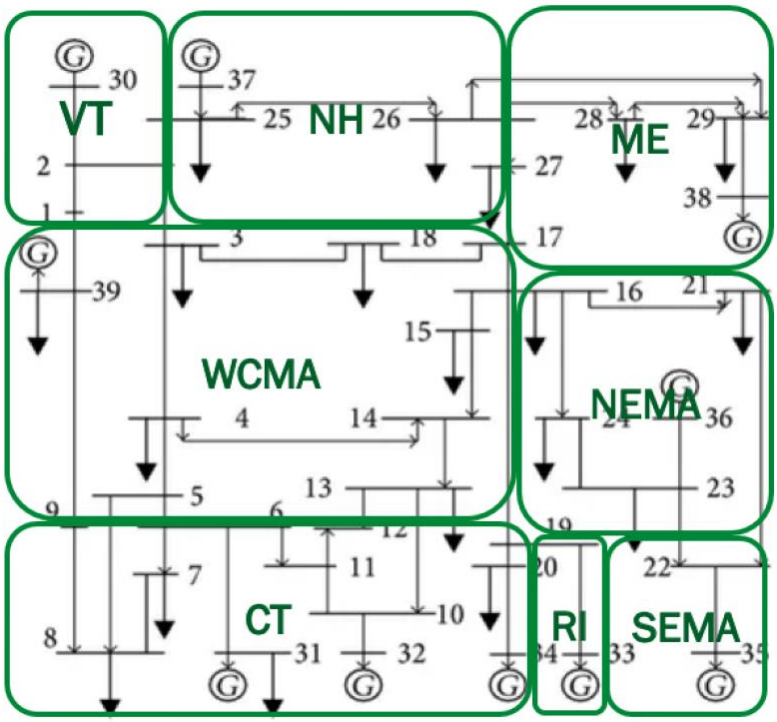

Figure 17: Load Zone Breakdown of IEEE 39 Bus Case System

Considering that the purpose of this project is to compare the impact of using different stochastic methods for wind speed on unit commitment, the penetration of wind generation will need to be quite high in order to see any differences between methods. ISO New England (ISO-NE), the New England interconnection, holds an annual Forward Capacity Auction (FCA) that breaks down the generation type for each of their eight load zones. The results of the FCA #12 are the most up to date and are used as a baseline for the generation mix in this research. Figure 18 displays the FCA #12 Generation Mix. The generation mix used in this particular model is a variation of the FCA #12 mix, as displayed in Figure 19.

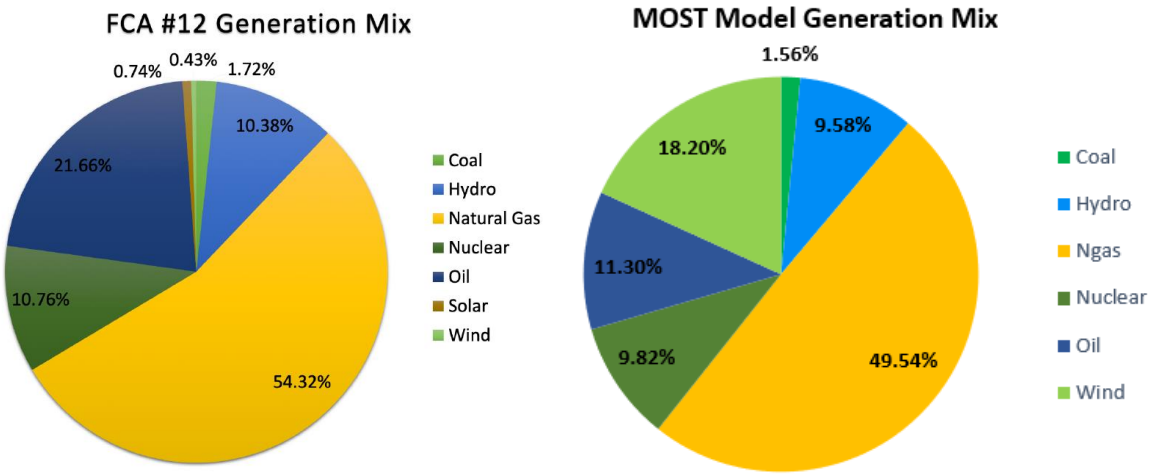

Figure 18 and Figure 19: FCA #12 Generation Mix Pie Chart (left) and the makeup used in this research (right)

The primary change made to the generation mix is a substantial increase in the penetration of wind to approximately 18% of the ISO-NE generation makeup in order to better understand the effects of using different wind speed distributions to generate stochastic wind power models. Although this is a significant increase from the current amount of installed wind capacity in the New England system, it is not unthinkable. Iowa, South Dakota, Kansas, Oklahoma, and North Dakota all produce over 20% of their electricity through wind power, with Iowa leading the pack by generating 37% of its electricity with wind.[19] Additionally, since the bus case used is not designed to handle the actual capacity of the New England grid system, all generation was scaled down to 20% of the actual generation. This results in a total installed capacity of approximately 6,800 MW. This results in about a 9% increase in installed capacity from the FCA #12 Generation Mix, which serves as a buffer to account for the intermittency of wind generation. The total wind installed capacity was equally divided to the four locations of data sets displayed in Figures 5 and 6, with 310 MW per location for a total of 1,240 MW of installed wind capacity. To reflect the increase in wind penetration, approximately 640 MW of oil generation is taken offline. This is based on the trend of the decreasing amount of oil and coal generation in the New England system, as seen in Figure 20, which displays the change in the New England grid's generation since 2000. The amount of coal in the FCA #12 mix is already very small and therefore the oil is partially displaced in favor of the increased wind power.

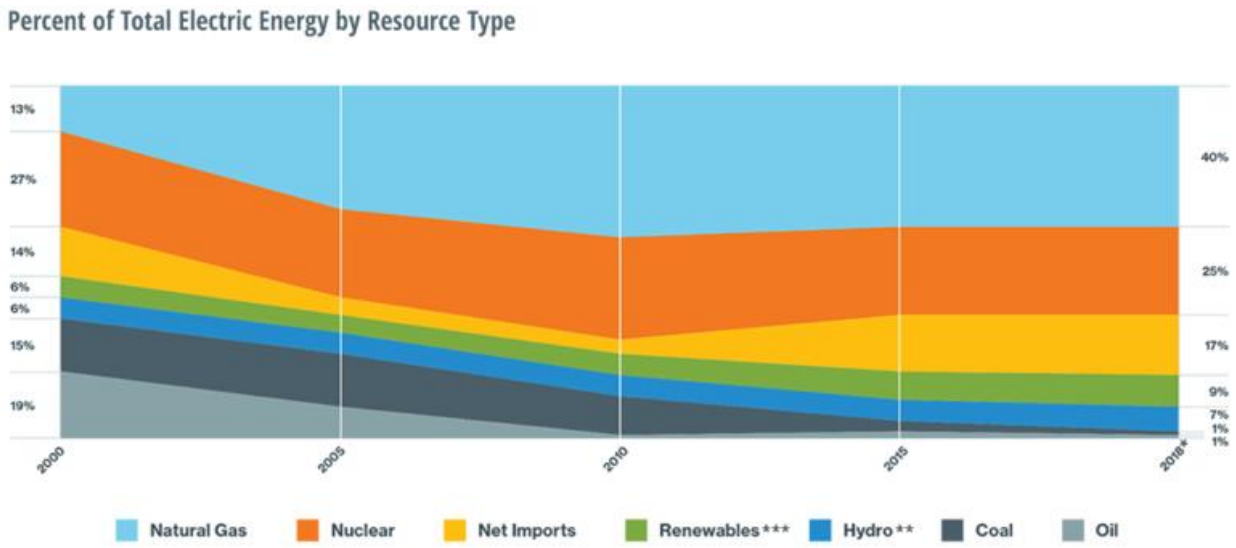

Figure 20: New England's energy mix changes since 2000[20]

The costs of generation for the model were taken from the United States Energy Information Administration (EIA).[21],[22] The EIA publishes average fossil fuel spot prices, which allows for costs to be calculated for all of the traditional generation. For the wind power and hydropower in the model, the generation costs were assumed to be zero, as the majority of generation cost comes from fuel. Figure 21 displays the average fossil fuel prices from January 2018 to January 2019.

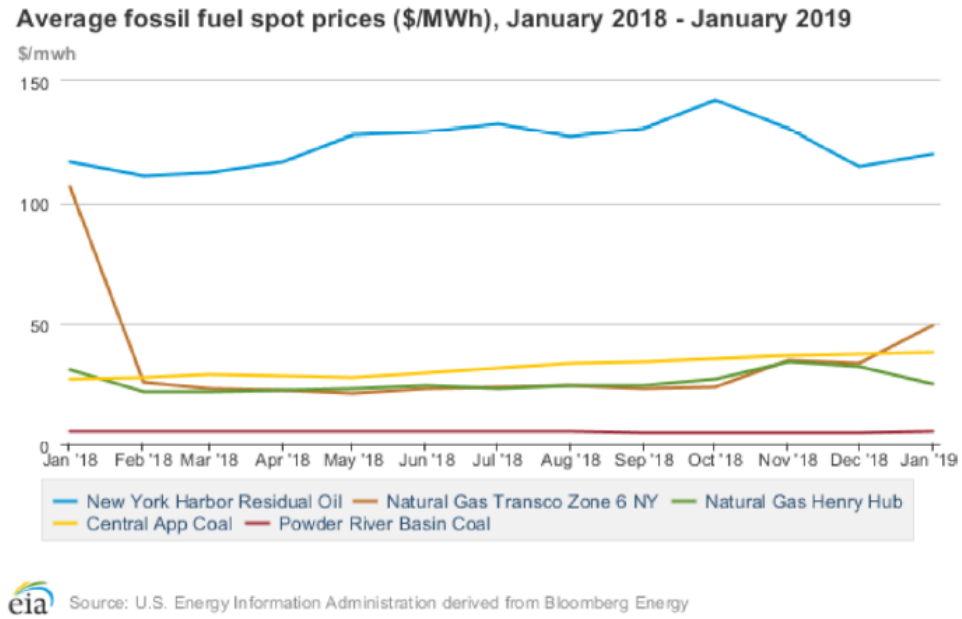

Figure 21: Average fossil fuel spot prices ($/MWh)[21]

As is evidenced by Figure 21, most fossil fuel prices stay relatively constant. The exception is natural gas, which spikes during winter. More natural gas is needed in the winter for heating, and therefore the price rises accordingly based on increased demand. The winter then represents the "worst case scenario" in terms of fuel prices and will therefore be used as the baseline for this model. This results in average fuel costs of $60/MWh for natural gas, $120/MWh for oil, $30/MWh for coal, and, from a different EIA source, $7.50/MWh for nuclear.[22]

The last component that needed to be modeled in the New England power grid, except for the stochastic wind element, which will be described in Section VII, is the load. To model the load, a profile needs to be created in order to represent the hourly changes in load. ISO-NE provides electricity load information, displaying the actual forecasted load in megawatts for each hour of the day.[23] For this model, the first day of January 2022 was used as a baseline for modeling the loads of the system. One full day, or 24-hour period, is be modeled in this study. The loads were scaled appropriately to ensure that the power flow problem would converge. The resulting 24-hour load profile in megawatts that is used in the model is displayed in Figure 22.

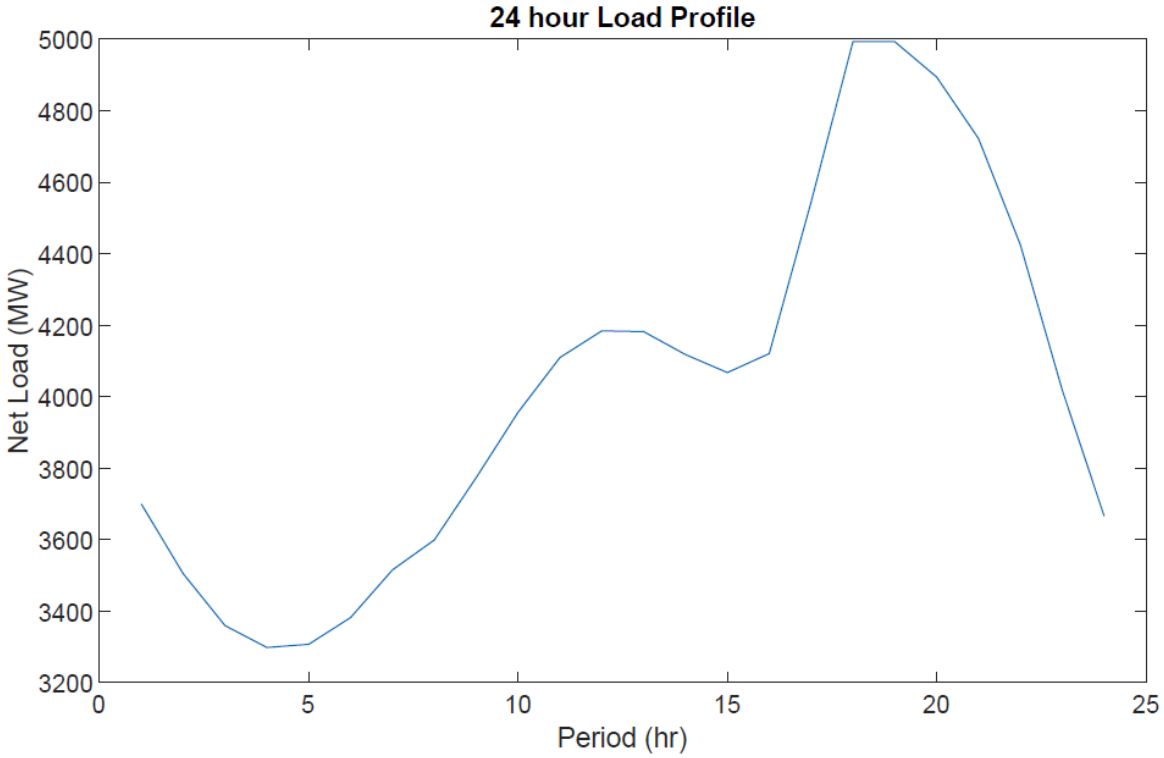

Figure 22: Scaled Load Profile of New England grid system used in model

VI. THE ECONOMIC DISPATCH PROBLEM

The power flow modeling for this research was achieved with MATPOWER, a MATLAB extension used for electric power system simulation and optimization. Much of the framework in the model was developed in previous semesters by Trisha Ray, a former researcher in the Anderson Lab. This section will summarize the model that was developed. The model was set up as a deterministic model, and the changes made to develop a stochastic model will be discussed in Section VII.

Using the MATPOWER Optimal Scheduling Tool (MOST), stochastic, multi-period economic dispatch and unit commitment problems can be solved. The input data for a MOST problem is loaded into a data struct with six inputs, not all of which are required. The inputs are as follows:

```
mdi = loadmd (mpc , transmat , xgd , sd , contab , profiles );
```

The only required input is the "mpc" file, which corresponds to the MATPOWER case. The case is the IEEE 39 Bus Case to correspond to the New England power grid, with several adjustments. The maximum real power output is based on the scaled down generation mix from Section V.[24]

For reactive power, it was assumed that all generators would be able to satisfy the typical requirement of 0.90 lag and 0.95 lead power factor.[25] This results in the maximum reactive power output of each generator being 48% of the maximum real power output and the minimum reactive power output being the negative of the 33% of maximum real power output.

The real power and reactive power demands are also updated. Real power demand is updated to reflect the load profile described in Section V, and reactive power is updated based on the assumption that the reactive power at each load bus is 30% of the real power. The final update made to the IEEE 39 Bus Case is adjusting the generation costs according to Section V.

The "xgd," or extra generation file, must also be updated. This file contains the commitment schedule and commitment key parameters to specify if a generator is available for unit commitment, as is the case for traditional fossil fuel generators, or if it must always be running, as is the case for nuclear and renewables.

The "profiles" must be included in the inputs to the MOST problem. There are two different profiles that must be included for this problem, load and wind. The load profile corresponds to the hourly system demand, as explained in Section V, and the wind profile will be discussed in detail in Section VII. The "transmat" input must also be included for stochastic MOST problems, and this input will be discussed further in Section VII.

It is worth noting that additional inputs can be included for more complex power flow problems, but they are not included in this model as this research focuses on solving an economic dispatch problem. These additional inputs include, but are not limited to, ramp rates for reserves, start-up and shutdown costs, minimum up and down times, and ramping constraints.

The economic dispatch problem is a short-term optimization problem that determines the optimal output from several generation sources to meet the load at the lowest cost. The economic dispatch problem ignores line constraints in the network and balances system load and generation supply. The formulation for a typical economic dispatch problem with wind generation is

$$Minimize \sum_{i}^{M} C_i(p_i) + \sum_{i}^{N} C_{w,i}(w_i)$$

$$Subject\ to\ constraints:$$

$$p_{i,min} \leq p_i \leq p_{i,max}$$

$$0 \leq w_i \leq w_{r,i}$$

$$\sum_{i}^{M} p_i + \sum_{i}^{N} w_i = L$$

where M is the number of traditional generators, N is the number of wind generators, C_i is the cost function of the i^{th} traditional generator, $C_{w,i}$ is the cost function of the i^{th} wind generator, p_i is the power from the i^{th} traditional generator, $p_{i,min}$ and $p_{i,max}$ are the minimum and maximum outputs for the i^{th} traditional generator as previously discussed, w_i is the scheduled power from the i^{th} wind generator, $w_{r,i}$ is the rated wind power from the i^{th} wind generator, and L is the total system load.[26] This formulation essentially states that the total cost of generation from traditional and wind generators shall be minimized subject to the constraints that each traditional generator

produces power between its minimum and maximum output, each wind generator produces power between zero and its rated output, and the sum of all power produced from traditional and wind generators equals the total system load. As mentioned in Section V, the generation cost from wind is assumed to be zero.

This framework in MOST allows for solving typical economic dispatch problems. For solving problems with stochastic wind generation, there must be additional complexity added, which is discussed in Section VII.

VII. INCORPORATING STOCHASTIC WIND POWER

The most crucial component of this project was the development of a stochastic wind component in the MATPOWER MOST model that could be uniquely generated by each of the four wind speed statistical distributions. In order to understand how this component was developed, it is first important to recognize how a stochastic wind model behaves in the MOST framework.

The most robust stochastic wind modeling that MOST offers is based on a maximum of three wind power trajectories, typically a low, average, and high power trajectory. There is a given power output for each of these three trajectories for each period. The stochasticity is introduced via a probability matrix that represents the probability of transitioning from one trajectory to another for each period studied. These transition probabilities include the probability of staying on the same trajectory as well as transitioning to either of the other two trajectories.[27] The result is a wind profile with full transition probabilities, a representation of which is displayed in Figure 23.

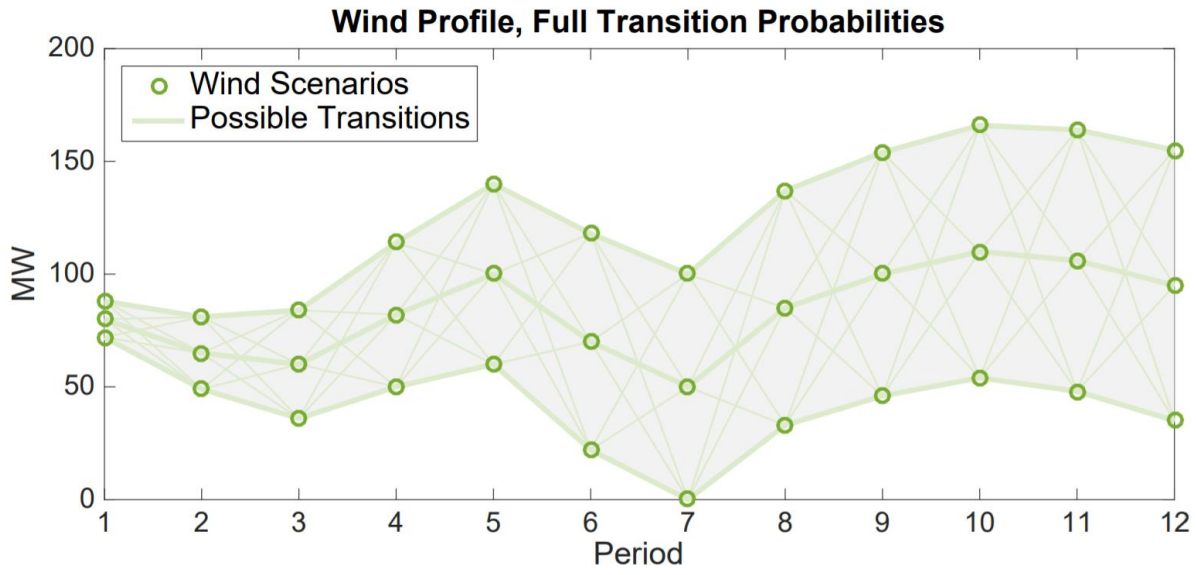

Figure 23: Example wind profiles with full transition probabilities[27]

The stochastic wind profiles are very dependent on the power produced during each period of the simulation. In order to reflect this in the development of the stochastic wind component for each model, the wind speed data was separated based on both the hour and the month. This results in 24 datasets, one for each hour in the day, for every month in the year. The monthly separation allows for any seasonal variation in the wind speed to be represented in the model. Given that this simulation is meant to represent a day in January, the 24 datasets for the month of January were used to develop the stochastic models. Several MATLAB functions were created that would input the hourly wind speed data for January and output the three wind power trajectories, which can be broken up into three steps.

The first step is to calculate the cumulative power distribution of the wind speed for a given distribution, and using this as well as the wind speed to wind power calculations presented in Section III, a cumulative power distribution of the generated wind power can be created, as seen in Figure 9. In order to do this, the parameters of the wind turbine must be known. Since this project is meant to represent the entirety of the New England grid and not a specific wind turbine, general averages were used, with a cut-in speed of 3 meters per second, a rated speed of 14 meters per second, and a cut-out speed of 25 meters per second.[28] The wind power within the cumulative

power distribution was represented as a percentage of maximum power such that it can be applied to various levels of wind power on the grid as a whole. With the hourly cumulative power distribution generated, the probability of producing at least a certain amount of power was modeled for all of the 24 hours of the day for the month of January.

The second step is to set probability cutoffs for the low, average, and high wind power trajectories. The choice of these cutoffs is somewhat subjective, but the options of 0.30 for low, 0.55 for average, and 0.80 for high were ultimately chosen. The reasoning was to allow the high cutoff to be at the rated power, the low to be at a fairly low but not unreasonably low level, and the average cutoff to fall directly in between. The cutoffs and their corresponding power proportions are displayed in Figure 24 for Location 1 using a Weibull distribution. Note that the corresponding power proportion will differ for each location and for each distribution used to generate the cumulative power distribution, but the cutoffs are the same for each of the four statistical distributions.

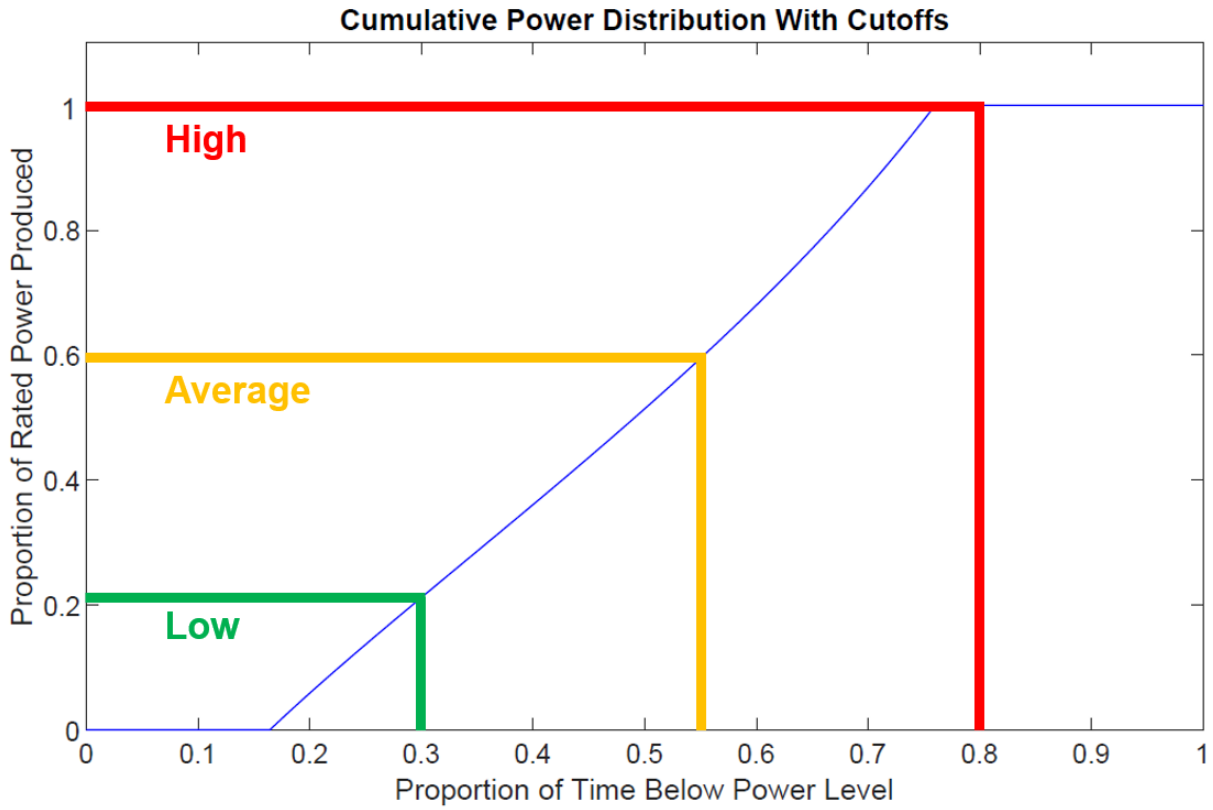

Figure 24: Probability cutoffs displayed on a cumulative power distribution generated using location one data and a Weibull distribution

The corresponding power proportions generated for each of these probabilities from the cumulative power distributions are then used to generate the wind profile. The result is three possible wind power generation levels for each hour of the day, represented as a proportion of the maximum installed wind capacity on the New England power grid. Figure 25 displays the differences in the profiles generated from each statistical distribution by displaying the average profile for each based on location one data.

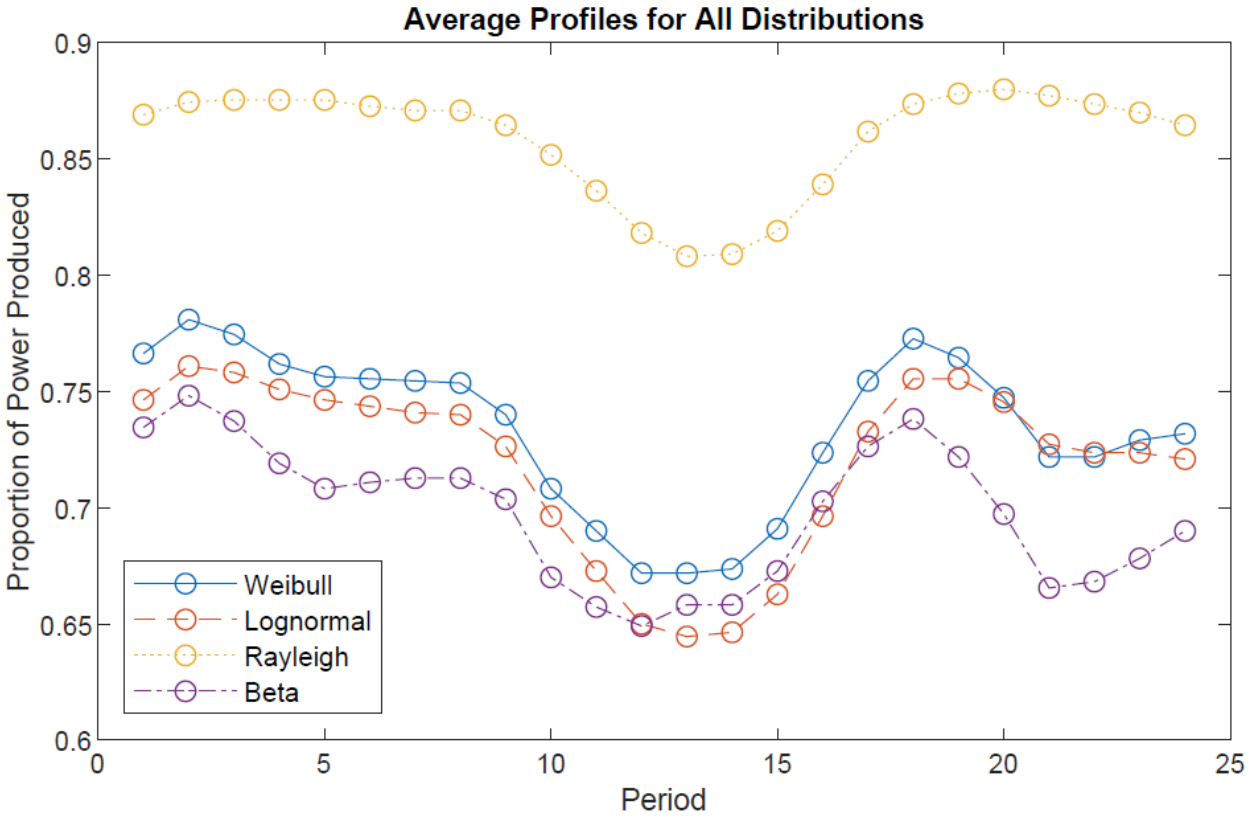

Figure 25: Average profiles for all four statistical distributions produced using location one wind speed data

The third and final step is to develop the transition probability matrix corresponding to the three profiles that were created. This matrix is the "transmat" input from Section VI. For each period in the simulation, the matrix is three by three, resulting in probabilities for the full range of transitions from each trajectory to each of the three trajectories for the next period. The exception is the initial matrix for the first period of the simulation, when the matrix is three by one, as there must be a starting point and there is no transition occurring. Calculating the initial period's three

by one matrix relies on the cumulative power distribution. For a given power generation in each profile for period one, the corresponding probability from the cumulative power distribution is collected. This is completed for the low, average, and high wind power profiles. A probability must then be associated with starting in each of the respective profiles. This is achieved by creating interval ranges for each of the profiles. The low profile probability range starts at zero and continues to the midpoint between its associated probability and that of the average profile. The average profile's range starts at the midpoint between the low profile and average profile associated probabilities and continues to the midpoint between the average profile and high profile associated probabilities. Lastly, the high profile's range starts at the midpoint between the average profile and high profile associated probabilities and continues to one. The result is three probabilities that sum to one. This creates the initial matrix in the full transition matrix, but the remainder of the hours are calculated differently since there will be transitions occurring. Figure 26 displays the interval ranges for each profile using the Weibull cumulative power distribution for Location 1.

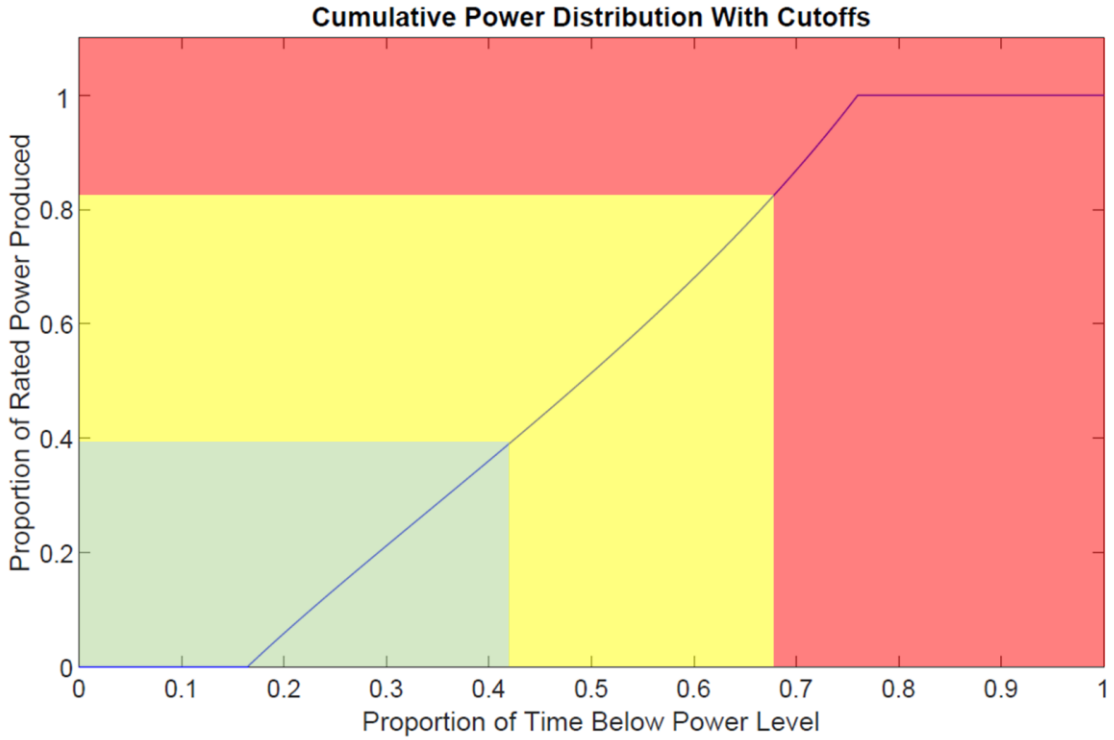

Figure 26: Interval ranges for the low, average, and high profiles. Low range is displayed in green, average in yellow, and high in red.

The three by three matrices for the remaining 23 hours in the day are based on the raw wind speed data. The average wind power is calculated for each hour of the day based on the wind speed. Then a simple counting algorithm is used. For each hour, the starting range is noted as being either low, average, or high power. Then, based on the range that the power changes to in the next hour, the transition is either recorded as going to the low, average, or high power region. This is done for each hour of each day in January for all 6 years of available data, based on the power proportion cutoffs for each respective distribution. The probability for each transition is therefore the number of times going from one range to another divided by the total number of times starting in that range. For example, if there are a total of 65 times that a period three begins in the low power range, and a total of 60 times that period four remains in the low range of these 65 instances, the probability of transitioning from the low region to the low region from period three to four would be 60/65 or 0.923. A sample transition matrix, with the first and second period matrices displayed, is shown. The general layout of each individual three by three matrix in the full matrix is displayed in the final 24th spot of the example transition matrix.

$$transition\ matrix = \left\{ \begin{bmatrix} 0.35 \\ 0.37 \\ 0.28 \end{bmatrix}, \begin{bmatrix} 0.81 & 0.08 & 0.01 \\ 0.09 & 0.87 & 0.03 \\ 0.10 & 0.05 & 0.96 \end{bmatrix}, \dots, \begin{bmatrix} low\ to\ low & average\ to\ low & high\ to\ low \\ low\ to\ average & average\ to\ average & high\ to\ average \\ low\ to\ high & average\ to\ high & high\ to\ high \end{bmatrix}_{24} \right\}$$

Note that the transitions from one region to that same region have the greatest probability. This makes sense, as it is less likely that a wind speed will change drastically from hour to hour than it is that the speed would remain roughly the same.

These three steps are completed for either the Weibull, Rayleigh, Lognormal, or Beta distribution before conducting a power flow simulation, depending on which distribution is being studied. The result is three unique wind power profiles and a unique transition probability matrix based on the respective statistical distribution. MOST calculates the expected dispatch based on the probability matrix and associated power profiles. With these stochastic elements in place,

economic dispatch simulations can be performed in MOST in order to compare the results based on which distribution is used to develop the stochasticity of the model.

VIII. RESULTS

The 24-hour January New England study period simulation was repeated to achieve results for each wind speed statistical distribution. The framework of the model was held constant as discussed in Section V and VI, with the only variable the stochastic behavior of the wind power as developed for each distribution described in Section VII. The expected wind generation for each location, expected traditional generation, traditional generator unit commitment, and expected generation cost was determined by the MOST simulation using the Weibull, Rayleigh, Lognormal, and Beta distributions for wind stochasticity. All results are displayed in this section for the Weibull distribution for background and to get an understanding of the general behavior of the results, but the traditional generation differences between each simulation will be the focus of comparison. Results for distributions absent in this section can be found in the Appendix in Figures A1 – A12. The results for the Weibull distribution are displayed in Figures 27 – 29.

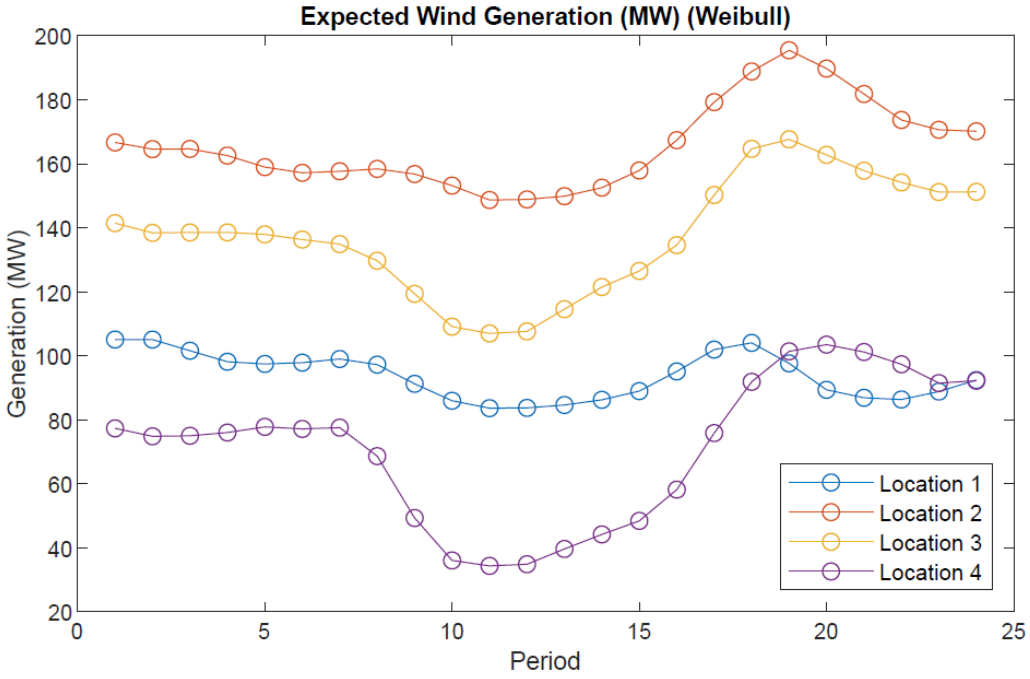

Figure 27: Expected wind generation for all locations during the 24-hour study period generated using a Weibull stochastic model

All locations follow a similar trend in terms of their expected generation, with a fairly constant power production until around hour 8 of the simulation when the power drops off during the daytime. The power then tends to rise again in the afternoon and into the evening, reaching a maximum power output around hour 18 to 20. Location 1 tends to have the least variation during the day, while location four has the most.

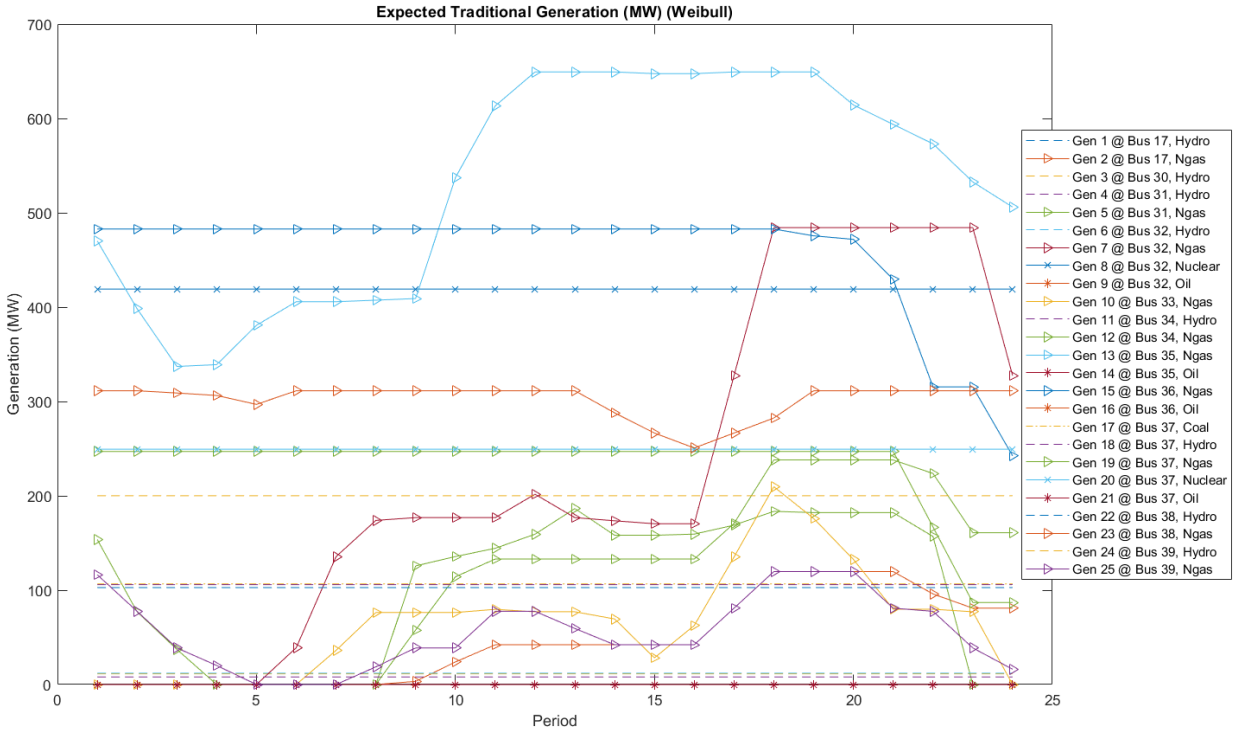

Figure 28: Expected traditional generation for all traditional generation sources during the 24-hour study period generated using a Weibull stochastic model

The traditional generator commitment tends to increase in the afternoon and into the evening when load increases, as evidenced by Figure 22. The cheaper forms of generation provide the base load (hydropower, nuclear, coal, some natural gas), and more expensive forms of generation add additional load during the peak. In this case, this additional generation is solely provided by natural gas, as the load increase is accompanied by a wind power increase around the same time, so no oil generation is required. This makes sense, as oil is the most expensive generation in this model and should therefore be avoided if it can.

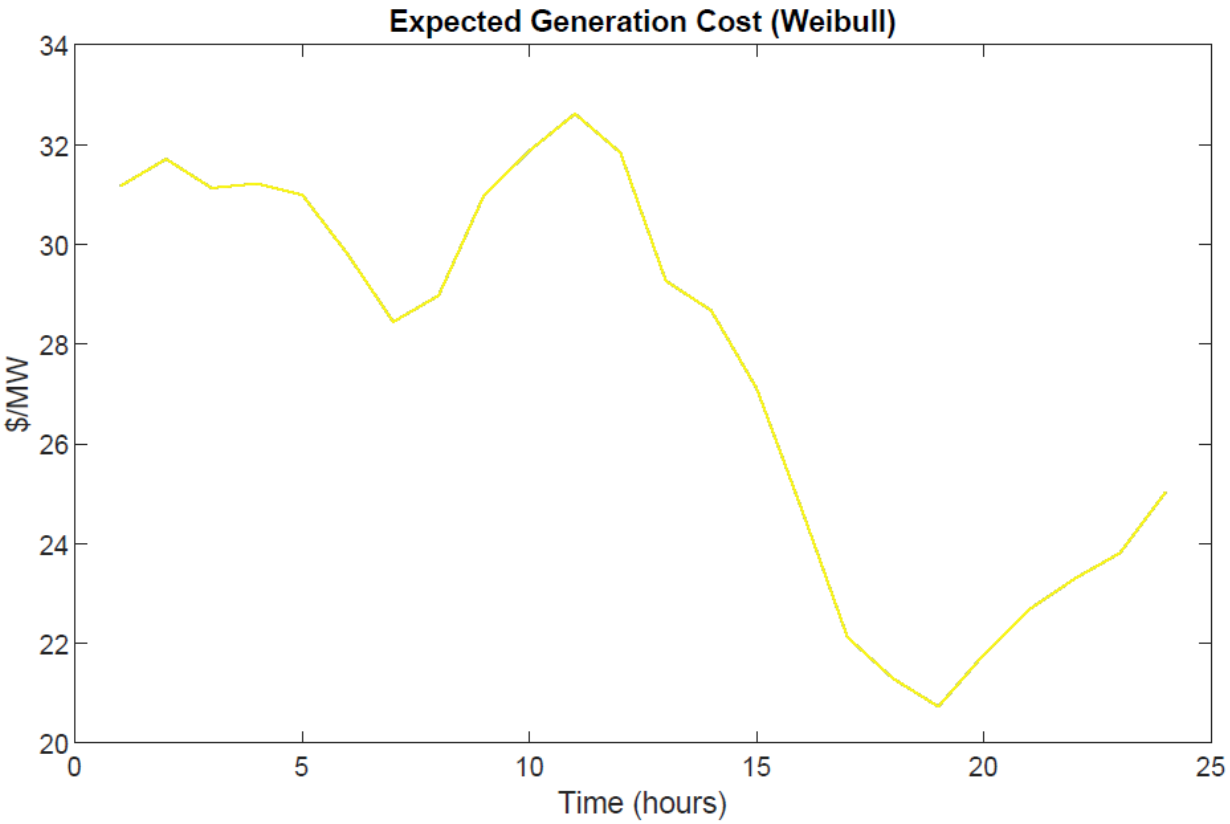

Figure 29: Expected generation cost in dollars per MW during the 24-hour study period generated using a Weibull stochastic model

The expected generation cost remains fairly low throughout the day due to the high amount of wind energy in the system. The generation cost per MW is at its highest in hour 11 during the first peak of load in the simulation, displayed in Figure 22, which corresponds to the minimum point of wind generation. Though load increases in the afternoon and evening, this corresponds with a large increase in wind generation that serves to decrease the cost per MW when averaged throughout the system.

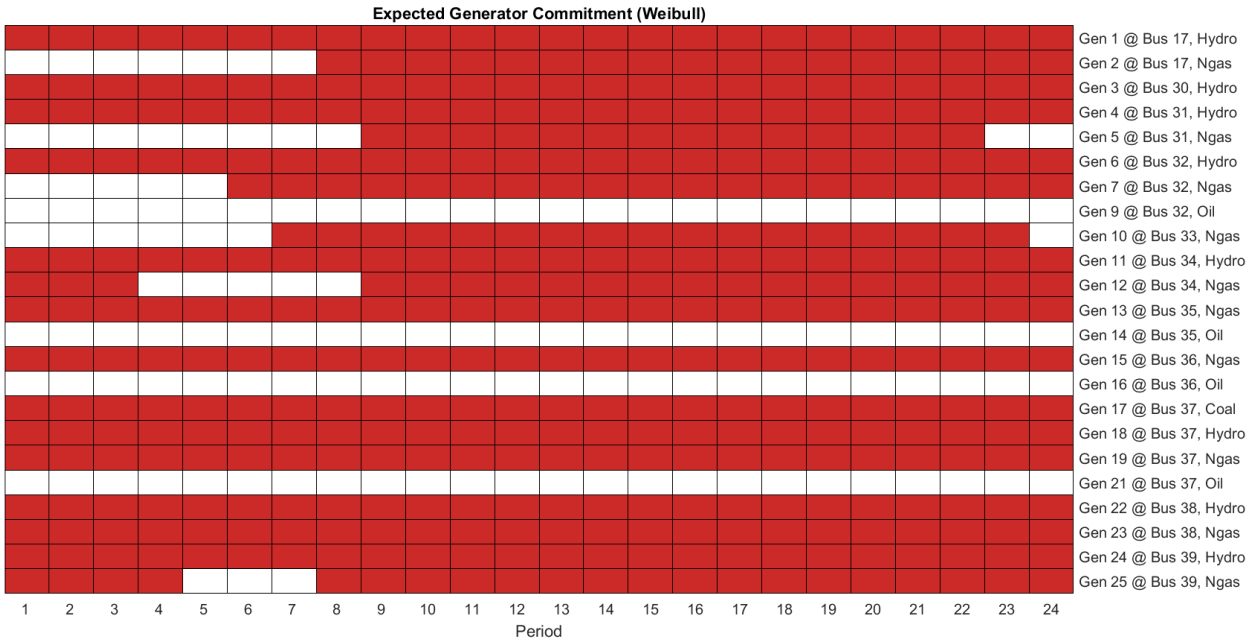

Figure 30: Expected generator unit commitment during the 24-hour study period generated using a Weibull stochastic model

Each distribution's stochastic model results in differing traditional generator unit commitment results, however there are some striking similarities. First, none of the simulations resort to using any oil (the most expensive generation in the model), meaning that the wind, nuclear, and hydropower generation, combined with natural gas and one coal generation source can alone meet the load requirement of this model. The lone coal generation source is always committed in all simulations as a result of its low cost in this model. In each simulation, hydropower and nuclear power are always committed as well as a result of them having the lowest cost of the traditional generation sources. Due to these similarities, the only difference between each distribution's traditional generation results is the required amount of the natural gas generation. Given that natural gas accounts for almost 50% of the installed capacity in the model with 10 available generators, this is a significant difference that will be analyzed further.

Figure 31 displays the total amount of natural gas generation (MW) that was required for each of the four simulations with differing stochastic models.

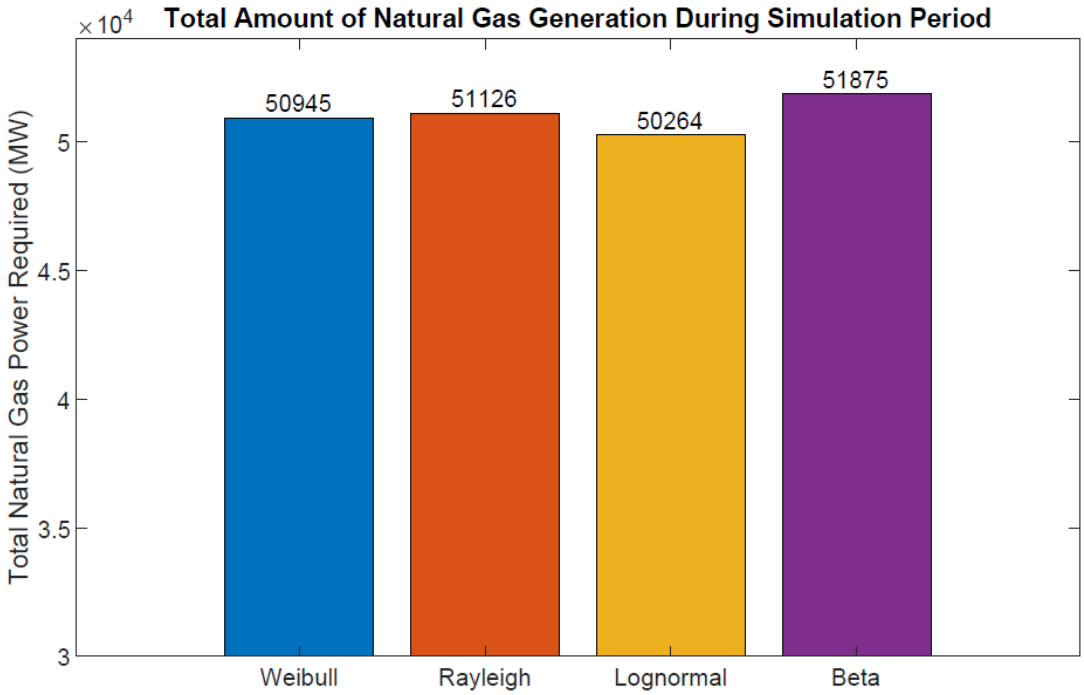

Figure 31: Total amount of natural gas generation (MW) required for each of the simulations with differing stochastic models

From Figure 31, the Beta simulation clearly has the most natural gas generation required, whereas the Lognormal simulation has the least. The Weibull and Rayleigh simulation results are the most similar, sitting in the middle, with the Rayleigh results requiring slightly more natural gas generation. Though this result alone gives some insight as to the behavior of each of the simulations depending on the stochastic model used, it is beneficial to analyze the natural gas generation required on an hour by hour basis throughout the simulation period. The simulation period is divided into three segments of eight hours for this analysis. Figure 32 displays hourly natural gas generation required for each of the simulations for the first eight hours of the simulation period. Table A1 in the appendix also displays numerical values for the hourly natural gas generation required for each simulation.

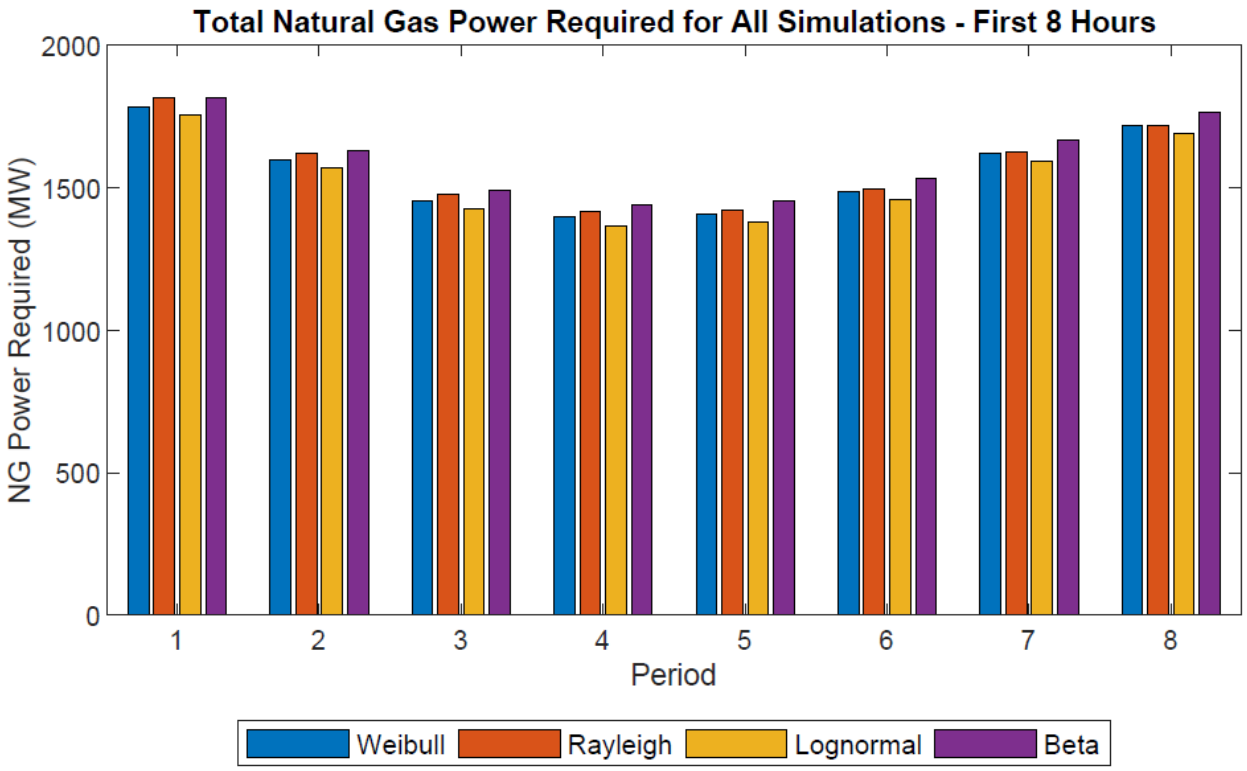

Figure 32: Total amount of natural gas generation (MW) required for each of the simulations with differing stochastic models for the first eight hours of the simulation period

For almost all hours in the early hours of the simulation, the Beta simulation requires the most natural gas, followed by the Rayleigh, Weibull, and Lognormal. The exception is hour eight, where the Rayleigh simulation requires slightly less natural gas than the Weibull. This exception coincides with a drop in expected wind generation, as seen in the expected wind power production in Figures 26 and A1 – A3. Figure 33 displays the results for the next eight-hour increment.

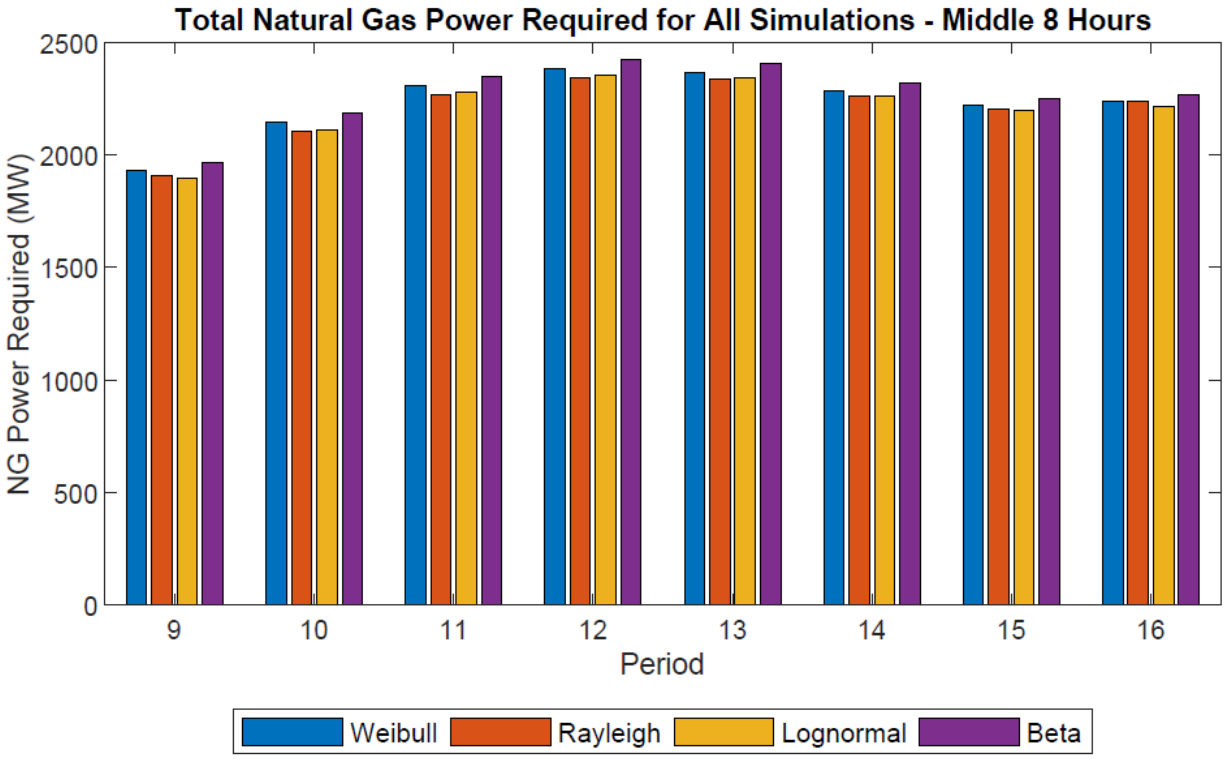

Figure 33: Total amount of natural gas generation (MW) required for each of the simulations with differing stochastic models for the middle eight hours of the simulation period

From Figure 33, there is a shift in the ranking of the simulations based on amount of natural gas generation required. For all hours, Beta still requires the most, and Lognormal requires the least for four of the hours while Rayleigh requires the least for the other four. The Weibull simulation requires more natural gas than Rayleigh for all hours in this simulation other than hour 16, a continuation of the trend that began in hour eight of the previous eight hours. This is the period of the day when wind power produced is at its lowest, as evidenced by Figures 27 and A1 – A3. There are also some hours when the Rayleigh simulation requires less natural gas generation than the Lognormal, as is the case in hours 10, 11, 12, and 13. These hours in particular are when wind power generation is at its absolute minimum in the simulation period, as seen in Figures 27 and A1 – A3. Figure 34 displays the results for the final eight hours of the simulations.

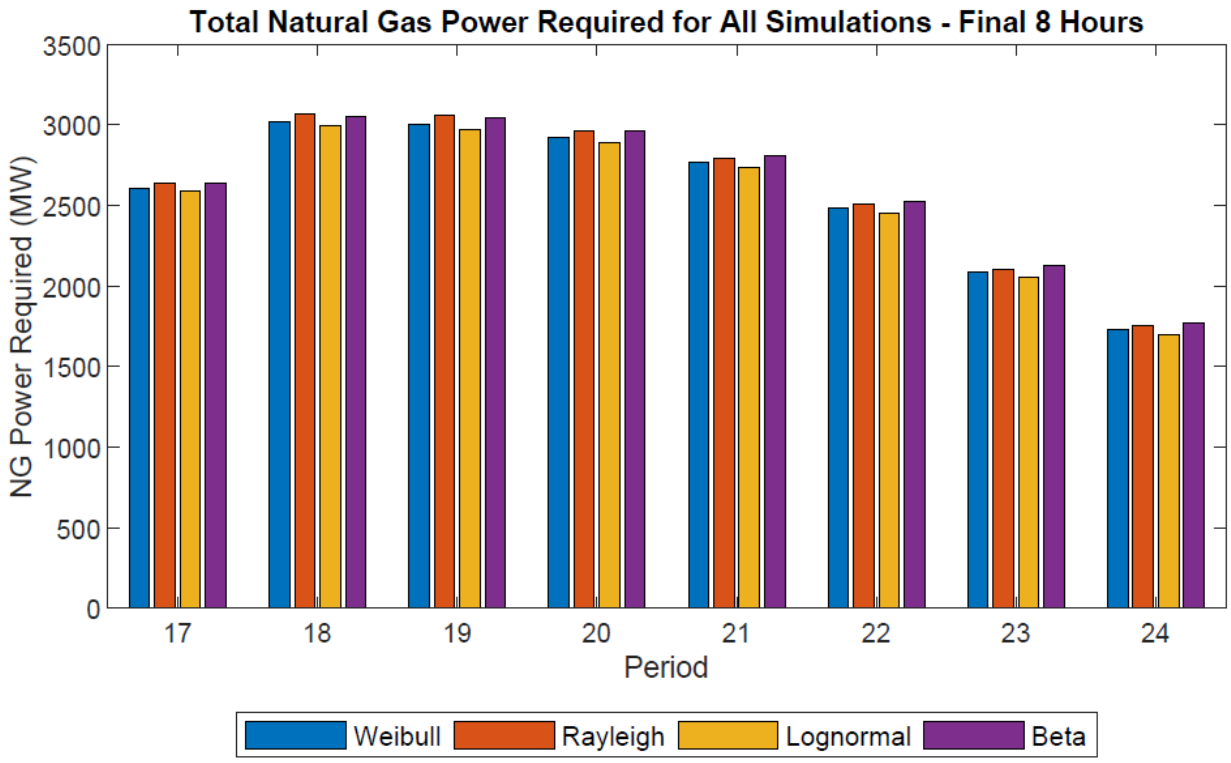

Figure 34: Total amount of natural gas generation (MW) required for each of the simulations with differing stochastic models for the final eight hours of the simulation period

In the final eight hours, there are again changes in the ranking of the simulations based on amount of natural gas generation required. The Beta simulation still requires the most natural gas for five of the eight hours, however, for hours 18, 19, and 20, the Rayleigh simulation requires more than the Beta. These three hours are when the wind generation is at its absolute highest point of the day, as seen in Figures 26 and A1 – A3. The Lognormal simulation again requires the least natural gas generation for all eight hours. In all hours in this segment as well, the Weibull simulation again requires less natural gas generation than the Rayleigh. These changes coincide with the amount of wind generation being at its highest point of the day, as is evidenced by Figures 27 and A1 – A3.

IX. DISCUSSION

From these results, there are some key takeaways. First, the general trend of the amount of natural gas required follows the shape of the load curve in Figure 22, however, there are fewer sharp changes due to the wind power generation. The Beta simulation consistently requires the

highest amount of natural gas generation for most hours, with the exception of hours 18, 19, and 20. The Lognormal simulation requires the least amount of natural gas generation for all hours except hours 10, 11, 12, and 13, when the amount of wind power produced is very low. In these hours, the Rayleigh simulation requires even less natural gas. The Weibull simulation requires more natural gas than Rayleigh when wind generation is lower, and less natural gas than Rayleigh when wind generation is higher. It can therefore be concluded that the Rayleigh simulation's natural gas generation requirements are the least sensitive to changes in expected wind power production, resulting in a more flat expected natural gas requirement. For very low wind power production, the Rayleigh simulation requires the least amount of natural gas generation as more load is met by wind power than the other simulations. For very high wind power production, the Rayleigh simulation requires the most amount of natural gas generation as less of the load is met by wind power than the other simulations. This is a result of the Rayleigh distribution being a one parameter model. It's predicted wind power is always more near the average than other models, which results in these conclusions.

The results for the Beta simulation consistently requiring the most natural gas or always producing the least wind power are likely a result of it having the highest probabilities for very low wind speeds, as is evidenced by the probability density function in Figure 8. The reason that the Lognormal simulation almost always requires the least natural gas or produces the most wind power is likely a result of the opposite. The Lognormal model results in significantly lower probabilities for very low wind speeds compared to other models, as seen in the probability density functions in Figure 8. The Weibull simulation tends to sit somewhere in the middle of all these distributions in terms of the results for the amount of natural gas required. From the cumulative power distribution in Figure 9, this makes sense, as the Weibull distribution generally sits in the middle in terms of power production, producing less than some distributions in some cases and more in others.

Overall, these results display that the choice of wind speed statistical distribution used to generate a stochastic model in a power flow simulation does impact the results. If these statistical

distributions are being used for system planning or even long-term forecasting, close attention should be paid to the distribution that most closely models the raw data from the geographic area of interest. Using a statistical model that does not model the data well can result in operators being unprepared for the amount of dispatchable generation they will need to commit at certain times during the day. The goodness of fit for each of the statistical models to the wind speed can be analyzed for the January data at each site, as well as for the lumped data across all locations. The goodness of fit of the theoretical statistical distribution to the empirical data can be analyzed using a Kolmogorov-Smirnov Test or KS Test. The KS Test quantifies a distance between the theoretical and empirical cdfs. This test produces a test statistic that can be used to determine the goodness of fit for each of the four wind speed statistical distributions to the empirical wind speed data. The test statistic is calculated as:

$$D = sup_x|F_n(x) - F(x)|$$

where D is the test statistic $F_n(x)$ represents the empirical cdf from raw wind speed data, $F(x)$ represents the theoretical cdf from a certain distribution, and sup_x is the supremum of the set of distances.[29] Therefore, lower test statistics correspond to a better fit. Tables 1 – 4 display the KS Test Statistic for goodness of fit for each location and distribution.

Table 1: KS Test Statistics for goodness of fit at Location 1

Location 1	
Distribution	KS Test Statistic
Weibull	0.05942
Rayleigh	0.08108
Lognormal	0.11832
Beta	0.05398

Table 2: KS Test Statistics for goodness of fit at Location 2

Location 2	
Distribution	KS Test Statistic
Weibull	0.04883
Rayleigh	0.10089
Lognormal	0.12115
Beta	0.02371

Table 3: KS Test Statistics for goodness of fit at Location 3

Location 3	
Distribution	KS Test Statistic
Weibull	0.04203
Rayleigh	0.05577
Lognormal	0.10081
Beta	0.04350

Table 4: KS Test Statistics for goodness of fit at Location 4

Location 4	
Distribution	KS Test Statistic
Weibull	0.02712
Rayleigh	0.08669
Lognormal	0.05984
Beta	0.03169

Table 5 displays the KS Test Statistic for goodness of fit for all locational data combined. Figures A13 – A16 also display graphical comparisons of the empirical and theoretical cdfs for all locational data combined.

Table 5: KS Test Statistics for goodness of fit for all location data combined

All Locations Combined Data	
Distribution	KS Test Statistic
Weibull	0.01753
Rayleigh	0.02656
Lognormal	0.07040
Beta	0.02765

Upon analyzing the goodness of fit of each of these distributions, it becomes clear that certain distributions have better fits at certain locations. Though the Weibull has the best fit of the lumped locational data, as evidenced by Table 5, the Weibull does not have the best fit at each individual location. The Beta distribution has the best fit at Locations 1 and 2, and the Weibull distribution has the best fit at Locations 3 and 4. Given the variation of the results of natural gas generation requirements depending on which statistical distribution was used to generate a stochastic model of wind power, this emphasizes the importance of using a statistical model with the best fit at a certain location in terms of characterizing corresponding wind speed data. As the results in Table 5 display, the distribution that best fits the entirety of a data set may not be the best

fit for certain locational components within the data set, so the best fitting model for each location should be selected in order to ensure the most accurate stochastic wind power models. For the case study presented in this paper, it is therefore most accurate to derive stochastic wind power models with a Beta distribution at Locations 1 and 2, and with a Weibull distribution at Locations 3 and 4. Using worse-fitting statistical distributions to generate stochastic wind power models could result in a lower or higher expected traditional generation dispatch when compared to using the statistical model determined to have the best fit.

X. CONCLUSION

The results from this research are useful for long-term system planning in terms of estimating the amount of dispatch from various generation sources that will be required. Though there becomes less uncertainty with planning in the short-term, a long-term approach such as the one used in this research can be made more accurate if a statistical model is used that closely fits the raw wind speed data. There are many wind speed statistical models available, and therefore choosing the proper one for a given data set is very important for long-term system planning and developing an understanding of traditional generator commitment.

Extensions of this research could involve increasing complexity to develop a more realistic understanding of the differences of these stochastic models in the New England power grid. This could involve turning the economic dispatch problem into an optimal power flow problem with more constraints, such as ramping, minimum time up and down, and startup and shutdown costs. More wind resource locations could also be added in order to develop a more distributed reference of the behavior of wind power production in the New England power grid. Continued research into these stochastic models and their effects on system behavior in the power grid will help to prepare for the increasing amount of intermittent wind generation across the globe.

REFERENCES

[1] Barthelmie, R. J. and Pryor, S. C. "Potential contribution of wind energy to climate change mitigation." *Nature Climate Change*, 4, 8, 684-688, Jun., 2014.

[2] Weibull, W. "A Statistical Distribution Function of Wide Applicability." *Journal of Applied Mechanics,* 233-297, Sept., 1951.

[3] Saxena, B.K. and Rao, K.V.S. "Comparison of Weibull parameters computation methods and analytical estimation of wind turbine capacity factor using polynomial power curve model: case study of a wind farm." *Renewables: Wind, Water, and Solar,* 2, 3, Jan., 2015.

[4] Montelpare, S. "An innovative wind-solar hybrid street light: Development and early testing of a prototype." *International Journal of Low-Carbon Technologies,* 10, 4, Jan., 2014.

[5] Manwell, J. F., et al. *Wind Energy Explained: Theory, Design and Application.* John Wiley & Sons, 2011.

[6] Galton, Francis XII. "The geometric mean, in vital and social statistics." *Proc. R. Soc. Lond,* 29, Jan., 1879.

[7] Morgan, Eugene C., et al. "Probability Distributions for Offshore Wind Speeds." *Energy Conversion and Management,* 52, 1, 15–26, Jul., 2010.

[8] Pearson, K. *Mathematical Contributions to the Theory of Evolution.* Cambridge University Press, 1904.

[9] Efthimiou, G.C., et. al. "A Statistical Model for the Prediction of Wind-Speed Probabilities in the Atmospheric Surface Layer." *Boundary-Layer Meteorol,* 163, 179-201, Nov., 2016.

[10] Jung, C., et al. "Introducing a system of wind speed distributions for modeling properties of wind speed regimes around the world." *Energy Conversion and Management,* 144, 181-192, Jul., 2015.

[11] Hosking, J.R.M. "The four-parameter kappa distribution." *IBM Journal of Research and Development,* 38, 3, 251-258, May, 1994.

[12] Morgan, E.C. et. al. "Probability distributions for offshore wind speeds." *Energy Conversion and Management,* 52, 15-26, Jan., 2011.

[13] Houghton, J.C. "Birth of a parent: The Wakeby Distribution for modeling flood flows." *Water Resources Research,* 14, 6, 1105-1109, Dec., 1978.

[14] Hosking, J.R.M. *Regional Frequency Analysis: An Approach Based on L-Moments*. Cambridge University Press, 1997.

[15] Marquardt, D.W. "An Algorithm for Least-Squares Estimation of Nonlinear Parameters." *Journal of the Society for Industrial and Applied Mathematics,* 11, 2, 431-441, Jun., 1963.

[16] Zephyr, L., et. al. "Multiarea Wind Scenario Generation for Planning Under Uncertainty."

[17] NREL Wind Prospector. https://maps.nrel.gov/wind-prospector/

[18] USGS U.S. Wind Turbine Database. https://eerscmap.usgs.gov/uswtdb/viewer/

[19] Unwin, Jack. "US Wind Energy by State: Ranking the Top 10." *Power Technology | Energy News and Market Analysis*, 18 Apr. 2019, www.power-technology.com/features/us-wind-energy-by-state/.

[20] Gellerman, Bruce. "New England's Electric Power Grid Is Undergoing A Transformation." New England's Electric Power Grid Is Undergoing A Transformation | Earthwhile, WBUR, 4 June 2019, www.wbur.org/earthwhile/2019/06/04/region-energy-future-climate-change

[21] U.S. Energy Information Administration Analysis and Projections. https://www.eia.gov/analysis/
[22] Sas output, https://www.eia.gov/electricity/annual/html/epa 08 04.html
[23] ISO New England. www.iso-ne.com
[24] Gonzalez-Salazar, M., et. al. Review of the operational flexibility and emissions of gas- and coal-fired power plants in a future with growing renewables. *Renewable and Sustainable Energy Reviews*, 82:1497-1513, 2018.
[25] McDowell, J., et. al. Reactive power interconnection requirements for PV and wind plants: Recommendations to NERC. 2012.
[26] Hetzer, John, et al. "An Economic Dispatch Model Incorporating Wind Power." *IEEE Transactions on Energy Conversion*, 2008.
[27] D. Ray Zimmerman and E. Carlos Murillo-Sanchez. Most 1.0 user's manual.
[28] "Wind Turbine Power Ouput Variation with Steady Wind Speed." Wind Turbine Power Curves., www.wind-power-program.com/turbine_characteristics.htm
[29] Massey, Frank J. "The Kolmogorov-Smirnov Test for Goodness of Fit." *Journal of the American Statistical Association*, 46, 253, 68-78, 1951.

APPENDIX

The simulation results for all distributions not displayed in the body of this paper are displayed in Figures A1 – A12. Table A1 displays numerical values for the required natural gas commitment for each hour and each simulation. Figure A13 – A16 display graphical comparisons for goodness of fit of the empirical and theoretical cdfs for all locational data combined.

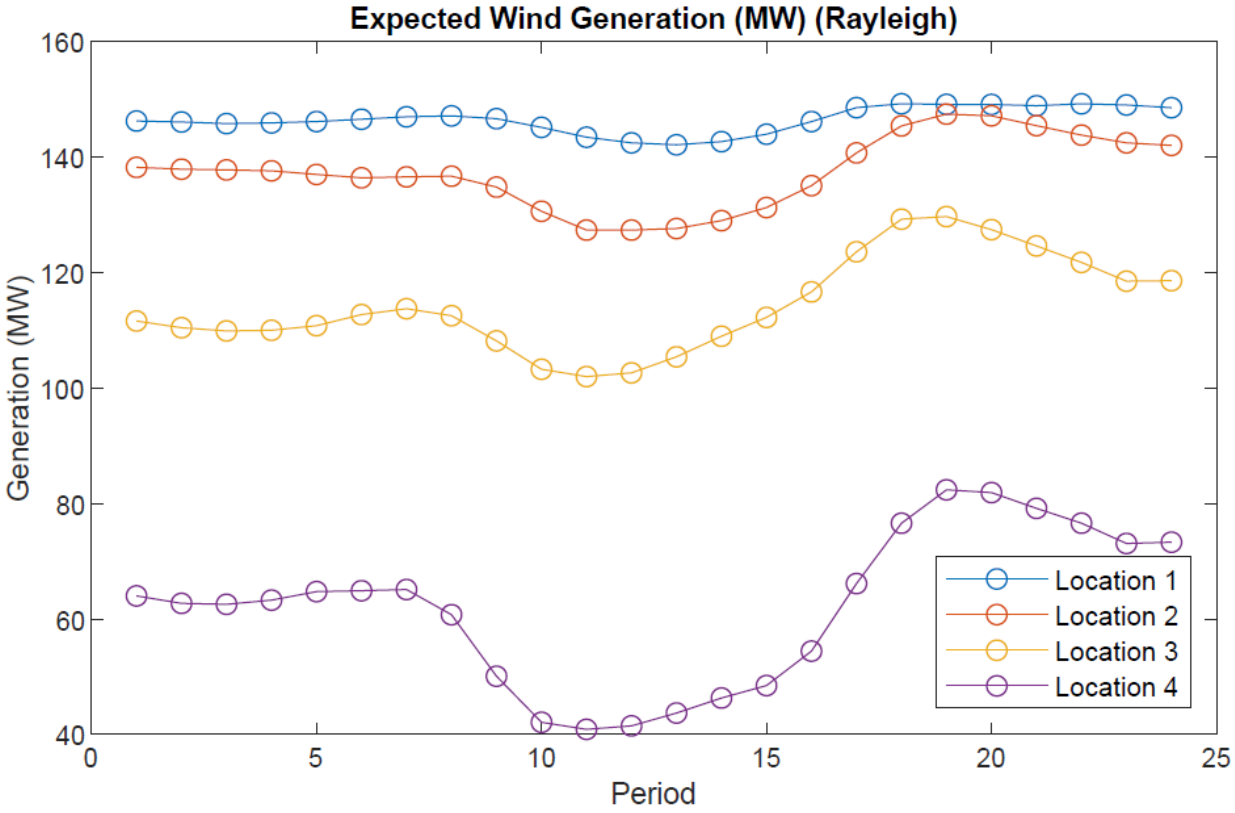

Figure A1: Expected wind generation for all locations during the 24-hour study period generated using a Rayleigh stochastic model

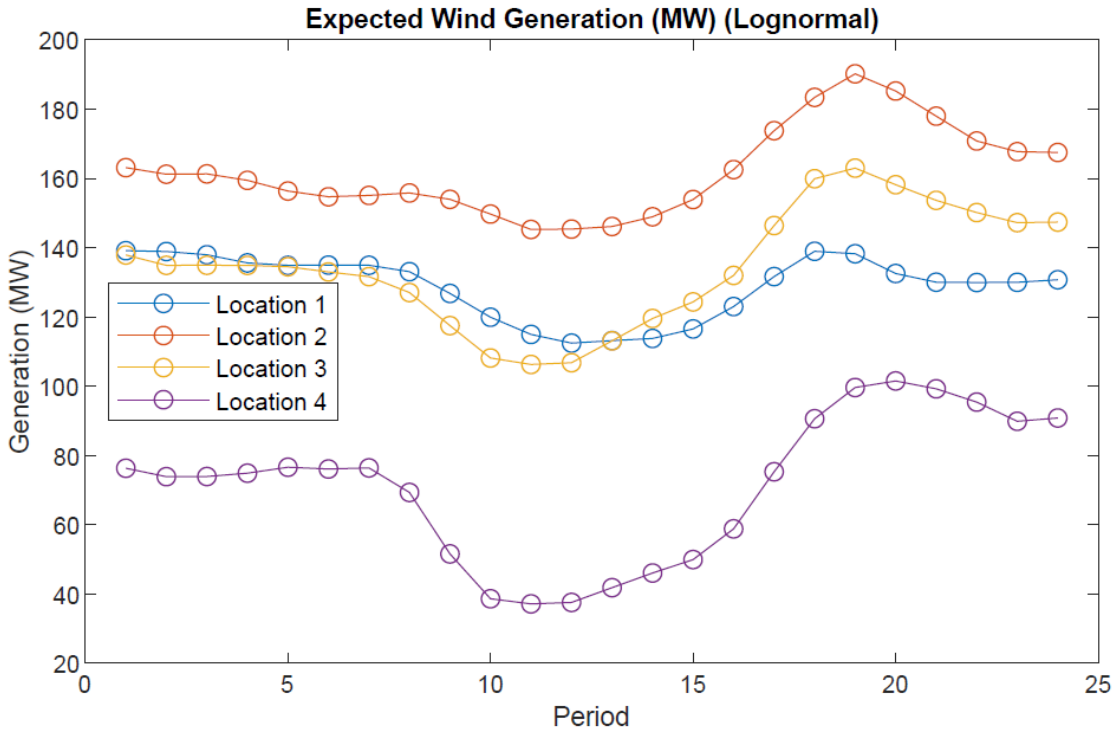

Figure A2: Expected wind generation for all locations during the 24-hour study period generated using a Lognormal stochastic model

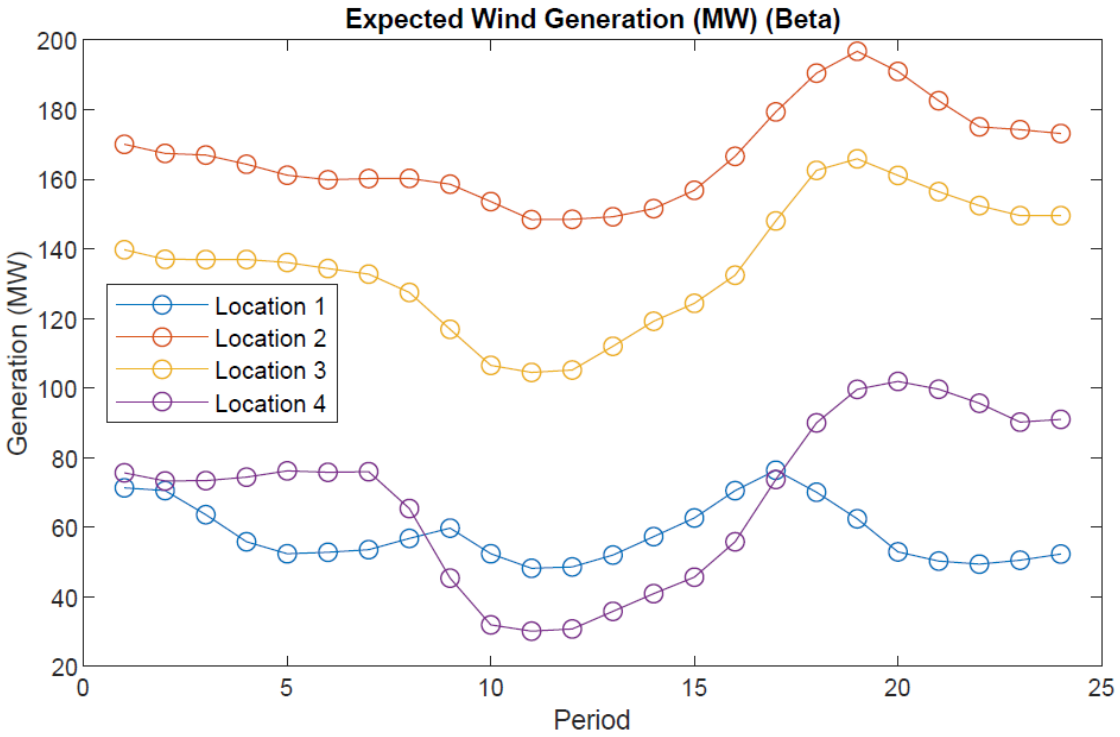

Figure A3: Expected wind generation for all locations during the 24-hour study period generated using a Beta stochastic model

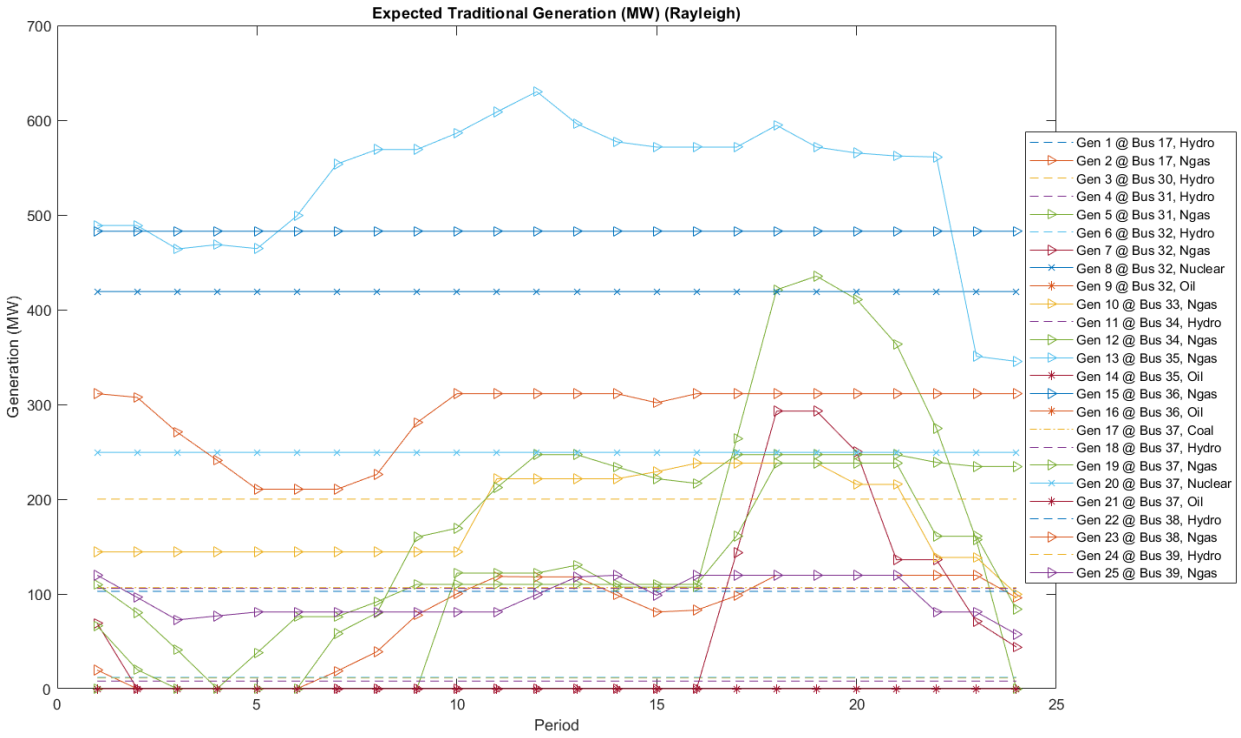

Figure A4: Expected traditional generation for all traditional generation sources during the 24-hour study period generated using a Rayleigh stochastic model

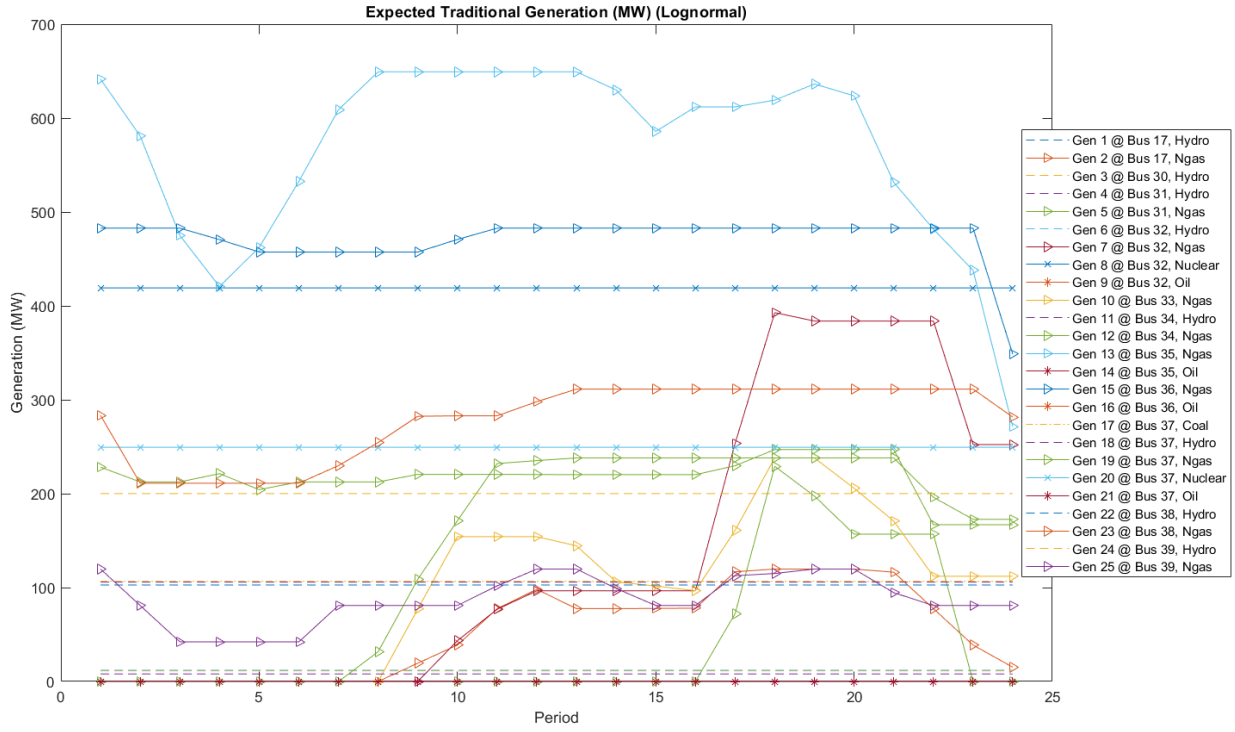

Figure A5: Expected traditional generation for all traditional generation sources during the 24-hour study period generated using a Lognormal stochastic model

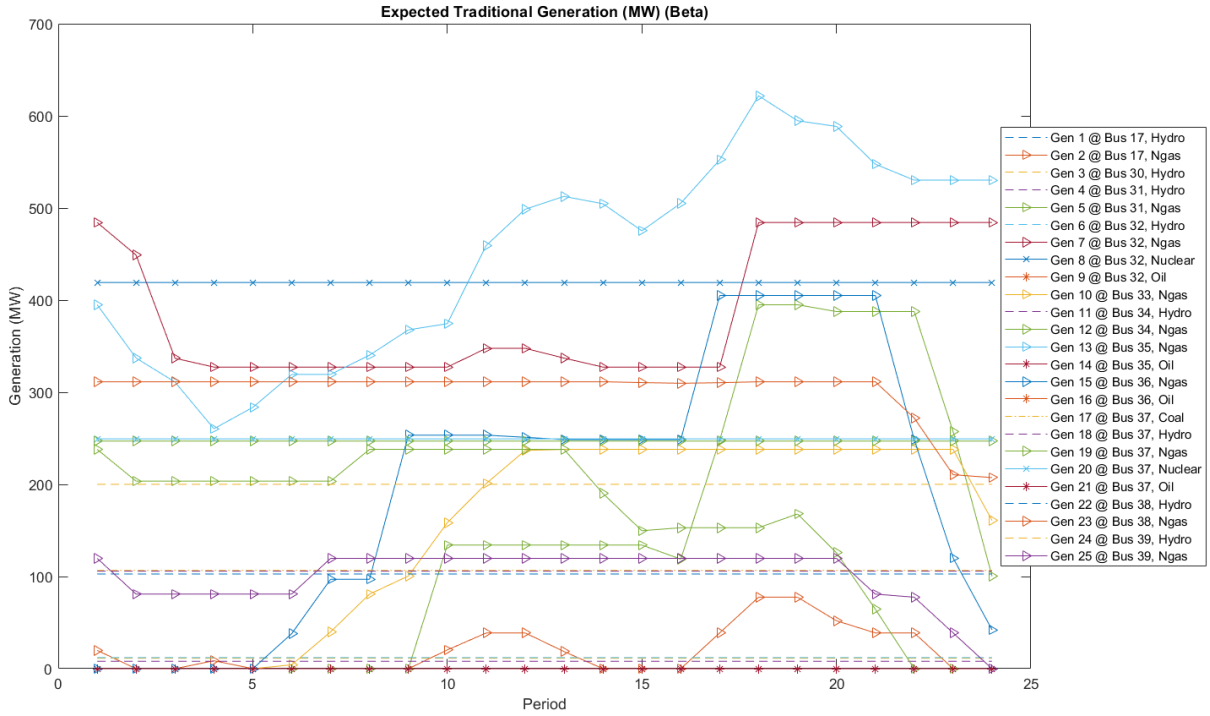

Figure A6: Expected traditional generation for all traditional generation sources during the 24-hour study period generated using a Beta stochastic model

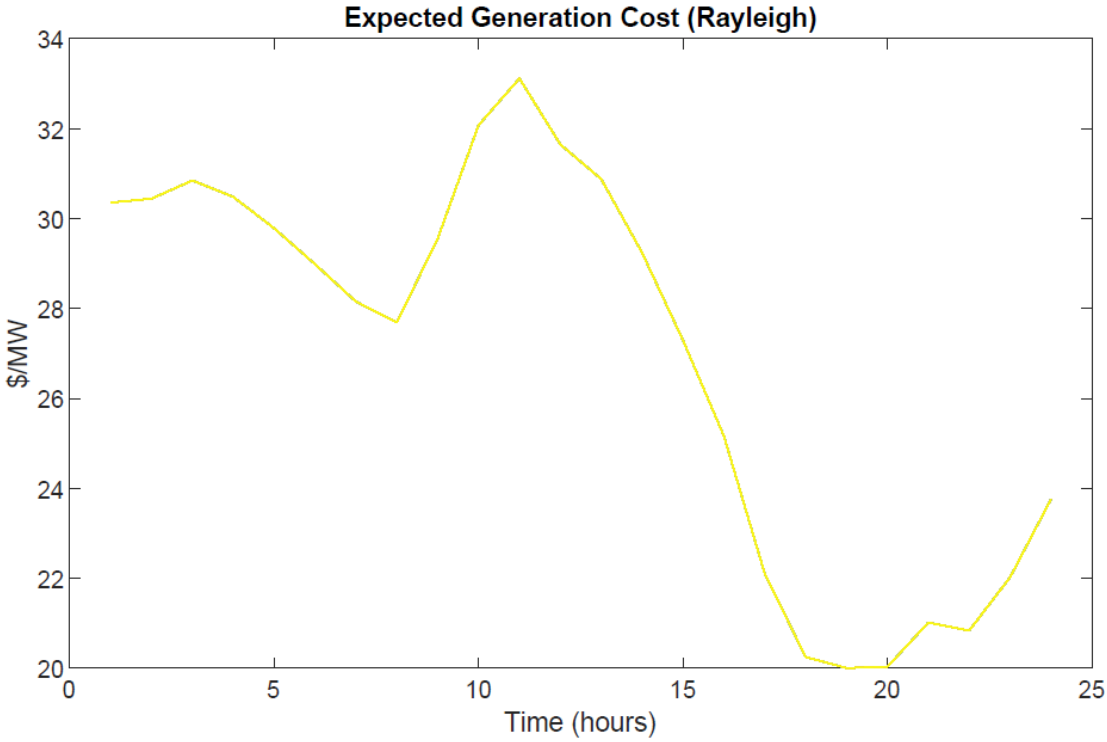

Figure A7: Expected generation cost in dollars per MW during the 24-hour study period generated using a Rayleigh stochastic model

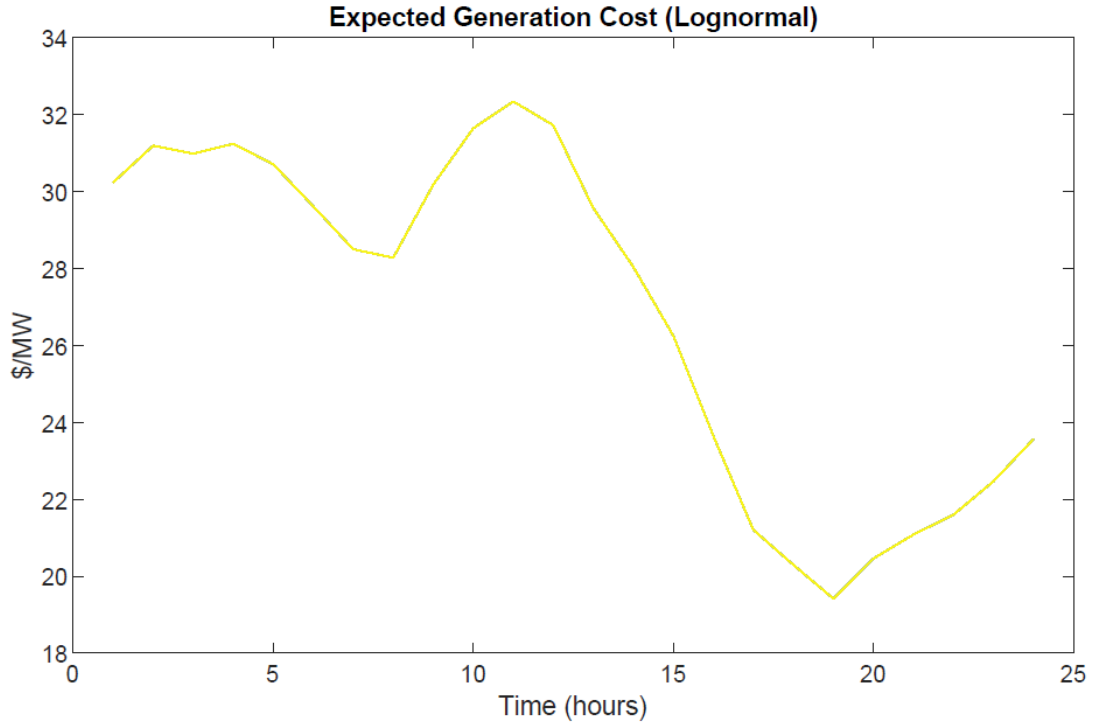

Figure A8: Expected generation cost in dollars per MW during the 24-hour study period generated using a Lognormal stochastic model

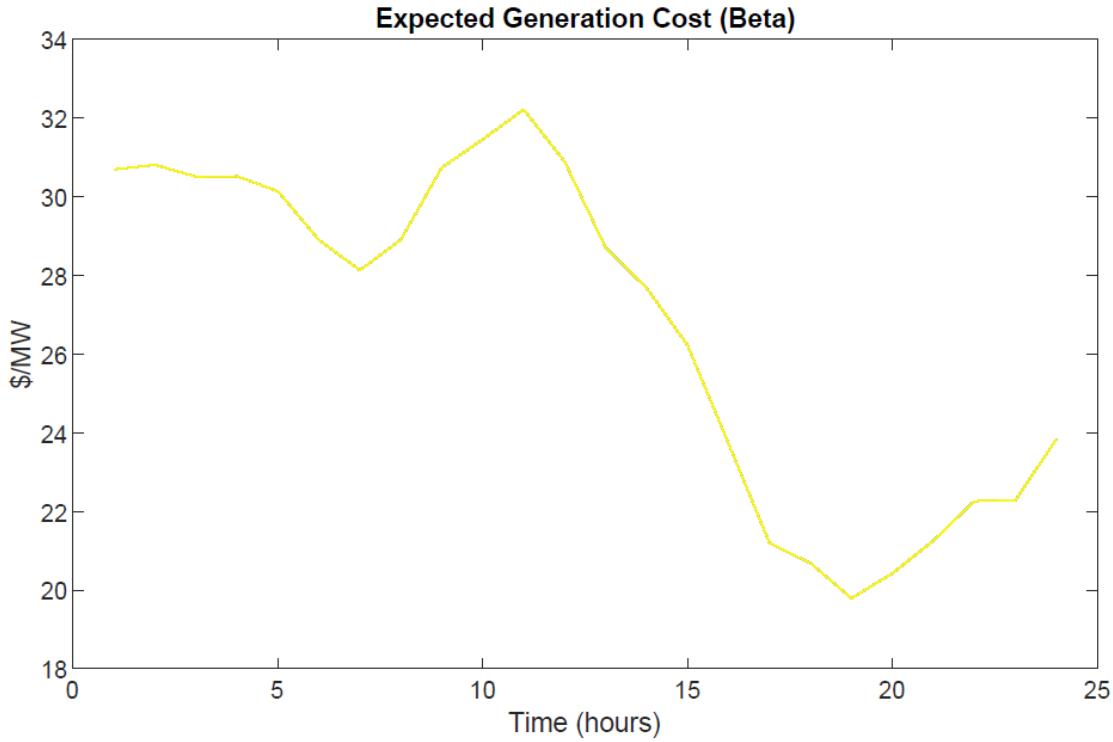

Figure A9: Expected generation cost in dollars per MW during the 24-hour study period generated using a Beta stochastic model

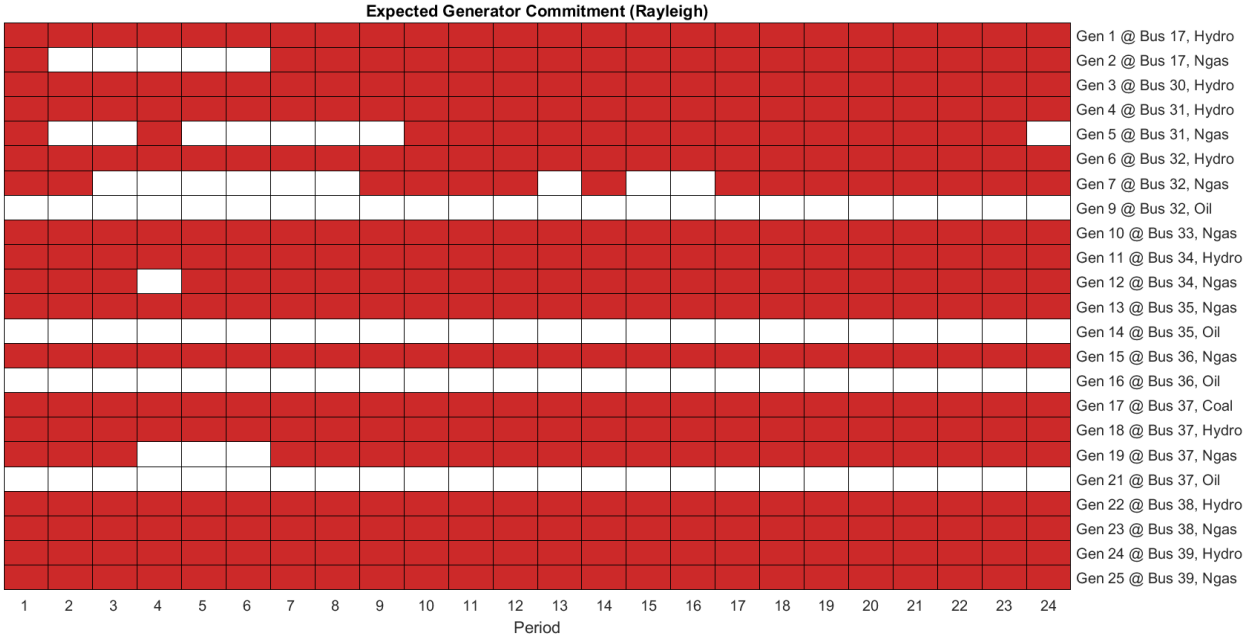

Figure A10: Expected generator unit commitment during the 24-hour study period generated using a Rayleigh stochastic model

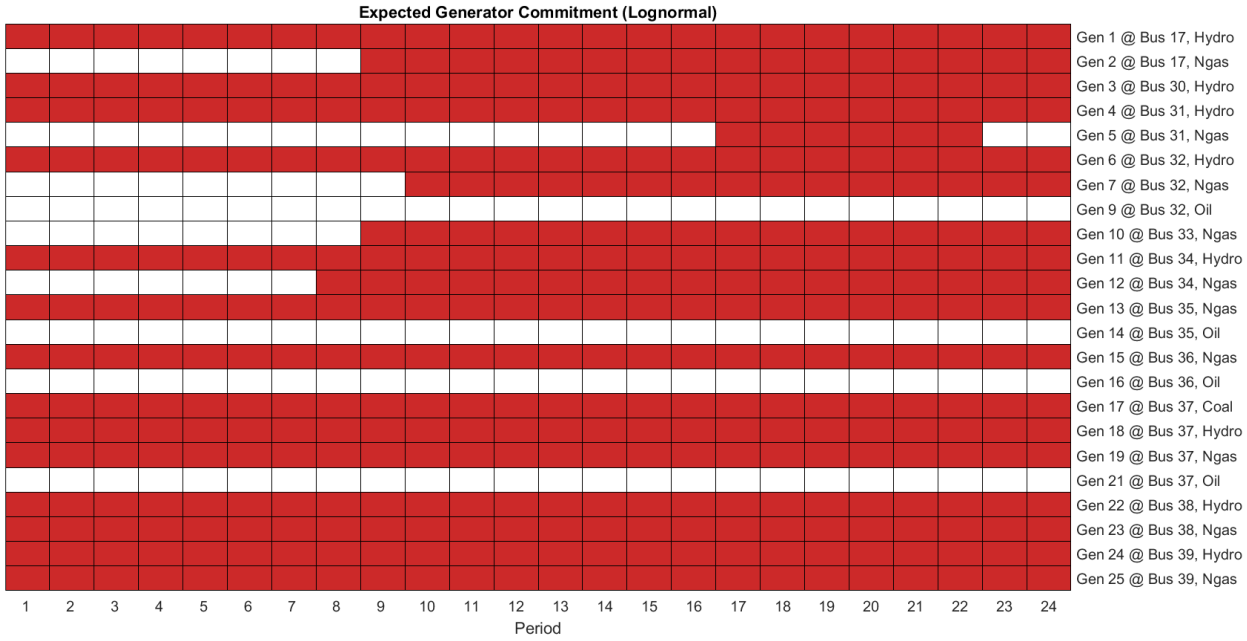

Figure A11: Expected generator unit commitment during the 24-hour study period generated using a Lognormal stochastic model

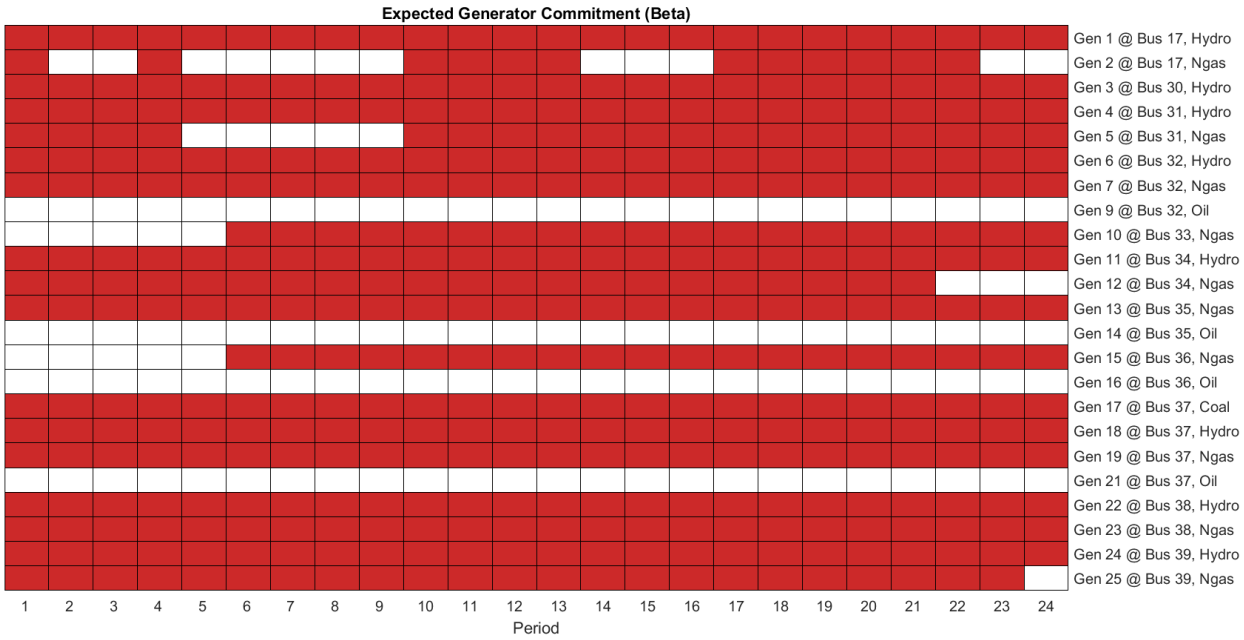

Figure A12: Expected generator unit commitment during the 24-hour study period generated using a Beta stochastic model

Table A1: Total amount of natural gas generation (MW) required for each simulation with differing stochastic models for all hours of the simulation period

Natural Gas Required Per Hour (MW)				
Hour	**Weibull**	**Rayleigh**	**Lognormal**	**Beta**
1	1,782	1,812	1,756	1,815
2	1,594	1,620	1,568	1,629
3	1,452	1,476	1,424	1,491
4	1,395	1,414	1,366	1,439
5	1,407	1,421	1,377	1,454
6	1,486	1,494	1,456	1,532
7	1,619	1,625	1,590	1,665
8	1,718	1,714	1,686	1,762
9	1,929	1,906	1,896	1,965
10	2,144	2,108	2,112	2,184
11	2,308	2,268	2,278	2,351
12	2,381	2,342	2,354	2,423
13	2,366	2,335	2,340	2,405
14	2,286	2,263	2,262	2,321
15	2,218	2,204	2,195	2,250
16	2,238	2,241	2,217	2,268
17	2,610	2,638	2,590	2,640
18	3,017	3,065	2,993	3,053
19	3,003	3,057	2,974	3,041
20	2,921	2,961	2,889	2,960
21	2,767	2,796	2,733	2,805
22	2,485	2,505	2,450	2,524
23	2,088	2,107	2,056	2,126
24	1,732	1,756	1,702	1,772

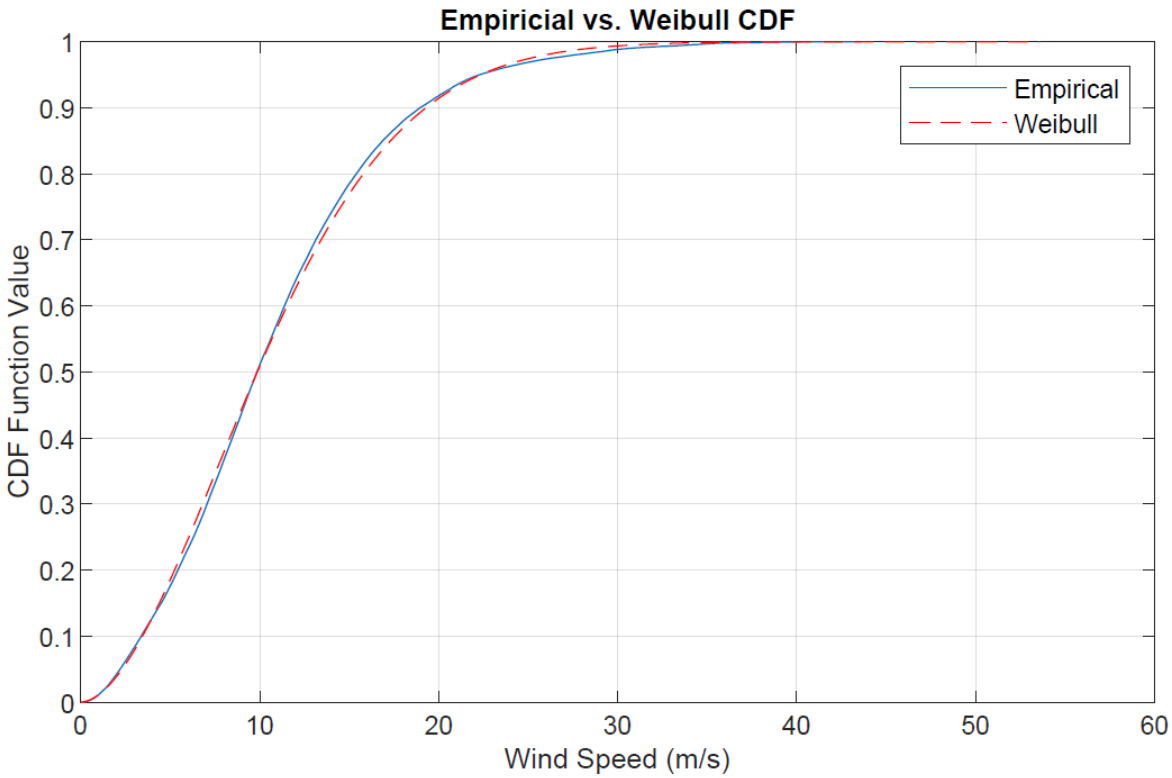

Figure A13: Graphical comparison of the empirical and Weibull cdf for all locational data combined

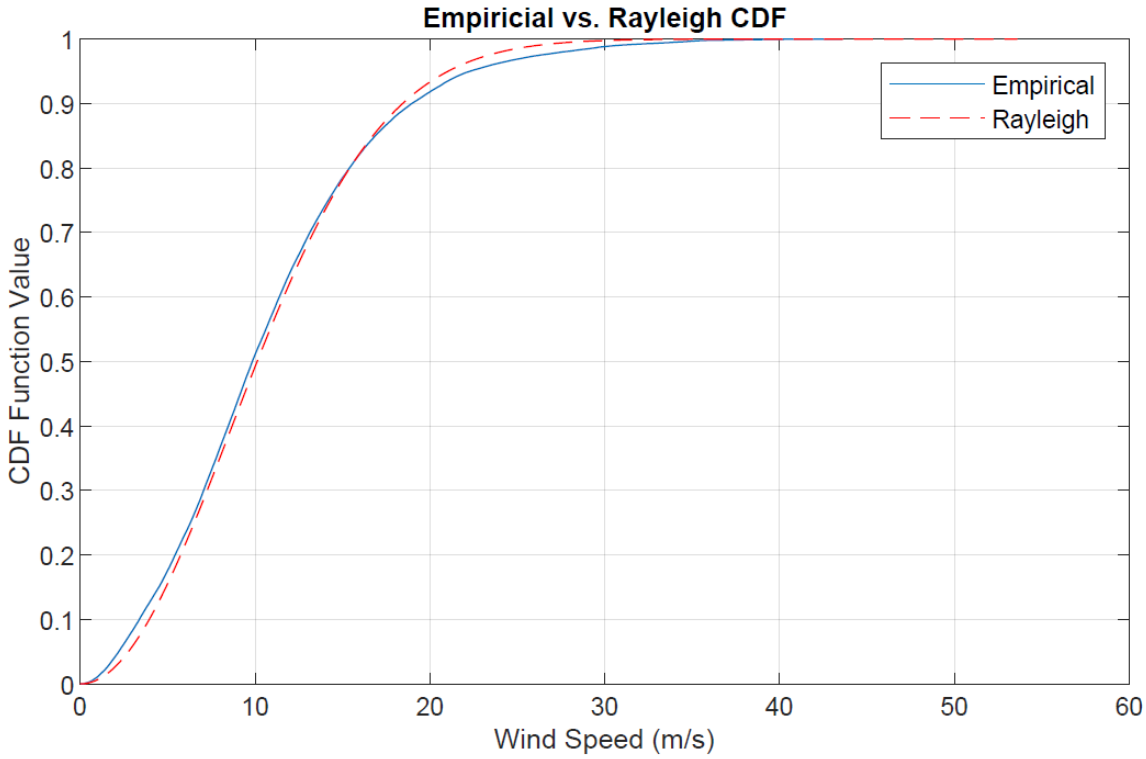

Figure A14: Graphical comparison of the empirical and Rayleigh cdf for all locational data combined

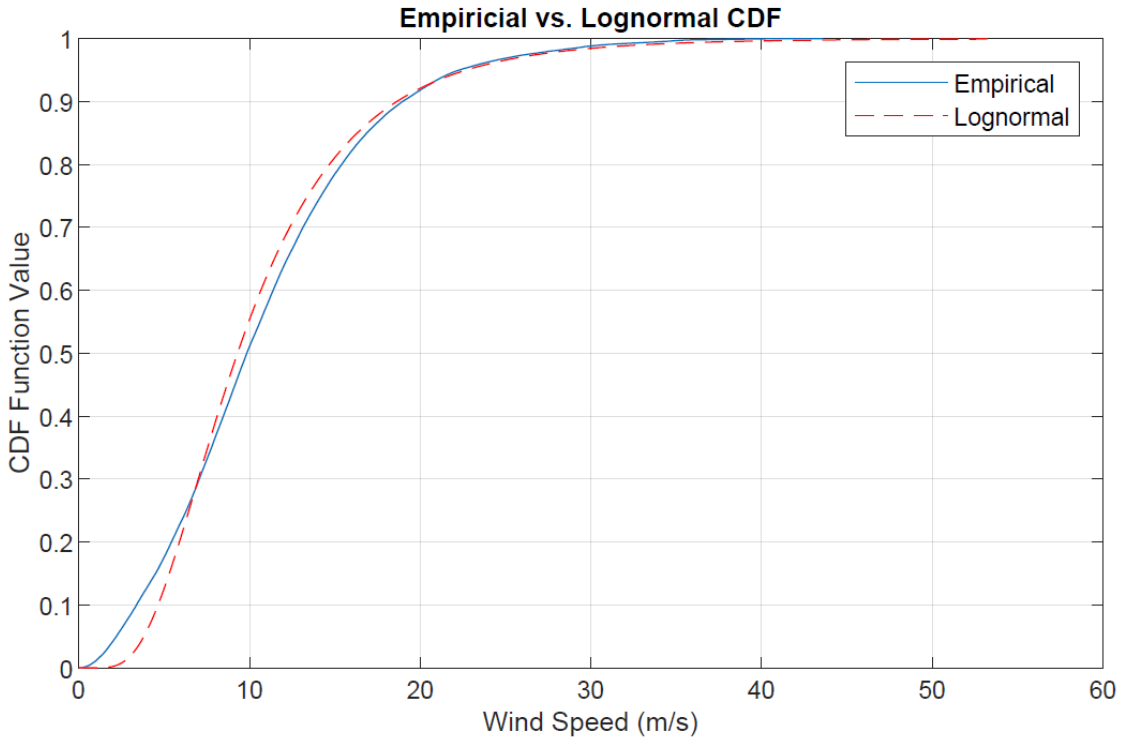

Figure A15: Graphical comparison of the empirical and Lognormal cdf for all locational data combined

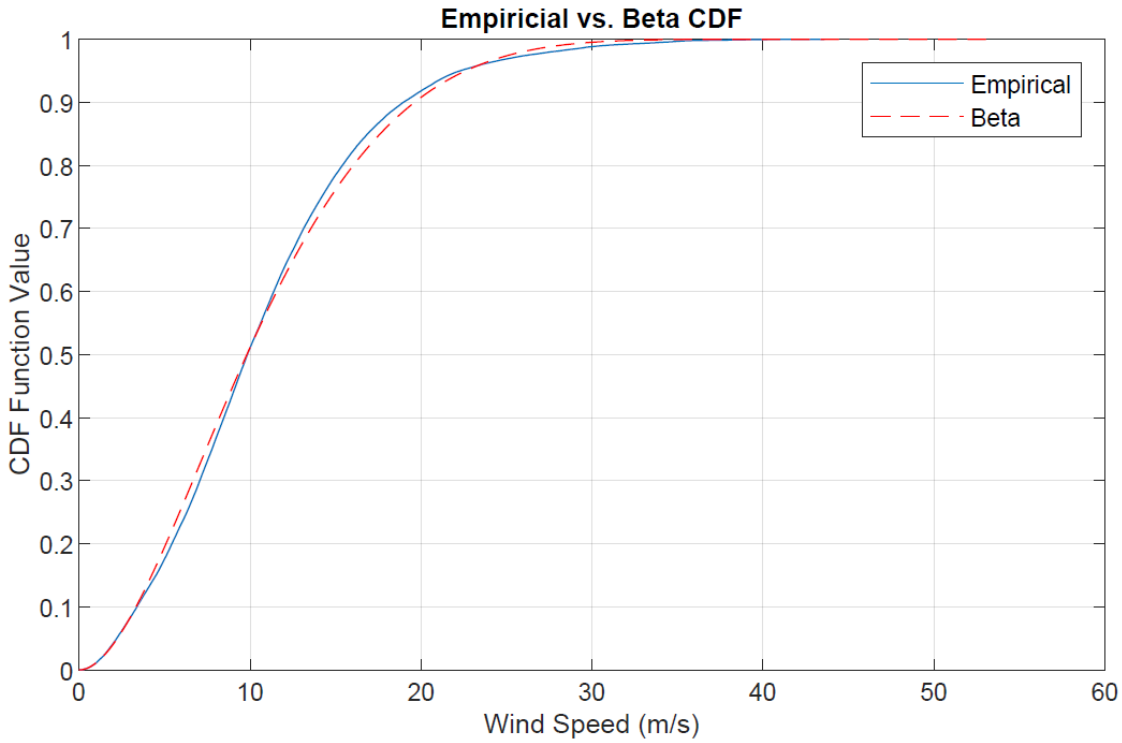

Figure A16: Graphical comparison of the empirical and Beta cdf for all locational data combined