Forecasting Centralized Wind Energy Generation: A Bayesian Approach

Tyler Valentine Nadir Siddiqui Theo Thormann

University of Virginia School of Data Science

1 Problem Description

As the energy industry focuses on the integration and expansion of variable renewable energy (VRE) as a viable alternative to fossil fuels, forecasting power generation is vital. When integrating VRE forecasts, up-ramps and down-ramps in VRE generation can be anticipated to "cost-effectively balance load and generation in intra-day and day-ahead scheduling," which leads to "reduced fuel costs, improved system reliability, and minimized curtailment of renewable resources" [7]. The most common sources of VRE are solar and wind power systems; both energy sources depend on highly variable weather conditions, such as cloud coverage and wind speeds, which introduce uncertainty to forecasting energy generation. According to the National Renewable Energy Laboratory, factors that affect forecast performance include "forecast time-horizon, local weather conditions, geographic scope, data availability (e.g., plant size, location, components), and data quality (e.g., consistency, accuracy, resolution)" [7].

In this report, we utilize a Bayesian approach to measure the uncertainty of centralized wind energy generation which is primarily dependent on variations in wind speed. The energy data are sourced from the ENTSO-E data portal and contain hourly measurements of wind energy generation (in megawatts) in Germany [5]. The wind speed data are sourced from NASA's MERRA-2 dataset and contain hourly measurements of wind speed 50 meters above ground (in m/s) from 256 locations across Germany [3], [6]. The data spans 2016 and 2017.

2 Probability Model

The objective of this model is to explain wind energy production given new observations of wind speed using a Bayesian approach. We perform this estimation using Bayesian linear regression with response y, intercept α, parameter β, data X, and error σ:

$$y = \alpha + \beta X + \sigma$$

The response and parameters of the model are generated from a probability distribution. The posterior probability of the model parameters is conditional on the training input and output data in the form of Bayes' theorem:

$$P(\beta|y,X) = \frac{P(y|\beta,X) * P(\beta|X)}{P(y|X)}$$

We estimate the posterior distribution, $P(\beta|y,X)$, which represents the probability distribution of wind energy generation conditioned on wind speeds. The posterior is estimated by multiplying the likelihood of the data, $P(y|\beta,X)$, by the prior probability of the parameters, $P(\beta|X)$, and dividing by the normalizing constant, $P(y|X)$.

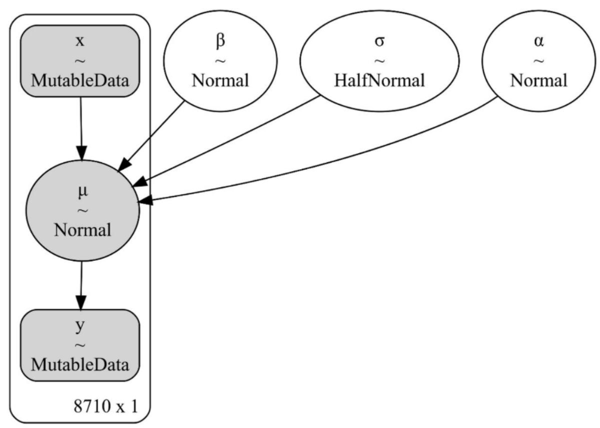

Figure 1: Probability Model Graph

Since the distribution of the data is approximately normal, our model assumes a Gaussian distribution for the priors of the intercept, coefficient, and response variable. To ensure only positive values, a Half-Normal distribution for the error term is assumed. The data, X, and the response variable, y, are treated as mutable to train the model on 2016 data and test it on 2017 data. The structure of the probability model is depicted in Figure 1.

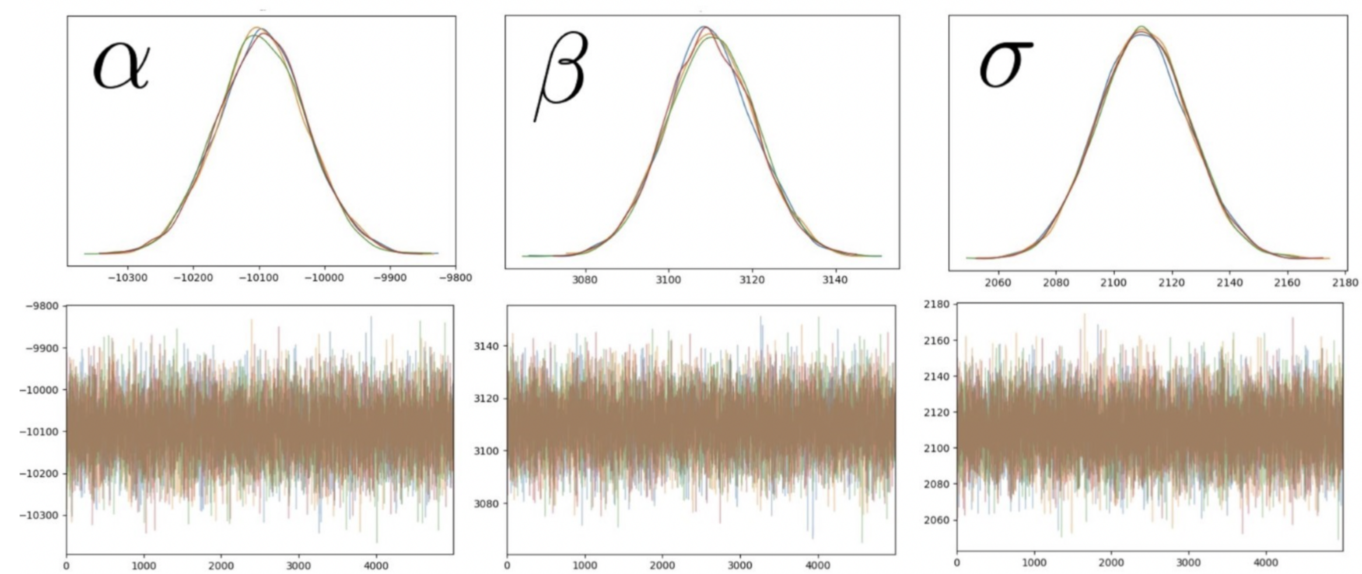

Figure 2: Trace Plot

3 Approach

We used Bayesian linear regression to predict wind energy generation from wind speed data. The advantage to using a Bayesian approach is the ability to estimate a probability distribution that explains uncertainty using prior information and evidence. While Ordinary Least Squares (OLS) produces point estimates of parameters, Bayesian linear regression produces an entire distribution for each parameter.

When calculating the posterior probability, a sampling method helps achieve a close approximation of the actual distribution. The algorithm samples observations using the No-U-Turn Sampler (NUTS), an extension of Hamiltonian Markov Chain sampling. NUTS uses a recursive algorithm to build a set of likely candidate points that spans a wide swath of the target distribution, stopping automatically when it starts to double back and retrace its steps [4]. NUTS is a preferable sampling method because it converges to the target distribution faster by searching along a gradient and eliminates the need to specify step size and desired number of steps. The algorithm is implemented using the PyMC [1] package in Python. The trace plots in Figure 2 demonstrate convergence of 5,000 samples to the stationary distribution, meaning our model generates the correct distribution.

4 Results

Our model calculates the mean for α as -10134.13, the mean for β as 3115.74, and the mean for σ as 2112.94. Since β represents the slope, we infer that, on average, every 1 unit increase in wind speed increases energy generation by about 3115.74 megawatts. The error term implies that our predictions deviate, on average, by about 2112.94 megawatts from the actual energy generation. Considering that the values for energy generation in 2016 range from 135 to 33626 megawatts, the error term seems appropriate.

Using the model trained on 2016 data, we calculated wind energy generation predictions based on wind speeds in 2017. Figures 3-6 represent two weeks of predictions in four different quarters of 2017. We chose to display the middle two weeks of each quarter instead of the entire quarter to declutter our graphs. The red points on the graphs represent daily predicted values of wind energy generation while the black line represents the actual wind energy production in 2017. The gray shading around the predictions represents the 95% high density interval (HDI) of the predictions. Comparing the actual energy output in 2017 with the 95% HDI of the predictions, the model makes close predictions. However, the actual energy output tends to be higher than the corresponding prediction, especially later in the year.

Our model measures uncertainty through the distributions corresponding to the model parameters. None of these parameters are a single measurement, but rather a distribution that measures the uncertainty relating to each of them. This measure of uncertainty is useful because wind speeds can be highly variable day to day. Accounting for uncertainty is crucial when forecasting wind energy generation because it allows wind power companies to approximate likely energy outputs for each day with a level of confidence.

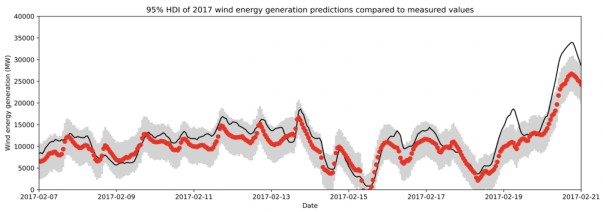

Figure 3: Confidence of Energy Predictions, February 2017

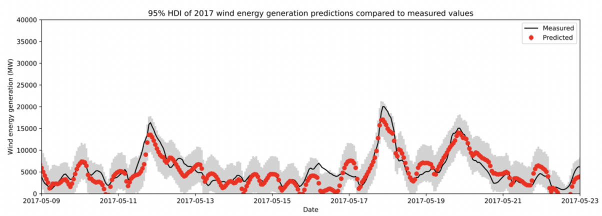

Figure 4: Confidence of Energy Predictions, May 2017

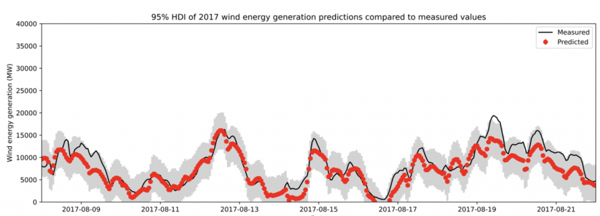

Figure 5: Confidence of Energy Predictions, August 2017

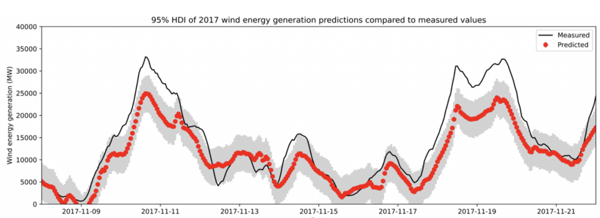

Figure 6: Confidence of Energy Predictions, November 2017

5 Conclusions

Plotting the predicted values for 2017 energy generation using the model trained with 2016 data shows that the model was effective at marking predictions using hourly wind data. However, the model slightly under-predicts energy generation. Fortunately, from a logistical perspective, it is better to under-predict energy generation than to over-predict it.

While these results are encouraging, there are limitations to this model. The predictions toward the end of 2017 were less accurate than expected based on predictions earlier in the year. One reason for this may be the variation in seasonal wind speed. Additionally, according to Bundesverband WindEnergie [2], there was an increase of 1792 wind turbines in Germany during 2017 which may also explain the large increases in energy generation toward the end of the year.

Another limitation of this model centers around the intercept of the model. Since the intercept is negative, energy predictions for wind speeds under about 3.25 m/s are likely to be negative, which is not possible. In general, the model suffers from a common pitfall of linear regression with poorer performance at lower and higher wind speeds and better performance at average wind speeds.

Further exploration of this topic can be pursued. The future of energy production will rely heavily on VRE if society is to avoid the catastrophic consequences of climate change. While energy production modeling's predictive power is getting more accurate, there will be a need for more accurate and robust models. As computers increase their processing power, more advanced methods can be applied to larger data sets, resulting in more accurate weather predictions. These weather predictions will be crucial to determine energy generation and assist energy production companies in predicting how much energy renewable energy sources will be producing at any given time.

References

[1] "API," *(API - PyMC 4.4.0 documentation*. [Online]. Available: https://www.pymc.io/projects/docs/en/stable/api.html. [Accessed: 21-Nov-2022].

[2] "German Wind Energy in Numbers," *Bundesverband WindEnergie*, 2022. [Online]. Available: https://www.wind-energie.de/english/statistics/statistics-germany/. [Accessed: 06-Dec-2022].

[3] Global Modeling and Assimilation Office (GMAO) (2015), *inst3 3d asm Cp: MERRA-2 3D IAU State, Meteorology Instantaneous 3-hourly (p-coord, 0.625x0.5L42), version 5.12.4*, Greenbelt, MD, USA: Goddard Space Flight Center Distributed Active Archive Center (GSFC DAAC), Accessed 11/16/2022 at doi: 10.5067/VJAFPLI1CSIV.

[4] M. D. Hoffman and A. Gelman, “The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo,” *arXiv*, Nov. 2011.

[5] Open Power System Data. “Data Platform Time Series,” 09-Jul-2017. [Online]. Available: https://data.open-power-system-data.org/time_series/2017-07-09. [Accessed: 15-Nov-2022].

[6] Open Power System Data. “Data Platform Weather Data,” 05-Jul-2017. [Online]. Available: https://data.open-power-system-data.org/weather_data/2017-07-05. [Accessed: 15-Nov-2022].

[7] Tian and I. Chernyakhovskiy, “Forecasting Wind and Solar Generation: Improving System Operations, Greening the Grid,” *National Renewable Energy Laboratory*, 01-Jan-2016. [Online]. Available: https://www.nrel.gov/docs/fy16osti/65728.pdf. [Accessed: 21-Nov-2022].