Forecasting Renewable Energy Generation

Harshitha Palegar, Farris Atif, Nasser Al-Rayes, Samrat Acharya

***Abstract*—In this paper, we will be analyzing energy supply and demand in order to allow energy companies to accurately ramp up production when needed. This research is important because it will allow for power utility companies to appropriately plan and match energy supply to demand, which will ultimately prevent negative consequences in the power grid utility operation. In the following, we will discuss our methods of data collection, data pre-processing and cleaning, model evaluation, and end deployment. Ultimately, this analysis will result in a clearer depiction of energy supply and demand patterns which can then be used to plan for more efficient energy distribution.**

I. Business Understanding

POWER grid utilities have been increasing the integration of renewable energy resources such as solar and wind in their generation portfolio; backed by global and regional decarbonization and clean energy goals. The global renewable generation capacity increased by 7.4% in 2019, with the lead increase being in solar (20%) and wind (10%) [1]. By 2025, the International Energy Agency (IEA) projects that solar and wind generation will cover 15-18% of the global electricity generation portfolio. Although the high penetration of renewables is advantageous from a clean and sustainable energy viewpoint, the power grid faces numerous challenges pertaining to reliability, economy, and power system security as a result [2]. In particular, these challenges get adverse once the renewables levels exceed 10% of the energy mix [3]. These challenges primarily surface due to the intermittent renewables supply, resulting in stochasticity of power generation.

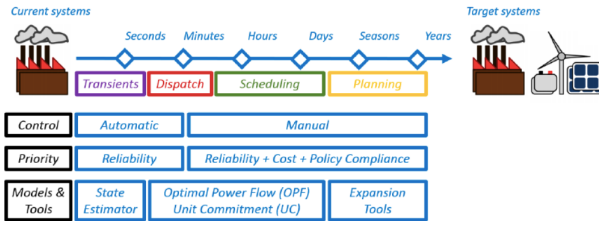

Fig. 1. Schematic diagram of power grid planning and operation.

As Fig. 1 shows, power grid utilities have four distinct phases of operation: i) Planning, ii) Scheduling, iii) Dispatch, and iv) Transients. In the planning phase, power grid utilities forecast their operations well ahead of time, typically by seasons and years in order to meet long-term policy goals such as decarbonization. The scheduling phase then deploys tools such as Optimal Power Flow (OPF) and Unit Commitment (UC) to schedule generators obtained from the planning phase and is usually scheduled one day ahead. The scheduled generators are then activated in the dispatch and transient phases in real-time on the day of. At times, while in the dispatch and transient stages, utilities must perform emergency actions which include load-shedding and turning on costly fossil-fuel generators during unexpected circumstances of energy generation-demand mismatch. However, this can come at a cost.

The mismatch between the generation and demand in real-time incurs technical and financial loss to the power grid utilities such as deviation in nominal operating frequency and voltage, causing wear and tear in power

grid infrastructures [4]. To avoid such consequences, power grid utilities turn on costly emergency fossil-fueled generators and/or cut out the customer demand during the generation-demand mismatch. In either case, the power grid utilities suffer cost and technical implications. The safe, reliable, and cost-effective operation of the power grid depends on the accuracy of generation and demand forecasting. With the high-penetration of renewable energy sources, the generation forecasting becomes more uncertain.

To visualize the aforementioned problem, we take the use case of Pennsylvania-New Jersey-Maryland (PJM) Interconnection, a power utility serving 65 million people in 13 states and Washington D.C. [5]. As seen in Fig. 2, the generation portfolio is significantly dominated by the conventional non-renewable energy resources while solar and wind share is just 0.4 %. However, this is expected to increase in the future. Fig. 2 shows the average hourly variation in solar and wind generation. Due to this intermittent nature of generation from solar and wind, once the use of these energies increase, the power utility faces various operational challenges. One of the significant challenges is a *duck curve* phenomenon, which occurs when the generation from solar farms suddenly starts and stops. As Fig. 4 shows, the net

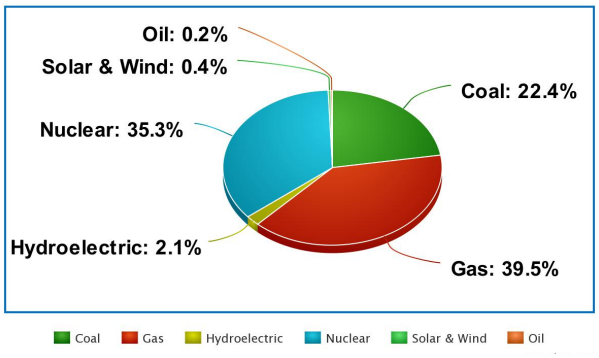

Fig. 2. Average generation portfolio observed in PJM.

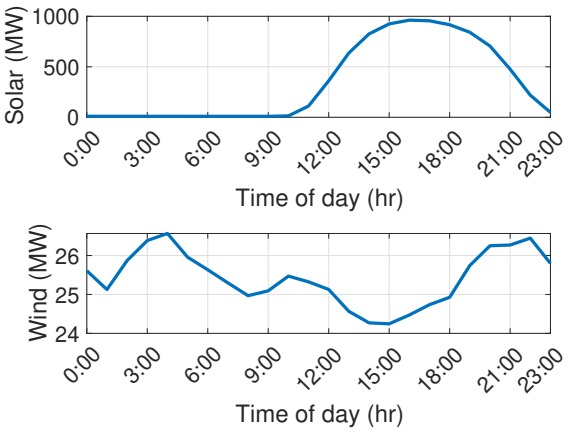

Fig. 3. Average hourly (in Universal Time Coordinated) Solar and Wind generation profile in PJM .

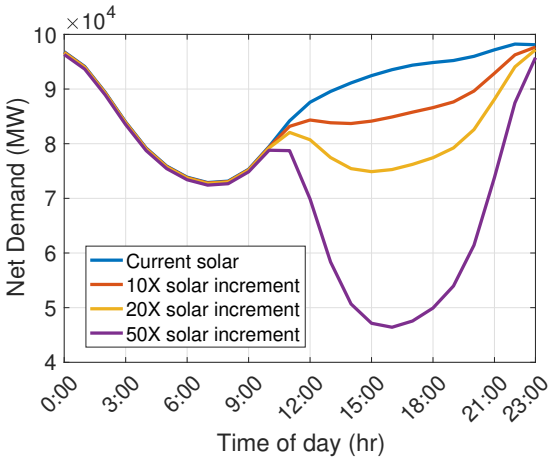

Fig. 4. Average Duck curve phenomenon with increase in Solar generation in PJM, time in Universal Time Coordinated.

demand is given as:

$$\text{Net Demand (MW)} = \text{Demand - Solar Generation}, \tag{1}$$

where MW is power measured in megawatts and net demand is the leftover demand that must be met by traditional energy sources. Currently, the low solar and wind penetration (Fig. 2) results in the absence of the duck curve phenomenon seen in Fig. 4. However, as the solar penetration increases, the net demand ramps up and down drastically within a short period of time. As seen in Fig. 4, a sudden decrease in net demand due to excess solar generation over 4 hours (hours 11:00 to 15:00 in Fig. 3) would require a severe shut down in conventional fuel

generators in order to balance mismatch of generation and demand. The opposite problem occurs when the net demand increases or the solar generation suddenly stops during sunset hours (hours 18:00 to 23:00). However, the start-up and shut-down time of these conventional generators is 2-10 hours on average [6] and the cost of doing so is more than $100 per MW. In both cases, the utility either cannot match the generation-demand instantly or, in the best case scenario, ends up paying high-cost for the generation-demand balance. Such variability issues can be extended to wind generation as well. To avoid the issues stemming from generation-demand mismatch, demand and renewable generation from solar and wind need to be accurately forecasted so that the utility schedules necessary nonrenewable generators beforehand and avoid incurring any costs.

Therefore, in this paper we attempt to forecast the net demand given the historical data on demand, generation, and weather conditions. Current research efforts [3], [7], [8] attempt to forecast either demand or generation using models ranging from neural networks to time series regression analysis, however few attempt to forecast net demand as we plan on doing.

II. Data Understanding

In order to forecast on the target variable net demand, we will be updating Eq. (1) to include wind generation, thus net demand will be represented as:

$$\text{Net Demand (MW)} = \text{Demand} - \text{Solar Generation} - \text{Wind Generation}. \quad (2)$$

Publicly available data on electricity demand and generation provided by PJM was examined. As discussed in Section I, PJM is a regional transmission organization that supplies energy for over 65 million people in the Eastern United States. From its headquarters, PJM coordinates this flow of electricity from generators to local utilities across a web of transmissions and distribution lines via a centralized supervisory control and data acquisition (SCADA) system. This particular RTO was chosen because of two reasons: 1) publicly available granular time-series data on electricity demand and generation and 2) increasing penetration of renewables such as wind and solar in its generation portfolio. Moreover, generation data is available for a multitude of energy sources, including conventional non-renewables such as nuclear, natural gas, and coal as well as renewables.

TABLE I
PJM Defined Regions and Corresponding States and Weather Data Location

PJM Region	Corresponding States	Chosen Area for Weather Data
MIDATL	DC, DE, MD, NJ, PA, VA	Baltimore
WEST	IL, IN, KY, MD, MI, OH, PA, TN, VA, WV	Cincinnati
SOUTH	NC, VA	Richmond

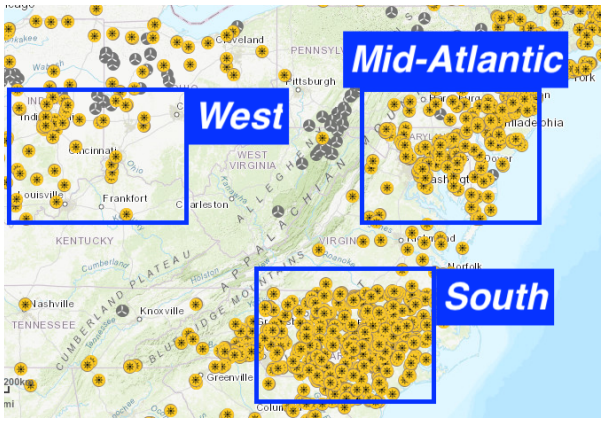

Fig. 5. Distribution of Solar and Wind Farms across the PJM network. Blue boxes indicate the specific area chosen to represent a region.

Using PJMS's data miner tool [9], individual datasets for solar generation, wind generation, and demand were pulled from 2019-01-01 00:00:00 to 2020-11-08 18:00:00. Additionally, each instance of this time series data equates to 1 hour, providing a total of 16,265 instances allowing for an adequate test/train split. Further data exploration showed that PJM has divided its network into six regions. Out of the six regions, three regions - "MIDATL", "WESTERN", and "SOUTHERN"

have significant penetration of solar and wind generation. The regions have various local power distribution authorities, but are coordinated with each other via PJM's centralized SCADA system and grid. Moreover, each of these regions provides a unique solar and wind generation capacity to the coordinated grid of PJM, which depends on their location, weather condition, and socio-economic factors such as renewable policy and population. In order to quantify these dynamic factors, solar generation, wind generation, and demand data were all grouped by region. Furthermore, a representative city for each of the 3 regions is chosen in Table I such that weather conditions (which were to be added as features) in a region remain uniform. Referencing U.S. Energy and Administration data [10], Fig. 5 shows the distribution of solar farms and/or wind generators within the PJM network.

To reiterate, up to this point three data frames for each region had been initialized on one uniform time series index. Each instance of the data corresponds to one hour with the following features: solar generation, wind generation, and energy demand. Moving forward, weather data was pulled from the National Centers for Environmental Information (NOAA) for each city then merged on each regional dataframe, providing 3 conclusive data structures suitable for analysis. Table II displays all the features used in our model before any feature engineering was conducted. It only displays features for the MIDATL region, however all features shown in Table II were also applied to the remaining two regions when modelling.

III. Data Preparation

The net demand, which we are attempting to forecast, is a function of renewable energy generation and demand. As a result, a minimum of 9 models (of the same algorithm) needed to be initialized - predicting solar

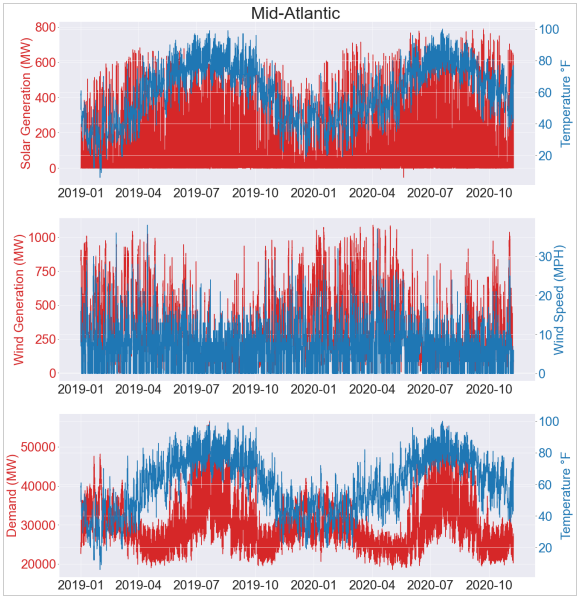

Fig. 6. Distribution of Target Variables (Solar Generation, Wind Generation, and Demand) & Subsequent Feature Exploration for Mid-Atlantic Region.

generation, wind generation, and demand for each of the 3 regions defined. Fig.6 shows a preliminary feature exploration from the Mid-Atlantic region (see Appendix for other regions) in which solar generation was plotted against temperature (Hourly Dry Bulb Temperature), wind generation against wind speed, and energy demand vs temperature (Hourly Dry Bulb Temperature). As a result, wind speed was shown to be highly correlated with wind generation, while solar generation and demand were less so with temperature. Both the West region, and the South region followed a similar trend (see Appendix). While this naive feature exploration proved informative, this approach made clear that in order to achieve a more comprehensive and accurate understanding of the feature space, a model would need to be fit. Before this could be done, the data needed to be cleaned and scaled appropriately. Features with uninterpretable values (data points with non-numeric/non classifiable values), and features with NaN values were removed. Additionally, it was understood that the nature of our goal, which is to forecast

TABLE II
SAMPLE FEATURE SPACE (PRE - FEATURE ENGINEERING) FROM MIDATL REGION

Hourly Altimeter Setting	Hourly Dew Point Temperature	Hourly Dry Bulb Temperature	Hourly Precipitation	Hourly Present Weather Type
Hourly Pressure Change	Hourly Pressure Tendency	Hourly Relative Humidity	Hourly Sea Level Pressure	Hourly Sky Conditions
Hourly Station Pressure	Hourly Visibility	Hourly Wet Bulb Temperature	Hourly Wind Direction	Hourly Wind Gust Speed
Hourly Wind Speed	MIDATL_solar	MIDATL_wind	MIDATL_demand	

net demand, requires us to evaluate our predictive model on some previously unseen testing data. In the context of this problem, the unseen data/features is forecasted weather data. Consequently, historical weather features which had no corresponding future/forecasted weather data were also removed. Lastly, 4 time index features were added to each dataframe: hour, day of the week, week number, and month. These features internalize the variability of the generation and demand across various seasons, weekdays, time of the day, and months. This is important as the generation and demand profile is very unique across these features. For example, the peak demand occurs at different times in summer vs winter, weekend vs weekday, and day vs night. Similarly, solar and wind generation vary significantly across a day as shown in Fig. 3.

All of the features, post-cleaning, were normalized using *sklearn preprocessing* in order to allow for generalizations in the evaluation of predictions for each of the target variables, irrespective of region. Table III summarizes this final feature space.

This aforementioned union of data engineering, naive feature selection, and domain based knowledge; then formed a sufficient basis allowing for supervised learning approaches to better understand the feature space. So, regression trees with 'Mean square error' criterion were chosen to predict each of three target variables for each of the three regions. Doing so was an attempt to capture non-linear trends in data (further discussed in the next section), and because of the inherent interpretability of the results. Moreover, this allows us to circumvent any

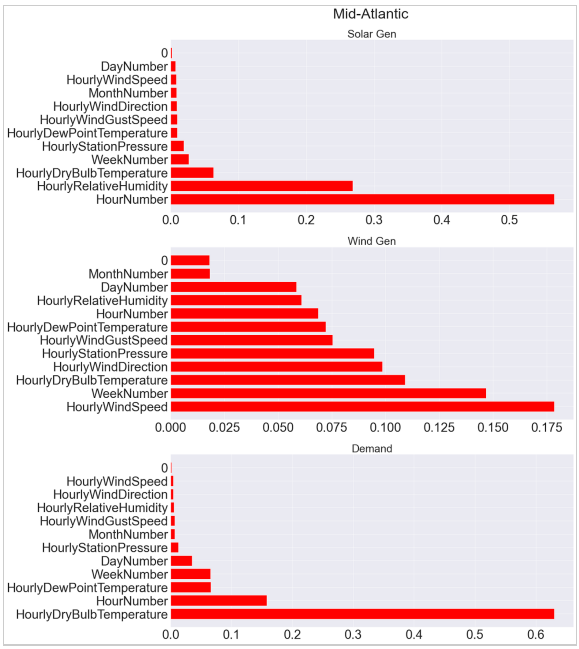

Fig. 7. Regression Tree Feature Importance for Mid-Atlantic Region.

co-linearity issues that would arise between weather features if linear regression coefficients were to be interpreted. The figure above shows a visualization of the feature importance for the Mid-Atlantic region (see Appendix for West and South regions), while Table IV summarizes the top 3 features for each region.

The results from the trees (shown in Table IV) verify our initial hypothesis that wind speed plays a major role in generation, since each of the three trees for wind generation had 'Hourly Wind Speed' as a root node. In the context of regression trees, the root node provides the most information since this initial node is the feature which reduces the residuals of the data the most. Moreover, it is evident that the wind data for each unique region is relatively similar in structure as each of the three regions are providing almost the same top 3 features (albeit Hourly Station Pressure in the Southern

TABLE III
FINAL FEATURE SPACE FROM MID-ATLANTIC REGION. HIGHLIGHTED FEATURES ARE EXCLUSIVE TO SOLAR GENERATION (BLUE), WIND GENERATION (GREEN), AND DEMAND (RED).

Hourly Dew Point Temperature	Hourly Dry Bulb Temperature	Hourly Relative Humidity	Hourly Station Pressure	Hourly Visibility
Hourly Wind Direction	Hourly Wind Gust Speed	Hourly Wind Speed	Month Number	Week Number
Day Number	Hour Number	Mid-Atlantic_solar	Mid-Atlantic_wind	Mid-Atlantic_demand

TABLE IV
REGRESSION TREES FEATURE IMPORTANCE FOR EACH REGION & EACH TARGET VARIABLE.

Target	West	Mid-Atlantic	South
Solar	1) Hour Number 2) Hourly Humidity 3) Hourly Dry Bulb Temp	1) Hour Number 2) Hourly Humidity 3) Hourly Dry Bulb Temp	1) Hour Number 2) Hourly Humidity 3) Hourly Dry Bulb Temp
Wind	1) Hourly Wind Speed 2) Week Number 3) Hourly Dry Bulb Temp	1) Hourly Wind Speed 2) Week Number 3) Hourly Dry Bulb Temp	1) Hourly Wind Speed 2) Hourly Station Pressure 3) Hourly Dry Bulb Temp
Demand	1) Hourly Dry Bulb Temp 2) Week Number 3) Hourly Number	1) Hourly Dry Bulb Temp 2) Hour Number 3) Hourly Dew Point Temp	1) Hourly Dry Bulb Temp 2) Hour Number 3) Hourly Dew Point Temp

Region). Alternatively, solar generation and demand are both heavily linked to temperature and hour number across all regions. Ultimately this addition of time indexed features (hour, week, etc.) proved beneficial in forecasting the demand and generation as they validated our prior assumptions that these factors internalize the variability and uniqueness in demand and generation.

IV. MODEL SELECTION AND EVALUATION

A simple linear regression model, decision tree, and random forest were tested and evaluated in order to see which would be able to most accurately predict solar and wind energy production, allowing for utility companies to optimally allocate energy generation from their appropriate sources. As we are dealing with a regression problem, the most sensible error metrics are r^2 and Root Mean Squared Error (RMSE). To ensure our RMSEs can be compared across regions and data types, all of our data were normalized on a min/max scalar. As mentioned earlier, we are attempting to optimize the usage rates out of renewable energy sources by forecasting their production 2 days ahead, allowing utility companies to adjust operations accordingly. To evaluate our models' ability to forecast into the future, we split our training and test set on time-based intervals. Our data ranged from 01/01/2019 to 11/08/2020. In order to simulate forecasting and be able to apply error metrics to understand how well our models are performing, we trained our models on all but the last week of data. By keeping the test set as our final week of data, we're allowing ourselves to simulate the forecasting capabilities of our models, while also ensuring that we can pull the necessary accuracy metrics of our models. Our baseline model was a simple linear regression, which we also compared against a simple decision tree and random forest. This section will be broken down by model type, in which we discuss the modeling evaluation for each region that we studied. Our metrics of evaluating the baseline, as mentioned above, are r^2 and RMSE; our final model of choice will be the one that yields the

TABLE V
ERROR METRICS FOR A LINEAR REGRESSION MODEL

Target	West		Mid-Atlantic		South	
	r^2	RMSE	r^2	RMSE	r^2	RMSE
Solar	0.05	0.140	0.32	0.057	0.27	0.090
Wind	-2.08	0.126	0.33	0.058	0.76	0.021
Demand	-1.03	0.016	-0.64	0.011	-1.44	0.016

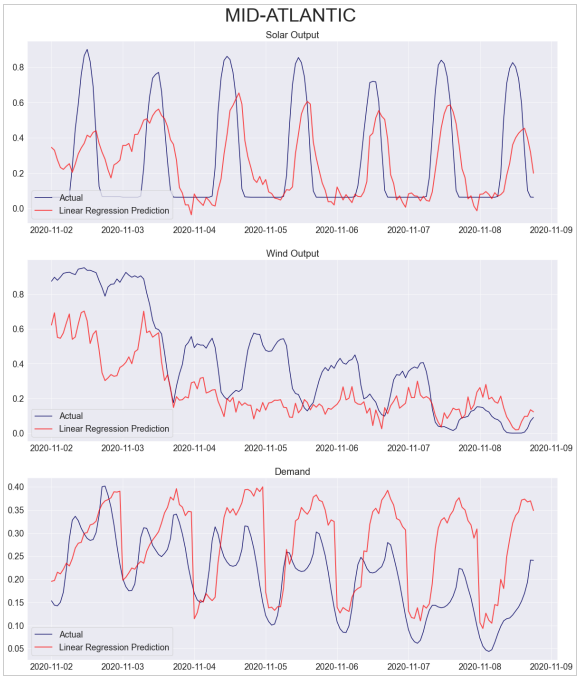

Fig. 8. Actual and Predicted Outputs Across the Mid-Atlantic Region with a Linear Regression Model.

highest r^2 value and lowest RMSE.

A. Linear Regression

We used a linear regression model as our baseline due to its simplicity and ability to give us clear information into the relative magnitudes at which each feature impacts the target variable. We kept all features as linear estimators to the target and decided to not add any polynomial characteristics to our data. Table V highlights the evaluation metrics of the linear regression model.

It is worth noting that our model's ability to predict solar, wind, and demand in the Western region of PJM's utility network is very poor. However, this is reasonable as the Western region encompasses many states, ranging from Illinois to parts of Pennsylvania and West Virginia, as seen in Table I. This carries with it inherent challenges, as weather from a single region will not carry significant predictive value for the output of our renewable energy sources. This explains why the predictive value of our linear regression model on PJM's West data was so poor. However, we're able to see a significant increase in the predictive capabilities of our model when we look at both our South and Mid-Atlantic data. As the energy data coming out of these regions is far more localized, the weather data we used carried far more predictive value. Fig. 8 shows the predictive abilities of our model for the Mid-Atlantic region. The linear regression plots for both the West and South regions can be found in the Appendix.

B. Decision Tree

We used a decision tree regressor to see if any nonlinear relationships within our data could be captured more adequately given the feature space. A linear regression model was unable to capture many of the nonlinear relationships that existed in our data, resulting in the poor fit that we saw in the previous section. We hoped that a decision tree would be able to account for this. However, decision trees have a tendency to overfit, which appears to be happening based on the poor modeling metrics shown in Table VI.

As seen, the r^2 values for our decision tree in both the wind and demand test sets are fairly low. When looking at Table VII, we are able to see how much

TABLE VI
ERROR METRICS FOR A DECISION TREE

Target	West		Mid-Atlantic		South	
	r^2	RMSE	r^2	RMSE	r^2	RMSE
Solar	0.28	0.1070	0.74	0.022	0.60	0.0490
Wind	-1.838	0.116	0.30	0.061	0.14	0.0779
Demand	-0.45	0.011	0.38	0.004	0.38	0.004

TABLE VII
PERCENT DIFFERENCE FROM BASELINE TO DECISION TREE

Target	West		Mid-Atlantic		South	
	r^2	RMSE	r^2	RMSE	r^2	RMSE
Solar	460	23.60	131	61	122	45
Wind	12	8	-10	-5	-82	-270
Demand	56	31	159	64	126	75

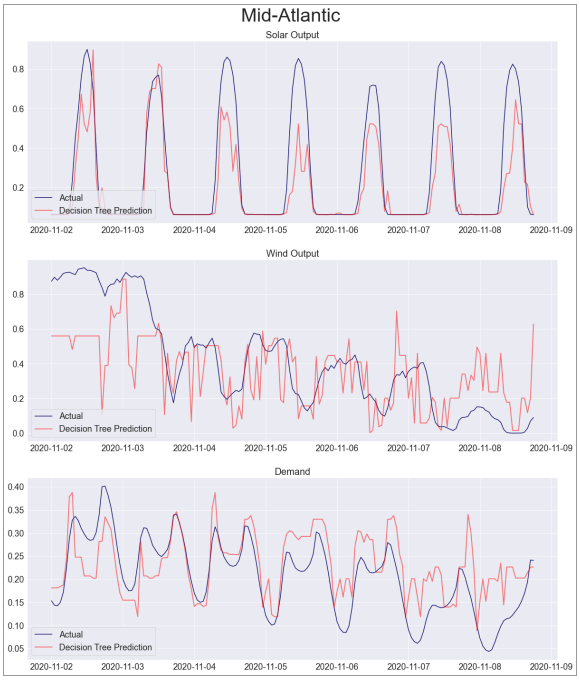

Fig. 9. Actual and Predicted Outputs Across the Mid-Atlantic Region with a Decision Tree.

better these graphs perform, excluding the predictions of wind generation in the Mid-Atlantic and South region. A positive percentage indicates an improvement over the baseline model. When we look at Fig. 9, we see that these trees are likely overfitting, causing the variance in the output. Nonetheless, it appears as though a decision tree does capture the nonlinear relationships with higher accuracy than a simple linear regression model can. Once again, this plot focuses on data from the Mid-Atlantic. West and South plots can be found in the Appendix.

C. Random Forest

In order to account for both the nonlinear relationships in our data and the overfitting tendencies of a decision

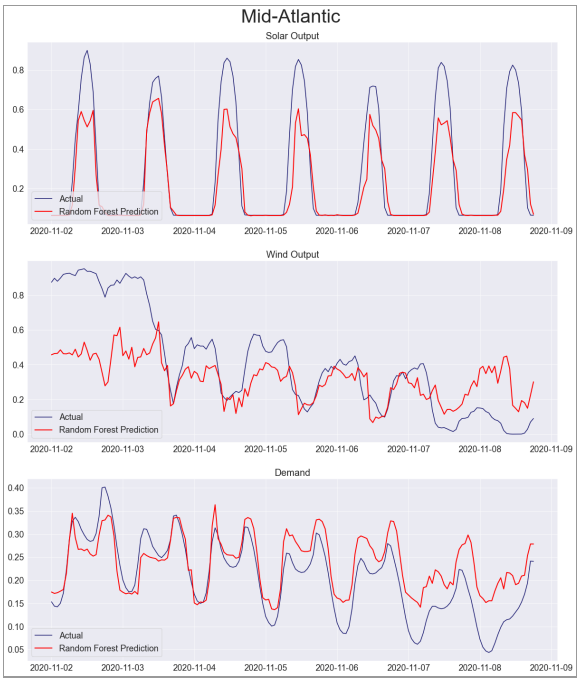

Fig. 10. Actual and Predicted Outputs Across the Mid-Atlantic Region with a Random Forest.

TABLE VIII
ERROR METRICS FOR A RANDOM FOREST

Target	West		Mid-Atlantic		South	
	r^2	RMSE	r^2	RMSE	r^2	RMSE
Solar	0.25	0.110	0.81	0.016	0.56	0.055
Wind	-1.25	0.092	0.31	0.060	0.66	0.030
Demand	0.22	0.006	0.56	0.003	0.71	0.002

tree regressor, a random forest was used. As seen from Table VIII, this model offered the overall best performance. Looking at Table IX, we once again see that the random forest regressor does significantly better than the linear regression model, excluding the wind prediction in the Mid-Atlantic and South regions. Overall, this model does better than the decision tree as well, suggesting that this is the best model for us to use. Fig. 10 visualizes the predictions from the random forest.

V. DEPLOYMENT

Deploying this model involves feeding our trained model forecasted weather data in order to project renewable energy supply, which would in turn allow utility

TABLE IX
PERCENT DIFFERENCE FROM BASELINE AND RANDOM FOREST

Target	West		Mid-Atlantic		South	
	r^2	RMSE	r^2	RMSE	r^2	RMSE
Solar	400	21	153	72	107	39
Wind	40	27	-6	-3	-13	-42
Demand	126	63	188	73	149	88

companies to optimize their load schedules. In order to do so, we collected forecast weather data from each of the regions we studied and fed this data into our models to see what its renewable energy forecasts would be. To evaluate the accuracy of these predictions, we compared real solar, wind, and demand with the projected values from our model. The results are presented in Fig. 11. From our results in Fig. 10 and Table VIII, we know that our model is able to accurately predict solar generation, wind generation, and demand when fed real weather data. However, looking at Fig. 11, we notice that the model's predictive abilities are not as strong. This is likely due to poor forecast wind data. Solar output and demand predictions are still fairly strong, but our model's ability to predict wind output is almost nonexistent.

With information about future net demand available well ahead of time, grid operators will have ample time to choose how to fill the gap and can create alert systems that dispatch orders to ramp up traditional energy production when needed. Such planning will also make re-allocating resources more efficient. As with any system deployment, our predictive model will be periodically tested against the true value we are estimating: net demand. By checking our model performance this way, we can monitor any variability of the predictions and detect any patterns within the variability. Doing so will also show when updates need to be made and can indicate the need to further test the model with more scrutiny. However, one thing to be aware of when deploying the system is extreme weather phenomena

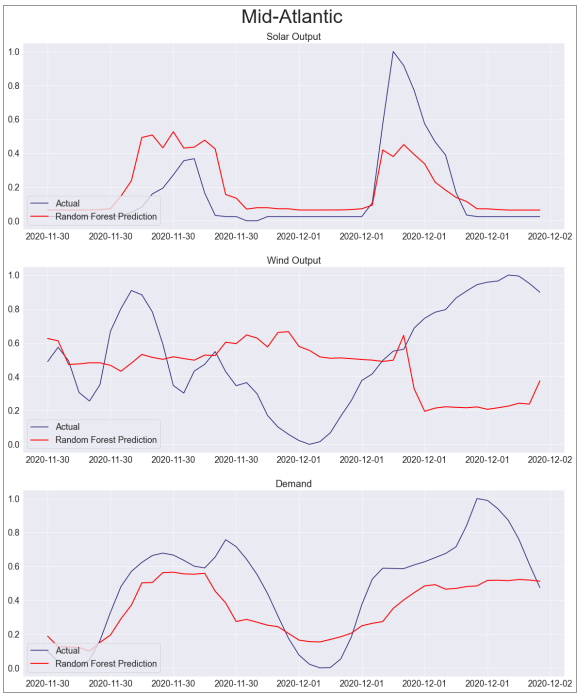

Fig. 11. Actual and Forecast-Based Predicted Outputs Across the Mid-Atlantic Region.

such as long droughts, hurricanes and heat waves as they will impact energy demand and supply. The time frame in which our data was collected may not have had these weather events occur but when they do, data can be collected and the model can be updated accordingly. An ethical concern to take into consideration throughout deployment and after would be any concept drift in terms of energy laws and regulations. The United States EPA oversees renewable energy generation, change in technology, and climate change in its Energy Policy Act. Consequently, any change in the act pertinent to our model would require needed updates to uphold ethical standards.

VI. CONCLUSION

As institutions and businesses strive towards the goal of decarbonization and clean energy in efforts to aid in the fight against climate change, it is important that power utilities adapt their planning phases to accommodate for this change. With renewable energy, specifically

solar and wind, growing in the energy generation portfolio, power utilities must prepare for any mismatch in terms of generation and demand. By utilizing our model, power utilities can forecast the net demand and plan accordingly; avoiding any serious consequences to the power grid and infrastructure.

Of course, there are other alternatives to our model including Facebook's Prophet which is also a time series based model. Prophet's ability to forecast data takes into account seasonal effects and works well in instances where large seasonal historical data is available-which is true in our case. Moving forward, Facebook's Prophet could be tested and deployed against our model to see which is cheaper to deploy and is more accurate. However, the model we have constructed is a good starting point into the insight of forecasting net demand and can offer information to aid power utilities in the age of fighting climate change.

REFERENCES

[1] (2020) IRENA renewable capacity highloghts. [Online]. Available: https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2020/Mar/IRENA_RE_Capacity_Highlights_2020.pdf

[2] E. J. Coster, J. M. Myrzik, B. Kruimer, and W. L. Kling, "Integration issues of distributed generation in distribution grids," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 28–39, 2010.

[3] E. B. Ssekulima, M. B. Anwar, A. Al Hinai, and M. S. El Moursi, "Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: a review," *IET Renewable Power Generation*, vol. 10, no. 7, pp. 885–989, 2016.

[4] S. Acharya, M. S. El Moursi, and A. Al-Hinai, "Coordinated frequency control strategy for an islanded microgrid with demand side management capability," *IEEE Transactions on Energy Conversion*, vol. 33, no. 2, pp. 639–651, 2017.

[5] Who we are. [Online]. Available: https://www.pjm.com/about-pjm/who-we-are.aspx

[6] (2019) Flexibility in conventional power plants. [Online]. Available: https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2019/Sep/IRENA_Flexibility_in_CPPs_2019.pdf?la=en&hash=AF60106EA083E492638D8FA9ADF7FD099259F5A1

[7] K. Muralitharan, R. Sakthivel, and R. Vishnuvarthan, "Neural network based optimization approach for energy demand prediction in smart grid," *Neurocomputing*, vol. 273, pp. 199–208, 2018.

[8] L. Suganthi and A. A. Samuel, "Energy models for demand forecasting—a review," *Renewable and sustainable energy reviews*, vol. 16, no. 2, pp. 1223–1240, 2012.

[9] Data miner 2. [Online]. Available: https://dataminer2.pjm.com/list

[10] U.S. EIA. [Online]. Available: https://www.eia.gov/state/maps.php

VII. CONTRIBUTION

Harshitha led the Business Understanding and also contributed to the Data Understanding, Deployment, and Conclusion sections.

Farris led the Data Understanding and Data Preparation sections and also contributed to most of the figure generation.

Nasser led the Modeling & Evaluation and Deployment sections and contributed to the figure generation.

Samrat brought significant domain expertise and contributed to all sections. He also led the Business Understanding and LaTeX file write-up.

APPENDIX

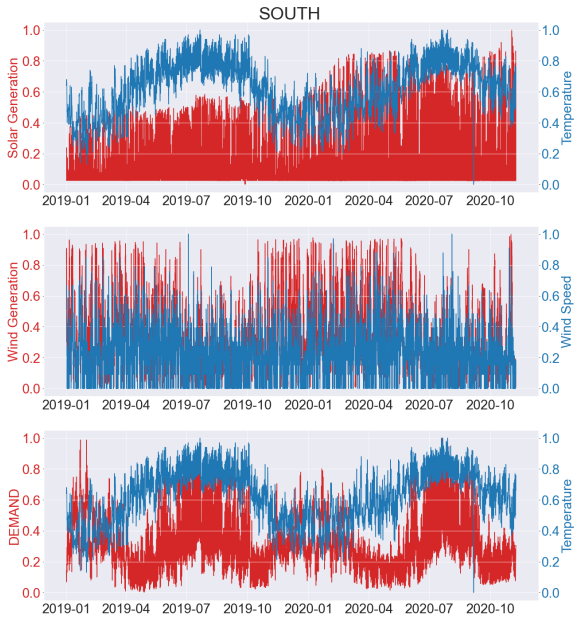

Fig. 12. Distribution of Target Variables (Solar Generation, Wind Generation, and Demand) & Subsequent Feature Exploration for South Region.

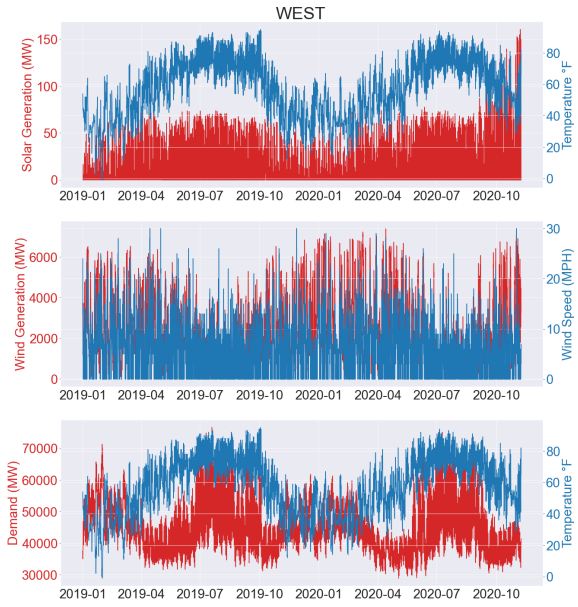

Fig. 13. Distribution of Target Variables (Solar Generation, Wind Generation, and Demand) & Subsequent Feature Exploration for West Region.

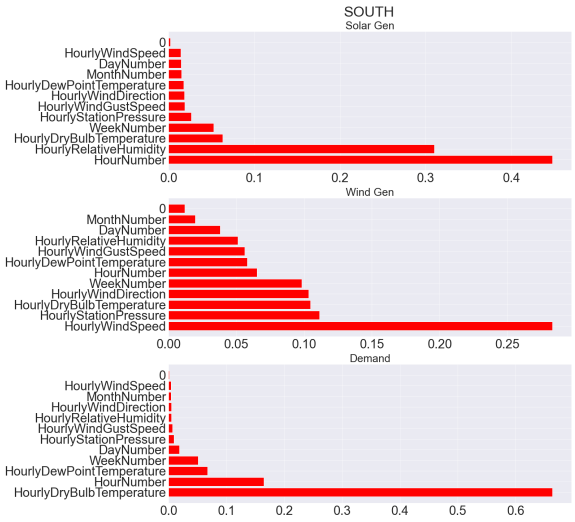

Fig. 14. Regression Tree Feature Importance for South Region.

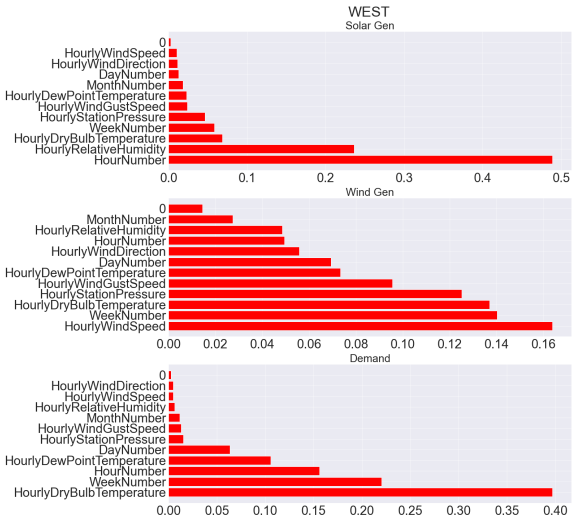

Fig. 15. Regression Tree Feature Importance for West Region.

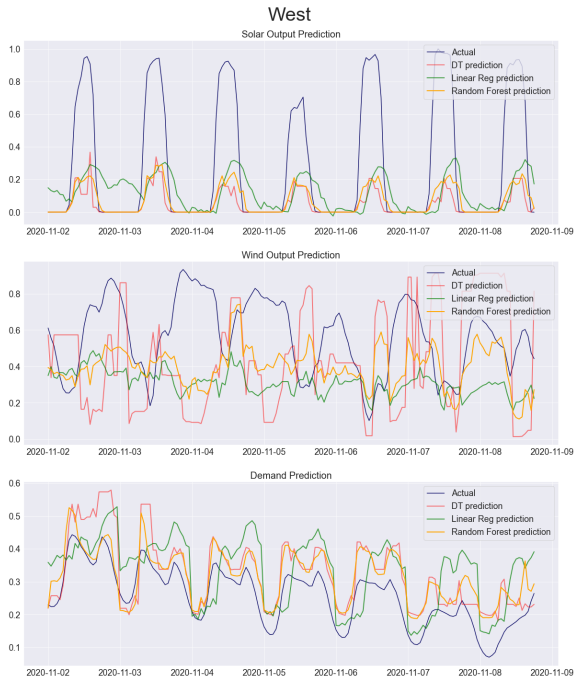

Fig. 16. Actual and Predicted Outputs Across the West Region

Fig. 18. Actual and Forecast-Based Predicted Outputs Across the PJM West Region

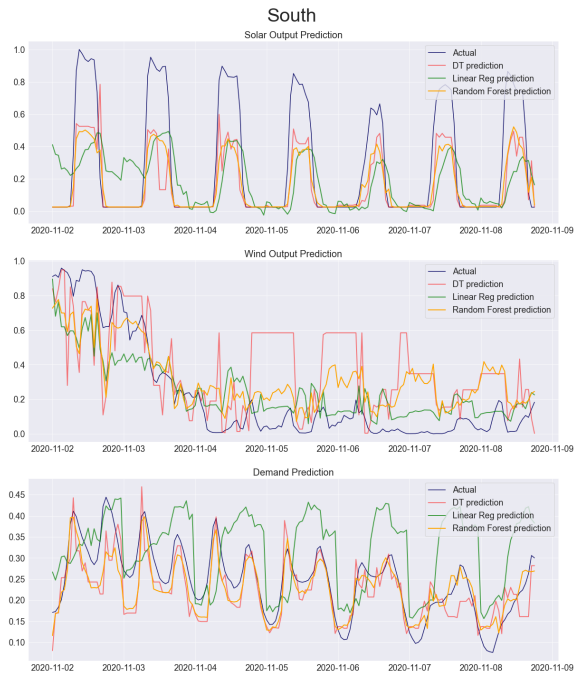

Fig. 17. Actual and Predicted Outputs Across the West Region

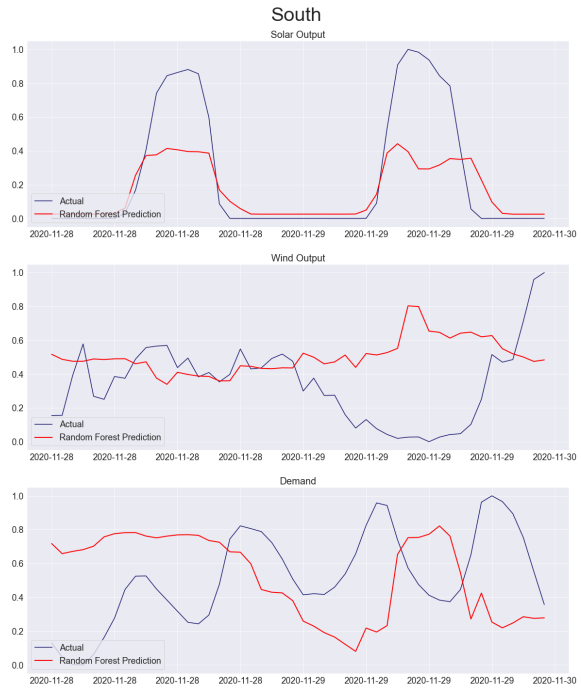

Fig. 19. Actual and Forecast-Based Predicted Outputs Across the PJM Southern Region