The effect of weather data on probabilistic electricity load forecasts

Manuel Sage [*1] **Thomas Wright** [*1] **Tara Akhound-Sadegh** [*1]

Abstract

Electricity generation is the foremost contributing sector to the global anthropogenic carbon footprint. Load forecasting plays an important role in reducing inefficiencies in the grid, helping to decrease unnecessary greenhouse gas (GHG) emissions. Given their effect on demand and the generation of renewables, weather patterns have varying consequences with respect to the grid, including causing blackouts and wasting resources. These effects are expected to grow in numbers and severity as climate change's (CC) influence on weather increases. While probabilistic machine learning methods are gaining traction in this domain, little research has been done to understand the effect of weather specifically on these probabilistic forecasts. In this paper, we investigate this impact through a case study on two grids, Alberta and the Netherlands, applying quantile random forests (QRF), quantile gradient boosting (QGB) and Bayesian long short term memory (LSTM) networks. Experiments are conducted on two time horizons, namely 1 hour and 24 hours, both with and without weather data. The nature of these forecasts and their means of evaluations limit us in definitively asserting that the models improve in absolute terms as a result of the incorporation of weather data. That being said, the trends observed indicate that including weather data, in particular forecast weather data, more often than not improves probabilistic forecasts, and always improves at least one of the two main metrics used in evaluation. This work is meant to highlight preliminary results, as well as to provide insight as to how further developments can be leveraged in practice and contribute to the decarbonization of the electric grid, encouraging further research on the topic.

[*]Equal contribution [1]McGill University. Correspondence to: Manuel Sage <manuel.sage@mail.mcgill.ca>, Tara Akhound-Sadegh <tara.akhound-sadegh@mail.mcgill.ca>, Thomas Wright <thomas.wright2@mail.mcgill.ca>.

COMP767 final project report - McGill University

1. Introduction

1.1. Electricity Grids and Climate Change

The generation of electricity is the leading contributor to anthropogenic global warming, accounting for 25% of global GHG emissions according to IPCC's 2014 report on CC (Pachauri et al., 2014). Efforts to transition power generation to renewable energies are aggravated by steadily increasing demand. In 2018, global electricity demand grew by 4%, nearly twice the rate of overall energy demand (IEA, 2019). Between 2018 and 2050, power generation is expected to increase by 79% (Kahan, 2019), making the decarbonization of electric grids an essential challenge to overcome in order to meet the goal of the Paris Climate Agreement of holding global warming below 1.5 degrees Celsius compared to pre-industrial era.

Due to a lack of large-scale energy storage, electricity supply must constantly and instantaneously match electricity demand (Rolnick et al., 2019; Pierpont et al., 2017). Accurately forecasting generation as well as demand is therefore a vital aspect of grid operations. For deregulated electricity markets, system or transmission operators balance supply and demand to ensure safe and stable operation. The operators schedule different types of reserves typically distinguished by the time-horizon of their deployment. Moment-to-moment changes are compensated by automatic generation control (AGC), a system that allows generators in power plants to adjust to the system frequency (Glavitsch & Stoffel, 1980). On a timescale of minutes, the spinning reserve stabilizes electric grids. Generators that are already synchronized to the grid increase or reduce their power output to level demand and supply. For intraday and daily balancing, the ability to ramp-up additional generators is important in case of misalignments.

A drawback of most renewables is their limited suitability as reserves. Wind and solar power are generated in non-synchronous generators. Additionally, the intermittency of renewables increases the need for reserves (Denholm et al., 2016). For instance, when demand peaks in the evening, solar energy is phasing out as the sun sets (Pierpont et al., 2017).

Precise demand forecasting is becoming increasingly important to facilitate grid operation, reduce the need for (often

GHG emitting) reserves, and allow for a higher penetration of renewable energy into electricity grids (Islam et al., 2020).

1.2. Forecasting Demand in Electricity Grids

As mentioned in section 1.1, electricity demand forecasting plays an essential role in managing the supply-demand system of the electricity grids and their markets. Additionally, long-term demand forecasting can play an important role in the sustainable development of countries, allowing them to plan for resources and better schedule for reserves (Ghalehkhondabi et al., 2017). The nature of the electricity usage, however, is extremely complicated and dynamic. It depends upon a spectrum of variables, such as the weather and socioeconomic factors, necessitating the use of sophisticated algorithms to predict the future energy consumption accurately (Weron, 2014; Ghalehkhondabi et al., 2017). However, even in the most sophisticated models, forecasting errors are unavoidable. These errors can cause significant social and economic problems and even be disastrous in extreme circumstances; underestimating these forecasts can lead to power outages while overestimating them can lead to a waste of resources (Son & Kim, 2020). Current methods of load forecasting rely on providing point forecasts, meaning they only return a single numerical value and cannot give any information regarding the uncertainty associated with it. Understanding the uncertainty in these forecasts can help electricity grids operate more efficiently and might be useful in avoiding some of the adverse consequences of forecast errors mentioned above. As an alternative to point forecasts, probabilistic load forecasting (PLF) can help provide a more comprehensive understanding of the trends in the electricity consumption.

1.3. Impact of Weather on Electricity Demand

The electricity use by residential, commercial and industrial sectors is heavily dependent on the climactic conditions (Auffhammer et al., 2017). For example, the electricity required for heating and cooling of residential spaces, which is estimated to account for half of the residential electricity load, is highly contingent on the ambient temperature (Allen-Dumas & Cunliff, 2019). Weather itself, however, is chaotic and difficult to predict, which makes predicting its impact on future electricity consumption equally difficult. Additionally, as CC-induced shifts in the weather patterns become more severe in the coming years, we can expect volatile and unforeseen trends in global energy consumption (Auffhammer et al., 2017). Consequently, understanding the weather-dependent uncertainty in the predictions of electricity demand forecasting models, will become even more important in the future.

Moreover, the increasing integration of renewables into the grid will present additional challenges for forecasting models. Due to weather-related intermittency of power generation via renewable sources, having a better understanding of the uncertainty associated with predictions becomes more important: it will allow grid operators to perform better risk analyses incorporating it into their management strategies (Hong & Fan, 2016).

As we shift towards the large scale deployment of smart grids (electricity grids that allow for two-way communication and incorporate sophisticated monitoring systems), consumers modulating their demand, a practice known as demand response, will begin to play an increasingly important role in operation of efficient grids. Additionally, we are beginning to see a growing incorporation of renewable generation methods (Hong & Fan, 2016). As renewable energy is highly dependent on favourable weather conditions, its impact on the supply of electricity will become even more prominent. This effect on the supply, and hence on the price, will encourage consumers to exploit renewable generation by engaging in the practice of demand response at larger scales, leading to a direct and growing impact of weather on load (Siano, 2014).

Given the fundamental importance of decarbonizing the electric grid in the pursuit of mitigating GHG emissions, we seek to better understand the important influence weather conditions have on PLF. The additional information obtained via accurate probabilistic forecasts that incorporate weather data can then be leveraged by the entities involved to help increase grid efficiency.

The paper is organized as follows: Section 2 provides an overview of related work researched before and throughout the completion of the study. In section 3, we introduce the design of our study, including a description of the datasets used, required preprocessing and specific algorithms tested. An overview of our results is provided in section 4, examining the performance of the models at different time horizons separately. These results are further analyzed and discussed in section 5. In section 6, the pathways to impact of this work are illustrated, advocating for further work to be pursued. Some general considerations regarding the work done are elaborated in section 7. Finally, given the promising preliminary results obtained, suggestions for future work are touched on in section 8.

2. Related Work

As demand forecasting plays an important role in power-system planning, there has been a substantial amount of research done in forecasting models. The majority of the models used for this purpose are statistical models, which rely on historical demand data to perform time series forecasting. Among these models, the most widely used are autoregressive integrated moving average (ARIMA) models (Vivas et al., 2020). These models are more sophisticated than classical regression models and can be applied to non-stationary data (Ghalehkhondabi et al., 2017). In addition to classical statistical and mathematical methods, in recent years, machine learning models have also been used for electricity load forecasting due to their ability to model complex non-linear relationships and handle large amounts of data as well as their generalization ability (del Real et al., 2020). For example. González-Briones et al. (2019) perform a comparison between standard machine learning models, such as support vector regression (SVR), K-nearest neighbours (KNN) and random forest (RF) for point forecasting of electricity demand. In addition to these models, deep learning models which have shown considerable success in other time series forecasting tasks, such as stock-market price forecasting (Son & Kim, 2020). These include are broadly categorized into two groups: simple feed-forward networks (FFNs) and recurrent neural networks (RNNs). As an example, Muzaffar & Afshari (2019) develop a LSTM network for load forecasting over multiple horizons (daily, weekly and monthly). They also perform a comparison with standard forecasting models such as ARMA and SARIMA and show that especially for short term forecasting, LSTM outperforms the other models.

As mentioned in section 1.3, most of the energy consumption forecasting models are only able to provide point forecasts. However, in recent years, there has been a growing interest in PLF over point forecasts. For example, in 2014, the Global Energy Forecasting Competition (GEFCom2014) was held where the contestants were required to provide one-month ahead hourly probabilistic demand forecasts (Hong et al., 2016). Hong & Fan (2016) provide a tutorial review of the PLF methodologies and evaluation metrics.

As with the point forecasting methods, the literature on PLF include a mixture of traditional statistical and machine learning models as well as simulations. For example, Wijaya et al. (2015) use Generalized Additive Models (GAM), a computationally-inexpensive to model the time-varying mean and variance of the national electricity demand of France. Other methods include using ensemble approaches with quantile regression forests (QRF) and gradient boosting (GB) (Kong et al., 2019). For instance, Nagy et al. (2016) apply gradient boosting decision trees (GBDT) and QRF to obtain probabilistic solar and wind power forecasts (as opposed to demand forecasts) and show that these ensemble models are able to obtain outstanding performance. However, they also mention their drawback of being computationally expensive and time consuming to train; a cost that can be especially problematic for short term load forecasting (STLF), where hour-ahead predictions are required. (Kong et al., 2019) address the computation and time cost of the ensemble models by developing an improved weighted extreme learning machine (IWELM) for STLF.

As Hong & Fan (2016) mention, a downside of current PLF models is that they fail to address the weather-related uncertainty in their forecasts and how it affects their prediction intervals (PI). Weather forecast errors, however, can be significant and as a result can have a serious effect on the accuracy of demand forecasts (Taylor & Buizza, 2002). Furthermore, weather models have been previously used in input-scenario models for PLF. In these models a point forecasting model is fed with various input scenarios (for example weather patterns) and the outputs are used to form a probabilistic predictions (Hong & Fan, 2016). For instance, Taylor & Buizza (2002) show that the incorporation of weather ensemble models can improve upon the uncertainty measurements of standard artificial neural networks (NNs) compared to those measured exclusively with historical load forecast errors. Ignoring the impact of weather forecasts on PIs might prove to be especially problematic in the future: due to CC related changes in the weather, future weather variables might be far from the variables used in training these models which could result in them becoming unreliable (Hong & Fan, 2016).

3. Methodology

In this section, we describe our work on the two case studies we conducted. We choose the electric grids of the Netherlands and the province of Alberta, Canada as application examples. Both grids are wholesale markets, operating in a deregulated fashion on a zonal level (AESO, 2021; TenneT, 2021). Moreover, the dimensions of electricity demand are comparable between both grids. For the investigated period of 2016-2019, the average hourly loads in Alberta and the Netherlands where approximately 9,500 MW and 13,000 MW, respectively.

The predictive tasks to be learned by the ML models are 1h and 24h demand forecasts. 24h forecast refers to predicting a single demand value for $t + 24$, where t is the last time step of the queried input sequence. Initial experiments with 24h sequence and 24h auto-regressive predictions result in a significant performance decrease for point forecasts and probabilistic forecasts. We therefore focus on single value predictions for the 1h and 24h forecast horizons. As mentioned in section 1.1, those time intervals are relevant in practice for the scheduling of operating, intraday or daily reserves. For both grids and forecast lengths, we compare how including weather data affects predictions of probabilistic models. All electricity and weather data used for these experiments is freely available online.

3.1. Datasets

3.1.1. Electricity data

For Alberta, we obtain historic electricity demand data from Alberta Electric System Operator (AESO)[1]. For the Netherlands, we extract the desired data from Open Power System Data[2], a package of several time-series datasets for power system modeling. The downloaded datasets for both grids consist of averaged hourly demand values (in MW) for the years 2016 to 2019.

3.1.2. Weather data

For Alberta, historic weather data was obtained from Alberta Climate Information Service (ACIS)[3]. For the Netherlands, data was obtained via the Royal Netherlands Meteorological Institute (KNMI)[4]. The datasets were pre-processed to to achieve parity, ending up with datasets consisting of temperature in deg. Celsius, relative humidity in %, precipitation in mm, wind speed in km/h and wind direction in degrees, covering 2016 through 2019. Given Alberta faces slightly more variance in weather across population dense areas in comparison to the Netherlands, these variables were aggregated across multiple weather stations, weighted by population density.

We were unable to obtain high quality, historic weather forecast data and had to compromise by using historic weather data that had been shifted in its place.

After combining demand and weather variables, we expand the date time information conveyed with every instance of the dataset into the following time-related features: hour, day of week, day of month, and month.

3.2. Data Preprocessing

We utilize the first 2.5 years (01.01.2016 – 30.06.2018) of the dataset as training set. The following six month (01.07.2018 – 31.12.2018) serve as validation set to identify the best configurations of preprocessing techniques and model-specific hyperparameters. We report the final performance of the three models on every task using the entire last year (2019) as test set. To account for the notion of similarities in time-related features, we experiment with sine-cosine encoding. For example, 5pm and 6pm have the same distance as 23pm and 0am but learning these relations from ordinarily encoded features can impose a challenge for ML models. Analyzing the demand time-series via autocorrelation and discrete Fourier transform shows that the most significant embedded frequencies have wave lengths of 12h, 24h, and 168h. We encode these frequencies, representing half day, day, and week, via a sine and cosine function for each (London, 2016).

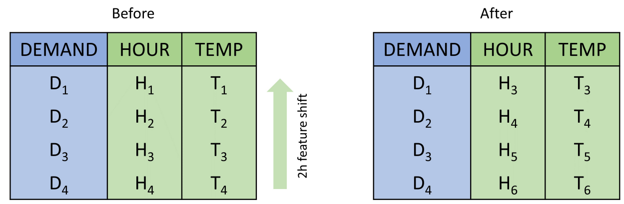

Figure 1. Example for a feature shift by 2 hours.

We further test two normalization techniques for demand and weather variables. These are standardization to zero mean and unit variance, and min-max scaling to a range of (0, 1). To allow our models to incorporate future time and weather information into demand forecasts, we experiment with shifting these features backwards in time, relative to the demand values. This technique is illustrated in figure 1. While it is not problematic to shift time variables, it treats historic weather data as "perfect forecasts". As explained above, we chose this approach due to suitable weather forecasts matching the time horizon of intended demand forecasts being unavailable.

The last step of the preprocessing pipeline is sample creation with a sliding window. We treat the length of input sequences as a hyperparameter and experiment with lengths between 1h and 120h. For each sample, the output consists of a single value, depending on the task reflecting either the true demand of the next hour or in 24 hours.

3.3. Algorithms

We deploy three basic probabilistic machine learning algorithms for our experiments: QRF, QGB, and a Bayesian LSTM, a type of RNN. QRF and QGB are implemented with the ML library scikit-learn and based on the work of Dataman (2020). The Bayesian LSTM was built using PyTorch. Following our literature review, discussed in section 2, we chose these algorithms due to their popularity for probabilistic predictions, performance in many areas, and straightforward implementation.

3.3.1. Quantile Random Forest

QRFs are a probabilistic interpretation of the widely used random forest algorithm. Random forests train multiple decision trees on bootstrapped versions of the dataset and perform a random feature selection to split each node. For regression tasks, the algorithm returns the average over the numerical outputs of all trees (Breiman, 2001).

[1] http://ets.aeso.ca
[2] https://data.open-power-system-data.org
[3] https://acis.alberta.ca
[4] https://www.knmi.nl

Meinshausen & Ridgeway (2006) showed that random forests do not only provide information about the conditional mean of the target variable, but also about its full conditional distribution. Instead of computing the average over every leaf of every tree, the QRF computes an estimate of the distribution function. According to the authors, QRFs provide an accurate way of estimating conditional quantiles for high-dimensional predictor variables.

3.3.2. Quantile Gradient Boosting

Similar to random forest, gradient boosting is an ensemble approach, typically based on decision trees. A gradient boosting regressor is an additive model that learns by fitting trees on pseudo-residuals, i.e. the predictive errors made by previous trees (Friedman, 2001; 2002). By using asymmetric loss functions, this recursive process can be used for quantile regression. Then, instead of predicting mean or median, QGB predicts the α quantile for a selected α (Kriegler & Berk, 2007). Consequently, other than for QRF, k models must be trained to obtain predictions for k quantiles.

3.3.3. Bayesian LSTM

LSTM models are popular for time series forecasting due to their ability to learn temporal dynamics of data, model complex and non-linear relations. Similar to RNNs they can internally maintain a memory of the input but also do not have the vanishing gradient problem of RNNs (Yang et al., 2020). Bayesian LSTM is an LSTM model that applies Bernoulli dropout of rate p at each layer, which is equivalent to placing a prior distribution over the parameters of the LSTM (Charnock et al., 2020; Zhu & Laptev, 2017). The network is then ran multiple times and the outputs are interpreted as the distribution of targets that the model is able to obtain. (Charnock et al., 2020). An advantage of these models over other methods, such as variational inference is that they are very easy to implement as they don't require changes to the architecture of the point forecasting models and hence can be integrated into existing models very easily (Zhu & Laptev, 2017).

3.3.4. Training and Evaluation

We begin model training by experimenting with different preprocessing techniques and model-specific hyperparameters on deterministic versions of the algorithms. Other than probabilistic outputs, the predictions of deterministic models can plausibly be evaluated using single metrics. We choose the root-mean-squared error (RMSE) on the validation set for these experiments. We then switch to probabilistic models and further experiment with hyperparameters based on the best configurations found before. For QRF and QGB, the number of trees in the ensemble was found to be most influential, whereas the number of hidden layers, units per layer, and dropout rate p affected the Bayesian LSTM the most. More specifically, the Bayesian LSTM model is constructed using a 2-layer LSTM network, followed by a fully-connected layer for the final output. We also apply dropout after each hidden layer. During training, we use a dropout rate of 0.5 and for the probabilistic predictions, this rate is determined by plotting the reliability of the PI for various dropout values. More precisely, we plot the coverage as a function of the PI for different dropout values and choose the value which neither underestimates nor overestimates the uncertainty. An example of this plot for the 24h predictions on Alberta's grid is shown in figure 2.

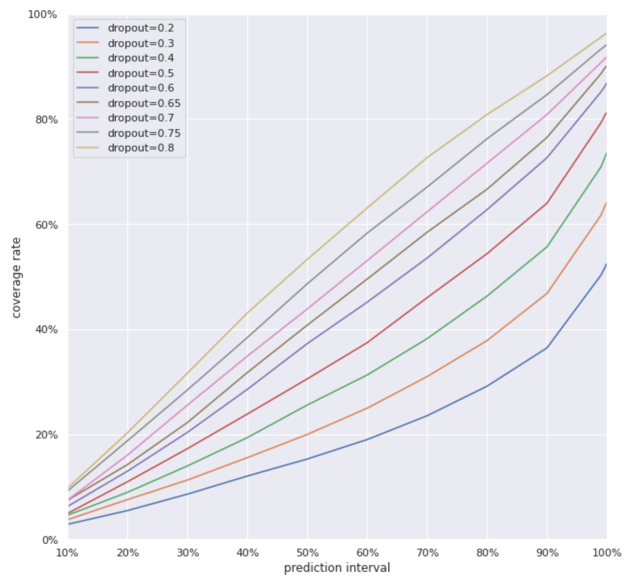

Figure 2. Reliability plots for various dropout values for the LSTM model.

After tuning, we measure the performance for three prediction intervals on the test set. For 95% (quantiles 0.025 – 0.975), 98% (0.01 – 0.99) and 99.8% (0.001 – 0.999) we report coverage, i.e. rate of true values laying within the predicted prediction interval in per cent, and average interval width in MW. Furthermore, we record the mean-absolute-percentage error (MAPE) of the deterministic versions to facilitate a comparison between both case studies.

4. Results

This section introduces the results obtained for the case studies on the grids of Alberta and the Netherlands. We first present the results for 1h forecasts, followed by those for 24h forecasts. For the sake of simplicity, we solely report scores for 98% prediction intervals.

4.1. 1 Hour Forecasts

Table 1 and 2 show 1h forecasting results for Alberta and the Netherlands, respectively. For Alberta, QRF outperforms QGB and LSTM on both probabilistic metrics. Only the LSTM clearly profits from additional weather data with and increased coverage and on average smaller prediction intervals. For QRF and QGB, weather data increases coverage at the cost of growing interval widths. A similar behavior can be observed for the Netherlands. Here, with values below 90%, the coverage of the LSTM is significantly worse compared to the tree-based models. The QGB on the other hand, achieves highest coverage with weather data but nearly doubles the size of its prediction interval compared to QRF. The MAPE obtained for the deterministic versions indicates that predictions on Netherlands' grid are more challenging for all models.

Table 1. 1h forecasting results on the test set for Alberta

Model	Weather	MAPE (%)	98% PI coverage rate (%)	98% PI avg. width (MW)
QRF	×	0.5	98.75	346.5
	√	0.5	99.40	400.1
QGB	×	0.4	97.91	642.6
	√	0.5	98.05	777.0
LSTM	×	0.6	92.73	474.6
	√	0.7	96.08	462.5

Table 2. 1h forecasting results on the test set for Netherlands

Model	Weather	MAPE (%)	98% PI coverage rate (%)	98% PI avg. width (MW)
QRF	×	1.2	95.81	991.5
	√	1.3	96.69	1071.33
QGB	×	0.8	92.93	1235.0
	√	0.9	97.19	2073.7
LSTM	×	1.4	86.20	1142.8
	√	1.5	89.40	1134.4

4.2. 24 Hour Forecasts

Table 3 and 4 present the results of 24h forecasts for Alberta and the Netherlands, respectively. Several trends from 1h forecasts are observable for 24h forecasts as well. Again, the LSTM covers the least true values in its 98% prediction interval, and again, the QGB produces the widest intervals. For both grids, weather data helps QRF and LSTM to increase coverage measurably in combination with a marginal increase of the interval width. The QGB behaves inconsistently when comparing the grids. Weather data decreases both probabilistic measures for Alberta, whereas the coverage profits in the case of Netherlands.

Table 3. 24h forecasting results on the test set for Alberta

Model	Weather	MAPE (%)	98% PI coverage rate (%)	98% PI avg. width (MW)
QRF	×	1.4	95.20	726.6
	√	1.3	98.30	814.2
QGB	×	1.4	97.27	947.2
	√	1.1	96.48	1130.4
LSTM	×	1.7	87.99	820.5
	√	1.3	90.46	849.1

Table 4. 24h forecasting results on the test set for Netherlands

Model	Weather	MAPE (%)	98% PI coverage rate (%)	98% PI avg. width (MW)
QRF	×	2.9	94.70	2110.0
	√	2.6	96.12	2207.1
QGB	×	3.1	94.75	2476.7
	√	2.6	97.50	4050.4
LSTM	×	3.2	86.78	2214.4
	√	3.2	90.25	2417.5

5. Discussion

Based on the results shown in tables 1, 2, 3, 4 we can draw some general conclusions regarding the impact of weather data on both point and probabilistic load forecasts. For point forecasting of demand, we note that the incorporation of weather forecast data is more beneficial for 24h forecasting compared to 1h forecasting. This is evident from the MAPE of the 24h predictions, where including the weather data results in a lower MAPE in all models (excluding LSTM on NL). We speculate that the reason the 1h predictions do not benefit from the incorporation of weather data is that the next-hour weather is more predictable than the next-day weather, implicitly including this information in the current demand. The data for both the NL and AB grids also suggest that the incorporation of weather forecasts resulted in an increase in the coverage rate of the PI for both forecast horizons. However, we also note that for most of the models, this increasing PI coverage comes with the cost of an increasing width of the confidence interval. This can be seen in figure 5, where we show an example of the PIs for the 24h forecast both with and without the weather forecast data. Further investigations are required to see whether improving the point predictions of these models might lead to a better coverage of the actual load values, without compromising the width. Additionally, even though these results are using historical data as forecasts, we believe that they strongly suggest that weather data impacts the uncertainty of predictions in PLF. It is important to investigate this effect when actual weather forecasts are used in the models.

Apart from studying the effect of weather data as a whole on predictions, a feature importance analysis can provide insight into the relevance of each variable has on the predictions. We use Gini importance for the random forest models,

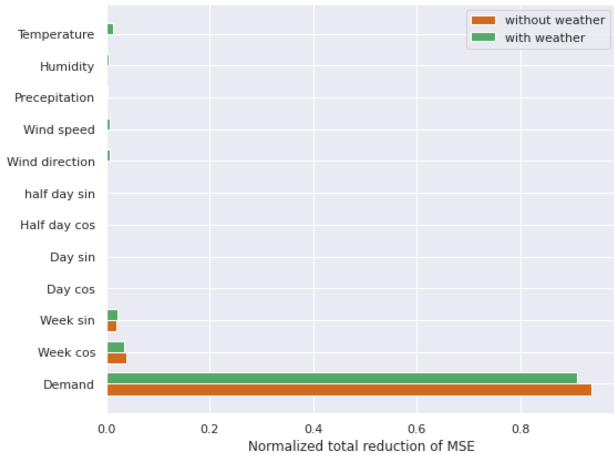

Figure 3. Feature importance with random forest: 24h forecasts for Alberta

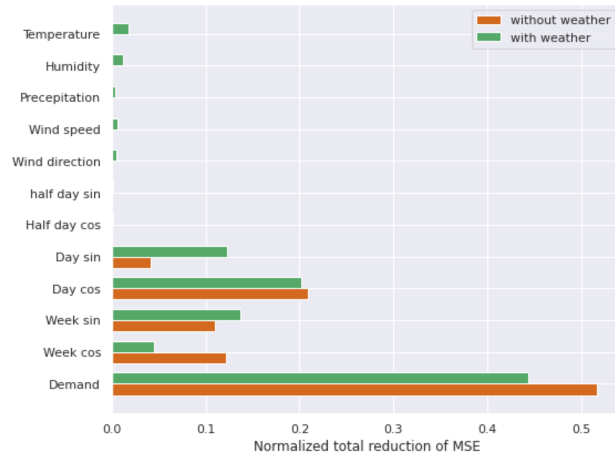

Figure 4. Feature importance with random forest: 24h forecasts for Netherlands

measuring the mean impurity decrease of the split criterion (MSE in our case), to compare predictions with and without weather data. Figure 3 and 4 display the corresponding results for 24h predictions for Alberta and Netherlands, respectively. Interestingly, both grids show distinct results. For Alberta, the demand itself is by far the most important feature, followed the sine and cosine waves encoding weekly information. Adding weather data only marginally changes this behavior, with a slight importance increase for air temperature. For predictions on the Dutch grid, figure 4 indicates a higher relevance for daily and weekly features. When added, all five weather variables show higher importance compared to Alberta.

The reasoning behind for this observation could stem from many factors, ranging from data quality to the limitations of the deployed feature importance technique. One hypothesis is that there exists differences in the behavior of consumers between the two grids. In Alberta, energy-intense industry is a major consumer, accounting for 75% of the province's total energy demand. In the Netherlands, the industrial sector consumes only 25% of the total energy demand (Government of Canada, 2021; IEA, 2020). While these numbers do not directly reflect the electricity demand, it is likely that Alberta's industry requires a higher share of electricity as well. Heavy consumers in industry potentially show less dependence on daytime, weekday, or weather than residential ones that are more important in the Netherlands.

6. Pathways to Impact

Electric grids are complex networks involving many different parties in order to ensure they continue functioning reliably. Each party has its own set of priorities and possesses unique means of reducing the carbon intensity of the grid. Our work can be leveraged by many of these parties, leading to multiple avenues of increased grid efficiency.

6.1. Grid Authorities

Grid authorities are organizations tasked with the management and operation of electric grids. While the exact responsibilities may vary across different grids and countries, they are in control of the grid and ensuring its reliability. An essential responsibility of grid authorities is the coordination of operating reserves. As already discussed in the introduction, these reserves can be extremely pollutant and inefficient.

By incorporating probabilistic forecasts, operating reserves can be maintained up to the peak of a chosen α-level prediction interval (PI). In doing so, we can mitigate the use of superfluous operating reserves and directly reduce unnecessary GHG emissions.

6.2. Power Producers

There are many different methods power producers can use for generating electricity. Carbon intensive methods make up the majority of generation due to the reliability and accessibility of fossil fuels (EESI, 2018). Renewable generation is intermittent, heavily dependant on weather and often geographically limited in distribution. If any of these barriers can be addressed, we can increase the incorporation of renewables into the grid, limiting carbon intensive power generation. As renewable generation increases and becomes more profitable, more players can become involved in generation and R&D (research and development), leading to a positive feedback loop.

By better understanding how weather data is affecting probabilistic forecasts of demand, we can plan in advance by

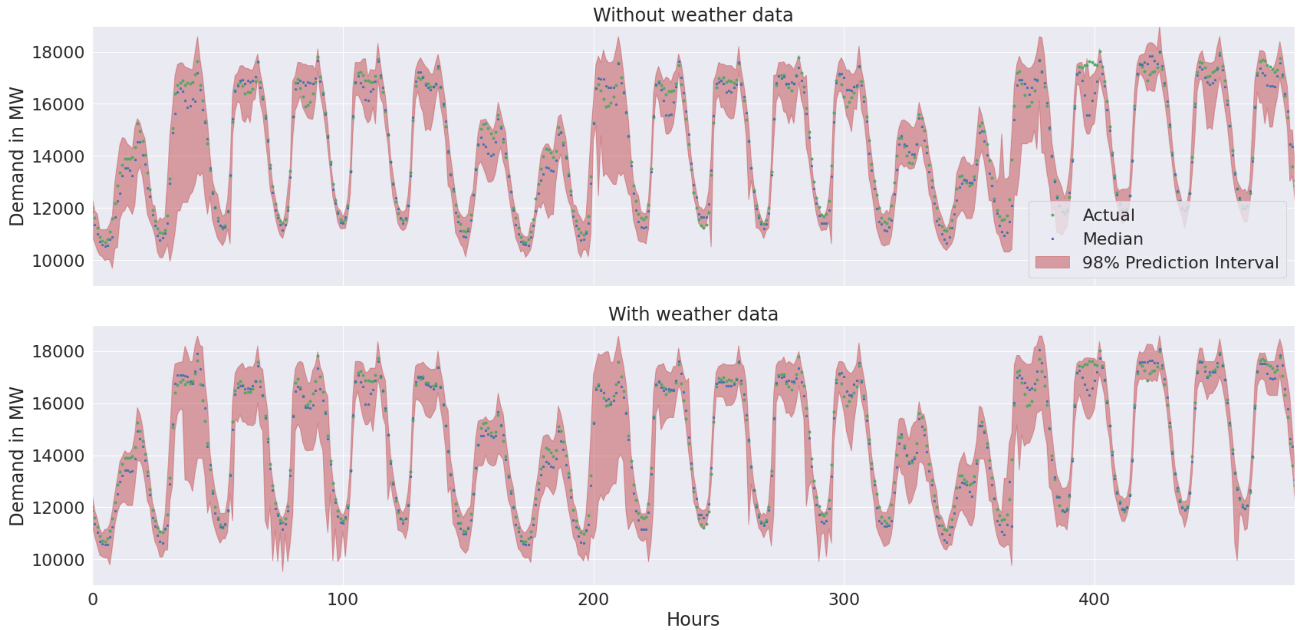

Figure 5. Prediction example for the Netherlands using random forest on the first 500h of the test set

observing whether or not forecast renewable generation can cover up to the peak of our PI. If so, we reduce the need for operating reserves. Since renewables are heavily dependent on weather, this work could be extended to further understand how weather data affects probabilistic forecasts of supply or price, which would give insight into how renewables are currently incorporated into the grid. With increased understanding of how renewables incorporate into the grid and affect the market, we can continue to scale renewable generation methods with confidence.

6.3. Storage Operators

Energy storage as a service (ESaaS) refers to organizations deploying energy storage and management systems at a cost. These storage solutions are vital to grids running reliably and optimally. They can incentivize demand response by allowing companies to buy and store electricity at periods of low demand. Differences in supply and demand could be accounted for using stored energy, reducing the need for operating reserves. Given the intermittency of many forms of renewable energy generation, this technology could be used to increase the incorporation renewables into the grid.

By extending this work to price forecasting, these ESaaS organizations can increase their profits through better managed buying and selling. If being a storage provider becomes more profitable, more organizations will be encouraged to engage in this, increasing storage capacity and potentially R&D for efficient energy storage. This can lead to increasing the presence of renewables in the grid (ESA, 2019), reducing the use of operating reserves as well as improving demand response at scale.

6.4. Consumers

As consumers, to alleviate stress on the grid, we can engage in demand response. This means we decrease or modulate our demand, to reduce dramatic peaks in demand. This reduce the need for operating reserves, making it a cost-effective strategy for reducing unnecessary GHG emissions.

While point estimates are able to give indications of when peaks are to be expected, their ability to convey forecasts about peak load is limited. Given probabilistic forecasts that incorporate weather data, we are not only able to capture these same trends, but the PIs can indicate potential peaks in demand where point estimates fail to do so. As such, using these models, can help consumers better engage in demand response to reduce emissions and help avoid system emergencies.

6.5. Market Participants

Many grids, mainly those in the US, that operate with a wholesale electricity market employ a tactic referred to as ”virtual bidding”. This allows financial entities to make bids to buy or sell in the day-ahead market, with a commitment to close their position with a sell or buy respectively in the real-time market. Given these virtual bids compete with physical bids, they contribute to price determination, help reduce in inefficiencies by converging the day-ahead and real-time markets (Kim, 2018).

If this work is extended to price forecasting, these financial entities can leverage these probabilistic forecasts to better inform their bidding strategies. With improved forecasts, virtual bidding can become more profitable, encouraging more entities to become engaged in this market. This leads to increasing the price convergence of the day-ahead and real-time markets, further reducing inefficiencies in the grid.

7. General Considerations

7.1. Perfect Forecasts

Though efforts were made to obtain forecast data through weather modelling organizations such as European Centre for Medium-Range Weather Forecasts (ECMWF) and National Oceanic and Atmospheric Administration (NOAA), as well as via private vendors, we were unable to obtain high quality historic weather forecasts under the time constraints associated with the project. As a result, we compromised by shifting historic weather data, assuming perfect short term forecasts.

The consequence of this cannot be measured precisely, however this implies our results can be assumed to be marginally optimistic. Nevertheless, it is to be noted that the intent of this project is not to improve upon state of the art point forecasts, nor to develop optimal probabilistic models for this domain. The objective of the research conducted is to understand how weather can impact PLFs, as a step towards probabilistic methods becoming standard in this domain. In that vein, using these ”perfect forecasts” are still able to give valuable insight into how a better understanding of weather patterns can lead to improved probabilistic forecasts.

7.2. Evaluation Metrics

A natural question that arises when dealing with probabilistic forecasts is how to best evaluate performance. We can compare the percentage of actual values that fall within the range of our given PI. We can also measure the width of such PIs. Both these metrics give some insight into the quality of a probabilistic estimate, however, neither do so definitively. For instance, a 98% PI that contains 99% of the actual values sounds great in theory, however if the interval width is massive, we are given very little information of value. Similarly, a narrow interval is ideal, unless it contains only a small percentage of the true values.

There also exists other metrics, such as reliability, sharpness and resolution, which all capture different information about the uncertainty of forecasts. They, however, also require to be studied in parallel to provide more complete analyses (Hong & Fan, 2016).

As a result, even though we can gauge which models generally perform better than others for a given α-level PI using different metrics, we are unable to form any kind of objective or definitive ranking of the models developed and tested.

7.3. Optimal α-level

Once high-performing probabilistic forecasts are fully developed, a new set of questions surface. Maybe the most prominent of which is how confident should models be in this domain? While standard $(1 - \alpha)$% PIs (90%, 95% and 99%) may be adequate in most work, in the context of electricity grids, this equates to the miss-management of generation and supply anywhere from three to thirty-seven days of the year. As previously mentioned, these forecast errors can result in power outages and wasted resources.

An important consideration is that as we increase the confidence of a PI, the width grows and can explode as we begin to include the tails of the distribution. Bearing this in mind, information may be lost at a certain level of confidence rather than gained and must be considered when considering what level of confidence is optimal in this domain.

8. Future Work

8.1. Sources of Uncertainty

In the work of Zhu & Laptev (2017), researchers were able to break down the uncertainty associated with Bayesian Neural Networks (BNNs) into its components, as well as develop estimators to capture the uncertainty originating from each source. Future work on this topic could try to follow their suit in identifying all theoretical sources of uncertainty in a given model and developing accurate estimators to capture them. This would lead to more accurate and less optimistic forecasts, helping to eliminate forecasting errors.

8.2. State of the Art Methods

Many of the methods of generating probabilistic forecasts involve using many point estimates that have been predicted within the pipeline to build posteriors. Developing a framework that can be added to current state of the art point forecasting techniques and models could not only help improve these probabilistic forecasts, but would make them more accessible as they could easily be incorporated into existing point forecast models. This could help probabilistic methods become standard in the domain of electricity grids.

8.3. Variations of Weather Data

Weather data can come in many forms including historical weather data, historical weather forecasts and sophisticated weather models such as ensemble models and climate reanalysis models. In particular, climate reanalysis has been shown to find errors across temporal and spatial scales (Slivinski, 2018), while ensemble models can incorporate uncertainty estimates into their forecasts, providing a clearer picture regarding future weather scenarios (Buizza & Richardson, 2017; Taylor & Buizza, 2002). Applying more sophisticated weather models could better capture the uncertainty and non-linear relationships that may exist, helping to improve probabilistic forecasts.

8.4. Price Forecasting

A natural next step is to extend this work to more complex challenge of price forecasting, as it further increases the potential impact of this work. Many of the pathways to impact examined consider the application of increased understanding of weather's impact on probabilistic price forecasts. Given the impact weather has on the generation of renewables, affecting not only demand but supply as well, we believe the impact of weather to be equally, if not more important in this context.

9. Conclusion

In this paper, we investigated the impact of weather forecasts on probabilistic demand forecasting. We proposed three methods for performing PLF and obtained some preliminary results on two different grids over two time horizons. We believe that these results suggest further investigation into the effect of weather data is an important step for PLF in this domain and can have a substantial contributions to climate change mitigation strategies.

References

AESO. Guide to understanding Alberta's electricity market. 2021. URL https://www.aeso.ca/aeso/training/.

Allen-Dumas, K. and Cunliff. Extreme weather and climate vulnerabilities of the electric grid: A summary of environmental sensitivity quantification methods, 2019. URL https://www.energy.gov/sites/prod/files/2019/09/f67/Oak%20Ridge%20National%20Laboratory%20EIS%20Response.pdf.

Auffhammer, M., Baylis, P., and Hausman, C. H. Climate change is projected to have severe impacts on the frequency and intensity of peak electricity demand across the united states. *Proceedings of the National Academy of Sciences*, 114(8):1886–1891, 2017.

Breiman, L. Random forests. *Machine learning*, 45(1): 5–32, 2001.

Buizza, R. and Richardson, D. 25 years of ensemble forecasting at ecmwf. *ECMWF*, 2017. URL https://www.ecmwf.int/en/newsletter/153/meteorology/25-years-ensemble-forecasting-ecmwf.

Charnock, T., Perreault-Levasseur, L., and Lanusse, F. Bayesian neural networks, 2020.

Dataman. From quantile regression to quantile random forests. 2020. URL https://github.com/dataman-git/codes_for_articles/blob/master/From%20Quantile%20Regression%20to%20Quantile%20Random%20Forests.ipynb.

del Real, A. J., Dorado, F., and Durán, J. Energy demand forecasting using deep learning: Applications for the french grid. *Energies*, 13(9), 2020.

Denholm, P., Clark, K., and O'Connell, M. On the path to sunshot-emerging issues and challenges in integrating high levels of solar into the electrical generation and transmission system. Technical report, EERE Publication and Product Library, 2016.

EESI. Fossil fuels, 2018. URL https://www.eesi.org/topics/fossil-fuels/description.

ESA. 25 years of ensemble forecasting at ECMWF. *ESA*, 2019. URL https://www.ecmwf.int/en/newsletter/153/meteorology/25-years-ensemble-forecasting-ecmwf.

Friedman, J. H. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pp. 1189–1232, 2001.

Friedman, J. H. Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4):367–378, 2002.

Ghalehkhondabi, I., Ardjmand, E., and Weckman, G. e. a. An overview of energy demand forecasting methods published in 2005–2015. *Energy Systems*, 8:411–447, 2017.

Glavitsch, H. and Stoffel, J. Automatic generation control. *International Journal of Electrical Power & Energy Systems*, 2(1):21–28, 1980.

González-Briones, A., Hernández, G., Corchado, J. M., Omatu, S., and Mohamad, M. S. Machine learning models for electricity consumption forecasting: A review. In *2019 2nd International Conference on Computer Applications Information Security (ICCAIS)*, pp. 1–6, 2019.

Government of Canada, C. E. R. Canada energy regulator, Mar 2021. URL https://www.cer-rec.gc.ca/en/data-analysis/energy-markets/provincial-territorial-energy-profiles.

Hong, T. and Fan, S. Probabilistic electric load forecasting: A tutorial review. *International Journal of Forecasting*, 32(3):914–938, 2016.

Hong, T., Pinson, P., Fan, S., Zareipour, H., Troccoli, A., and Hyndman, R. J. Probabilistic energy forecasting: Global energy forecasting competition 2014 and beyond. *International Journal of Forecasting*, 32(3):896–913, 2016.

IEA. CO2 status report 2019. *International Energy Agency*, 2019.

IEA. The Netherlands - total final consumption by sector, 2020. URL https://www.iea.org/countries/the-netherlands.

Islam, M., Che, H. S., Hasanuzzaman, M., and Rahim, N. Chapter 5 - energy demand forecasting. In Hasanuzzaman, M. and Rahim, N. A. (eds.), *Energy for Sustainable Development*, pp. 105–123. Academic Press, 2020.

Kahan, A. EIA projects nearly 50% increase in world energy usage by 2050, led by growth in Asia. *International Energy Outlook, 2019*, 2019. URL https://www.eia.gov/todayinenergy/detail.php?id=41433.

Kim, H. *Welfare Impacts of Optimal Virtual Bidding in a Multi-settlement Electricity Market with Transmission Line Congestion*. PhD thesis, Purdue University, January 2018.

Kong, Z., Xia, Z., Cui, Y., and Lv, H. Probabilistic forecasting of short-term electric load demand: An integration scheme based on correlation analysis and improved weighted extreme learning machine. *Applied Sciences*, 9 (20), 2019.

Kriegler, B. and Berk, R. Boosting the quantile distribution: A cost-sensitive statistical learning procedure. *Department of Statistics, UCLA, working paper*, 2007.

London, L. Encoding cyclical continuous features - 24-hour time. 2016. URL https://ianlondon.github.io/blog/encoding-cyclical-features-24hour-time/.

Meinshausen, N. and Ridgeway, G. Quantile regression forests. *Journal of Machine Learning Research*, 7(6), 2006.

Muzaffar, S. and Afshari, A. Short-term load forecasts using lstm networks. *Energy Procedia*, 158:2922–2927, 2019. Innovative Solutions for Energy Transitions.

Nagy, G. I., Barta, G., Kazi, S., Borbély, G., and Simon, G. Gefcom2014: Probabilistic solar and wind power forecasting using a generalized additive tree ensemble approach. *International Journal of Forecasting*, 32(3): 1087–1093, 2016.

Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., Church, J. A., Clarke, L., Dahe, Q., Dasgupta, P., et al. *Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change*. IPCC, 2014.

Pierpont, B., Nelson, D., Goggins, A., and Posner, D. The path to low-carbon, low-cost electricity grids. 2017.

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A., Maharaj, T., Sherwin, E. D., Mukkavilli, S. K., Kording, K. P., Gomes, C., Ng, A. Y., Hassabis, D., Platt, J. C., Creutzig, F., Chayes, J., and Bengio, Y. Tackling climate change with machine learning, 2019.

Siano, P. Demand response and smart grids—a survey. *Renewable and Sustainable Energy Reviews*, 30:461–478, 2014.

Slivinski, L. C. Historical reanalysis: What, how, and why? *Journal of Advances in Modeling Earth Systems*, 10(8): 1736–1739, 2018.

Son, H. and Kim, C. A deep learning approach to forecasting monthly demand for residential–sector electricity. *Sustainability*, 12(8), 2020.

Taylor, J. W. and Buizza, R. Neural network load forecasting with weather ensemble predictions. *IEEE Transactions on Power Systems*, 17(3):626–632, 2002.

TenneT. The Dutch market. 2021. URL https://www.tennet.eu/electricity-market/dutch-market/.

Vivas, E., Allende-Cid, H., and Salas, R. A systematic review of statistical and machine learning methods for electrical power forecasting with reported mape score. *Entropy*, 22(12), 2020.

Weron, R. Electricity price forecasting: A review of the state-of-the-art with a look into the future. *International Journal of Forecasting*, 30(4):1030–1081, 2014.

Wijaya, T. K., Sinn, M., and Chen, B. Forecasting uncertainty in electricity demand. *AAAI-15 Workshop on Computational Sustainability*, Jan 2015. URL https://github.com/tritritri/uncertainty.

Yang, Y., Li, W., Gulliver, T. A., and Li, S. Bayesian deep learning-based probabilistic load forecasting in smart grids. *IEEE Transactions on Industrial Informatics*, 16 (7):4703–4713, 2020.

Zhu, L. and Laptev, N. Deep and confident prediction for time series at uber. *2017 IEEE International Conference on Data Mining Workshops (ICDMW)*, Nov 2017.