PREDICTING HOURLY ELECTRICITY DEMAND IN ONTARIO

Mathilde Bourget and Gabriel Morin

UQAM

OBJECTIVES

The objective of our project was to create a model that would predict the hourly electricity demand in Ontario for the residential sector.

METHOD

A random forest model was created with the h2o[3] package in R using 500 trees and 12 randomly sampled variables at each split (mtry). We found that a mtry of 12 was optimal by using a tuning algorithm. We started by keeping every variables available to us and then adding more when we saw a potential correlation with the electricity demand. Those added variables are discussed later. Then, we got rid of variables one at a time, always checking with the MAE that our model was gaining accuracy.

Changes to the hourly aggregated electricity demand

The database used for the model is a merge of the **Hourly Demand** database and the **Weather** database. Since the data for the hourly electricity demand of the residential sector was not available in the data, changes have been made to the aggregated data. From the **Annual Demand** database, a ratio of the annual residential demand on the total annual demand was calculated for each year. Then, within the residential annual demand, a ratio of the yearly demand on the total demand was calculated. These two ratios were then used to obtain an estimate of the hourly residential demand.

RESULTS

The goal was to identify patterns in the predictors that would help increase the model's accuracy in order to predict the hourly residential electricity demand with precision. The variable that has the most impact on the hourly residential electricity demand was mean of the electricity demand for the weekday and the hour of the day. This is to capture cyclical consumption patterns related to workdays and weekends.

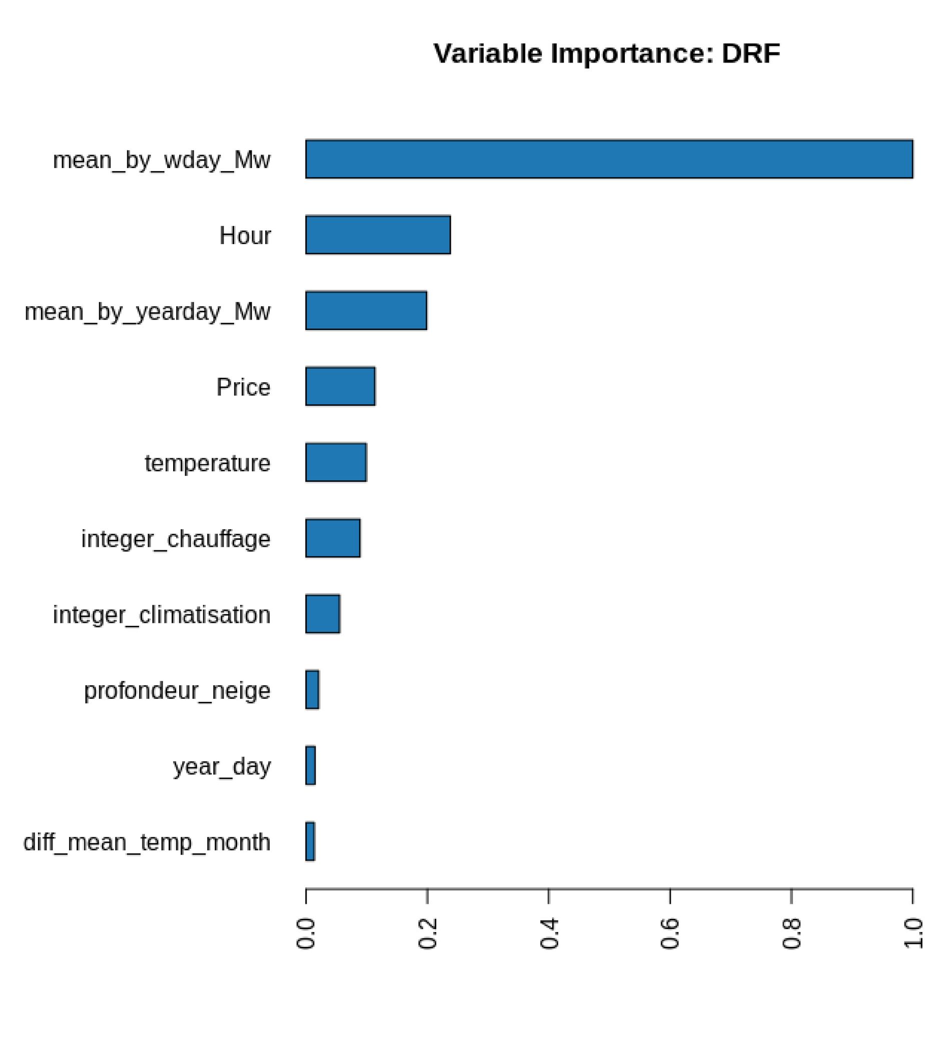

Fig. 1: Variable Importance graphic

CHOICE OF VARIABLES

Variables removed

All the variables from the original databases were kept, expect Date, irradiance surface, Year and air density. These variables were either irrelevant for the model or redundant with other variables.

We also had to remove the Year variable. This was our best predictor, but because we only had 13 years in the training data, we decided to remove it since the model would not be able to make any useful predictions in the future. To capture any additional trend on annual electricity consumption, we decided to add electricity prices in the model (see below).

The blackout of 2003 (August 14 and August 15, 2003) was also removed from the database because it was clearly an outlier and had nothing to do with the normal electricity demand.

Variables kept

We kept the following variables from the original databases because of their relevance towards the electricity demand variability.

Month	Hour	snowfall	temperature
cloud cover	snow depth	precipitation	

Variables added

After some data exploration, we saw that electricity consumption was highly correlated to the temperature (heating and cooling patterns), to the hour of the day (e.g. heating, cooling, appliances while at home rather than at work) and to day of the year (patterns of day and night lighting, seasons, etc.). Thus, we tried to identify patterns in the data that could explain these relationships. Based on the residential consumption variables provided by the annual demand database, we looked at our electric consumption habits and tried to identify patterns in our data that could apply to the databases. High electricity demand in summer and winter were explained by the usage of heat and air conditioning. Heating and cooling demand depends on how long temperature is intensely low or high, which is why we created the integer_climatisation and integer_chauffage variables that are explained below. We used 18°C as a threshold at which heating or cooling need is null. Since we had to take out the Year variable, but still wanted to showcase some influence from that variable, we added some price variables that explain well why is the electricity demand changing between each years. The following variables were calculated and added to the database:

- Day : The day of the month.
- mean_by_wday_Mw : The mean electricity demand for each weekday and hour.
- mean_by_yearday_Mw : Then mean electricity demand for each day of the year.
- interger_chauffage : The integer of the difference between 18 °C and the observed temperature, when the following was under 18 °C for more than 3 consecutive hours.
- integer_climatisation : The integer of the difference between the observed temperature and 18 °C , when the following was over 18 for more than 3 consecutive hours (similar to the heating degree hours).
- diff_mean_temp_month : The difference between the temperature observed and the mean temperature for the month for which it is observed.
- diff_mean_temp_month : The difference between the temperature observed and the mean temperature for the month for which it is observed.
- holiday : A dummy variable indicating 1 when the data observed is on a holiday (TSX calendar).
- Rate : Rate category for the smart meters (either A for off-peak, B of mid-peak and C for on-peak)[2]
- Price : The price of electricity according to the rate category.[1]

MODEL ACCURACY

To validate the goodness of fit for the model, the mean absolute error (MAE) was compute for each year with a leave one out cross validation (LOO-CV) technique. The following figure shows those values for each year that was left out of the model and used as validation, in blue is the MAE.

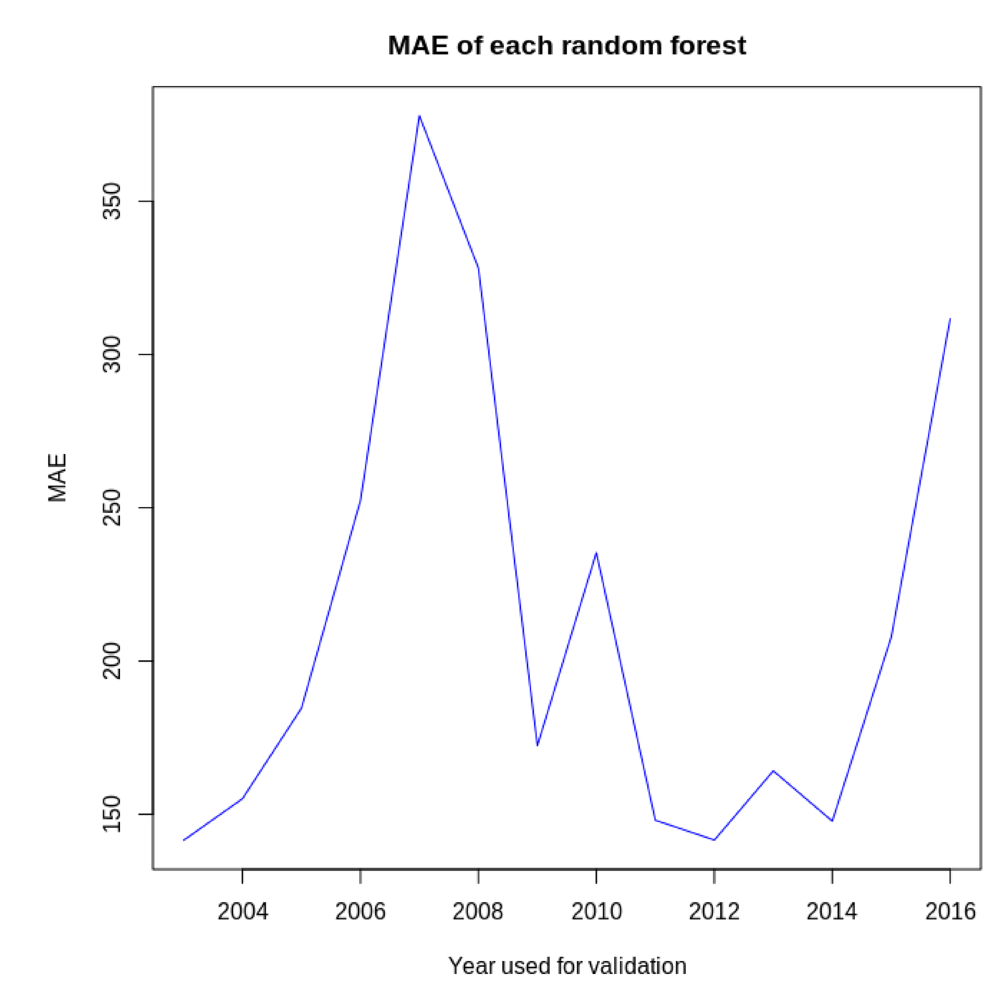

Fig. 2: Model performance using the MAE criteria

ACKNOWLEDGEMENT

We want to thank our supervisor Mathieu Boudreault, Ph.D., F.S.A., F.I.C.A., for his precious help and patience. We also want to thank the Statistical Society of Canada for organizing this case study.

References

[1] Ontario Hydro. en. 2020. URL: `http://www.ontario-hydro.com` (visited on 05/22/2020).

[2] Ontario Hydro. *Ontario hydro rates*. en. 2020. URL: `http://www.ontario-hydro.com` (visited on 05/22/2020).

[3] Erin LeDell et al. *Package 'h2o'*. en. 2018. URL: `https://cran.r-project.org/web/packages/h2o/h2o.pdf` (visited on 05/22/2020).