A regression model estimating the impact of natural, socio-economic regional characteristics on hourly electricity consumption in Japan

Yuki HIRUTA, National Institute for Environmental Studies, 029-850-2227, hiruta.yuki@nies.go.jp
Lu GAO, National Institute for Environmental Studies, 029-850-2227, gao.lu@nies.go.jp
Shuichi ASHINA, National Institute for Environmental Studies, 029-850-2227, ashina.shuichi@nies.go.jp

Overview

We propose a series of methods with which to construct a single regression model that explains hourly electricity consumption based on various regional characteristics such as hourly climatic conditions and the socio-economic and geographical conditions of the region.

Given that oil thermal power and pumped-storage hydroelectric power are mainly used during periods of peak demand for electricity in Japan, fossil fuels tend to be consumed to larger degrees when there are more fluctuations in electricity demand. Global warming, and a declining and aging population in Japan, mean that the measures undertaken to date (such as peak shifts) are insufficient. Comprehensive measures such as land-use planning should be applied concurrently. We need to develop a method to evaluate the effect of multiple concurrent factors, such as regional climate, and the socio-economic and geographical characteristics of the region, on hourly electricity demand. In this research, we propose a series of methods with which to construct a single regression model that explains hourly electricity consumption based on various regional characteristics such as hourly climatic conditions and the socio-economic and geographical conditions of the region. We seek to overcome three technical problems for this purpose. The first is how to consider the complicated combination of region-specific factors and time-varying factors. Given that Japan has a variety of regions, from subarctic to subtropical, as well as four distinct seasons, electricity consumption is affected by multiple time-varying factors and region-specific factors. The second question is how to select an effective variable (predictors) from a wide range of possibilities. The third is how to model complex relationships such as the nonlinear relationship between power consumption and factors, and at the same time to construct generalizable models using as few parameters as possible. We propose a unique algorithm using RandomForests (Breiman 2001) to narrow down the small number of effective predictors from more than 6,000 possibilities, including predictors for region-specific, time-varying and interaction factors between the region-specific factor and time-varying factor. We apply MARS (Friedman 1991) for the modeling, which rendered it possible to construct a flexible yet highly generalizable model, while selecting important predictors. We constructed a regression model that explains hourly power consumption nationwide. The constructed model exhibits high quality in terms of both fitness and generalization, but it could still be improved. We experimentally applied the simulation to understand the effect of the spatial distribution of households on hourly power demand using the model we constructed. The proposed method may offer a means to assess the comprehensive measures, including land-use policies, to lower the fluctuation of hourly electricity demand, even there is a room for improvement.

Methods

We use the power supply-demand data provided by major electric power companies for the analysis. There are 10 major electric power companies in Japan (Figure 1), hereinafter referred to as EPC. This research targets the 9 EPC's jurisdictional area excluding Okinawa. We used the hourly power supply-demand data which are recorded every hour for a period from April 1, 2016, to March 31, 2018, in each EPC's jurisdictional area as the explained variable. If we simply count the possible determinants of power demand including interaction variables, the candidate predictors would reach over 14,855. From such numerous predictors, we need to narrow down those that explain power consumption more precisely. We developed the original algorithm for selecting the predictors that have a greater impact on power

Figure 1 Target Area

demand. We constructed the MARS using the selected predictors, then evaluated model performance in terms of both fitting and generalization, and the relationship between temperature and estimated power consumption. Using the constructed model, we experimentally applied the simulation to understand the effect of the spatial distribution of households on hourly power demand.

Results

One region-specific predictor and 9 interaction predictors were adopted for the final model from 20 predictors, and 36 (including the intercept) terms were used for the model. Figure 2 compares actual observations to the estimated value by the constructed MARS. The figure on the left hands shows the result using the training data to build the model (in sample result, fitting), and the figure on the right shows the result using the test data which is not used for the model construction (out of sample result, prediction). The coefficient of determination by linear regression (OLS) is 0.99026 in the in-sample result and 0.98909 in the out-of-sample result. The determination coefficient of the in-sample result is higher, but the determination coefficient of the out-of-sample result is also sufficiently high, so it can be said that the constructed model is a high-quality model in terms of both fitting and generalization. We also examined the relationship between temperature and estimated power consumption in order to evaluate the model. In some regions, there is a noticeable difference between observation and estimation. This suggests that the predictors adopted in the constructed model are not sufficient to express the actual power consumption determinants. There is still room for improving the model by finding more appropriate predictors.

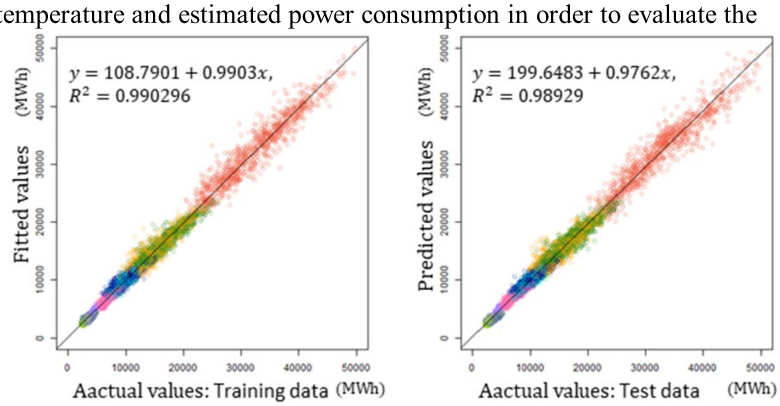

Figure 2 Plot of actual values against predicted values for electricity demand

Conclusions

We proposed a series of methods to construct a single regression model that explains hourly electricity consumption based on both region-specific factors and time-varying factors. Although the performance of the whole model was appropriate from the viewpoint of fit and generalization, there is still room to improve it. As we were able to test a general simulation for estimating the effect of the spatial distribution of household on electricity demand, however, further improvement of the constructed model will offer a useful tool, not only for understanding the relationship between climate and power consumption but for discussing various measures such as land-use planning for power management. In order to build a model that is applicable to practical use, it is necessary to develop a more efficient variable selection method that can deal with multi-dimensional interactions in the early stage of the variable selection process. It is also necessary to discover new variables that can represent the characteristics of the Okinawa region, so as to include the region in the model. It may necessary to consider adopting more complex (flexible) regression models that retain a generalization ability than MARS. Although MARS is excellent in flexibility, generalization, and interpretability, the number of predictors that can be adopted for the model is hard to tune, and it is difficult to incorporate small effects of factors of electricity demand into the model.

There have been many models built for predicting power, or for identifying the determinants of power demand, however, there are few models to evaluate the effect of changes in regional factors such as land use on hourly power demand, taking both region-specific factors and time-varying factors into consideration. Global warming and a declining and aging population in Japan mean that comprehensive measures should be studied. Although there is a room for improvement, this research proposes a means of modeling that may contribute to planning policy.

References

- Breiman, Leo. 2001. "Random Forests." Machine Learning 45 (1). Kluwer Academic Publishers: 5–32. doi:10.1023/A:1010933404324.
- Energy, Agency for Natural Resources and. 2017. "Japan's Energy White Paper 2018." http://www.enecho.meti.go.jp/about/whitepaper/2018pdf/.
- Friedman, Jerome H. 1991. "Multivariate Adaptive Regression Splines." The Annals of Statistics 19 (1): 1–67. doi:10.1214/aos/1176347963.