IMPACTS OF TRAFFIC DATA ON DAY-AHEAD RESIDENTIAL LOAD FORECASTING

Aksornchan Chaianong, Center for Energy and the Environment, School of Management and Law, Zurich University of Applied Sciences, +41 (0) 76 634 29 47, aksornchan.chaianong@zhaw.ch
Christian Winzer, Center for Energy and the Environment, School of Management and Law, Zurich University of Applied Sciences, +41 (0) 58 934 49 45, christian.winzer@zhaw.ch
Mario Gellrich, Institute of Business Information Technology, School of Management and Law, Zurich University of Applied Sciences, +41 (0) 58 934 78 90, mario.gellrich@zhaw.ch

Overview

Accurate load forecasting is essential for the power sector's planning and management. This applies during normal situations as well as phase-changes such as the Coronavirus (COVID-19) pandemic due to variations in electricity consumption that has posed challenges to system operators to accurately forecast load. So far, few studies have used traffic data to improve load prediction accuracy [1,2]. Although the integration of traffic or mobility data can improve load prediction accuracy, different combinations of traffic data and other features are rarely tested for load prediction. Especially, the previous works did not assess to what extent this prediction accuracy held if time variables were included in the forecasting models. Therefore, this analysis aims to (1) investigate whether traffic data can improve short-term residential load forecasting accuracy during phase changes such as COVID-19 , and (2) test different combinations of feature set (historical load, weather, time, and traffic). This will help system operators to accurately forecast electricity demand when phase changes occur, or when historical load is not available, such as in case of predicting the electricity demand of competitors in a liberalized market.

Methods

To forecast target variable (aggregate, residential electricity load), we used the following four groups of features (independent variables): (1) historical residential electricity load, (2) weather-related, (3) time-related, and (4) traffic (road traffic and train traffic) time-series. All time-series were aggregated into an hourly resolution, from 2016-2020.

During the **first** step, some of the individual features, mainly weather and individual road traffic data that contain many missing values, were dropped from the feature list. In the **second** step, weather and time data were pre-screened to remove variables that are strongly correlated with each other, and then were standardized and used as inputs to forecast day-ahead residential electricity demand. There were four different feature sets, which were tested ("Base", Time", "Traffic" and "All"[1]) in this analysis. Moreover, there was an assumption that historical electricity demand might already contain most of the information included in traffic data. To test this hypothesis, the cases with and without using historical load as an independent variable were also tested in the modelling.

In the **third** step, for each feature set, the model was trained using pre-COVID-19 data for each feature set. To determine the impact of each individual feature, the features were added iteratively during the calibration process (stepwise addition) in order of increasing importance. During the first iteration, the model was calibrated with one feature. The importance of each feature was measured by root mean square error (RMSE). At the end of the first iteration, the feature, which achieved the lowest RMSE during the four phases (pre-COVID-19, lockdown, post-lockdown, and strict regulation[2]), was retained for inclusion in the subsequent iterations. This process was repeated until all features from the feature set were added to the model.

Finally, during the **fourth** step, the model performance (RMSE) of each iteration of the stepwise addition was reported by calibrating the data using pre-COVID-19 and calculating RMSE based on training dataset (pre-COVID-19) as well as test datasets from three different phases (lockdown, post-lockdown, and strict regulation). Focusing on the forecasting technique, in this analysis, random forest was used to calibrate the load prediction model.

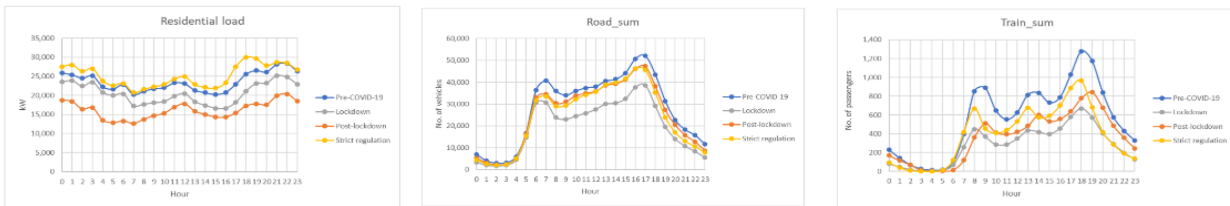
---

[1] "Base" includes historical load and weather; "Time" includes historical load, weather and time; "Traffic" includes historical load, weather and traffic; "All" includes historical load, weather, time and traffic.
[2] 4 different phases were based on the pandemic situation in Switzerland: (1) Pre-COVID-19 (2016–March 2020), (2) Lockdown (March–April 2020), (3) Post-lockdown (May–October 2020), and (4) Strict regulation (November–December 2020).

Results

Descriptive comparison of residential load and traffic patterns

Fig. 1 shows the average values of the residential electricity consumption and traffic variables (road and train) for each hour of the day during each of the four phases. For both electricity demand and traffic data, we can see that each phase has different magnitudes. However, daily patterns of each variable during each phase show almost no change except the patterns of the total number of train passengers during the post-lockdown and strict regulation phases. Focusing on the residential load for both DSOs, the different magnitudes of the four phases seem to be partly linked to the respective season. Regarding traffic data, the total number of vehicles and passengers during COVID-19 decreased due to the introduction of stay-at-home policies.



(i) Residential load (ii) Total number of vehicles (iii) Total number of passengers

Fig. 1 Comparison of average residential load and traffic data during four phases.

Electricity load forecasts

The results from this study (Table 1) show that traffic data – as well as weather and historical load data – improved prediction accuracy both before and during COVID-19. However, time variables have a much more significant impact on prediction accuracy than traffic data. Adding traffic data to time, weather, and historical load data can only improve forecasting accuracy to a small degree. However, traffic data still improves load prediction when historical load information is not available.

Table 1 The lowest RMSE (for day-ahead) of each feature set during each phase.

Lowest RMSE	**With historical load**				**Without historical load**			
	Base	*Time*	*Traffic*	*All*	*Base*	*Time*	*Traffic*	*All*
Pre-COVID-19	0.14	0.07	0.10	0.07	0.23	0.04	0.15	0.04
Lockdown	0.38	0.32	0.37	0.32	0.59	0.52	0.52	**0.46**
Post-lockdown	0.37	0.27	0.34	0.27	0.50	0.28	0.40	**0.26**
Strict regulation	0.40	0.33	0.70	0.33	0.57	0.53	0.57	0.53

Conclusions

The findings suggest that the impact of traffic data on load forecasts is much smaller than the impact of time variables. However, traffic data could improve load forecasting where information on historical load is not available. The inclusion of traffic data as a feature could be justified for two main reasons: First, improving prediction accuracy in situations where historical load data is unavailable in real-time (such as for neighboring grid area predictions), and second, deriving further insights regarding the phenomenon of interest (the behavior of individuals in relation to electricity demand).

References

[1] Y. Chen, W. Yang, and B. Zhang, “Using Mobility for Electrical Load Forecasting During the COVID-19 Pandemic,” ArXiv200608826 Phys., Jun. 2020, Accessed: Dec. 08, 2021. [Online]. Available: http://arxiv.org/abs/2006.08826.

[2] Z. Wang and H. Wang, “Improving load forecast in energy markets during COVID-19,” in Proceedings of the 8th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, New York, NY, USA, Nov. 2021, pp. 168–171. doi: 10.1145/3486611.3486648.