Zurich University
of Applied Sciences

School of
Management and Law

Impact of traffic data on day-ahead residential load forecasting

Building Competence. Crossing Borders.

Aksornchan Chaianong and **Christian Winzer**, Center for Energy and the Environment

Mario Gellrich, Institute of Business Information Technology

aksornchan.chaianong@zhaw.ch , christian.winzer@zhaw.ch, mario.gellrich@zhaw.ch

27.01.2023

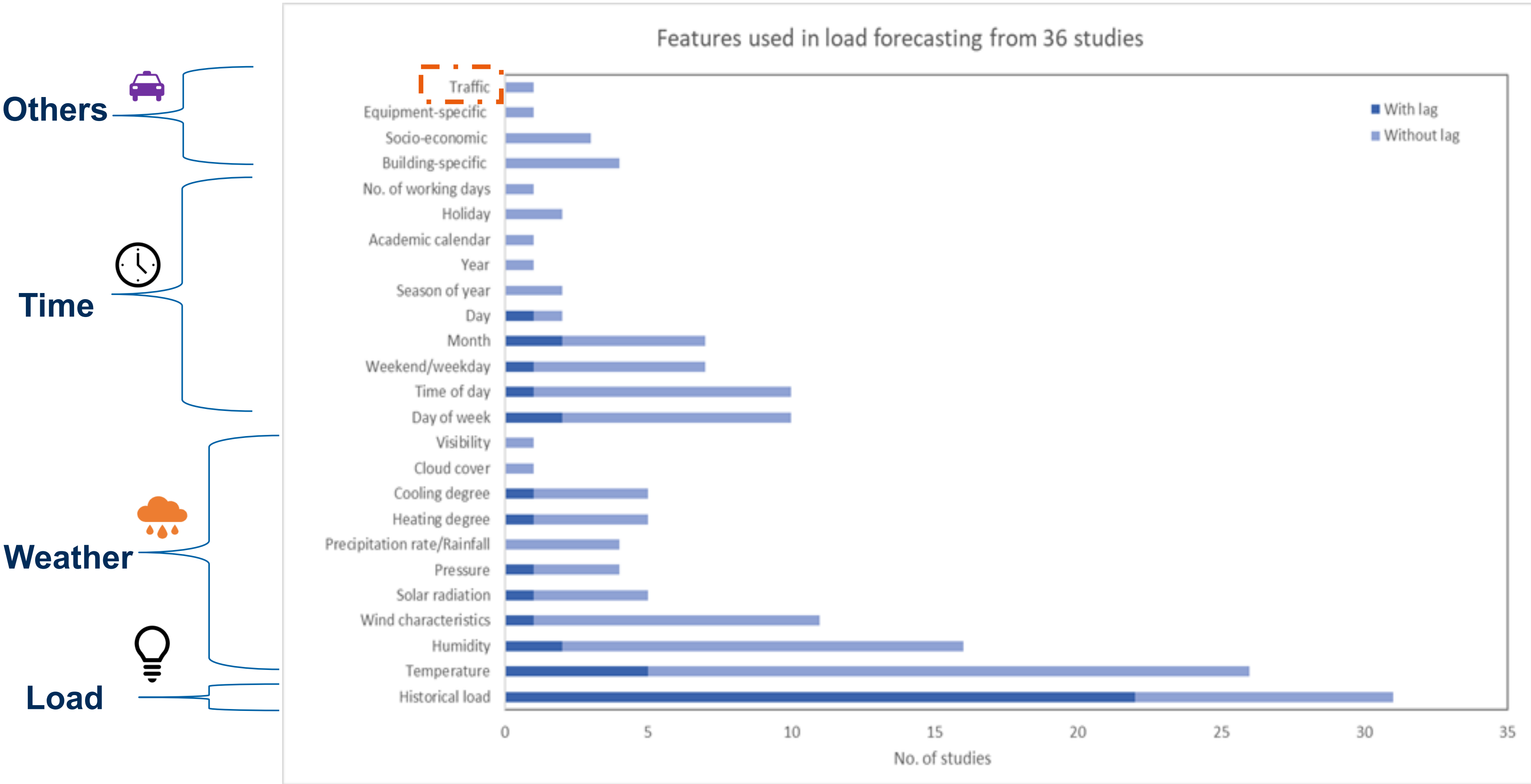
Acknowledgement

- This work was a part of the Machine Learning for Demand Response (ML4DR) project, which was supported by Swiss National Science Foundation (SNSF), Spark Grant Number: CRSK-1_196304.

- We would also like to thank two DSOs, MeteoSchweiz, FEDRO and SBB for data support.

Background, Data and Research steps

Few literatures used transportation data for load forecasting



- 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 the change of electricity consumption that has posed challenges to system.
- So far, few studies have used traffic data to improve load prediction accuracy.

zh aw School of Management and Law

Objectives

Load forecasting *before COVID-19*

- Previous works usually combined several features together (historical load, weather, time, and others) in different combinations to predict load.
- The concept of “co-mobility" was introduced.

→

Load forecasting *during COVID-19*

- Traffic data, together with other features (***but no time variables***) were used to improve the prediction accuracy load forecasting.

→

Research gaps

- **Time variables were not included in previous works**, unlike the literatures of load forecasting before COVID-19.

→

Objectives

(1) Investigate whether **traffic data can improve load forecasting** accuracy during COVID-19.

(2) **Test different combinations** of traffic data and other features, including time variables.

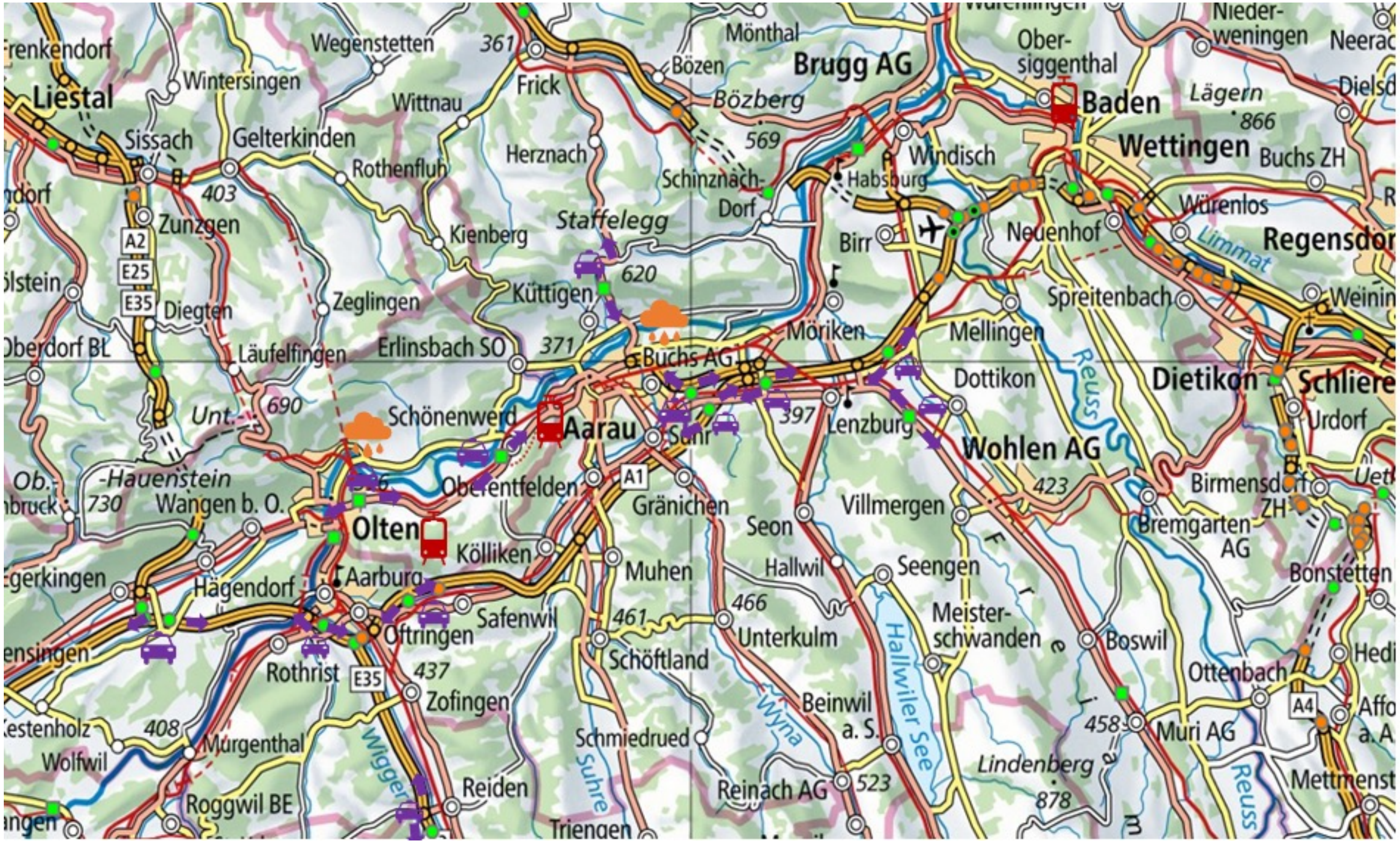
Features/Independent variables

Features

- **Road traffic**: Number of buses, trucks, passenger cars, etc.
- **Train traffic**: Number of visitors at train stations
- **Weather**: Temperature, radiation, rainfall, etc.)
- **Time-related**: Hour-of-day, day-of-week, month-of-year, etc.
- **Historical load**

Traffic data

Example: Map of traffic/train/weather station



- Weather stations
- Road counters
- Train stations

Research steps

1) Data collection (2016-2020)

- Four different groups of feature: historcial load , weather-related, time-related, traffic-related (total road traffic and total train traffic).

2) Data pre-processing

- Drop some individual features due to missing data.
- Pre-screening using Spearman correlation.
- Standardization.
- Adjust data.
  - Eight dataset combinations.

3) Stepwise addition of features to forecasting model

- *Stepwise addition* of each feature based on the lowest RMSE.
  - Train using pre-COVID-19* and test using training dataset (pre-COVID-19) as well as test datasets from three different phases* of COVID-19.
  - *Random Forest* as a forecasting model.

4) Performance evaluation of forecasting model

- Calibrate forecast using pre-covid training data.
- Report performance metric (RMSE) based on training dataset (pre-COVID-19) and three test datasets (lockdown, post-lockdown, and strict regulation).

*Pre-COVID (Before Mar 2020); Lockdown (Mar-Apr 2020); Post-lockdown (May-Oct 2020); Strict-regulation (Nov-Dec 2020)

Eight dataset combinations

	Feature combinations	Historical load	Weather	Time	Traffic
WITH historical load	Base	✖	✖		
	Time	✖	✖	✖	
	Traffic	✖	✖		✖
	All	✖	✖	✖	✖
WITHOUT historical load[1]	Base		✖		
	Time		✖	✖	
	Traffic		✖		✖
	All		✖	✖	✖

***To test the assumption that historical electricity demand might already contain most of the information included in traffic data**

Results and Key findings

Day-ahead: Traffic data could better improve prediction where load is not available

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

- **Traffic** data could slightly improve load prediction accuracy when comparing to **weather** data (“Base” vs. “Traffic” dataset).
- But when adding **time and traffic** variables together (“All” dataset), RMSE of all phases **have almost no change** from using only **time** variables alone (“Time” dataset).
- In the case ***with*** historical load, traffic could not show positive effects on prediction accuracy, unlike the case ***without*** historical load.
- Although RMSE is higher if historical load is excluded, this case could still be beneficial, for example in order to predict load of **neighboring grid areas**, where historical load data is essential for load prediction, but it might be not available close to real-time.
- Another advantage of using traffic data is that it is closer to explaining the **phenomenon of interest** (behavior of individuals in relation to electricity demand) than historical load.

Key findings

- **Traffic**, **on top of weather and historical load**, data could **improve** prediction accuracy both before and during COVID-19. However, **time variables** have a **much larger impact** than traffic data.
- Traffic data could **better improve** load prediction where information on **historical load is not available**.
- Inclusion of traffic data as a feature could thus be warranted for two main reasons. ***First***, to improve prediction accuracy in situations, where **historical load data is not available in real-time**, such as in case of predictions for **neighboring grid areas**. And ***second***, to derive further insights regarding the **phenomenon of interest** (behavior of individuals in relation to electricity demand).
- The methodology framework of this analysis could also **contribute to broader impacts**.
  - For instance, when predicting the electricity demand of competitors in **a liberalized market and historical load is not available**, this framework could help system operators to accurately forecast electricity demand **using traffic, weather, and time features.**
  - It could also be extended to analyze load forecasting in **some specific areas such as those under limited communication (island/mountain regions).** It is, however, important to note that our proposed method would **require near real-time traffic information**, which may not be available in some of these areas.

Energy Strategy Reviews 43 (2022) 100895

Contents lists available at ScienceDirect

Energy Strategy Reviews

journal homepage: www.elsevier.com/locate/esr

Impacts of traffic data on short-term residential load forecasting before and during the COVID-19 pandemic

Aksornchan Chaianong [a,*], Christian Winzer [a], Mario Gellrich [b]

[a] Center for Energy and the Environment, School of Management and Law, Zurich University of Applied Sciences, Winterthur, Switzerland
[b] Institute of Business Information Technology, School of Management and Law, Zurich University of Applied Sciences, Winterthur, Switzerland

ARTICLE INFO

Keywords:
Load forecasting
Traffic
COVID-19
Random forest
Machine learning

ABSTRACT

Accurate load forecasting is essential for power-sector 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 made it difficult for system operators to forecast load accurately. So far, few studies have used traffic data to improve load prediction accuracy. This paper aims to investigate the influence of traffic data in combination with other commonly used features (historical load, weather, and time) – to better predict short-term residential electricity consumption. Based on data from two selected distribution grid areas in Switzerland and random forest as a forecasting technique, 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. Another benefit of using traffic data is that it might explain the phenomenon of interest better than historical electricity demand. Some of our findings vary greatly between the two datasets, indicating the importance of studies based on larger numbers of datasets, features, and forecasting approaches.

Thank you.