

AUTOMATED DETECTION OF ELECTRIC VEHICLES IN HOURLY SMART METER DATA

Volker HOFFMANN
SINTEF AS – Norway
volker.hoffmann@sintef.no

Bjørn Ingeberg FESCHE
SINTEF AS – Norway
bjornife@student.matnat.uio.no

Karoline INGEBRIGTSEN
SINTEF Energy Research AS – Norway
karoline.ingebriksen@sintef.no

Ingrid Nytnun CHRISTIE
Eidsiva Nett AS – Norway
ingridnytnun.christie@eidsiva.no

Morten PUNNERUD ENGELSTAD
Systek AS – Norway
morten.punnerud@eidsivaenergi.no

ABSTRACT

Automated detection of EVs from smart meter data can provide important insights for DSOs about spatiotemporal EV charging patterns. However, smart meters typically provide only hourly measurements of consumption while most load disaggregation techniques require at least minute level data. We use machine and deep learning methods to detect EV signatures in hourly smart meter data. Models are trained and evaluated on labelled data, before being tested on unlabelled field data. While balanced models catch about 75% of EVs at false positive rates of 35%, tuned models detect up to 90% of EVs with 10% false positives. When using models to detect EVs on unlabelled Norwegian smart meter data, detections are in line with EV fractions from the national registry as well as expected spatiotemporal patterns. However, models may be confused by baseline consumption patterns. Collection and inclusion of labelled EVs is therefore the next step.

INTRODUCTION

Norway is among the countries with the highest per-capita ownership of electric vehicles (EVs) with further increases projected in the coming decade. By 2019, Norway will also complete the national roll-out of smart meters to all electricity customers. While the meters sample three-phase voltage as well as active and reactive power every 10 seconds, technical and regulatory constraints require aggregation of measurements into hourly blocks.

By establishing a method for distribution system operators (DSOs) to use smart meter data to detect charging patterns of existing EVs, the affected geographical areas, and their demand, DSOs can cheaply and readily obtain insights to support future planning and operation of their grids.

Depending on local energy consumption habits, detection of EVs may be as simple as observing extended periods (a few hours) of kW level demand [16]. If, however, the baseline household energy consumption is already high (such as in Norway or the US), extraction of EV signatures from hourly demand measurements becomes an exercise in Nonintrusive Load Monitoring (NILM) [1,2,3]. While there exists a rich literature of NILM methods for high frequency (kHz) sampling [4,14], more recent efforts have

begun to focus on disaggregation at low (seconds) sampling rates [5,6,7,10]. Effort at very low rates (minutes to hours) is sparser [8,9].

The problem of locating charging EVs may also be recast as a detection problem. In this guise, templating methods operating over a range of sampling frequencies have been explored [11,12,13], although these require libraries of charge profiles. This can be sidestepped through machine learning and deep learning approaches where signatures are inferred automatically [10,15].

In this contribution, we use a publicly available labelled dataset of electricity consumption, downsample it to very low sampling rates (hourly), evaluate the capability of a variety of machine learning models to detect EV charges, and attempt to detect EVs in unlabelled data. In order, we describe data sources, summarize the models and evaluation metrics, and assess performance on the labelled dataset. Afterwards, we explore how models perform when detecting EVs in Norwegian smart meter data and address difficulties of such a model transplantation.

DATA SOURCES

Our goal is to develop a model that can detect charging EVs in the Norwegian grid. Unfortunately, no labelled data of electricity consumption exists in Norway. Therefore, we use two different datasets – (a) the Pecan Street Dataport set of labelled consumption data [17], and (b) electricity consumption measurements from smart meters provided by the Norwegian DSO Eidsiva.

Pecan Street Dataport (US, Labelled)

This dataset consists of second-level measurements of household active power including submetering points for electric vehicles. To simulate smart meter conditions, we downsample the data to hourly resolution. Geographically, the dataset covers houses in the continental United States (primarily Texas, Colorado, and California). We use measurements from 81 households with EV submeters and collected data in the period from January 1st to December 31st in 2017. Ten of these households had no EV charging, while the rest had between 20 and 712 such events. To evaluate model performance, we use data from the first five months of 2018. In this interval, 41 of the houses from the training set are also present.

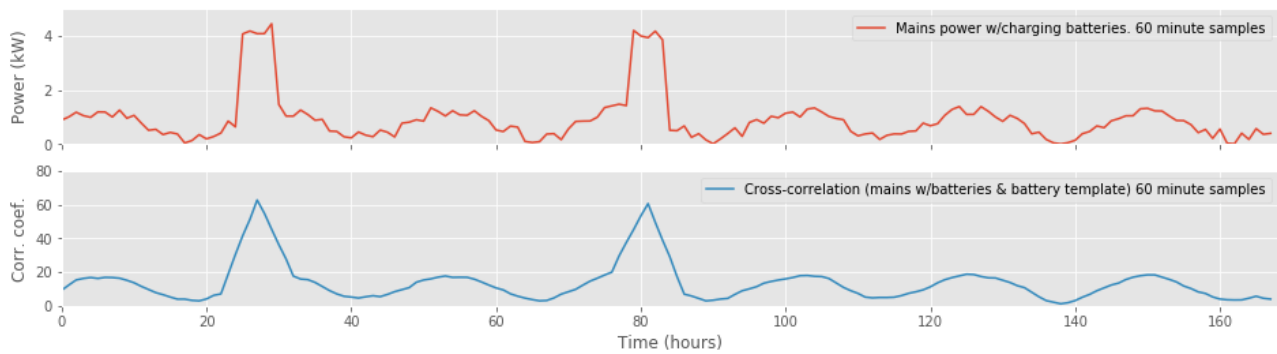


Figure 1: Illustration of EV detection by matched filtering. The upper panel shows a demand time series with two EV charges at the 30- and 80-hour marks. By matching a library of signatures (“filters”), we calculate the cross-correlation (bottom panel) indicating how similar the consumption signal is to the charge signature. The peaks coincide well with the charging event

Eidsiva Smart Meter Data (Norway, Unlabelled)

This dataset consists of hourly total power usage data from 116'679 customers for between two years and a few months. We have used data from one week in April 2017 and one week in April 2018. The location of every customers' nearest electrical substation is known. There are 5412 substations in total and those substations with five or fewer customers were excluded for privacy.

For the weeks in 2017 and 2018, data from 16'019 and 103'244 residential households is available. The data is unlabelled (we do not know whether EVs are charging), but the number of EVs registered in the area per December 31st, 2017 is known.

MODELS & METRICS

Models

Three different models have been implemented, one using a matched filter and two using neural networks – a convolutional neural network (CNN) and a recurrent neural network (RNN). We also implemented a stack that combines the above by way of logistic regression.

We found that both relative and absolute magnitude is important when discovering charging events, and therefore did not normalize the data. Both neural networks were trained on different window lengths, and the length giving the best results was chosen. Models were trained and tested on 90% of the data, while 10% was held out for evaluation.

For the matched filter model, a library of signatures was inferred from the measurements of EVs in the training dataset. The signatures were first partitioned according to their length, giving 6 groups of 2 to 7-hour long charging signatures (one-hour signals were excluded). We then used k-means to yield clusters of similar signatures for each signal length. The number of clusters giving the best clustering results was 6 (based on the Silhouette score), thus taking the average of each cluster gives 6 typical charging signatures for each signal length, giving 42 typical charging signatures in total.

With signatures available, a matched filter now attempts to locate them in the consumption data. Specifically, the procedure computes the cross-correlation between the charge signatures and the consumption signal. If the cross-correlation exceeds some thresholds, an EV detection is marked. This is illustrated in Figure 1.

The CNN also slides filters across the signal, but instead of predefining the filters, they are learned during training (where filter weights are adjusted). The input data to the CNN is the total consumption signal divided into overlapping time slices of 24 hours. To increase the generality of the dataset, a synthetic dataset was created to complement the original data. In periods with no charging events, an EV charging signal was added at a random point in these time periods with a probability of 50%.

RNNs include previous outputs in the next prediction. The model used in this work consists of long short-term memory (LSTM) layers. The input data to the RNN consisted of time series which only contained real data (and no synthetic charging events as for the CNN), in 672-hour sequences (28 days).

Performance Metrics

After training, models accept a smart meter timeseries as input and return a probability of an EV charging at each sample. To assess classifier performance, we hold out a test set, ask the classifiers to make predictions on this set, and use the known ground truth to determine how often the classifier predicted (in-)correctly.

The data has 26 times fewer samples without charging EVs than samples with charging EVs. In other words, it is imbalanced. In imbalanced sets, evaluation of predictive performance using metrics like accuracy or ROC curves are difficult to interpret (the large number of true negatives tends to exaggerate performance).

Measures like precision and recall do not use true negatives (“no EV is charging”) but focus on performance of the model with respect to its capacity for detecting true

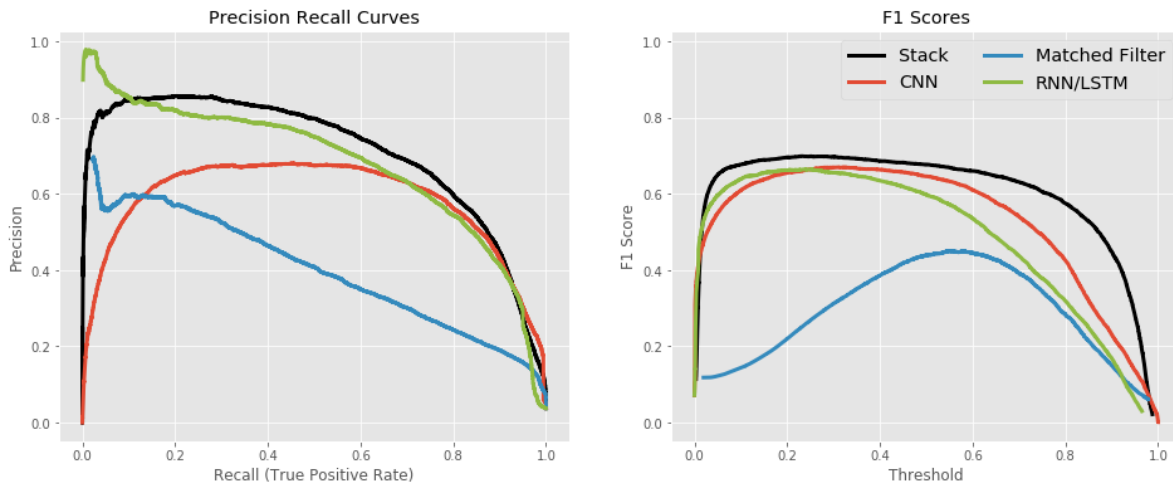


Figure 2: Precision-recall curves (*left*) and F1 scores over all thresholds (*right*) for the three models as well as the stack (see legend).

positives ("an EV is charging"). Intuitively, recall is how likely a positive prediction is to correspond to a true label, whereas precision indicates how likely a true label is to be predicted positively. Precision and recall are frequently combined through the F1 score (their harmonic mean).

As classifiers yield probabilities, we must choose a threshold above which the probability is interpreted as a charging EV. Intuitively, a higher threshold should only keep predictions in which the classifier is very confident in (high precision), but this comes at the price of potentially missing more EVs (lower recall). By varying the threshold, precision and recall can be traded-off. We can generate precision-recall curves (as well as calculate the average precision) by evaluating these metrics over a range of thresholds. Note that even if the classifier is very confident in a prediction, it may not be correct. If a classifier routinely misclassifies predictions at the highest confidence, precision drops to zero at high thresholds.

We first evaluate the performance of the different modes on labelled data. Afterwards, we use the stacked model to detect EVs in unlabelled data in Norway, but only ask whether a given this household owns (regularly charges) an EV owner. A customer is labelled if the model predicts at least seven hours of EV charging during the week in 2018. We then compare the number of predicted EV owners with the actual number of owners in the area. We also examine the geographical location of the predicted charging events (at the level of the secondary substation).

PERFORMANCE, LABELLED DATA

Figure 2 shows the precision-recall curves as well as F1 scores for four different predictive models. Together with the average precision and maximum F1 scores (see Table 1), we find the following:

1. Except at very small recall (high thresholds), the filter model performs worst. This indicates that the manual

feature extraction step is inferior to allowing neural networks to select features.

2. The CNN and LSTM have similar maximum F1 scores, but at different thresholds. However, the CNN rapidly loses precision at high thresholds (where recall is low). This means that the CNN tends to be wrong with its most confident predictions. While the LSTM and filter models do not suffer from this problem, it does bleed into the stack model.
3. Overall, the stack performs best, except at high thresholds, where the overconfidence of the CNN bleeds in. Here, the LSTM outperforms all models.
4. In general, if misdetections at high thresholds can be tolerated, the model stack is the best model. If correctness at high thresholds is essential issue, the LSTM model should be preferred.

To understand the implications of these findings in terms of how many charging EVs such models will detect, we must examine recall and precision separately. The maximum F1 score of the stacked model (0.70) occurs at a threshold of 0.22. Here, the precision is 0.65 and the recall 0.75. In other words, using this classifier on unlabelled data, we expect it to discover about three in four time slots of a charging signal, and that about one in three slots predicted as corresponding a charge are false positives.

If we optimize for precision or recall separately, we can improve these numbers. If confidence in the discovered signals trumps the cost of missing some, we can reduce the

Table 1: Summary of model performance parameters.

Model	Maximum F1 Score	Average Precision
Filter	0.45	0.40
CNN	0.67	0.58
LSTM	0.67	0.68
Stack	0.70	0.71

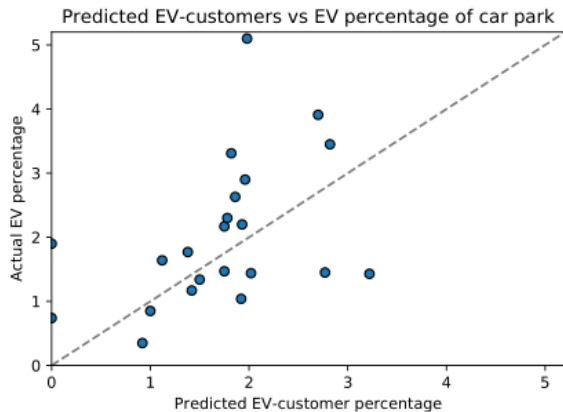


Figure 3: Predicted vs. actual percentage of registered EVs for different municipalities in Hedmark county.

false-positive probability to almost one in ten. Conversely, if we want to capture as many signals as possible, models can be tuned to discover nine out of ten signals.

Models were assessed on measurements collected from the same houses as those used to train the models (although at a later time). This induces a bias which may be reduced by withholding houses for the test set. However, as all houses see the same EV charge signatures (almost all cars in the data were of two makes), there was little to generalize on (besides the baseline consumption). For the CNN models, we attempted to push models towards better generalization by introducing synthetic charges. Unfortunately, we suspect this may be causing the overconfidence (precision collapsing at high thresholds).

PERFORMANCE, UNLABELLED DATA

We now execute our models on unlabelled data. To develop some understanding of detection performance, we (a) compare to publicly available data on EVs, and (b) explore charging patterns on a map.

Figure 3 shows the percentage of predicted EV owners versus the actual percentage of EV owners for various municipalities. Usually, the predicted percentage of EV owners is close to the actual percentage, but some municipalities have large errors. The municipality with the largest discrepancy is Hamar, which the principal city in the county. There are two municipalities in which our models do not detect any EVs. Both are rural areas.

Overall, we identified four key reasons why the numbers of detections are out of step with the aggregate numbers from the vehicle registry, viz.

1. Hamar is home to many companies, which may own EV fleets that may not necessarily charge in Hamar. Similarly, leasing companies may have registered EVs in Hamar that are used in other municipalities.

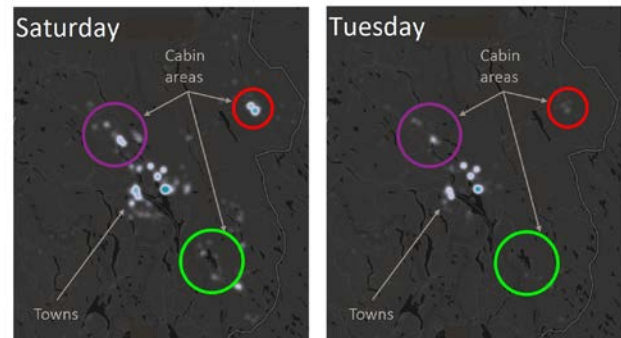


Figure 4: Heatmap of EV detections in Hedmark county for two days. We indicate urban and rural areas.

2. Households have individual meters, even in large apartment complexes. However, some complexes have common charging solutions which typically have a specific metering point. In this case, EVs are detected on the common charging point. Similarly, for businesses with charging stations for their own EVs, only a subset of EVs may be detected.
3. Households with several cars per metering point would not be processed correctly.
4. Hamar (and other urban areas), have been left out for the last part of the AMS rollout due to the anticipated technical challenges. Metering data from such urban areas may also have quality issues due to early operations challenges. In other words, the dataset may be incomplete.

Figure 4 shows a map of detected EV charges for the week in 2018 (at secondary substation level). Overall, the charging patterns coincide with expectations – more Saturday evening charges in the cabins, more Tuesday night charges in the more urban areas. Note, however, that it is entirely possible that the model merely picks up on typical consumption patterns of the general public – cabin areas consuming more electricity on the weekend.

Finally, we note three key challenges when transplanting models across geographies. Firstly, there are differences in electricity usage habits (Norway uses electric heating the winter, the US uses AC units in the summer). Secondly, there are difference car park composition (models sold in the US are not used on Norway and vice versa). Thirdly, the overall percentage of EVs in the datasets differ. All these issues must be addressed carefully.

Without definite ground-truth (labelled data) in Norway, however, none of the above issues can be addressed and quantified adequately. Therefore, collection and analysis of such a dataset should be a priority.

CONCLUSION

To support future network and operations planning for Norwegian DSOs, we have addressed two questions in this

contribution. Firstly, can hourly smart meter data be used to detect charging EVs? Secondly, can a model trained on data from the continental United States be transplanted to give accurate predictions in Norway?

We find that detection of EVs from smart meter data is feasible. Balanced models can detect about 75% of charging events, but with a false positive rate of about 35%. Models tuned to maximize either precision (be certain about detections) or recall (find as many EVs as possible) can achieve false positive rates of 10% and successfully locate 90% of charging events.

Application of models to unlabelled data, on the other hand, is challenging. While the amount of EV detections is roughly in line with the national registries and matches up with expected spatiotemporal patterns, models may be confusing EVs with normal household activities. To properly assess whether models can be transplanted, ground-truth must be collected to establish a labelled dataset of Norwegian consumption data.

ACKNOWLEDGEMENTS

The authors would like to thank the "ENERGYTICS" project consortium for supporting the project, and the Norwegian Research Council for support through the ENERGIX program. We would also like to thank Eidsiva Nett (DSO) for supporting data acquisition and analysis.

REFERENCES

- [1] G. W. Hart, 1992, "Nonintrusive Appliance Load Monitoring", *Proceedings of the IEEE*, vol. 80, 1870—1891
- [2] M. Zeifman, K. Roth, 2011, "Nonintrusive appliance load monitoring: Review and outlook", *IEEE Trans. Consum. Electron.* vol. 57, 76—84.
- [3] A. Zoha, A. Gluhak, M. A. Imran, S. Rajasegarar, 2012, "Non-Intrusive Load Monitoring Approaches for Disaggregated Energy Sensing: A Survey", *Sensors* vol. 12, 16838-16866.
- [4] J. M. Gillis, S. M. Alshareef, W. G. Morsi, 2016, "Nonintrusive Load Monitoring Using Wavelet Design and Machine Learning", *IEEE Trans. Smart Grid* vol. 7, 320—328.
- [5] K. He, L. Stankovic, J. Liao, V. Stankovic, 2018, "Non-Intrusive Load Disaggregation Using Graph Signal Processing", *IEEE Trans. Smart Grid* vol. 9, 1739—1747.
- [6] C. Dinesh, B. W. Nettasinghe, R. I. Godaliyadda, M. P. B. Ekanayake, J. Ekanayake, J. V. Wijayakulasooriya, 2016, "Residential Appliance Identification Based on Spectral Information of Low Frequency Smart Meter

Measurements", *IEEE Trans. Smart Grid* vol. 7, 2781—2792

[7] H. Kim, M. Marwah, M. Arlitt, G. Lyon, J. Han, 2011, "Unsupervised Disaggregation of Low Frequency Power Measurements", *Proceedings of the 2011 SIAM International Conference on Data Mining*, SIAM, 747—758.

[8] K. Basu, A. Hably, V. Debusschere, S. Bacha, G. J. Dirven, A. Ovalle, 2016, "A comparative study of low sampling non-intrusive load disaggregation". *42nd Annual Conference of IEEE Industrial Electronics Society*, Florence, Italy.

[9] B. Zhao, L. Stankovic, V. Stankovic, 2018, "Electricity usage profile disaggregation of hourly smart meter data", *Poster Session at the 4th International Workshop on Non-Intrusive Load Monitoring*, Austin, United States

[10] J. Kelly, W. Knottenbelt, 2015, "Neural NILM: Deep Neural Networks Applied to Energy Disaggregation", *BuildSys*, ACM

[11] P. Zhang, C. Zhou, B. G. Brian Stewart, D. M. Hepburn, W. Zhou, J. Yu, 2011, "An Improved Non-Intrusive Load Monitoring Method for Recognition of Electric Vehicle Battery Charging Load", *Energy Procedia*, vol. 12, 104--112

[12] Z. Zhang, J. H. Son, Y. Li, M. Trayer, Z. Pi, D. Y. Hwang, J. K. Moon, 2014, "Training-free non-intrusive load monitoring of electric vehicle charging with low sampling rate", *IECON 2014 - 40th Annual Conference of the IEEE Industrial Electronics Society*, Dallas, USA

[13] A. Shaw, B. P. Nayak, 2017, "Electric vehicle charging load filtering by power signature analysis", *ICDMAI*, Pune, India

[14] R. Jia, Y. Gao, C. J. Spanos, 2015, "A fully unsupervised non-intrusive load monitoring framework", *SmartGridComm*, Miami, USA

[15] M. Valenti, R. Bonfigli, E. Principi, S. Squartini, 2018, "Exploiting the Reactive Power in Deep Neural Models for Non-Intrusive Load Monitoring", *IJCNN*, Rio de Janeiro, Brazil

[16] Q. Dang, Y. Huo, C. Sun, 2018, "Privacy Preservation Needed for Smart Meter System: A Methodology to Recognize Electric Vehicle (EV) Models", *ISGT Asia*, Singapore

[17] Pecan Street Inc., "Dataport", [Online], Available: <https://dataport.cloud/>.