Energy Informatics

Detection of heat pumps from smart meter and open data

Check for updates

Andreas Weigert[1]* , Konstantin Hopf[1] , Nicolai Weinig[1] and Thorsten Staake[1,2]

*Correspondence:
andreas.weigert@uni-bamberg.de
[1]Information Systems and Energy
Efficient Systems, University of
Bamberg, Kapuzinerstraße 16,
DE-96047 Bamberg, Germany
Full list of author information is
available at the end of the article

Abstract

Heat pumps embody solutions that heat or cool buildings effectively and sustainably, with zero emissions at the place of installation. As they pose significant load on the power grid, knowledge on their existence is crucial for grid operators, e.g., to forecast load and to plan grid operation. Further details, like the thermal reservoir (ground or air source) or the age of a heat pump installation renders energy-related services possible that utility companies can offer in the future (e.g., detecting wrongly calibrated installations, household energy efficiency checks). This study investigates the prediction of heat pump installations, their thermal reservoir and age. For this, we obtained a dataset with 397 households in Switzerland, all equipped with smart meters, collected ground truth data on installed heat pumps and enriched this data with weather data and geographical information. Our investigation replicates the state of the art in the area of heat pump detection and goes beyond it, as we obtain three major findings: First, machine learning can detect the existence of heat pumps with an AUC performance metric of 0.82, their heat reservoir with an AUC of 0.86, and their age with an AUC of 0.73. Second, heat pump existence can be better detected using data during the heating period than during summer. Third the number of training samples to detect the existence of heat pumps must not be necessarily large in terms of the number of training instances and observation period.

Keywords: Heat pump detection, Smart meter data, Machine learning

Introduction

Heat pumps are modern systems that effectively, and sustainably, heat and cool rooms and domestic hot water. They use electricity to convert natural energy from ground water, the earth or air into usable heat energy. This energy comes with zero emissions at the installation. Heat pumps are not only attractive for residential homes due to their efficient energy generation, they also require little maintenance (Karytsas and Choropanitis 2017) and have a long service life, which usually amortizes their higher purchase price over their time of operation. In addition to these basic characteristics, such heating systems

Weigert *et al. Energy Informatics* 2020, **3**(Suppl 1):21

Page 2 of 14

enable the effectively combination with local photovoltaic installations to realize self-supply and storage concepts on a micro level on various scales, from a single residency up to industrial environments (Lorenzo and Narvarte 2019).

Grid operators can benefit from a greater diffusion of heat pumps—under their control—in four ways (Fischer and Madani 2017): First, they can use heat pumps for grid easing (e.g., voltage control, congestion management, and as operating reserve), to integrate renewable energies (e.g., wind, photovoltaic, smoothing of residual load) by coupling the sectors electricity and heat, and to better manage electricity prices (e.g., time of use, day ahead, and dynamic pricing).

The diffusion of heat pumps shows a strong increase. The European Heat Pump Association (EHPA 2019) estimates 11.8 million installed units in 2018, whereas only 1.14 million units were installed in 2005. This significant investment in sustainable technologies pleases climate policymakers, but causes headaches for energy suppliers, especially in terms of load forecasting and grid planning. They rely on accurate forecasts to determine needed resources to maintain the energy balance between supply and consumption constantly. Heat pumps represent a significant load on the power grid and show different load curves than households that have other heating installations. Grid planning for private households is still often carried out with standard load profiles, especially for consumers who have not yet installed a smart meter (Fischer and Madani 2017; Pflugradt and Muntwyler 2017). Significant additional load, as heat pumps generate, can have a negative impact on the stability of the grid, when grid operators do not know the needed energy. For known heat pump installations, energy utilities use special load profiles in their planning, but energy utilities do not necessarily know all heat pump installations, given that homeowners have no obligation to report them to the grid operator in the case of small installations. In addition to the problem context of load forecasting and grid operation, energy utilities want to develop new services around the topic of energy efficiency, partially because they are mandated to do so (EU 2012). Further information, in addition to the existence of a heat pump installation, such as the heat reservoir or the age of a system, enables novel services. For example, when providers know the reservoir (ground source or air source) and the age of a heat pump, they can detect wrongly calibrated ones or conduct energy efficiency checks for homes to offer retrofit options or support consumers to avoid rebound effects (Winther and Wilhite 2015). Besides efficiency improvements, old heat pumps often rely on gases that are harmful to the environment, e.g., like hydrofluororcarbon as refrigerant that has a global warming effect up to 23,000 times greater than carbon dioxide if it leaks into the atmosphere (EC 2016).

To some extent, energy utilities know installed heat pumps from their grid data and use separate electricity meters for such installations because consumers can then choose a special tariff for heat pumps. The ability to detect heat pumps is nevertheless relevant, as small heat pumps do not require notification to the utility company. In addition, energy suppliers must optimize their meter to cash processes and reduce the number of separate electricity meters. Hence, it is helpful to extract the existence of heat pumps and other details from the available data. This paper therefore explores the following research question: *How well can machine learning extract information on installed heat pumps in residential homes from data available to grid operators (i.e., electricity smart meter data, weather data, open data)?*

We structure the remainder of the paper as follows: The following section summarizes the related works in adjacent areas. The third section describes our research method together with the dataset available for our study. Thereafter, we describe our findings. We close the paper with a summary and name implications for research and practice.

Related work

Analytics of smart meter data is a vivid field of research. Many studies exist that aim to recognize electric appliances (Hart 1992; Zeifman and Roth 2011) and to predict household characteristics (Albert and Rajagopal (2013); Beckel et al. (2013, 2014); Hopf et al. (2014, 2018)), in order to realize load forecasting or demand shifting potentials. When limiting the scope to 15-min data that standard smart meter installations record, just two works investigate the detection of heat pumps: Fei et al. (2013) test the predictability of heat pumps in a marketing context in the U.S. using daily electricity consumption and weather data from a 21-month period. The applicability of the results in Central Europe is questionable, because building characteristics such as the typical age and insulation standards of buildings are different (Hu and Qiu 2019). In addition, air conditioning systems are more widespread in North America and have on average a lower energy efficiency level (IEA 2020). This should influence the results. Hopf et al. (2018) and Hopf (2019) investigate the predictability of, in total, 38 household characteristics based on a dataset with 12 months of 15-min electricity consumption and weather data from Switzerland. Their work does not dive into detailed characteristics of heat pumps, like the used heat reservoir or the age of a heat pump.

In our study, we replicated and extend the existing studies with a newly collected dataset. We further tested additional public available data and investigated the predictability of the heat pump reservoir and age which are relevant information for the development of energy efficiency services.

Method

We employ a data science approach to answer the research question and use machine learning to investigate the predictability of heat pump characteristics (see Fig. 1). Below, we describe the dataset, the feature extraction, the application of machine learning algorithms, and the evaluation approach.

Fig. 1 Data science approach to evaluate heat pump predictions
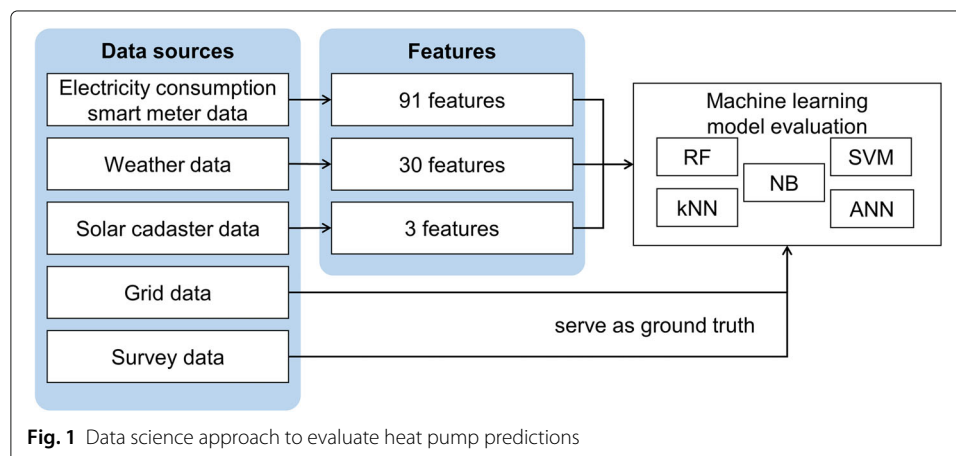
Data

We use data from four different sources: A dataset with electricity smart meter data and information what is measured on the meter, a dataset with weather observations, a solar cadaster dataset, and a survey of residential customers.

The electricity smart meter dataset and the information on what is measured on the meter stems from a large electricity retailer in central Switzerland. During our study, the utility company was rolling out the smart meter infrastructure and, in spring 2020, the company had installed such meters in 8,389 residential households. We received data (kWh consumption in 15-min measurement intervals) from all residential customers with such a meter, together with a short description of each meter that stems from grid operation. This description contains the information whether a heat pump is connected and reported to the grid operator. However, this information might be incomplete because installers report this information late or not at all, because smaller heat pumps do not need to be reported to the grid operator. The data in total covers a time span between January 2012 and March 2020 with an increasing number of metering points (873 in January 2012 to 13,176 in March 2020, including also meters in commercial places).

In order to obtain additional information on existing heat pumps and verify the information reported to grid operation, we conducted a survey in February 2018. We see customer surveys as a reasonable method to collect training data for machine learning applications, especially when it comes to collect objective technical information about housing (i.e., the heating type). Surveys are also a popular data collection method for existing machine learning applications on smart meter data, for example to detect household characteristics (Albert and Rajagopal 2013; Beckel et al. 2014).

We invited all 3,636 residential customers whose metering points where equipped with a smart meter at that time. Given that the cooperating electric utility operates in a monopoly market at the time of the study (Switzerland), a sufficient regional coverage is given. We asked survey participants what heating system they use in general (e.g. oil heater, gas heater, heat pump), if they had a heat pump, what reservoir it uses (e.g. ground source or air source), and we obtained a consent for the use of their smart meter data and address in our study. In total, 589 households participated in that survey, and 397 households provided data on their heating installation. For this study, we used this information to construct the dependent variables that we list in Table 1. We found a mismatch between the heat pump information that was stored in the utility's grid data and the reported existence of heat pumps: 90 customers stated that there was a heat pump,

Table 1 Heat pump characteristics and available ground truth data

Dep. variable	Class	Class size	Relative
Heat pump existence	Heat pump	93	23%
	No heat pump	304	77%
Heat pump type (reservoir)	Ground source	63	16%
	Air source	24	6%
	No heat pump	304	78%
Heat pump age	< 10	49	54%
(binary)	≥ 10	41	46%
Heat pump age	< 10	49	54%
(three classes)	10 ≤ X < 20	31	34%
	≥ 20	10	11%

but the utility company was only aware of 51 installations. There were also three installations listed at the utility where customers did not report any heat pump in the survey. For the training dataset, we counted all houses in the class "Heat pump", where either in the survey or in the grid data a heat pump was specified.

We enrich the training dataset with weather information because the outside temperature influences the consumption pattern of heat pumps strongly. We expected that this additional information improves the models, as the thermal energy demand of a heat pump, required to keep a house on a comfortable temperature level increases with a decreasing outside temperature. Conversely, low outside air temperatures decreases the coefficient of performance a heat pump because electrical power consumed by the compressor must increase to compensate the lower air temperature. Weather data were also used in related work (Fei et al. 2013; Hopf et al. 2018). We used NOAA (2020) weather data for four weather variables (temperature, wind speed, air pressure, and precipitation) from the six nearest stations within the area of the distribution grid of the utility company. The most obvious approach to assign a weather station to a metering point is to use flat distance between both sites. However, due to special mountainous landscape in Switzerland not only the flat distance but also the altitude to the next weather station must be considered. For this reason, we decided to calculate for each variable the average value of the six nearest weather stations together, instead of using only the nearest weather station. The weather data has a measurement interval of 60 min and we completed missing values through linear interpolation.

Finally, we use geographic information to account for heating system related household characteristics that are otherwise not available for grid operators or utilities and might be beneficial to detect the existence of heat pumps. We found the Swiss solar cadaster[1] as a helpful dataset in this case, as it provides data on the living area of a house that must be heated and contains an estimation of the thermal energy demand (heating and domestic hot water generation) for 3,677,970 individual houses in Switzerland (Klauser and Schlegel 2016). In order to assign the solar cadaster information to the households in the smart meter dataset, we selected the nearest building at the given the customer address.

Feature extraction

In order to prepare the data for further analyses, we computed 91 features for each week of the smart meter dataset. This time window is one instance of the natural working days and weekend cycle and is sufficiently large to detect household characteristics (Beckel et al. (2014); Hopf et al. (2014, 2018); Hopf (2019)). We used features on the smart meter electricity consumption data for one week that earlier works found effective to the detect household characteristics (Hopf et al. (2014, 2018); Hopf (2019)). These features describe the smart meter data from four directions: consumption features (e.g., mean consumption during times of the day), ratios of consumption measurements (e.g., ratio between consumption on weekdays and the weekend), statistical values (e.g., standard deviation, auto-correlation), and time-series related features (e.g., seasonal trend decomposition). The full list of features can be found in Hopf et al. (2018) and the implementation is available in the R package `SmartMeterAnalytics`.

---

[1]https://www.uvek-gis.admin.ch/BFE/sonnendach/, last access March 31, 2020; the data was provided by the Swiss Federal Office of Energy

Weigert *et al. Energy Informatics* 2020, **3**(Suppl 1):21

Page 6 of 14

Table 2 Tested combinations of feature sets and available entries in the dataset for heat pump existence

	Sample size	Smart meter	Weather	Solar cadaster
Num. features	–	91	30	3
Model 1 (SMD)	397	X		
Model 2 (SMD + solar cadaster)	397	X		X
Model 3 (SMD + weather)	387	X	X	
Model 4 (all features)	387	X	X	X

For each of the four weather variables, we computed eight features that describe the correlation between electricity consumption data and weather data (e.g., overall correlation, during different daytimes, and days of the week). Two correlations for the precipitation could not be calculated because of missing values in the weather data, therefore we obtained 30 features from smart meter and weather data. A full list of features is given in Hopf et al. (2018).

From the solar cadaster dataset, we computed three features: The basal area of the building, the energy demand of hot water (in kWh per year), and the energy demand for room heating (in kWh per year). Details on the estimation of these numbers can be obtained from the technical report (Klauser and Schlegel 2016). For two observations in our sample, we had no geo-reference, thus, we interpolated the missing values of the solar cadaster features (32 values in total) with the respective column mean values. A list of the number of variables calculated for each data source is given in Table 2.

Application of machine learning algorithms

We apply machine learning for the detection of installed heat pumps in residential homes in order to create prediction models from the ground truth data on heat pumps, following earlier studies (Beckel et al. 2014; Fei et al. 2013; Hopf et al. 2018). We test five machine learning algorithms from different categories:

- Random Forest (RF) as an ensemble learner generates multiple low correlated decision trees and uses majority vote to decide which example belongs to which class.
- Support Vector Machine (SVM) searches for a hyper plane in the vector space that separates all training examples with a maximal margin (Vapnik 1998).
- Naïve Bayes (NB) is a classifier that predicts the class membership based on a probability that a given data point belongs to a class by applying the Bayes' theorem.
- k Nearest Neighbor (kNN) as distance-based approach infers the class-membership by considering the k training instances with the lowest (e.g., Euclidean) distance.
- A simple feed-forward Artificial Neural Network (ANN) was used which consists of a single layer of outputs.

For a detailed description of the used algorithms we refer to Kuhn and Johnson (2013). We used a standard set of parameters and packages in R[2].

---

[2] RF: $ntree = 500, mtry = 9, nodesize = 1$;, implementation of (Liaw and Wiener 2002) SVM: radial kernel, $cost = 1, \gamma = 0.01, \epsilon = 0.1, coef = 0$, implementation of (Meyer et al. 2019) kNN: Euclidean distance, $k = 1$, implementation of (Kuhn 2020) NB: implementation of (Meyer et al. 2019) ANN: One hidden layer, rprop+ algorithm, SSE error function, logistic activation function, $threshold : 0.01, repetitions : 1, stepmax = 1e + 05$, implementation of (Fritsch et al. 2019)

Weigert *et al. Energy Informatics* 2020, **3**(Suppl 1):21

Page 7 of 14

Model evaluation

We evaluated the prediction results by comparing predicted with true labels. Thereby, we used the measures:

$$precision = \frac{true\ positives}{predicted\ positives}$$

$$recall = \frac{true\ positives}{actual\ positives}$$

$$F_1 = \frac{2 * precision * recall}{precision + recall}$$

These three measures are well known, but they are biased by the class distribution. Consequently, a comparison of the results between different dependent variables is difficult. Therefore, we use the Receiver Operating Characteristic (ROC) curve. This curve is a two-dimensional figure with true positive and false positive rates on vertical and horizontal axes (Fawcett 2006). Area under the ROC curve (AUC) is a performance metric derived from the ROC portion of the area of the unit square, and its value varies between 0 and 1. Random guessing produces a diagonal line between (0, 0) and (1, 1), which has an AUC of 0.5. Effective prediction models are therefore expected to achieve values above 0.5 (Fawcett 2006). For the performance evaluation, we apply 10-fold cross-validation and present the mean values of measures.

Results

We organize the result presentation in three sections. We start with the detection of heat pumps from smart meter electricity consumption data. We tested different machine learning algorithms and combinations of feature sets. This analysis also helped us to compare our work with the state of the art and to select the best performing model for the consecutive analyses. Second, we analyzed the prediction performance over time to get an impression of the model stability as well as times of the year in which data collection for real applications is most helpful. Third, we tested how well heat pump characteristics such as the type of the heat reservoir or age of the device can be predicted by our model.

Prediction of heat pump existence

In the first analysis, we predicted the existence of a heat pump in the form of a binary classification problem. The ground truth data for this analysis stem from grid information and survey data that we used to define the dependent variable heat pump existence (see Table 1). We tested the different machine learning algorithms and feature sets with data from one typical week in spring 2020 that has no school or public holidays included, and is still within the typical heating period in Switzerland (ISO week 10, March 02-08).

Table 2 gives an overview to the different combinations of feature sets and the respective dataset sizes. The first model contained only features extracted from the smart meter data (91 features in total), the second model also included the solar cadaster features (94 features in total), the third model considered the smart meter in combination with the weather features (121 features in total), and the last model included all features (124 features in total). Due to missing values in the weather features, the number of observations is reduced by 10 to 387 in the third and fourth model, respectively.

Weigert *et al. Energy Informatics* 2020, **3**(Suppl 1):21

Page 8 of 14

Table 3 shows the Mean (M) prediction performance and the Standard Deviation (SD) in brackets of these models with all tested machine learning models and performance metrics. The best result for each performance metric is marked bold. Using AUC as the central performance metric, RF leads to the best results compared to the other algorithms. Combining either the solar cadaster or the weather features with the smart meter features increases the performance, but the model with all features is worse than the combination of smart meter and weather features (model 3). Thus, the best model (smart meter and weather data) achieved an AUC of $M = 0.822$ ($SD = 0.07$), which is slightly higher than an AUC of $M = 0.807$ ($SD = 0.07$) of the model with all features (model 4), but the difference was not statistically significant $t(17.99) = 0.50, p = 0.624, d = 0.22$. Based on these results, we have excluded the ANN algorithm from the following analyses and only use smart meter and weather data (model 3). This helped to reduce the complexity of the following steps.

For the RF algorithm, we also illustrate the four models as ROC curves in Fig. 2. It is visible that model 3 has the strongest curvature, but the difference to the other models is not large.
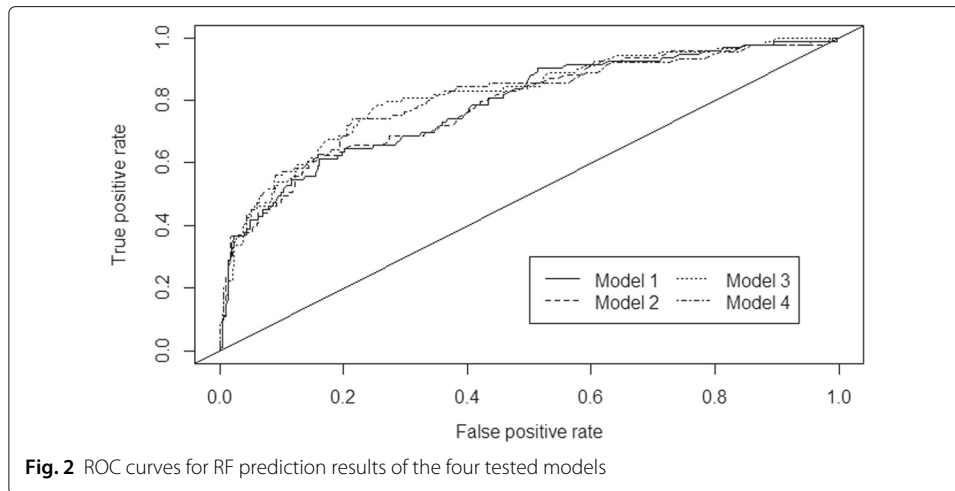
Seasonal impact on the classification performance

We further tested whether the time of the year—and respectively changing heating behavior—affects the classification performance. For this analysis, we used the RF algorithm with smart meter and weather features (model 3). We calculated the classification performance for each week between January 1, 2017 and March 31, 2020 and visualize the AUC results in Fig. 3.

Table 3 Mean and standard deviation of prediction performance for heat pump existence with different machine learning algorithms (Data: ISO week 10, 2020)

Algo.	Models	AUC	Precision*	Recall	F1
RF	Model 1	0.794 (0.10)	0.733 (0.17)	0.401 (0.19)	0.691 (0.23)
RF	Model 2	0.797 (0.11)	0.728 (0.17)	0.412 (0.19)	0.689 (0.23)
RF	Model 3	**0.822** (0.07)	0.737 (0.21)	0.449 (0.13)	0.723 (0.21)
RF	Model 4	0.807 (0.07)	0.750 (0.22)	0.426 (0.12)	0.715 (0.21)
SVM	Model 1	0.769 (0.11)	0.822 (0.16)	0.433 (0.19)	0.724 (0.22)
SVM	Model 2	0.773 (0.11)	**0.847** (0.16)	0.443 (0.18)	0.733 (0.21)
SVM	Model 3	0.792 (0.10)	0.803 (0.23)	**0.449** (0.15)	**0.736** (0.21)
SVM	Model 4	0.786 (0.11)	0.790 (0.23)	**0.449** (0.15)	0.734 (0.21)
kNN	Model 1	0.637 (0.06)	0.524 (0.20)	0.420 (0.13)	0.641 (0.22)
kNN	Model 2	0.663 (0.12)	0.447 (0.22)	0.481 (0.26)	0.650 (0.26)
kNN	Model 3	0.641 (0.09)	0.486 (0.13)	0.426 (0.18)	0.639 (0.23)
kNN	Model 4	0.605 (0.11)	0.374 (0.20)	0.394 (0.19)	0.599 (0.26)
NB	Model 1	0.689 (0.12)	0.267 (0.05)	0.830 (0.08)	0.413 (0.11)
NB	Model 2	0.693 (0.11)	0.263 (0.05)	0.831 (0.08)	0.393 (0.12)
NB	Model 3	0.701 (0.10)	0.272 (0.04)	0.821 (0.14)	0.446 (0.09)
NB	Model 4	0.704 (0.11)	0.278 (0.04)	0.843 (0.13)	0.452 (0.09)
ANN	Model 1	0.529 (0.09)	-	0.030 (0.09)	0.456 (0.43)
ANN	Model 2	0.545 (0.05)	-	0.033 (0.10)	0.455 (0.43)
ANN	Model 3	0.500 (0.02)	-	0.000 (0.00)	0.435 (0.45)
ANN	Model 4	0.536 (0.05)	-	0.000 (0.00)	0.435 (0.45)

*The ANN model could not predict any positive example

Weigert *et al. Energy Informatics* 2020, **3**(Suppl 1):21

Page 9 of 14

Fig. 2 ROC curves for RF prediction results of the four tested models

The average AUC is significantly higher and predictions are more stable during heating times: The weeks 1–12 (roughly the first three months of the year) together with week 40–52 (roughly the last two months of the year) have an AUC of $M = 0.774$ ($SD = 0.13$) and the weeks 13–39 $M = 0.674$ ($SD = 0.12$). This difference is statistically significant with $t(164.85) = 5.18, p < .001, d = 0.80$.

Heat pump type (reservoir)

For a sample of 87 installations, we have survey data on the reservoir of the heat pump available. Based on this data, we tested whether this detail can be predicted based on smart meter and weather data (which was the feature set that performed best in our first analysis). We set up this prediction in two ways. First, we used a three-class problem with the classes "Ground source", "Air source", and "No heat pump". Second, we used only the subset of data with the known heat pump types ($n = 87$) and predicted the type as a two-class problem.

The results of both prediction problems are shown in Table 4. The RF algorithm performed best for the three-class problem. In contrast to the performance of our initial prediction problem (where we just predicted the existence of a heat pump), the
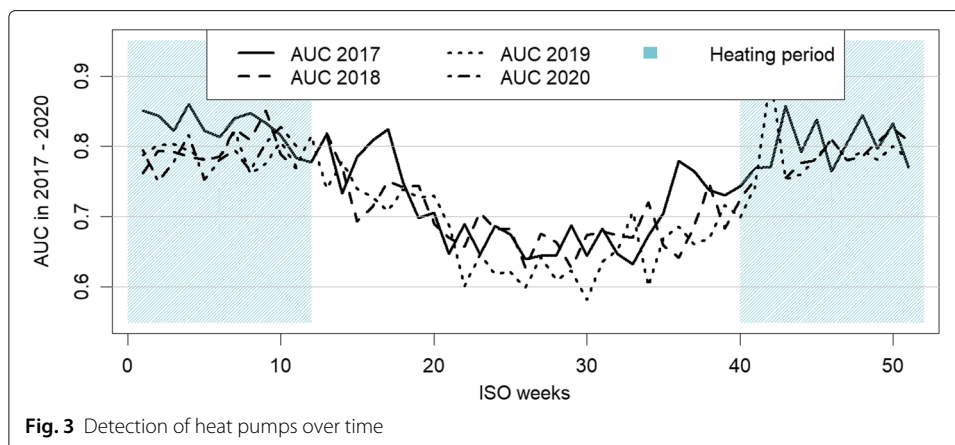

Fig. 3 Detection of heat pumps over time

Table 4 Heat pump type (reservoir) with data from week 10, 2020 using smart meter and weather features

Algo.	AUC (ground source)	AUC (air source)	AUC (no heat pump)
Three classes {*ground source, air source, no heat pump*} using complete dataset ($n = 381$)			
RF	0.732 (0.08)	**0.859** (0.21)	**0.811** (0.05)
SVM	**0.734** (0.09)	0.835 (0.25)	0.803 (0.06)
kNN	0.548 (0.07)	0.640 (0.17)	0.670 (0.06)
NB	0.494 (0.14)	0.802 (0.20)	0.728 (0.08)
Two classes {*ground source, air source*} using known installations ($n = 87$)			
RF	0.559 (0.30)		
SVM	0.529 (0.27)		
kNN	0.593 (0.16)		
NB	**0.649** (0.30)		

prediction of air heat pumps AUC of $M = 0.859$ ($SD = 0.21$) is not statistically significantly different ($t(10.77) = 0.51, p = 0.620, d = 0.23$) from the initial prediction problem $M = 0.822$ ($SD = 0.07$). In case of ground source heat pumps the AUC of $M = 0.732$ ($SD = 0.08$), the prediction is worse ($t(17.41) = -2.69, p = 0.008, d = -1.20$) compared to the initial prediction problem $M = 0.822$ ($SD = 0.07$). However, we can predict more information (three classes instead of two) with a considerable performance loss.

The two-class problem, where we tested whether the reservoir can be predicted when the existence is known, achieves lower performance values than the three-class problem. We attribute this lower prediction AUC to the lower number of training examples in this experiment.

We conclude that, based on the data, knowledge about the existence of a heat pump does not contribute significantly to a better prediction of the heat pump reservoir, probably because households with heat pumps show a considerably different consumption pattern compared to households without heat pumps and thus can be easily discriminated. However, a combined prediction of the existence and reservoir of a heat pump can lead to more detailed information that is also more accurate.

Heat pump age

Finally, we investigated the predictability of the age of heat pumps in our sample. Table 5 shows the performance of the different classification algorithms with a two- and a three-class classification problem. We observe that the RF algorithm, again, shows better results in the three-class problem compared to the two-class problem. However, the NB model shows better results in detecting heat pump installations that are newer than ten years, but not in detecting older systems. Here, RF is better, but all results are affected to high variations.

Discussion and conclusion

This paper investigated the application of machine learning algorithms to detect the existence of heat pumps as well as characteristics about such installations from 15-min smart meter data. We draw on two earlier works on heat pump detection (one covers installations in the U.S. (Fei et al. 2013) and the other installations in Switzerland (Hopf et al. 2018; Hopf 2019)), replicate their results and pursued further analyses.

Table 5 Heat pump age in years with data from week 10, 2020 using smart meter and weather features

Algo.	AUC (< 10)	AUC (≥ 10)	AUC (≥ 20)
Three classes {< 10, 10 ≤ X < 20, ≥ 20}			
RF	0.651 (0.09)	0.617 (0.15)	**0.726** (0.23)
SVM	0.631 (0.17)	**0.626** (0.14)	0.395 (0.34)
kNN	0.417 (0.11)	0.517 (0.20)	0.439 (0.05)
NB	**0.656** (0.14)	0.623 (0.19)	0.366 (0.29)
Two classes {< 10, ≥ 10}			
RF	0.654 (0.18)		
SVM	0.671 (0.15)		
kNN	0.491 (0.13)		
NB	**0.695** (0.17)		

We collected a dataset that covers 3.5 years of smart meter electricity consumption data together with ground truth data. This time span is larger than those of previous work (with 1 or 2.5 years). Drawing on survey data and data from electricity grid operations, we identify true class labels. This combination of the two data sources allowed us to create a stable training data set that was only to a very small extent inaccurate (out of 397 households that provided information on heating, only 3 were implausible because a heat pump was reported in the grid data but not in the survey data). To this comparably comprehensive dataset, we applied five machine learning algorithms and tested whether geographical information can help—in addition to weather and smart meter data which was already tested in earlier studies (Fei et al. 2013; Hopf et al. 2018)—to predict the existence of heat pumps. We could predict the existence of heat pumps with an AUC of up to 0.82 ($F_1 \leq 0.74$). These results are on a comparable level to the results of existing studies. Fei et al. (2013) achieved a performance of ($F_1 \leq 0.86$), whereas (Hopf et al. 2018) could predict the existence of heat pumps with a performance of ($AUC = 0.677$). Thus, our work replicates their findings with a novel dataset. This suggests that these studies do not report dataset-specific findings and that the prediction models are relatively stable.

In addition to earlier studies, we assessed whether the use of further data sources (i.e., geographic information) can improve the prediction performance. There, we tested different combinations of feature sets that stem from three for grid operators available data sources (smart meter data, weather data, geographical data). Our results show that smart meter data alone allow good prediction results for the detection of the existence of a heat pump. Additional geographical information such as the solar cadaster dataset, containing basic building characteristics and estimations on the heat energy demand, improves the prediction marginally, whereas weather information considerably improves the prediction.

Finally, we tested the predictability of heat pump characteristics that are particularly interesting for energy efficiency campaigns. The *heat pump reservoir* (ground source vs. air source) could be predicted with an AUC of 0.86 and the heat pump age with an AUC of 0.73. Both prediction performances are significantly higher than AUC of 0.5, which means that the predictions are clearly better than random. A priori knowledge on the heat pump existence (two class prediction of known installations) could not improve the prediction of the heat pump reservoir. One reason for this could be that the number of training examples (n=87) was to small to build a reliable model in the two class prediction

whereas the availability of a larger set of negative examples in the three class prediction increases the ability to detect heat pump specific characteristics. The *age of the heat pump* could be predicted with an AUC uf up to 0.726 (heat pumps older than 20 years), but not reliably, because of the large standard deviation of AUC based on our training sample. We expect that a training dataset with more heat pump observations will provide more stable results.

Implications for research and practice

We can derive two implications from our study regarding the detection of existing heat pumps. First, the number of ground truth data necessary to train prediction models does not need to be large (a dataset with less than 400 households was sufficient in our study to achieve a successful prediction). Second, selecting the right period of smart meter data matters. This finding is in line with previous research (Hopf et al. 2018), indicating that the detection of heat pumps would perform better in the winter months than in the summer, given the data from one year. With our data, we can confirm this finding with data over a period of more than three years. We explain this higher performance with the typical consumption patterns of heat pumps attributable to space heating in the winter. Only a small portion of heat energy is consumed during the summer for hot water production. This decreases the predictability in the respective times in summer. We conclude that the right choice of the observation period (e.g., month November to February) allows to reduce the size of the required dataset.

Limitations and future research

There are some limitations that should be addressed in future studies. First, we have not carried out a detailed parameter tuning to optimize the machine learning algorithms for heat pump detection. This has the potential to significantly boost the performance but conveys the risk of overfitting the models to the data. Although our dataset is comprehensive in terms of the time span of the observations, it has only a moderate number of different households what makes parameter tuning difficult. Consequently, our results are conservative in terms of the maximum achievable performance. Second, this work considers only data from residential households in central Switzerland where heat pumps are primarily used for heating. These results could be transferable to countries with similar climatic conditions but not to regions where heat pumps are used in a dual-use mode to heat and cool buildings, e.g. climatic hot regions in Greece (Karytsas and Choropanitis 2017). Third, further household information, like building characteristics or the number of residents in a household that have a large impact on the heat energy used, could improve the prediction performance of the investigated models significantly. This could also be included in future works.

Weigert *et al. Energy Informatics* 2020, **3**(Suppl 1):21

Page 13 of 14

About this supplement

This article has been published as part of *Energy Informatics Volume 3 Supplement 1, 2020: Proceedings of the 9th DACH+ Conference on Energy Informatics*. The full contents of the supplement are available online at https://energyinformatics. springeropen.com/articles/supplements/volume-3-supplement-1.

Authors' contributions

AW and KH collected the data. AW, NW, and KH conducted the statistical analysis and wrote the manuscript. TS provided critical review and wrote parts of abstract and introduction. All author(s) have read and approved the final manuscript.

Funding

Availability of data and materials

The ground truth data and a subset of seven weeks of electricity consumption data is available in the R package *ResidentialEnergyConsumption* (Hopf et al. 2020) and the implementation of the feature extraction procedures is available in the R package *SmartMeterAnalytics* (Hopf et al. 2020) that we both publish along with this paper. We thank the U.S. NOAA and the Swiss Federal Office of Energy for providing weather and solar cadaster data.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Information Systems and Energy Efficient Systems, University of Bamberg, Kapuzinerstraße 16, DE-96047 Bamberg, Germany. [2]Department of Management, Technology and Economics, ETH Zurich, Weinbergstrasse 5, CH-8092 Zurich, Switzerland.

Published: 28 October 2020

References

Albert A, Rajagopal R (2013) Smart meter driven segmentation: what your consumption says about you. IEEE Trans Power Syst 28(4):4019–4030

Beckel C, Sadamori L, Santini S (2013) Automatic socio-economic classification of households using electricity consumption data. In: Culler D, Rosenberg C, Keshav S, Kurose J (eds). ACM Press, Berkeley, California, USA. p 75

Beckel C, Sadamori L, Staake T, Santini S (2014) Revealing household characteristics from smart meter data. Energy 78:397–410

EC (2016) Fluorinated greenhouse gases. Library Catalog: ec.europa.eu. https://ec.europa.eu/clima/policies/f-gas_en. Accessed 28 May 2020

EHPA (2019) Market Report 2019 - Executive Summary. Technical report, European Heat Pump Association, Brussels. https://www.ehpa.org/fileadmin/red/07._Market_Data/2019/2019_Executive_Summary.pdf. Accessed 14 May 2020

EU (2012) Directive 2012/27/EU of the European Parliament and of the Council of 25 October 2012 on energy efficiency, amending Directives 2009/125/EC and 2010/30/EU and repealing Directives 2004/8/EC and 2006/32/EC Text with EEA relevance. http://ec.europa.eu/energy/en/topics/energy-efficiency/energy-efficiency-directive. Accessed 28 Apr 2015

Fawcett T (2006) An introduction to ROC analysis. Pattern Recogn Lett 27(8):861–874

Fei H, Kim Y, Sahu S, Naphade M, Mamidipalli SK, Hutchinson J (2013) Heat pump detection from coarse grained smart meter data with positive and unlabeled learning. In: Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery And Data Mining. KDD '13. ACM, New York. pp 1330–1338

Fischer D, Madani H (2017) On heat pumps in smart grids: A review. Renew Sust Energ Rev 70:342–357

Fritsch S, Guenther F, Wright MN (2019) Neuralnet: Training of neural networks. R package version 1.44.2. https://CRAN.R-project.org/package=neuralnet. Accessed 28 May 2020

Hart GW (1992) Nonintrusive appliance load monitoring. Proc IEEE 80(12):1870–1891

Hopf K (2019) Predictive analytics for energy efficiency and energy retailing, 1st edn. Contributions of the Faculty Information Systems and Applied Computer Sciences of the Otto-Friedrich-University Bamberg, vol. 36. University of Bamberg, Bamberg

Hopf K, Sodenkamp M, Kozlovskiy I, Staake T (2014) Feature extraction and filtering for household classification based on smart electricity meter data. In: Computer Science-Research and Development, vol. (31) 3. Springer, Zürich. pp 141–148

Hopf K, Sodenkamp M, Staake T (2018) Enhancing energy efficiency in the residential sector with smart meter data analytics. Electron Mark 28(4):453–473 https://link.springer.com/article/10.1007/s12525-018-0290-9

Hopf K, Weigert A, Kozlovskiy I, Staake T (2020) Smart Meter Data Analytics, 1st edn. Information Systems and Energy Efficient Systems Group, University of Bamberg. Information Systems and Energy Efficient Systems Group, University of Bamberg. R package, forthcoming. https://CRAN.R-project.org/package=SmartMeterAnalytics. Accessed 28 May 2020

Hopf K, Weigert A, Weinig N, Staake T (2020) Residential Energy Consumption Data, 1st edn. Information Systems and Energy Efficient Systems Group, University of Bamberg. Information Systems and Energy Efficient Systems Group, University of Bamberg. R package, forthcoming. https://CRAN.R-project.org/package=ResidentialEnergyConsumption. Accessed 28 May 2020

Hu M, Qiu Y (2019) A comparison of building energy codes and policies in the USA, Germany, and China: progress toward the net-zero building goal in three countries. Clean Techn Environ Policy 21(2):291–305. https://doi.org/10.1007/s10098-018-1636-x. Accessed 14 Aug 2020

IEA (2020) Cooling - Analysis. https://www.iea.org/reports/cooling. Accessed 13 Aug 2020

Karytsas S, Choropanitis I (2017) Barriers against and actions towards renewable energy technologies diffusion: A Principal Component Analysis for residential ground source heat pump (GSHP) systems. Renew Sust Energ Rev 78:252–271

Klauser D, Schlegel T (2016) Sonnendach.ch - Datenmodell. Dokumentation v 1.4, Bundesamt für Energie, Bern

Kuhn M (2020) caret: Classification and Regression Training. R package version 6.0-86. https://CRAN.R-project.org/package=caret. Accessed 28 May 2020

Kuhn M, Johnson K (2013) Applied predictive modeling. Springer, New York, NY. https://doi.org/10.1007/978-1-4614-6849-3 http://link.springer.com/10.1007/978-1-4614-6849-3. Accessed 25 Apr 2019

Liaw A, Wiener M (2002) Classification and Regression by randomForest. https://CRAN.R-project.org/doc/Rnews/. Accessed 28 May 2020

Lorenzo C, Narvarte L (2019) Performance indicators of photovoltaic heat-pumps. Heliyon 5(10):e02691

Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F (2019) E1071: Misc functions of the department of statistics, probability theory group (Formerly: E1071), TU Wien. R package version 1.7-2. https://CRAN.R-project.org/package=e1071. Accessed 28 May 2020

NOAA (2020) Global Surface Hourly Global. Technical report, NOAA National Centers for Environmental Information. https://data.nodc.noaa.gov/cgi-bin/iso?id=gov.noaa.ncdc:C00532. Accessed 28 May 2020

Pflugradt N, Muntwyler U (2017) Synthesizing residential load profiles using behavior simulation. Energy Procedia 122:655–660

Vapnik VN (1998) Statistical learning theory. Adaptive and learning systems for signal processing, communications, and control. Wiley, New York

Winther T, Wilhite H (2015) An analysis of the household energy rebound effect from a practice perspective: spatial and temporal dimensions. Energy Efficiency 8(3):595–607

Zeifman M, Roth K (2011) Nonintrusive appliance load monitoring: Review and outlook. IEEE Trans Consum Electron 57(1):76–84

Publisher's Note