Building power demand forecasting using *K*-nearest neighbours model – practical application in Smart City Demo Aspern project

Oleg Valgaev[1] ✉, Friederich Kupzog[1], Hartmut Schmeck[2]

[1]*Center for Energy, Austrian Institute of Technology, Vienna, Austria*
[2]*FZI Research Center for Information Technology, Karlsruhe Institute of Technology, Karlsruhe, Germany*
✉ *E-mail: oleg.valgaev@ait.ac.at*

Abstract: Following the ongoing transformation of the European power system, in the future, it will be necessary to locally balance the increasing share of decentralised renewable energy supply. Therefore, a reliable short-term load forecast at the level of single buildings is required. In this study, we use a forecaster, which is based on *K*-nearest neighbours approach and was introduced in an earlier publication, on three buildings of Smart City Demo Aspern project. The authors demonstrate how this forecaster can be applied on different buildings without any manual setup or parametrisation, showing that it is viable to replace load-profiling solutions for predicting electricity consumption at the level of single buildings.

1 Introduction

The increasing share of renewable energy sources within electricity generation leads to new challenges for the power system infrastructure. Establishing smart grids helps addressing these challenges. The concept seeks to improve the transparency and controllability of the distribution grid by installing advanced information and communication technology at the level of single buildings. It allows applying advanced control schemes to use load flexibilities present in such smart buildings [1]. Many of these control capabilities require an accurate building-level short-term load forecast (STLF). However, appropriate techniques that can be universally applied to various different buildings without their explicit knowledge are yet to be developed.

Until recently, forecasting the load in the local distribution system was of limited interest, and buildings were modelled as low-voltage end-consumers with uncontrollable load. Electricity demand of such consumers is commonly forecast using standardised load profiles (SLP) predefined for general consumer groups. As long as power demand and supply are balanced at the high-voltage level, only large aggregations of end-consumers have to be considered for the load forecasting, where SLPs deliver reliable accuracy.

With the increasing share of decentralised supply connected to the distribution grid, balancing of power generation and consumption will have to be done locally, at the low-voltage level. For this purpose, SLPs are inappropriate as they only rudimentarily reflect the diversity and highly stochastic nature of the building electricity demand. Additionally, newly commissioned smart buildings can no longer be assumed to have a relatively constant load profile as it was done in the past. Such a building is often equipped not only with photovoltaic (PV) panels producing power, but also with electrical heating, ventilation, air-conditioning (HVAC) systems, electrical and thermal storages. Its electricity demand can change considerably over the days, depending on PV-generated power and weather. At the same time, it has load flexibilities, which can be used for balancing purposes locally or at higher domains.

To do so, STLF at the building level is required. Wide-area introduction of smart meters in the European Union [2] will allow to use individual load profiles (ILPs) [3] and to apply more

advanced STLF concepts to the buildings. While there are numerous approaches for the STLF in higher domains (such as system and balancing group levels [4]), techniques specialised for wide-area application at numerous different buildings still present a big task. In general, it is easier to forecast large aggregated loads present in higher domains [5] and, until now, there have been only few attempts [6–8] to adopt those approaches at the building level, where the load aggregation is much smaller. Such propositions are, mostly, done for a specific building and it is still an ongoing challenge to develop a universally applicable forecaster that can be applied on various different buildings disregarding their size or purpose, delivering reliable accuracy.

In a recent publication [9], we have proposed an initial approach for such forecaster based on the *K*-nearest neighbours (KNN) method. This technique is very common for data classification, but can also be extended to forecast functional time series. KNN forecasters, in general [10], are attractive because of their simplicity and the ability to predict complex nonlinear behaviour, such as the one expected of a building load. Our proposition required no manual setup and was applied to numerous different end-consumers, without any explicit knowledge about them.

In this study, we use our forecaster at the smart buildings erected for Smart City Demo Aspern (SCDA) project [11]. We start with a problem formulation where we provide the information about the buildings and define the forecasting problem. Afterwards, we explain the forecasting approach of our KNN forecaster and present the results concluding this study.

2 Problem formulation

Three buildings – a housing, a student dorm and a school campus including primary school and a kindergarten – constitute Aspern Smart City research testbed [11]. Together with the building energy management system and necessary information and communication technology infrastructure, these smart buildings are used within SCDA project to try various load-forecasting concepts which will allow to use the load flexibilities in those buildings in further stages of the research.

Table 1 Buildings overview including total energy (E_{total}) and average daily (E_{avg}) energy consumption during the period from April to September 2016, peak power demand measured during this period (P_{max}) as well as size (number of dwellings/inhabitants/net area), installed PV generation capacity (PPV), battery storage capacity, and installed electrical heating power

	Housing	Dorm	School
size	213 dw	300 inch	11.900 m²
E_{total}, MWh	418.7	55.5	277.6
E_{avg}, kWh	2273	297	1505
P_{max}, kWh	400	192	460
P_{PV}, kW$_p$	36	221	58
electrical heating, kW$_{th}$	800	16	510
battery, kW$_h$	2	120	none

Equipped with PV panels, solar thermal panels, hybrid panels, heat pumps and various thermal as well as electrical storage facilities, smart buildings of the SCDA project represent modern prosumer buildings which are being increasingly commissioned all over Europe. The overview of electrical equipment for each building is provided in Table 1. Due to this equipment, building load considerably deviates from the predefined SLPs which makes it harder to predict.

Metering infrastructure of each building delivers precise load measurements, and for this study we use the electricity consumption data between 1 April 2016 and 30 September 2016 obtained for each building with a 30 min resolution.

The load curves for three randomly selected days are depicted in Fig. 1. Due to the 221 kW of installed PV capacity, dorm often feeds electricity into the grid which results in negative net building load. School and housing appear to have more volatile curves with the former having a load peak during the day, while the latter features a distinguishable evening peak expected of a residential
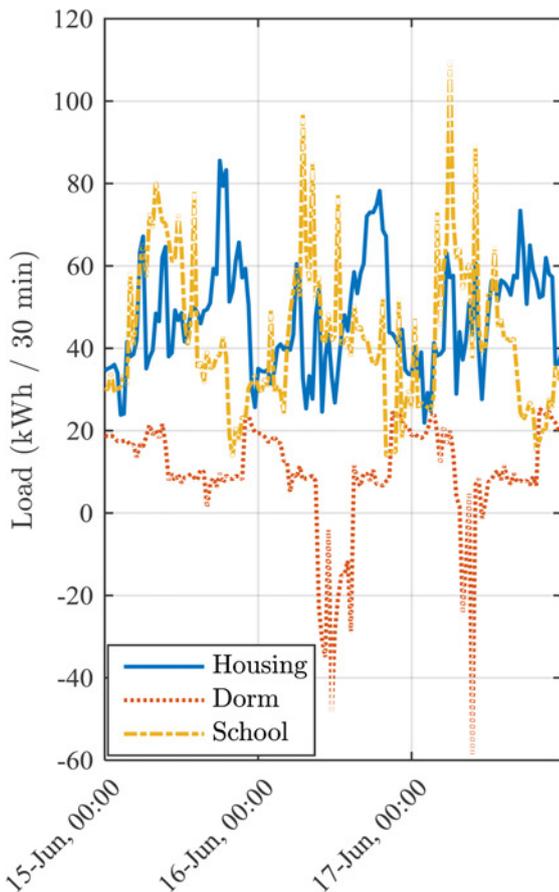
Fig. 1 *Electricity consumption during 30 min intervals of the SCDA buildings on three exemplary days in June 2016*

load. Note how the dorm load considerably depends on PV generation. On 15th of June, probably due to bad weather, the dorm had significantly increased overall consumption around midday.

At this point, we introduce the following notation. Each metering device delivers continuous equidistant measurements that we divide into daily load curves $S_1,...,S_L$ where S_1 represents the first and S_L represents the last day of historic data. Given the 30 min smart meter resolution, each curve Si with $i=1,...,L$ represents a set of 48 equidistant measurements:

$$S_i = [S_i(t_1), \ldots, S_i(t_{48})]. \tag{1}$$

For our buildings, we formulate the forecasting problem as follows. Given historic daily load curves $S_1,...,S_L$, where S_L corresponds to the most recently measured load curve, we want to predict the curve

$$\hat{S}_{L+1} = [\hat{S}_{L+1}(t_1), \ldots, \hat{S}_{L+1}(t_{48})], \tag{2}$$

representing the load for the next 24 h. In this study, we assume that forecast is done at midnight predicting the load curve for the entire upcoming day.

3 Forecasting approach

To predict the load of buildings described above, we apply the forecaster based on the KNN technique. Its overview is provided in Fig. 2 and it is described in more detail in our recent publication [9]. After automatic parametrisation, it finds KNN of the most recently measured load curve S_L. The historic successors of KNN are, then, combined into a forecast using a predefined combination function.

Given a set of historic observations and their successors, a KNN model relies on the assumption that similar observations are also likely to produce similar successors. In our case, the observations represent historic daily load curves $S_1,...,S_L$ with their historic successors $S'_1, \ldots, S'_L$, where $S'_1 = S_2, \ldots, S'_{L-1} = S_L$. Let us assume that each observation S_k with $k=1,..., L-1$ represents an object with a set of features. Coordinates of the object in a 48-dimensional feature space $\mathbb{R}^{48}$ are the 48 data points comprising a daily load curve. For a univariate model, the only feature of each object is its historic successor, so an object can be described as a set $\{S_k, S'_k\}$. We call a collection of such objects a feature space.

Having measured load curve S_L in the last 24 h, we want to predict the curve $\hat{S}_{L+1}$ for the next 24 h. To do so, we place S_L in the feature space described above. We apply workday calendar to use only historic load curves whose successors have the same day type (workday, Saturday and holiday – representing public holidays and Sundays) as $\hat{S}_{L+1}$. For example, if next day is a workday, then only historic curves S_k whose successors S'_k were also workdays are considered in the feature space.
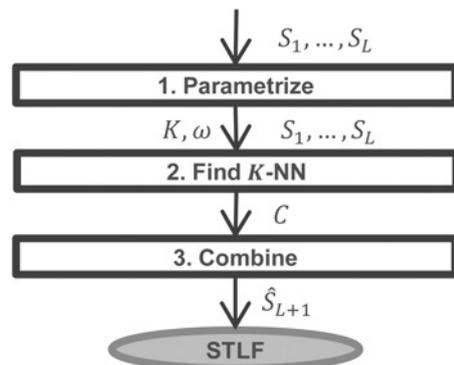
Fig. 2 *KNN forecaster algorithm overview. Description is provided in the text*

Next, a predefined distance notion $d(S_k, S_L)$ is used to measure the distances between S_L and other historic load curves in the feature space. According to the distance notion, KNN of S_L corresponding to the K most similar historic load curves are selected. Their historic successors are included into a set $C := \{S'_i\}$ with $i = 1,\ldots, K$.

Ultimately, the curves in C are combined into a forecast $\hat{S}_{L+1}$ using a predefined combination function.

3.1 Parametrisation

The forecaster parameters K and ω (described further in the text) are set automatically using k-fold cross-validation [12]. It is a common technique from machine learning which searches through a list of candidate parameters evaluating the resulting forecaster on a training set. We use the historic data to parametrise our model before finding KNN.

3.2 Find KNN

To find KNN of the curve S_L, we calculate its distances to the other historic load curves in the feature space. For this purpose, instead of using common Euclidian distance (L^2 norm) we define the distance using the concept of the, so called, permuted L^2 norm which has been developed specifically to quantify the similarity between the time series, rather than calculating a point-wise deviation.

Given time series S_k and S_L, distance $d(S_k, S_L)$ is calculated allowing small-time permutations of each point $S_k(t_i)$ comprising S_k, while comparing the time series and penalising the amplitude differences using L^2 norm:

$$d_\omega(S_k, S_L) := \min_{\pi \in P(\omega,48)} \left(\sum_{i=1}^{48} \left| \pi(S_k))_i - S_i(t_1) \right|^2 \right)^{1/2} \quad (3)$$

In this equation, $P(\omega,48)$ represents the set of all ω-local permutations $\pi(S_k)_i$ on $i = 1,\ldots,48$ points constituting the curve S_k. An ω-local permutation rearranges the time series points by moving each one forwards and backwards by up to ω-time units. The above equation is solved according to [13].

3.3 Combine

Having stored the KNN successors in the set C, we need to combine these load curves using a predefined combination function to a consensus representation of C. Such representation is a curve that has minimal cumulated distance to the curves in C, according to a predefined distance notion.

For our forecaster, we use permutation merge as a combination function. It is a technique that uses permuted L^2 norm as such notion. This technique determines a curve $\hat{S}_{L+1}$ that minimises the cumulated distance according to that norm with respect to all curves in C. The combined load curve $\hat{S}_{L+1}$ can be expressed as a solution to the following optimisation problem:

$$\hat{S}_{L+1} = \arg\min_x \sum_{i=1}^{K} \left(d_\omega(S_i, x) \right)^2. \quad (4)$$

This optimisation problem is solved according to [14].

4 Results

The KNN forecaster described above is used to forecast the electricity consumption of SCDA buildings during the time period between 1 May and 30 September. The forecast is done day after day, while the preceding three weeks are used as history for the KNN forecaster (sliding window approach). The history is also used for parametrisation which is repeated every 6 weeks.

There is an ongoing discussion about how to assess forecast accuracy at lower levels of load aggregation [13, 15]. In this study, we define the adjusted error (AE) based on the permuted L^2 norm as

$$E^{\text{AE}}(F, A) := d_{\omega=3}(F, A), \quad (5)$$

where A and F are actual and forecast daily load curves, respectively. With $\omega = 3$, we allow permutations of up to 90 min in total (given

Table 2 Median daily forecast error (kWh)

	$E_{\text{KNN}}^{\text{AE}}$, kWh	$E_{\text{ILP}}^{\text{AE}}$, kWh	E^{relAE}
housing	50.5	49.9	1.01
dorm	71.5	76.4	0.93
school	55.9	72.1	0.77
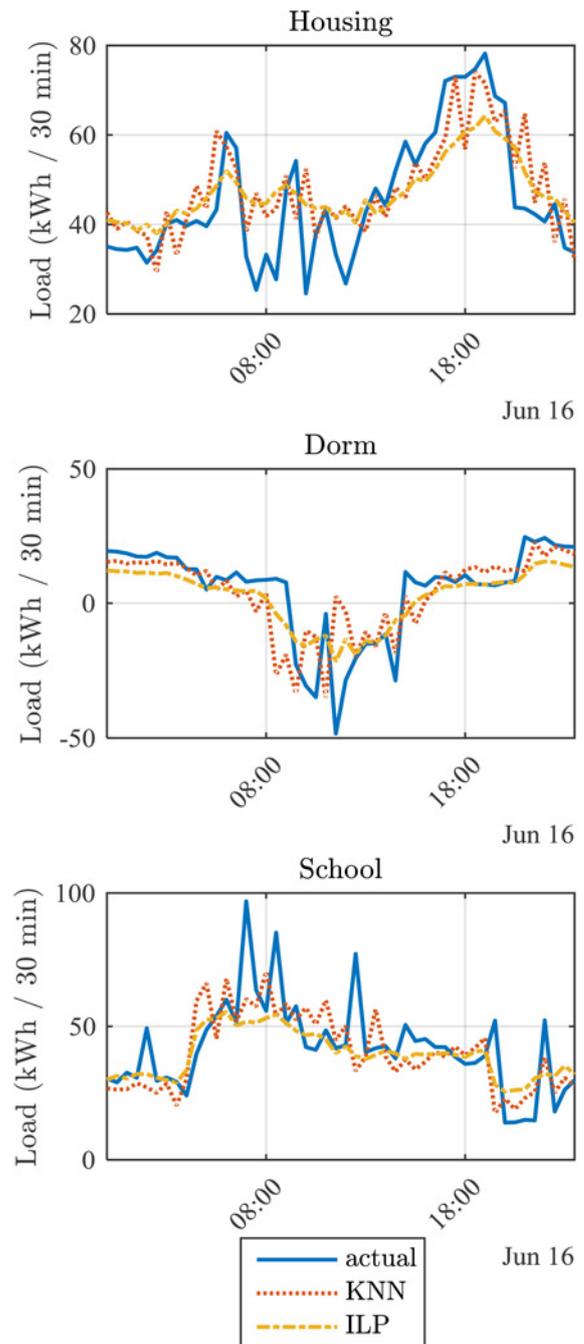
Fig. 3 *Forecast and actual electricity consumed during 30 min period on 16 June*

30 min smart meter resolution). To assess the entire forecast, we calculate median daily error over the forecast period.

To put KNN forecast accuracy into perspective, we also forecast the load using individual load profiles (ILP). Nine profiles are calculated for each building by averaging all the available historic data for every unique combination of attributes: seasons (winter, summer, transition) and day-type (weekday, Saturday, holiday). Similarly, to SLP approach, the forecast is obtained by using the appropriate ILP (e.g. profile for a holiday in summer). Using the absolute AE notion, we define a relative error that is scale independent, allowing us to compare forecast results among the building as

$$E^{\text{relAE}} = \frac{E_{\text{KNN}}^{\text{AE}}}{E_{\text{ILP}}^{\text{AE}}}. \tag{6}$$

Herewith, $E^{\text{relAE}} < 1$ indicates that KNN forecaster was more accurate than ILP and vice versa.

Forecast results are presented in Table 2. In case of housing, our forecaster and ILP had similar accuracy. However, substantial improvement is achieved for other buildings. KNN forecaster turned out to be 7 and 23% more precise than ILP used for the dorm and the school, respectively.

To illustrate these results, we exemplify the forecast and actual load curves in Fig. 3. In case of the housing and the dorm, KNN forecaster has its biggest deviations during the morning/midday where overall building energy consumption is substantially dependent on PV generation of the building. Therefore, our forecaster can be further improved by explicitly considering PV generation forecast for the next day – something we will address in our future work.

5 Conclusion

In this study, we have applied a forecaster based on the KNN technique to predict daily load curves of three different smart buildings participating in SCDA project, achieving convincing forecast accuracy. Without any manual setup or parametrisation, our forecaster turned out to be significantly more accurate than the forecast done using individual load profiles obtained for each building. However, we have seen that our forecaster can be further improved by explicitly considering the PV generation, electrical heating and ventilation systems as well as other load flexibilities installed in the building.

In the future research, we will place our forecaster in wider context, comparing it with others state-of-the-art forecasters for buildings. We will, also, extend our model to account for the PV-generation and scheduled usage of load flexibilities of the building, so that it can become a powerful tool for predicting electricity consumption of the smart buildings.

7 References

1 Baharlouei, Z., Hashemi, M.: 'Demand Side Management challenges in smart grid: a review'. Smart Grid Conf. (SGC), 2013, IEEE, 2013
2 Von Roon, S., Gruber, A., Buber, T., et al.: 'Das Smart Meter Pilotprojekt SM500-Einsparpotenziale, Nachhaltigkeit und weiterer energiewirtschaftlicher Nutzen' (Internationale Energiewirtschaftstagung an der TU Wien, Wien, 2013)
3 Baranek, D., Probst, A., Tenbohlen, S.: 'Optimierung der lastprognose mittels smart meter daten'. IEEE PES Bielefeld, 2013, 2013
4 Alfares, H.K., Nazeeruddin, M.: 'Electric load forecasting: literature survey and classification of methods', Int. J. Syst. Sci., 2002, 33, (1), pp. 23–34
5 Sevlian, R., Rajagopal, R.: 'Short term electricity load forecasting on varying levels of aggregation'. 2014, arXiv preprint arXiv:1404.0058
6 Kumar, R., Aggarwal, R.K., Sharma, J.D.: 'Energy analysis of a building using artificial neural network: a review', Energy Build., 2013, 65, pp. 352–358
7 Ahmad, A.S., Hassan, M.Y., Abdullah, M.P., et al.: 'A review on applications of ANN and SVM for building electrical energy consumption forecasting', Renew. Sust. Energy Rev., 2014, 33, pp. 102–109
8 Jetcheva, J.G., Majidpour, M., Chen, W.-P.: 'Neural network model ensembles for building-level electricity load forecasts', Energy Build., 2014, 84, pp. 214–223
9 Valgaev, O., Kupzog, F., Schmeck, H.: 'Low-voltage power demand forecasting using K-nearest neighbors approach'. Innovative Smart Grid Technologies – Asia (ISGT – Asia), 2016 IEEE, IEEE, 2016
10 Smith, B.L., Williams, B.M., Keith Oswald, R.: 'Comparison of parametric and nonparametric models for traffic flow forecasting', Transport. Res. Part C: Emerg. Tech., 2002, 10, (4), pp. 303–321
11 Aspern Smart City Research, 2017, Available at http://www.ascr.at/en
12 Arlot, S., Celisse, A.: 'A survey of cross-validation procedures for model selection', Stat. Surv., 2010, 4, pp. 40–79
13 Haben, S., Ward, J., Greetham, D.V., et al.: 'A new error measure for forecasts of household-level, high resolution electrical energy consumption', Int. J. Forecast., 2014, 30, (2), pp. 246–256
14 Charlton, N., Greetham, D.V., Singleton, C.: 'Graph-based algorithms for comparison and prediction of household-level energy use profiles'. 2013 IEEE Int. Workshop on Intelligent Energy Systems (IWIES), IEEE, 2013
15 Hyndman, R.J., Koehler, A.B.: 'Another look at measures of forecast accuracy', Int. J. Forecast., 2006, 22, (4), pp. 679–688