

# A Low-complexity Non-intrusive Approach to predict the Energy Demand of Buildings over Short-term Horizons

Athanasios Aris Panagopoulos<sup>a</sup>, Filippos Christianos<sup>bc</sup>, Michail Katsigiannis<sup>e</sup>, Konstantinos Mykoniatis<sup>e</sup>, Marco Pritoni<sup>f</sup>, Orestis P. Panagopoulos<sup>g</sup>, Therese Peffer<sup>h</sup>, Georgios Chalkiadakis<sup>d</sup>, David E. Culler<sup>h</sup>, Nicholas R. Jennings<sup>i</sup> and Timothy Lipman<sup>h</sup>

<sup>a</sup>California State University, Fresno, USA, <sup>b</sup>University of Edinburgh, UK, <sup>c</sup>Heriot-Watt University, UK, <sup>d</sup>Technical University of Crete, Greece, <sup>e</sup>Auburn University, Alabama, USA, <sup>f</sup>Lawrence Berkeley National Laboratory, USA, <sup>g</sup>California State University, Stanislaus, USA, <sup>h</sup>UC Berkeley, USA. <sup>i</sup>Imperial College, UK

## ARTICLE HISTORY

Compiled December 9, 2020

## ABSTRACT

Reliable, non-intrusive, short-term (of up to 12 hours ahead) prediction of a building's energy demand is a critical component of intelligent energy management applications. A number of such approaches have been proposed over time, utilizing various statistical and, more recently, machine learning techniques, such as decision trees, neural networks and support vector machines. Importantly, all of these works barely outperform simple seasonal auto-regressive integrated moving average models, while their complexity is significantly higher. In this work, we propose a novel low-complexity non-intrusive approach that improves the predictive accuracy of the state-of-the-art by up to ~10%. The backbone of our approach is a K-nearest neighbours search method, that exploits the demand pattern of the most similar historical days, and incorporates appropriate time-series pre-processing and easing. In the context of this work, we evaluate our approach against state-of-the-art methods and provide insights on their performance.

## KEYWORDS

Energy Consumption; Forecasting; Smart Buildings

## 1. Introduction

Transitioning to an energy sustainable future is one of the greatest challenges of our era. Intelligent energy management applications, such as smart thermostats (e.g., (Hilliard, Kavgic, & Swan, 2016; A. Panagopoulos, Maleki, Rogers, Venanzi, & Jennings, 2017)), smart electric vehicle charging stations (Liu, McNamara, Shorten, & McLoone, 2015), and smart solar tracking systems (A. A. Panagopoulos, Chalkiadakis, & Jennings, 2015), have been heralded as key means to achieve this goal (Dounis, 2010; Ramchurn, Vytelingum, Rogers, & Jennings, 2012). This is supported by the fact that such applications introduce energy efficiency improvements with minimum retrofitting

---

CONTACT A. A. Panagopoulos. Address: Science II Building 248, 2576 E. San Ramon Avenue, M/S ST90 Fresno CA 93740. Email: [apanagopoulos@outlook.com](mailto:apanagopoulos@outlook.com)

and installation cost. For instance, a smart thermostat can preheat a space to maximally exploit potential renewable energy availability and, hence, minimize the reliance on dispatchable energy generation with a potentially higher carbon footprint. Such energy management efficiency improvements can come with less cost when compared to improving insulation or upgrading the heating technologies employed (A. Panagopoulos, Alam, Rogers, & Jennings, 2015).<sup>1</sup>

That said, most intelligent energy management applications require reliable short-term (of up to 12 hours ahead) predictions of the demand of the non-controllable loads (i.e., loads that are typically not scheduled in the context of demand side management schemes (Hemanth, Charles Raja, Jeslin Drusila Nesamalar, & Senthil Kumar, 2020; Ramchurn et al., 2012), such as lights and hair dryers.) to perform adequately (e.g., (A. Panagopoulos et al., 2015; Zhao & Magoulès, 2012)). For instance, if in the previous example, the rest of the consumption of the building surpasses the renewable generation during the preheat period, then the energy used for preheating would come from the dispatchable energy generation. This renders preheating inefficient in this case.<sup>2</sup> Nevertheless, when a smart thermostat is aware of the future consumption of the non-controllable loads of a building, it can take this information into account. As such, reliable short-term predictions of the energy demand of non-controllable loads of a building is a crucial component in such settings.

In this context, a number of approaches to predict the energy demand of buildings over short-term horizons have been proposed (Amasyali & El-Gohary, 2018). The majority of these approaches are highly intrusive; they rely on the installation and calibration of sensor networks to monitor demand at various scales, such as at a per-appliance one. However, demand prediction should require minimum retrofitting and installations in order to efficiently support intelligent energy management applications. In any other case, the competitive advantage of energy management applications over deep retrofitting, such as improving the building insulation and installing advanced energy technologies, is, subsequently, also deteriorated. As such, intrusive approaches have narrower practical value in this respect when compared to non-intrusive ones. Acknowledging this fact, several non-intrusive approaches have also been proposed over time, utilizing various statistical or machine learning techniques, such as decision trees, neural networks and support vector machines (SVMs) (e.g., (Amasyali & El-Gohary, 2018; Edwards, New, & Parker, 2012; O. P. Panagopoulos, Pappu, Xanthopoulos, & Pardalos, 2016; O. P. Panagopoulos, Xanthopoulos, Razzaghi, & Şeref, 2019; Pappu, Panagopoulos, Xanthopoulos, & Pardalos, 2015; Tso & Yau, 2007)). Importantly though, all of these works barely outperform a simple seasonal auto-regressive integrated moving average (SARIMA) model (Amjady, 2001; Dagnely, Ruelle, Tourwé, Tsiporkova, & Verhelst, 2015), while their complexity is significantly higher (Dagnely et al., 2015; Edwards et al., 2012). In more detail, a SARIMA model is a straightforward extension of simple auto-regressive integrated moving average models for time series analysis to consider seasonality (Box, Jenkins, Reinsel, & Ljung, 2015). It has

---

<sup>1</sup>Such energy efficiency-improving approaches have long been investigated in the context of supervisory heating, ventilation and air conditioning (HVAC) control, due to their low cost compared to insulation and heating technology improvements (Wang & Ma, 2008). For instance, replacing an old thermostat with a smart one can be less costly than improving the insulation of a building, and has substantial energy efficiency improvement potential (Wang & Ma, 2008).

<sup>2</sup>In particular, preheating generally requires additional energy when compared to “on-time” heating strategies. This is the case, as more energy is needed to get the temperature above the desired level—to then let it drop to the desired one. However, the energy used for preheating can potentially be cheaper, such as energy coming from self-owned renewable generators. This, in turn, can lead to lower energy bills. Nevertheless, this is not the case in this setting and, as such, preheating would only increase the overall cost.

been repeatedly shown to outperform far more complex machine learning and statistical approaches in this task, including neural networks, regression models, and genetic algorithms (Blázquez-García, Conde, Milo, Sánchez, & Barrio, 2020; Dagnely et al., 2015; Edwards et al., 2012). Notably, machine learning approaches can outperform more straightforward statistical approaches when very-short-term horizon predictions are considered (of up to 60 minutes ahead), e.g., Li, Ding, Zhao, Yi, and Zhang (2017).

Against this background, we propose a novel low-complexity non-intrusive energy demand prediction approach suitable for short-term horizons. Our method utilizes a K-nearest neighbors approach and incorporates appropriate time-series pre-processing and easing (i.e., the iterative application of an appropriate function to smooth out the transition from observed consumption towards the predicted one, as detailed in Section 3). We provide a thorough evaluation to show that a simple K-nearest neighbors approach outperforms the benchmark SARIMA model. Furthermore, we show that our improved K-nearest neighbors approach (that incorporates appropriate time-series pre-processing and easing) further improves the predictive accuracy. In particular, we show that our improved approach improves the predictive accuracy by up to  $\sim 10\%$  when compared to the SARIMA model. As discussed, the SARIMA model has been shown to outperform far more complex statistical and machine learning approaches in this task (Dagnely et al., 2015; Edwards et al., 2012). As such, this considers a carefully selected benchmark that enables us to indirectly compare our approach with far more complex approaches. By doing so, in this work we also provide insights into the comparative effectiveness of different approaches in this domain. To sum up, our main contributions to the state-of-the-art are the following:

- We propose a novel low-complexity non-intrusive approach that utilizes K-nearest neighbors and incorporates appropriate time-series preprocessing and easing.
- We evaluate our approach against the state-of-the-art to show that it can improve the predictive accuracy by up to  $\sim 10\%$ .

The remainder of this paper is structured as follows: First, in Section 2 we provide an overview of background material and related work. In Section 3 we detail our approach. Then, in Section 4, we provide a thorough evaluation of our approach (also detailing our case study and the instantiation specifications). Finally, Section 5 concludes.

## 2. Background & Related work

Predicting energy demand is tightly connected with the task of predicting energy consumption (as for small enough intervals one approximates the other). As such, here we review the work on both energy demand and consumption prediction over short-term horizons. In general, existing energy demand and consumption prediction approaches can be mainly classified into either intrusive or non-intrusive, with respect to their reliance on a distributed sensing network or a single sensing point, respectively (Ridi, Gisler, & Hennebert, 2014). The former utilize an extensive sensor network to predict demand using numerous indicative signals, such as occupancy, indoor temperature and/or sub-metering demand values (Edwards et al., 2012; Ekici & Aksoy, 2009; Faisal et al., 2019; Shabani & Zavalani, 2017; Zhao & Magoulès, 2012). Since the additional information can significantly enhance their efficiency, most work on energy consumption prediction proposed in the literature is generally intrusive (Amasyali & El-Gohary, 2018).

For instance, the work of Solomon, Winter, Boulanger, Anderson, and Wu (2011), uses SVMs to provide hourly predictions of the energy demand of a building with the aim to improve the efficiency of the HVAC system. In this context, it utilizes temperature, dew point temperature, pressure, wind direction, wind speed, humidity, and monthly precipitation data to achieve a 0.71 - 0.95 regression coefficient. Furthermore, the work of Edwards et al. (2012), compares a wide range of machine learning techniques, including different artificial neural network architectures and SVMs to provide reliable hourly predictions of building energy hourly consumption. To this end, it utilizes 140 different input signals, including temperature and solar flux. In a similar line of research, the work of Kwok and Lee (2011) trained a neural network to predict the hourly building energy consumption pattern utilizing a number of input signals including occupancy area and occupancy rate to achieve a  $\sim 15\%$  root mean square percentage error and a  $\sim 0.98$  coefficient of correlation. Finally, the work of Farzana, Liu, Baldwin, and Hossain (2014) trained a neural network to predict the yearly energy consumption of residential buildings of a city in order to support effective supply-side management applications. By doing so, it achieved an impressive 0.09% mean relative prediction error and a  $\sim 0.97$  coefficient of determination, utilizing several detailed inputs including the number and type of lightning bulbs per household, and the space heating and cooling energy use of fuel.

Now, although such additional information can enhance prediction efficiency, the required extensive instrumentation is typically not available in most buildings today (Sioshansi, 2016). By contrast, non-intrusive approaches, rely solely on aggregated consumption data and, hence, can be readily deployed in domestic settings today by merely monitoring the existing house metering infrastructure. For this reason, in this work we focus on non-intrusive approaches.

Many non-intrusive prediction approaches have been proposed over time, utilizing various statistical and, more recently, or machine learning techniques, such as decision trees, neural networks, and support vector machines (e.g., Edwards et al. (2012); Shabani and Zavalani (2017); Tso and Yau (2007); Zhao and Magoulès (2012)). For instance, the work of Yun, Luck, Mago, and Cho (2012) proposes a statistical autoregressive model to predict the hourly consumption of the HVAC system of residential buildings taking into account only the previous consumption to achieve a  $\sim 20\%$  coefficient of variation. However, this work does not consider the cumulative consumption of the building; it solely focuses on HVAC consumption. Nevertheless, predicting the non-controllable loads is needed to support intelligent applications as discussed in Section 1. Furthermore, the work of Dagnely et al. (2015) implements three autoregressive baseline models to predict the hourly energy consumption of a building at time  $t$  considering the hourly energy consumption at times  $t - 1$  day,  $t - 7$  days and  $t - 14$  days respectively. This selection is supported by the fact that these periods consider the strongest auto-correlation lags in the data-set. By doing so, they achieve a performance similar to more advanced intrusive statistical approaches (i.e., support vector regression and ordinary least squares on several signals including occupancy and irradiance).

Machine learning approaches have been shown to outperform simpler statistical approaches when very-short horizon predictions are considered (of up to 60 minutes ahead), e.g., Li et al. (2017). Importantly, however, when longer but still short-term (up to 12 hours) non-intrusive predictions are considered, advanced machine learning approaches have been shown to barely outperform a simple seasonal autoregressive integrated moving average (SARIMA) model (Amjady, 2001; Dagnely et al., 2015), while their complexity is significantly higher (Dagnely et al., 2015; Edwards et al.,

2012). The work in Yang, Meng, Xia, Lu, and Yu (2011) proposes a similarity-based approach which has lower complexity than a SARIMA model. This work, does not come with a concrete evaluation, does not consider time-series pre-processing (that could potentially improve its accuracy), and does not account for the continuous dynamics of house energy demand (as further discussed in Section 3). Nevertheless, it is the most suitable point of departure for our work; it sets the foundations of our approach, and it also considers a benchmark in our evaluation.

For a thorough review of building energy consumption pattern and demand prediction approaches see the work of (Amasyali & El-Gohary, 2018).

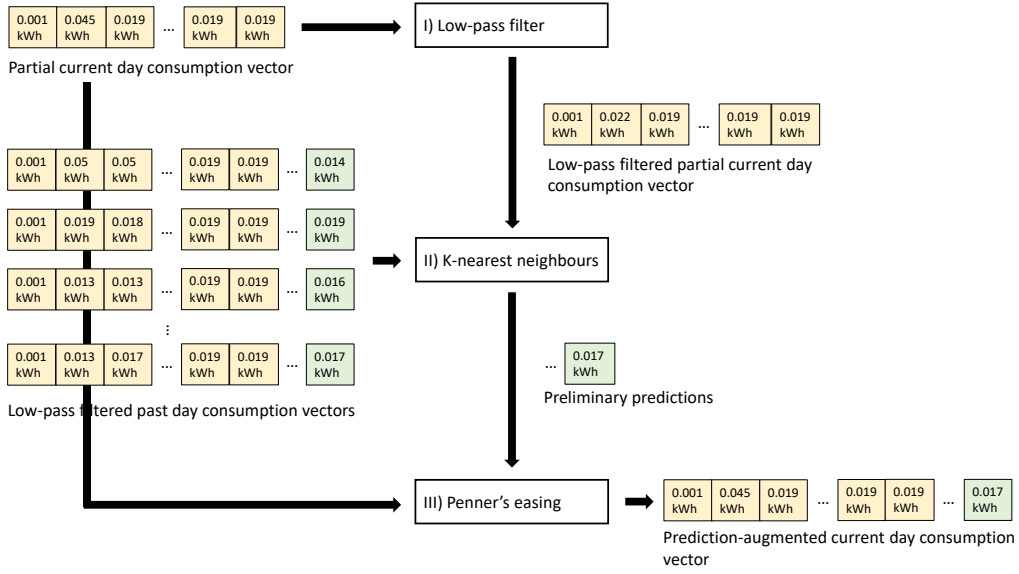
### 3. Our approach

As discussed in Section 1, in this work we present a non-intrusive *improved similarity-based approach* to predict the energy demand of buildings over a short-term predictive horizon. The backbone of our approach is a K-nearest neighbours search method that predicts the demand pattern of the building using the pattern of the K most similar historical days (i.e., the K-nearest neighbours). Essentially reflecting the household activity, the energy demand pattern is expected to demonstrate periodicity and to be similar among days. As such, an approach that relies on identifying the most similar days, in order to subsequently utilize them, in the prediction process of energy demand is expected to be highly effective. K-nearest neighbours (Bishop, 2006) is a straightforward choice in this respect. In addition, K-nearest neighbours is a far less complex approach when compared to advanced machine learning approaches, such as neural networks and support vector machines (Bishop, 2006). This, in turn, enables its utilization in systems with restricted computational and memory resources, such as those available in residential, as well as in small and average size commercial buildings (Fierro & Culler, 2015).

Now, we further improve on this approach by incorporating appropriate *pre-processing and easing*. In more detail, we *pre-process* our time-series utilizing a low-pass-filter to distinguish the base consumption from the spikes to predict the base consumption alone. This pre-processing is based on the assumption that a smooth base demand is easier to be predicted than a signal that includes highly dynamic energy spikes. Furthermore, our approach incorporates *easing* to account for the continuous dynamics of the energy demand. This is done to maximally account for the most recent time series observations in our prediction model and ease them as we transition to our predictions. By doing so, we propose a similarity based approach that has higher predictive accuracy than the state-of-the-art, as further discussed in the evaluation section. with low complexity.

Importantly, our approach is suitable for predicting both energy demand as well as small-interval consumption patterns (which can approximate the energy demand) over short-term horizons. In more detail, our approach predicts the future demand and/or low-resolution consumption pattern given historical demand samples or low-resolution consumption data, respectively. In the following paragraph we present our approach, predicting minute-wise-consumption as an approximation of the demand (demand samples can be used instead in a straightforward manner). Our approach follows a 3-step procedure (as illustrated in Figure 1):

- I Calculate a baseline vector by low-pass filtering the partial day consumption vector,  $\mathbf{c}$ . In particular,  $\mathbf{c}$  considers energy consumption values for every minute from



**Figure 1.** Flowchart of our improved similarity-based approach. The partial current day consumption vector is first passed through a low-pass filter. Subsequently, the  $K$  nearest past days (and, in particular, the  $K$  nearest week-days for a week-day or the  $K$  nearest weekend-days for a weekend-day) are selected, and the predictions are estimated. Finally, the predictions are refined using a series of Penner’s easing transformations and the partial current day vector.

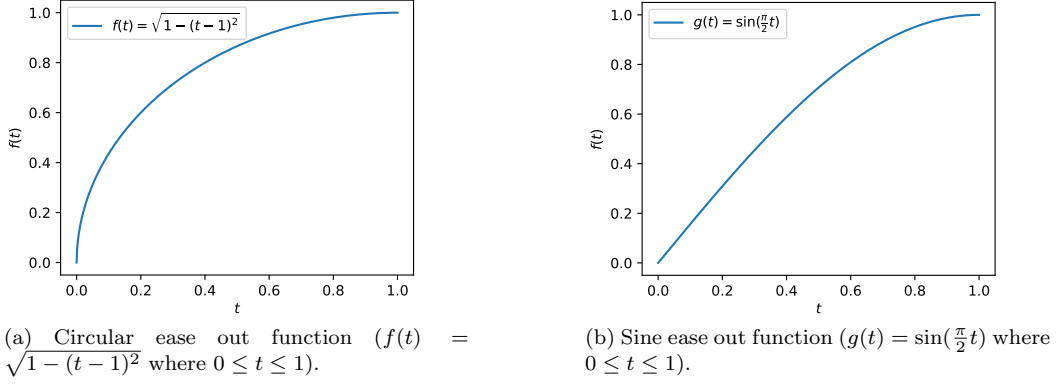
the beginning of the day until the current time (please note that demand samples can be used instead to directly predict demand if those are available as discussed above). Note that this vector extends to 3.5 hours in the past in order for our approach to perform adequately at the beginning of a day. Now,  $\mathbf{c}$  is passed through a centered moving average filter of  $\lambda$  minutes. This acts as a low-pass filter, estimating the base consumption pattern.  $\lambda$  considers an adjustable parameter.

- II Find the  $K$  most similar *filtered* historical days using the K-nearest neighbours based on the euclidean distance of the two vectors (where  $K$  is an integer). Notably, our approach distinguishes week from weekend days, i.e., predictions for week and weekend days are computed only from past week and weekend days, respectively (since the household activity pattern is generally quite distinct between these two). The average of the remainder of these filtered days is computed to estimate a future consumption vector,  $\hat{\mathbf{c}}$ .<sup>3</sup>  $K$  is also an adjustable parameter.
- III A series of easing steps are performed to smooth out the transition from the observed consumption towards the predicted one and derive the final prediction vector,  $\mathbf{c}'$ . Inspired by techniques used in animation design, we use Penner’s easing functions (Izdebski & Sawicki, 2016) and, in particular, the circular ease out function, i.e.,  $f(t) = \sqrt{1 - (t - 1)^2}$ ; and the ease out sine function, i.e.,  $g(t) = \sin(\frac{\pi}{2}t)$ ; where  $t \in [0, 1]$ .<sup>4</sup> The Penner’s easing functions used are illustrated in figures 2a and 2b, respectively.

In more detail, the circular ease out function is used to ease out the last N-

<sup>3</sup>We note that the remainder of the day can be appropriately augmented according to the predictive horizon needs of the application.

<sup>4</sup>This easing technique has been selected after experimenting with various other easing techniques and considering its low complexity (when compared, for instance, with Gaussian-processes-based easing (A. Panagopoulos et al., 2017)).



**Figure 2.** Penner's easing functions utilized.

minute average consumption  $m = \frac{1}{N} \sum_{i=|\mathbf{c}|-N}^{|\mathbf{c}|} \mathbf{c}_i$ , where  $N \leq |\mathbf{c}|$  and  $|\mathbf{c}|$  denotes the size of  $\mathbf{c}$ . The ease out sine function (operating on top of the circular ease out one) is used to ease out the last observed consumption value  $\mathbf{c}_{|\mathbf{c}|}$ . More formally, given an initial house consumption prediction during interval  $\tau$  of length  $\delta$ ,  $\hat{\mathbf{c}}_\tau$ , the final eased-out prediction  $\mathbf{c}'_\tau$  is calculated as:

$$\mathbf{c}'_\tau = \int_{t_\tau}^{t_\tau + \delta} \begin{cases} \mathbf{c}_{|\mathbf{c}|} - g\left(\frac{t-t^0}{t^a-t^0}\right)((\mathbf{c}_{|\mathbf{c}|} + m) + \\ + f\left(\frac{t-t^0}{t^b-t^0}\right)(m + \hat{\mathbf{c}}_\tau)) & t^0 \leq t < t^a \\ m - f\left(\frac{t-t^0}{t^b-t^0}\right)(m + \hat{\mathbf{c}}_\tau) & t^a \leq t < t^b \\ \hat{\mathbf{c}}_\tau & t^b \leq t \end{cases} dt$$

where  $t^0$  is the time at the beginning of the predictive horizon;  $t_\tau$  is the beginning time of interval  $\tau$ ;  $t^b$  is the time until when  $f(t)$  is used to ease out  $m$ ; and  $t^a$  is the time until when  $g(t)$  is used to ease out  $\mathbf{c}_{|\mathbf{c}|}$ . As such,  $0 \leq t^0 \leq t_\tau$  and  $t^0 < t^a < t^b$ ; and  $N$ ,  $t^a$  and  $t^b$  are also adjustable parameters to be calibrated.<sup>5</sup>

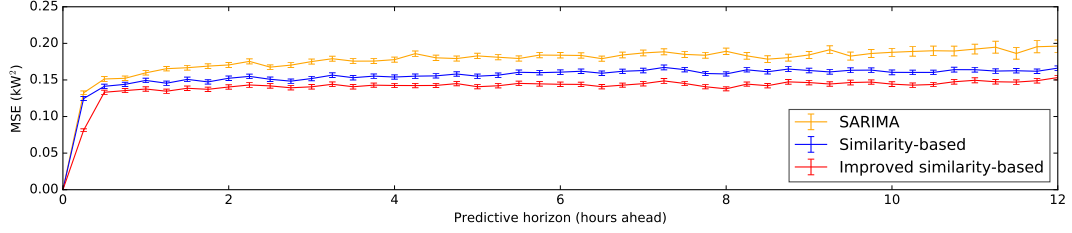
Given the above steps, our approach has a number of adjustable parameters, i.e.,  $\lambda$ ,  $K$ ,  $N$ ,  $t^a$  and  $t^b$  which can be selected with various means, such as via a targeted initial grid search using a dedicated training-set. In the following sections we detail our evaluation of the proposed approach and show its efficiency. There we also provide an instantiation of our approach for the case study considered.

#### 4. Evaluation

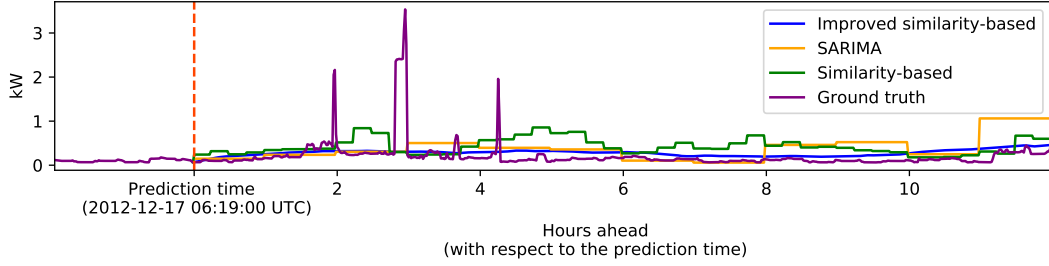
In this section we provide an evaluation of the proposed approach and show its efficiency. In the following paragraphs, we first detail our case study and how we instantiated our approach. Subsequently we discuss the evaluation results.

---

<sup>5</sup>A discussion on adjusting these parameters is provided in Section 4 where, essentially, a grid search approach is followed to instantiate our approach for our case study data-set.



**Figure 3.** Short-term (up to 12 hours ahead) house demand prediction evaluation results. The bars correspond to the standard error of the mean.



**Figure 4.** Short-term (up to 12 hours ahead) predictions for a typical day weekday based on SARIMA, Similarity-based and Improved similarity based methods.

#### 4.1. Case Study

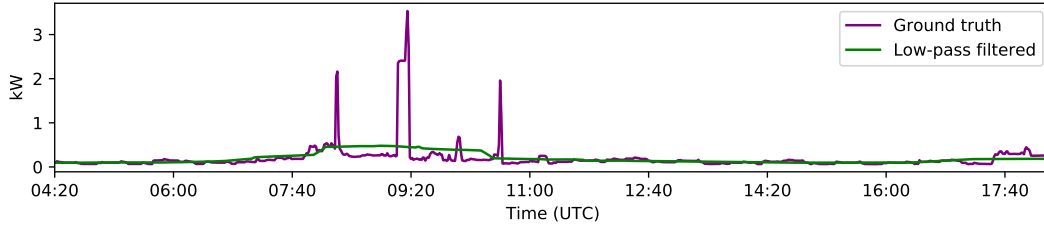
For our evaluation case study we consider 8 months of house consumption data (i.e., [6/2012-2/2013]) from the publicly available Electricity Consumption & Occupancy (ECO) data-set (Beckel, Kleiminger, Cicchetti, Staake, & Santini, 2014) of a typical detached, average-sized family house in Switzerland. The most suitable values for the adjustable parameters of our approach, i.e.,  $t^a$  and  $t^b$ ,  $\lambda$ ,  $K$  and  $N$  (defined above), were identified using 1 month of consumption data, as this amount has been shown to be adequate to this end. Based on a grid search we concluded on:  $\lambda=140m$ ,  $K=9$   $N=250m$ ,  $t^a=25m$  and  $t^b=250m$ . As a rule of thumb, we identified that smoothing intervals that are  $\sim 10$  times greater than the average spike intervals (which are typically  $\sim 15$  minutes in our data-set) perform best. Notably, this (training) month is excluded from our evaluation (testing) data-set.

Finally, we evaluate the approaches considered using  $\sim 30,000$  (uniformly drawn) instances within our training data-set and making predictions for all 15 minute intervals in the future for 12 hours ahead. Subsequently, the mean square error (MSE), for each prediction interval is calculated:

$$\mathbf{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

where  $n$  is the number of data points,  $Y$  is the vector of observed values of the variable being predicted, and  $\hat{Y}_i$  are the predicted values.





**Figure 5.** Illustration of the low-pass filter pre-processing effect.

#### 4.2. Evaluation results

In order to evaluate our improved similarity-based demand prediction approach, we compare against a simpler approach that does not include time-series preprocessing or easing, i.e., the simple similarity-based approach in (Yang et al., 2011), and a seasonal auto-regressive integrated moving average (SARIMA) model (Dagnely et al., 2015). Importantly, the latter performs equally well or better when compared to far more advanced machine learning approaches, as discussed in Section 2 (Dagnely et al., 2015). At the same time its complexity is significantly lower when compared to advanced machine learning approaches (Dagnely et al., 2015; Edwards et al., 2012). As such, evaluating against this approach allows us to get an understanding of where we stand with respect to the state-of-the-art. Given that our approach is also of low complexity, this is also a fair comparison from such a perspective. Figure 3 collectively reports our evaluation results, where the bars correspond to the standard error of the mean. Interestingly, the simple similarity-based approach outperforms the more-complex SARIMA model. This improvement is statistically significant as further confirmed via t-tests at the 0.05 significance level. Furthermore, our approach also improves on the simple similarity-based one in a statistically significant manner (confirmed by t-tests at a 0.05 significance level). Specifically, the predictive accuracy is improved by  $\sim 10\%$ . This confirms that filtering and easing can improve the predictive accuracy of similarity-based approaches.

In more detail, the improved similarity-based approach has a considerably lower MSE when very-short predictive horizons are considered. This improvement is attributed to easing, which appropriately takes into account the observed data. That said, the improved similarity-based approach further and consistently outperforms the simple similarity-based approach when all predictive horizon intervals are considered (in a statistically significant manner at the 0.05 significance level). Further investigation suggests that  $\sim 90\%$  of the overall improvement of the improved-similarity-based approach is due to time-series preprocessing and specifically due to the low-pass filtering. Easing seems to attribute for the rest smaller fraction of  $\sim 10\%$ .

To expand on the points raised above, Figure 4. focuses on short-term predictions for the SARIMA, similarity-based and improved similarity-based approaches for a typical weekday (a Monday) from our dataset. As can be seen, both SARIMA and the basic similarity-based approach predict demand peaks that overestimate the demand. These high peaks aim to predict demand spikes. However demand spikes are highly dynamic and easy to mis-predict, deteriorating the overall predictive accuracy. By contrast, our improved similarity-based approach is less sensitive to this effect due to the smoothing of demand spikes through appropriate low-pass filtering (as discussed in Section 2). The effects of low pass filtering for our typical day are also illustrated in Figure 5.

Figure 5 shows how the demand spikes are appropriately filtered resulting in a smooth "base" demand curve.

To conclude, Dagnely et al. selected the title "Predicting hourly energy consumption. Can you beat an auto-regressive model" for their work in Dagnely et al. (2015). By so doing, they highlighted the primary outcome of their line of research, namely that a simple auto-regressive model outperforms far more complex approaches (such as neural networks and support vector machines) in predicting the energy consumption of a building over short-term horizons when non-intrusive approaches are considered. Our results here (based on real data) suggest that an arguably even simpler similarity-based approach can outperform the auto-regressive model (i.e., SARIMA) proposed by Dagnely et al. (2015) in this task while appropriate preprocessing and easing can enhance its performance even further.

## 5. Conclusions

In this work, we proposed a novel non-intrusive approach to predict the energy demand of a building over a short-term prediction horizon (of up to 12 hours ahead). Our approach relies on a K-nearest neighbours search and incorporates time-series pre-processing filtering and easing. We provide a thorough evaluation of our approach to show that it outperforms a simple similarity-based approach as well as a SARIMA model (by up to  $\sim 10\%$ ). Importantly, since the SARIMA model outperforms far more complex statistical or machine learning approaches in this task, the fact that our approach outperforms SARIMA provides implicit support to the strength of our method.

Imminent future work includes incorporating our energy demand prediction approach in an advanced intelligent electric vehicle charging system. Moreover, since machine learning approaches have been shown to perform adequately when very-short-term prediction horizons are considered (as discussed in the Background section), we intend to incorporate them in our approach to enhance further the transition from real data to the initial predictions. Furthermore, our approach relies on several adjustable parameters which are calibrated during an initial calibration session. Defining a dedicated procedure to re-calibrate these parameters as new data become available, could potentially enhance the reliability of our approach by enabling the adaptation to (seasonal or other arbitrary) changes in the demand pattern. Therefore, developing an adaptive improved-similarity-based approach is an additional and promising work direction. Lastly, at this time our approach distinguishes weekend from week days. This can be extended to also account for vacation days. Investigating to what extent these enhancements improve the efficiency of the algorithm is an additional research direction.

## References

- Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, 81, 1192 - 1205. Retrieved from <http://www.sciencedirect.com/science/article/pii/S1364032117306093>
- Amjady, N. (2001, 08). Short-term hourly load forecasting using time-series modeling with peak load estimation capability. *IEEE Transactions on Power Systems*, 16(3), 498-505.

- Beckel, C., Kleiminger, W., Cicchetti, R., Staake, T., & Santini, S. (2014). The eco data set and the performance of non-intrusive load monitoring algorithms. In *Proceedings of the 1st acm conference on embedded systems for energy-efficient buildings* (pp. 80–89).
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. springer.
- Blázquez-García, A., Conde, A., Milo, A., Sánchez, R., & Barrio, I. (2020). Short-term office building elevator energy consumption forecast using sarima. *Journal of Building Performance Simulation*, 13(1), 69–78.
- Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
- Dagnely, P., Ruelle, T., Tourwé, T., Tsiporkova, E., & Verhelst, C. (2015). Predicting hourly energy consumption. can you beat an autoregressive model. In *Proceeding of the 24th annual machine learning conference of belgium and the netherlands, benelearn, delft, the netherlands* (Vol. 19).
- Dounis, A. I. (2010). Artificial intelligence for energy conservation in buildings. *Advances in Building Energy Research*, 4(1), 267–299.
- Edwards, R. E., New, J., & Parker, L. E. (2012). Predicting future hourly residential electrical consumption: A machine learning case study. *Energy and Buildings*, 49, 591–603.
- Ekici, B. B., & Aksoy, U. T. (2009). Prediction of building energy consumption by using artificial neural networks. *Advances in Engineering Software*, 40(5), 356 - 362. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0965997808001105>
- Faisal, H. M., Javaid, N., Sarfraz, B., Baqi, A., Bilal, M., Haider, I., & Shuja, S. M. (2019). Prediction of building energy consumption using enhance convolutional neural network. In L. Barolli, M. Takizawa, F. Xhafa, & T. Enokido (Eds.), *Web, artificial intelligence and network applications* (pp. 1157–1168). Cham: Springer International Publishing.
- Farzana, S., Liu, M., Baldwin, A., & Hossain, M. U. (2014). Multi-model prediction and simulation of residential building energy in urban areas of chongqing, south west china. *Energy and Buildings*, 81, 161 - 169. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0378778814004897>
- Fierro, G., & Culler, D. E. (2015). Xbos: An extensible building operating system. In *Proceedings of the 2nd acm international conference on embedded systems for energy-efficient built environments* (pp. 119–120).
- Hemanth, G., Charles Raja, S., Jeslin Drusila Nesamalar, J., & Senthil Kumar, J. (2020). Cost effective energy consumption in a residential building by implementing demand side management in the presence of different classes of power loads. *Advances in Building Energy Research*, 1–26.
- Hilliard, T., Kavgić, M., & Swan, L. (2016). Model predictive control for commercial buildings: trends and opportunities. *Advances in Building Energy Research*, 10(2), 172–190.
- Izdebski, L., & Sawicki, D. (2016). Easing functions in the new form based on bézier curves. In *International conference on computer vision and graphics* (pp. 37–48).
- Kwok, S. S., & Lee, E. W. (2011). A study of the importance of occupancy to building cooling load in prediction by intelligent approach. *Energy Conversion and Management*, 52(7), 2555 - 2564. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0196890411000744>
- Li, C., Ding, Z., Zhao, D., Yi, J., & Zhang, G. (2017, 10). Building energy consumption prediction: An extreme deep learning approach. *Energies*, 10, 1525.
- Liu, M., McNamara, P., Shorten, R., & McLoone, S. (2015, 06 01). Residential electrical vehicle charging strategies: the good, the bad and the ugly. *Journal of Modern Power Systems and Clean Energy*, 3(2), 190–202. Retrieved from <https://doi.org/10.1007/s40565-015-0122-2>
- Panagopoulos, A., Alam, M., Rogers, A., & Jennings, N. (2015, 01). Adaheat: A general adaptive intelligent agent for domestic heating control. , 2, 1295-1303.
- Panagopoulos, A., Maleki, S., Rogers, A., Venanzi, M., & Jennings, N. (2017, 05). Advanced economic control of electricity-based space heating systems in domestic coalitions with shared intermittent energy resources. *ACM Transactions on Intelligent Systems and*

- Technology*, 8, 1-27.
- Panagopoulos, A. A., Chalkiadakis, G., & Jennings, N. R. (2015). Towards optimal solar tracking: a dynamic programming approach. In *Twenty-ninth aaai conference on artificial intelligence*.
- Panagopoulos, O. P., Pappu, V., Xanthopoulos, P., & Pardalos, P. M. (2016). Constrained subspace classifier for high dimensional datasets. *Omega*, 59, 40–46.
- Panagopoulos, O. P., Xanthopoulos, P., Razzaghi, T., & Şeref, O. (2019). Relaxed support vector regression. *Annals of Operations Research*, 276(1-2), 191–210.
- Pappu, V., Panagopoulos, O. P., Xanthopoulos, P., & Pardalos, P. M. (2015). Sparse proximal support vector machines for feature selection in high dimensional datasets. *Expert Systems with Applications*, 42(23), 9183–9191.
- Ramchurn, S. D., Vytelingum, P., Rogers, A., & Jennings, N. R. (2012). Putting the 'smarts' into the smart grid: a grand challenge for artificial intelligence. *Communications of the ACM*, 55(4), 86–97.
- Ridi, A., Gisler, C., & Hennebert, J. (2014, 08). A survey on intrusive load monitoring for appliance recognition. In *2014 22nd international conference on pattern recognition* (p. 3702-3707).
- Shabani, A., & Zavalani, O. (2017, 05). Predicting building energy consumption using engineering and data driven approaches: A review. *European Journal of Engineering Research and Science*, 2, 44.
- Sioshansi, F. P. (2016). *Future of utilities - utilities of the future: How technological innovations in distributed energy resources will reshape the electric power sector*. Elsevier Science.
- Solomon, D., Winter, R., Boulanger, A., Anderson, R., & Wu, L. (2011, 01). Forecasting energy demand in large commercial buildings using support vector machine regression.
- Tso, G. K., & Yau, K. K. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, 32(9), 1761–1768.
- Wang, S., & Ma, Z. (2008, 01). Supervisory and optimal control of building hvac systems: A review. *HVAC&R Research*, 14, 3-32.
- Yang, Y., Meng, Y., Xia, Y., Lu, Y., & Yu, H. (2011). An efficient approach for short term load forecasting. In *International multiconference of engineers and computer scientists*.
- Yun, K., Luck, R., Mago, P. J., & Cho, H. (2012). Building hourly thermal load prediction using an indexed arx model. *Energy and Buildings*, 54, 225 - 233. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0378778812003933>
- Zhao, H.-x., & Magoulès, F. (2012). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews*, 16(6), 3586–3592.