Sensitivity analysis of data-driven building energy demand forecasts

Felix Bünning$^{1,2}$, Philipp Heer$^1$, Roy S. Smith$^2$, John Lygeros$^2$

$^1$Empa, Urban Energy Systems Laboratory, Überlandstrasse 129, 8600 Dübendorf, Switzerland
$^2$Automatic Control Laboratory, Department of Electrical Engineering and Information Technology, ETH Zürich, Switzerland
E-mail: felix.buenning@empa.ch

Abstract. Data-driven models of buildings could potentially reduce implementation barriers for demand forecasting and predictive control in the built environment. However, such models appear to be sensitive to the quality of the available input data. Here, we investigate the influence of sampling time, noise level and amount of available measurement data as well as the quality of the weather forecast on a heating demand forecast with online corrected Artificial Neural Networks. Based on a case study, we demonstrate that sampling time has a stronger influence on the prediction performance than noise level and the amount of available data. Furthermore, we show that using measured ambient temperatures for training appears to provide no benefit compared to using weather forecasts.

1. Introduction
Predictive control of building heating and cooling systems has the potential to significantly reduce the energy consumption and as a consequence the CO2 emissions of buildings [1]. Model Predictive Control (MPC) has been successfully demonstrated in many studies [2, 3]. However, as several authors point out [4, 5], the effort required to develop and maintain first principle models of buildings might be prohibitive for large-scale real life deployment of MPC in the building domain.

With increasing availability of high-resolution energy monitoring data in the built environment and the success of data-driven modelling in a variety of domains, the possibility to use these methods for demand forecasting [6] and predictive control [7] has arisen. As the methods are purely data-driven, the question regarding the influence of the quality of the available data on the prediction performance needs to be considered.

In [8] we introduced a method based on Artificial Neural Networks (ANN) and forecast correction methods based on online learning and error auto-correlation to make a day-ahead forecast of the heating demand of a multi-use building in Switzerland in a 15-minute time interval. The method showed superior performance compared to other Machine Learning methods and to a variety of fitted simple resistor-capacitor models. Furthermore, the correction methods allowed a significant reduction of variance in the prediction performance of ANN. In [9] the method was validated on four individual buildings with different uses, availability and quality of measurement data. The correction methods significantly improved the forecast quality and reliability and the approach outperformed a fitted 5R3C resistor-capacitor building model
in all cases. However, the results indicated that the prediction performance is dependent on the quality of the training data.

Here, we investigate the dependence of the prediction performance of ANN with forecast correction methods on the sampling time and the noise of the measurement data, as well as the available amount of data and the quality of the weather forecast that is used as an input in the demand forecast with a case study. Our results indicate that the sampling time of measurements has a bigger impact on prediction quality than noise levels, and that one week of historical training data is sufficient for meaningful demand predictions if online correction methods are applied. Moreover, using ambient temperature measurements for ANN training appears to offer no benefit compared to using weather forecasts.

2. Methodology
2.1. Forecasting task
The following forecasting task is assumed in this study: A heating energy demand forecast of a whole building is made once per day for the following 24 hours, sampled in a 15-minute interval. The measurement data used for training and validation are assumed to be sampled in the same interval. Different weather and time related measurements are available as inputs (see section Case study for more detail). We use ANN with online correction methods based on the error auto-correlation and on online learning for the heating demand forecast. Both methods make use of the persistence of error-inducing disturbances in building energy demand forecasting. For more details on ANN please refer to [10], for more details on the correction methods to [8] or [9].

2.2. Data preparation for varying sampling time
We assume that the building measurement data is sampled in a 15-minute interval. To simulate different sampling times, we sub-sample the data set to 30 minutes, 60 minutes and 120 minutes by taking the mean over all relevant samples. The window to compute the mean starts with the sample under consideration, for example: in the 30-minute sub-sampled set, the data point for 2017-04-01 00:00 is calculated by taking the mean of samples 2017-04-01 00:00 and 2017-04-01 00:15 from the 15-minute data set.

2.3. Data preparation for varying measurement noise
We assume that the reference building measurement data is noise-free. As the measurement tolerance of equipment to measure temperatures and mass flows in buildings is usually given in a maximum percentage of the measured value, we add noise to each sample value $s$ in the measurement data with

$$\bar{s} = s + \alpha \cdot R \cdot s \text{ with } R \in [-1, ..., 1],$$

in which $\bar{s}$ denotes the new sample value, $\alpha$ denotes the desired noise level (accuracy of the measurement equipment) and $R$ is random, independent (for different samples) and identically distributed uniformly between -1 and 1. For $\alpha$, values of 0.1, 0.2 and 0.3 are considered.

2.4. Data preparation for varying data set size
In the base case, the first 70% of the data set is used for training of the ANN and the remaining 30% for validation of the model. In order to ensure comparability between the different test cases, we keep the validation set the same throughout. To simulate limited availability of measurement data, the training set is shrunk from 41 weeks to one week by removing samples from the beginning, such that there is no temporal gap between the training and the validation set.
2.5. Use of weather forecasts
To investigate the effect of the quality of the weather forecast on the heat demand prediction accuracy, we compare prediction results obtained with measured weather data, which would represent a perfect forecast, to results obtained from a weather forecast. For the weather forecast, forecasts of the ambient temperature by MeteoSwiss are used. Linear interpolation is used to generate 15 minute forecast samples from the forecast with hourly resolution.

3. Case study
3.1. General description
To validate the neural networks and the correction methods and to investigate the effect of data quality and quantity, the NEST building at Empa in Switzerland is used as a case study. The building became operational in 2016 and comprises residential units, office units as well as meeting rooms and a fitness center. Measurement data of the whole building’s heating demand over a period of 13.5 months is used. As in [8] and [9], we use the coefficient of determination, \( R^2 \), as a key performance indicator. It is one if the forecast \( f(x_i) \) with inputs \( x_i \) exactly matches the validation data \( y_i \) and zero if the forecast is as good as taking the average \( \bar{y}_N \) of the data in the considered set \( N \). \( R^2 \) can become negative if the forecast is worse than taking the average.

\[
R^2 = 1 - \frac{\sum_{i\in N} (y_i - f(x_i))^2}{\sum_{i\in N} (y_i - \bar{y}_N)^2},
\]

3.2. Specific ANN models for the case study
As in [8] and [9], Artificial Neural Networks with an architecture that comprises two hidden layers, each containing twelve nodes are used. The networks are fitted with ten epochs on the training set, which means that their parameters are updated ten times based on each sample in the set. The chosen parameters for the architecture and the training process were found with a sensitivity analysis. The networks use the features ambient temperature, hour of the day, weekday/weekend, heating load of the previous day and heating load of the previous week as inputs. The feature hour of the day is one-hot encoded, which means that 24 inputs with binary values are used instead of one input with a continuous value. The feature weekday/weekend is also binary while all other features are modelled as continuous inputs. To implement and train the network, the Python package Keras [11] with the solver adam [12] was used. The rationale behind all design choices is discussed in [8].

4. Results
The prediction performance of ANN is dependent on their parameter values, some of which are set randomly. Therefore, 100 instances of ANN were trained and tested for each case to investigate the variance in prediction accuracy.

4.1. Influence of sampling time
Figure 1 shows the prediction performance of the ANN with varying sampling rate of 15 minutes, 30 minutes, 60 minutes and 120 minutes for uncorrected and corrected demand forecasts [8]. The forecasts for 30, 60 and 120 minutes were up-sampled by interpolation and validated against the test set for 15 minutes. In the uncorrected demand forecasts, a decreasing sampling time leads to subsequent lowering of the average coefficient of determination and also to a growing variance. A sampling time of 120 minutes does not lead to meaningful forecasts any more, which indicates that the Nyquist sampling rate is not respected any more. This effect could shift to higher or lower sampling times depending on the building inertia. In the corrected demand forecasts the same trend occurs, but the negative effects of bigger sampling times are lowered. The variance is
significantly decreased in all cases and the relative variance reduction increases with increasing sampling time (factors 4.9, 12.9, 14.5, 14.9 for sampling times 15, 30, 60, 120 subsequently). The average $R^2$ is significantly raised by the correction methods for all sampling times. With an average $R^2$ of 0.845 and an interquartile range of 0.008 a corrected forecast with a sampling time of 30 minutes leads to better results than an uncorrected forecast with 15-minute sampling time ($R^2 = 0.823$ and $IQR = 0.043$).

4.2. Influence of noise level
The results in Figure 2 show that the noise level does have significant influence on the variance in the uncorrected case. However, the average $R^2$ is lowered subsequently for higher noise levels. When online correction methods are applied, the variance appears to be constant for all noise levels. A noise level of 10% slightly increases the prediction performance for the corrected forecasts (average $R^2$ of 0.892 vs. 0.885). Although this might appear surprising, adding noise to data is a common way to avoid overfitting in ANN [13]. Further increase of the noise level reduces the prediction performance. By comparison with the results for varying sampling time, it can be seen that the performance is less sensitive to the noise level than to the sampling time.

4.3. Influence of training data amount
Figure 3 shows the forecast trajectory of an uncorrected ANN and a corrected ANN with a single week of historical data for offline training. While the uncorrected forecast appears to have learnt daily demand fluctuations, it does not catch the influence of the ambient conditions. The corrected forecast performs well from the start. It predicts daily demand fluctuations and shows dependency on the ambient conditions. After one month of operation, the performance does not differ from a model trained with 41 weeks of historical data. The achieved $R^2$ of the
Figure 3: Forecast trajectory for test set with both correction methods applied with one week of offline training

shown corrected forecast is 0.860 (the average for 41 weeks of training is 0.885). The shown trajectory is not a ‘best case’ example and the performance is repeatable.

4.4. Influence of weather forecast

Figure 4: Variance of the coefficient of determination for different uses of weather forecasts with (a) uncorrected demand forecasts, and (b) corrected forecasts. Legend and box-plot definition as in Figure 1.

Figure 4 shows the prediction performance of the networks for (a) uncorrected demand forecasts and (b) corrected demand forecasts with different inputs used in the training and testing of the ANN: measured ambient temperature used both for training and testing (perfect weather forecast), weather forecasts used for testing a model trained on measured ambient temperature, and weather forecasts used for both training and testing. The variance of the uncorrected ANN increases significantly when weather forecasts are used for forecasting but not for training. This is not the case if forecasts are used for both training and testing. However, the correction methods bring the variance to a constant level as can be seen in (b). The average $R^2$ reduces
from 0.885 to 0.866 and 0.863 when online correction methods are used. The comparison of networks that use ambient temperature measurements for training with those that use weather forecasts shows that the former appears to offer no significant benefit over the latter.

5. Conclusion

The performance sensitivity of heating demand forecasts with ANN and online correction methods was investigated, in particular with respect to sampling time, noise level, amount of data, and accuracy of weather forecast. In a case study, it was found that the sampling time of the inputs has significant impact. An input noise level of 10% increases prediction performance when online correction methods are used, whereas higher noise levels lead to reduced performance. Furthermore, one week of historical data suffices to train reliable ANN for forecasting in the demonstrated case. Moreover, using ambient temperature measurements for training of the ANN appears to offer little benefit compared to using weather forecasts directly for training when online correction methods are applied.

Acknowledgements

This research project is financially supported by the Swiss Innovation Agency Innosuisse and is part of the Swiss Competence Center for Energy Research SCCER FEEB&D. We would like to thank Kristina Orehouning and Viktor Dorer for their valuable help and support. We are also grateful to Ahmed Aboudonia, Andrew Bollinger, Annika Eichler, Benjamin Flamm, Reto Fricker, Marc Hohmann, Mohammad Khosravi, Ralf Knechtle, Sascha Stoller, Bratislav Svetozarevic, Christoph Waibel and Joseph Warrington for fruitful discussions.

References

[1] Halvgaard R, Poulsen N K, Madsen H and Jorgensen J B 2012 Economic Model Predictive Control for building climate control in a Smart Grid 2012 IEEE PES Innovative Smart Grid Technologies (ISGT) (IEEE) pp 1–6 ISBN 978-1-4577-2159-5
[2] Oldewurtel F, Pariso A, Jones C N, Gyalistras D, Gwerder M, Stauch V, Lehmann B and Morari M 2012 Use of model predictive control and weather forecasts for energy efficient building climate control Energy and Buildings 45 15-27 ISSN 0378-7788
[3] Sturzenegger D, Gyalistras D, Morari M and Smith R S 2016 Model Predictive Climate Control of a Swiss Office Building: Implementation, Results, and Cost–Benefit Analysis IEEE Transactions on Control Systems Technology 24 1–12 ISSN 1063-6536
[4] Jain A, Nghiem T, Morari M and Mangharam R 2018 Learning and Control Using Gaussian Processes 2018 ACM/IEEE 9th International Conference on Cyber-Physical Systems (ICCPSS) (IEEE) pp 140–149 ISBN 978-1-5386-5301-2
[5] Smarra F, Jain A, de Rubeis T, Ambrosini D, D’Innocenzo A and Mangharam R 2018 Data-driven model predictive control using random forests for building energy optimization and climate control Applied Energy 226 1252–1272 ISSN 03062619
[6] Amasyali K and El-Gohary N M 2018 A review of data-driven building energy consumption prediction studies Renewable and Sustainable Energy Reviews 81 1192–1205 ISSN 18790690
[7] Afram A, Janabi-Sharifi F, Fung A S and Raahemifar K 2017 Artificial neural network (ANN) based model predictive control (MPC) and optimization of HVAC systems: A state of the art review and case study of a residential HVAC system Energy and Buildings 141 96–113 ISSN 03787788
[8] Bünning F, Bollinger A, Heer P, Smith R S and Lygeros J 2019 Improved day ahead heating demand forecasting by online correction methods URL https://engrxiv.org/wwx6g
[9] Bünning F, Bollinger A, Heer P, Smith R S and Lygeros J 2019 Empirical validation of a data-driven heating demand simulation with error correction methods URL https://engrxiv.org/deyg8/
[10] Basheer I and Hajmeer M 2000 Artificial neural networks: fundamentals, computing, design, and application Journal of Microbiological Methods 43 3–31 ISSN 0167-7012
[11] Chollet F 2018 Keras: The Python Deep Learning library Astrophysics Source Code Library
[12] Kingma D P and Lei Ba J 2015 ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION 3rd Int. Conf. Learn. Representations (Preprint 1412.6980v9)
[13] Zur R M, Jiang Y, Pesce L L and Drukker K 2009 Noise injection for training artificial neural networks: A comparison with weight decay and early stopping Medical Physics 36 4810–4818 ISSN 00942405