Improved day ahead heating demand forecasting by online correction methods

Felix Bnning[1,2], Philipp Heer[1], Roy S. Smith[2], John Lygeros[2]

[1]*Urban Energy Systems Laboratory, Swiss Federal Laboratories for Materials Science and Technology, Empa, Dbendorf, Switzerland*

[2]*Automatic Control Laboratory, Department of Electrical Engineering and Information Technology, ETH Zrich, Switzerland*

Abstract

Novel control strategies to reduce the heating and cooling energy consumption of buildings and districts are constantly being developed. Control on higher system levels, for example demand side management, usually requires forecasts for the future energy demand of buildings or entire districts. Such forecasts can be done with Artificial Neural Networks. However, the prediction performance of Artificial Neural Networks suffers from high variance. This means that two parameter-wise identical networks fitted to the same training data set perform differently well in forecasting the testing set. Here, we use two correction methods, one based on the forecasting error autocorrelation, and one based on online learning, to obtain reliable forecasting models. The approach is tested in the frame of day-ahead sub-hourly heating demand forecasting in a case study of a complex building, which has properties of a district heating system. It is demonstrated that the methods significantly reduce variance in prediction performance and also increase average prediction accuracy. When compared to other grey-box and black-box forecasting models, the approach performs well.

Keywords:
Artificial Neural Networks, demand forecasting, online learning, error autocorrelation, building energy demand, district energy demand

1. Introduction

Building and district heating and cooling energy amounts to 70% of the total energy demand in households in Switzerland (BFE, 2014) and to more than 44% of the total primary energy consumption worldwide (Ölz et al., 2007). To reduce this demand, novel control strategies are constantly being developed and tested on the level of individual buildings (see for example Široký et al. (2011); Yudong Ma et al. (2012); Oldewurtel et al. (2010); Sturzenegger et al. (2016); Bünning et al. (2017), or Hameed Shaikh et al. (2014) and the references therein).

In addition to control on a building level, there are a number of higher-level control tasks with the objective of optimizing energy consumption in the building domain or in coupled sectors (i.e. the electricity grid). Such tasks are for example the optimal control of the network temperature in low temperature district heating and cooling networks (Bünning et al., 2018), frequency regulation with electrified heating systems and water storage tanks (Kondoh et al., 2011), or optimal

control of energy hubs with various heating, cooling and storage technologies (Geidl et al., 2007; Arnold et al., 2009; Arnold and Andersson, 2011; Darivianakis et al., 2017). Here, control tasks include minimizing energy costs, maximizing self-consumption, peak shaving and management of long-term and short-term energy storage devices for example.

For these tasks, it is necessary to have detailed forecasts about the future energy consumption of individually connected buildings or the whole district. This has motivated an increasing research interest in demand forecasting for buildings for heating, cooling and overall electricity consumption, see for example the survey of different methods and models under (Mat Daut et al., 2017; Zhao and Magoulès, 2012; Harish and Kumar, 2016; Wang and Srinivasan, 2017; Foucquier et al., 2013; Suganthi and Samuel, 2012; Amasyali and El-Gohary, 2018; Ahmad et al., 2018). According to Amasyali and El-Gohary (2018), 31% of related studies make yearly, monthly or daily forecasts, while 57% consider hourly and 12% sub-hourly forecasts respectively, which would be suitable for above named control tasks.

Artificial Neural Networks (ANN) have been proven to perform well in a variety of complex and difficult tasks, such as gaming (Silver et al., 2017), computer vision (Wojna et al., 2018) and robotics (Tobin et al., 2017), and several studies have also demonstrated that ANN perform well on the task of short-term heating demand forecasting with high temporal resolution in the building and district domain.

On a district level, Kato et al. (2008) use a Recurrent Neural Network to make 24 hour ahead predictions of the hourly heating load of a real district heating system and compare it to a three-layered feed-forward ANN. The Recurrent Neural Network has an overall better forecasting performance, especially in periods that the authors consider as non-stationary. Similarly, Park et al. (2010) compare feed-forward ANN, Support Vector Machines (SVM) and the Particle Least Squares (PLS) method to perform 24 hour ahead prediction of the heating load of a real district heating system. PLS slightly outperforms SVM and ANN. Johansson et al. (2017) compare Extra-Trees Regressors (ETR) with feed-forward ANN on a real district heating testing case in Sweden for 24 hour ahead hourly forecast. ANN outperform ETR in this study. In Suryanarayana et al. (2018) the same Neural Network approach is compared to linear regression, ridge regression and lasso regression with an automatized feature selection process on two different district heating test cases in Sweden. ANN outperform the other three approaches in both cases. Saloux and Candanedo (2018) use ANN, SVM, and Regression trees (RT) and compare them to Piecewise Linear Regression on six hour ahead and 48 hour ahead heating demand forecasts with a sampling time of ten minutes in the case of a real district heating system. While ANN, SVM and RT clearly outperform the Piecewise Linear Regression, no distinguished performance difference is found between the other methods.

On the building level, Paudel et al. (2014) use ANN together with a pseudo-dynamical model to make four-day predictions with a sampling time of 15 minutes for the heating demand of a French office building. Kwok and Lee (2011) and Leung et al. (2012) combine ANN with different occupancy prediction models to enhance the hourly load prediction of a large office building and a university building in Hong Kong respectively. The same forecasting task is performed for a commercial building in Mestekemper et al. (2013). Here, the authors improve the prediction accuracy of ANN by splitting the cooling demand in a trend and a periodic signal and by training the ANN with the global optimization method Modal Trimming.

Related to heating and cooling demand prediction, Bagnasco et al. (2015); Kamaev et al. (2012); Escrivá-Escrivá et al. (2011); Chae et al. (2016) use ANN to make sub-hourly demand predictions for the electricity consumption of hospitals, shopping centers and other commercial buildings.

While all of the above named studies have promising results for energy demand forecasting with ANN, none of them communicate whether the achieved prediction results are average results or best case results, or if they were obtained with a fixed random seed. This is highly problematic as certain Machine Learning approaches and especially ANN suffer from high variance (Henderson et al., 2018; Recht, 2018; Bengio, 2012; Jamieson and Talwalkar, 2016): Depending on the random seed that is used while training the ANN, the ANN will have different prediction accuracies for the same training and testing data and same model parameters (i.e. number of neurons, number of layers etc.). This is due to the initialization of the network's weights with random values and the following non-convex optimization that generally converges to a local minimum in the fitting process. Jovanović et al. (2015), Jetcheva et al. (2014) and De Felice and Yao (2011) use ensemble methods to overcome this problem in the context of building energy demand prediction. However, using ensembles also has disadvantages. For example, the prediction accuracy of an ensemble is not guaranteed to be better than the one of its best predictor. Moreover, using ensembles is computationally more expensive than using single predictors. The computational effort typically grows linear with the number of predictors in the ensemble.

1.1. Contribution

We therefore aim to make the following contribution with this work: A prediction model based on single ANN to make 24 hour ahead forecasts with a 15-minute sample rate of the heating demand for complex buildings and districts with individual uses, individual occupant patterns and unknown control systems, is developed. Two simple forecast correction methods, based on the error-autocorrelation and online learning, which make use of the history of previous forecasting errors during the online phase, are introduced. The methods are verified in a case study based on data from the NEST building, which comprises independent modular control zones that allow it to mimic a district heating network. We demonstrate that uncorrected ANN indeed show high variance in prediction performance and that the presented correction methods significantly reduce this variance, giving rise to reliable single predictors. Moreover, the methods remove prediction bias. The prediction results are compared to other machine

learning (or regression-based) approaches and to conventional resistor-capacitor building models (as used by De Coninck et al. (2015); Zhou et al. (2008)).

1.2. Structure

The remaining article is structured as follows: In Section 2 our ANN based methodology and the forecast corrections, which are based on the forecasting error autocorrelation and online learning, are described. In Section 3 the case study is defined. The energy system under consideration is introduced, and the structure of the applied ANN and the benchmarking grey-box and black-box methods are defined. In Section 4 the results regarding prediction accuracy and variance are discussed, as well as limitations of the study. Moreover, the ANN methods are compared to the benchmarking methods. In Section 5 the article is concluded.

2. Methodology

2.1. Forecasting task and correction methods

In this study the following forecasting task is considered: A heating demand forecast for a building or a district is made at midnight for the next 24 hours, sampled every 15 minutes. The training and validation data are assumed to be sampled at the same rate. Furthermore, perfect knowledge of the ambient temperature for the forecasting period is assumed. This is done as it avoids uncertainty in the inputs when comparing the different methods.

The forecast is made with a feed-forward ANN, which will be described in detail in Section 3. To model the closed-loop response of the building and its lower-level climate controllers to the ambient conditions, model inputs related to ambient conditions and time features are used. To reduce variance and to improve the forecasting accuracy, we introduce two methods that are used in the online phase of the forecasting task.

2.2. Error-autocorrelation correction

Empirical evidence suggests that forecast errors persist over a longer period of time than a single forecasting interval in the setting of building energy or district energy forecasting. For example, if windows are left open, a room temperature set point is changed or a fluid pump fails, this does not only have an effect on one single forecasting interval (15 minutes) but also on the following intervals, as the source of the error is usually not eliminated within one single interval.[1]

_____
[1]This is further demonstrated in section 4.3.

Motivated by this, we predict the error $\tilde{e}$ for day T and forecast interval $t \in [1, 96]$ by setting

$$\tilde{e}_{T,t} = \frac{1}{4} \sum_{\tau=93}^{96} e_{T-1,\tau} R_{ee}(t, E) \qquad (1)$$

with

$$R_{ee}(l, E) = \frac{\mathbb{E}[(E - \mu)(E_{+l} - \mu)]}{\sigma^2}. \qquad (2)$$

Here the first term in the right hand side of (1) denotes the average of the past day's forecasting errors e for the last four forecasting intervals; averaging over a few elements helps to denoise the error at the cost of a small decrease in the autocorrelation. The choice of four elements could be questioned, however, by further increasing the averaging interval, prediction power is lost because of the decreasing correlation of the error. $R_{ee}(l, E)$ denotes the autocorrelation for lag l, which is calculated with the set E which contains all past forecasting errors e that have occurred until day T - including the errors of the training set. E_{+l} is the same set shifted by lag l. The expected value $\mathbb{E}$, the mean μ and the standard deviation σ are empirical approximations for the stochastic process based on the set E.

Instead of including all past forecasting errors in E, one can also implement a shifting window and compute the autocorrelation using the forecast errors for the past few days. However, our experiments suggest that this does not improve the error prediction.[2]

Figure 1 shows a schematic of the forecast correction procedure; A demand forecast is made with an ANN (uncorrected forecast). Based on the database of all previous forecasting errors and the measured error of the previous day's forecast, the error estimation described in (1)-(2) is made. The estimated error is then added to the uncorrected forecast to give a corrected demand forecast. At the end of the day, the difference between demand realization and demand forecast is added to the database. At the first prediction day in the online phase, the correction is based on the errors in the training set.

_____
[2]In case of a coupled multiple output system, e.g. energy demands of different zones within a large building, measured errors from neighbouring zones $n \in N$ could also be used to make an correlation correction for the error in zone i. Equation 1 would be changed to

$$\tilde{e}_{i,T,t} = s_i R_{ee}(t, E_i) \sum_{\tau=93}^{96} e_{i,T-1,\tau} + \sum_{n \in N} [s_n R_{i,n}(t, E_i, E_n) \sum_{\tau=93}^{96} e_{n,T-1,\tau}],$$

in which $R_{i,n}(t, E_i, E_n)$ denotes the Pearson correlation between measured errors E_n in neighbouring zones and the error E_i in the zone under consideration, shifted by lag t. s_i and s_b are scaling factors that need to be learned to avoid overestimation of $\tilde{e}_{i,T,t}$.
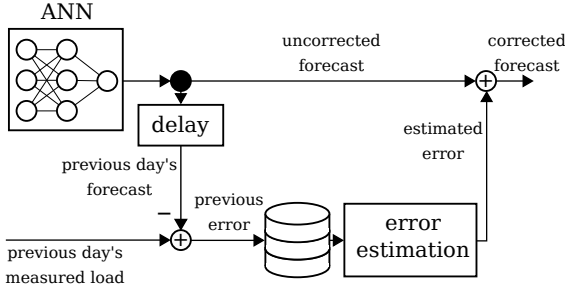
Figure 1: Forecast correction based on error-autocorrelation scheme

Figure 2: NEST building at Empa in Switzerland, Copyright: Zooey Braun - Stuttgart

The corresponding algorithm is shown in Algorithm 1. The results of this correction are shown in Section 4.

Algorithm 1

1: **procedure** ERROR-AUTOCORRELATION CORRECTION
2:     **for** *T in runtime* **do**
3:         **for** *t in* $[1, 96]$ **do**
4:             forecast[T,t]:=ANN.predict(inputs[T,t])
5:             error[T,t]:=AvError[T-1]*R_{ee}(t,E)
6:             correctedFcast[T,t] := forecast[T,t]
                + error[T,t]
7:         *wait until day passes*
8:         measuredError[T] := realization[T]
           - forecast[T]
9:         E:=E.add(measuredError[T])

2.3. Online learning

The second forecast correction method is online learning. Instead of using only the training set for training the model, after each day the newly collected data (network inputs and demand realization) is used to re-train the ANN. The inputs are identical to the ones used in offline training. The argument behind this approach is again that phenomena in the building domain persist for time constants longer than hours and that the assumption "tomorrow will be like today" (also referred to as "persistence model" or "naïve forecast") has some validity here. By retraining the network on the previous day, we intentionally bias the network towards a certain solution, which is usually undesirable in machine learning (Huang et al., 2006), but potentially beneficial when the model is persistent. We can adjust this biasing via the learning rate of the optimizer used for training. This balances the importance between the training set and the daily feedback from the realised prediction errors.

Combining this online training heuristic with the autocorrelation based error correction outlined above, leads to Algorithm 2.

It should be noted, that in line 10 a second forecast is made at the end of the day, after the ANN has been re-trained on the inputs and demand realization of the day. The error of this second forecast is then added to the error database. This is done because next day's error estimation needs to be performed for the retrained ANN. If this is not done, the error for the next day is overestimated. The results for this combined method are shown in Section 4.

Algorithm 2

1: **procedure** COMBINED CORRECTION
2:     **for** *T in runtime* **do**
3:         **for** *t in* $[1, 96]$ **do**
4:             forecast[T,t]:=ANN.predict(inputs[T,t])
5:             error[T,t]:=AvError[T-1]*R_{ee}(t,E)
6:             correctedFcast[T,t] := forecast[T,t]
                + error[T,t]
7:         *wait until day passes*
8:         ANN:=ANN.train(inputs[T],realization[T])
9:         **for** *t in* $[1, 96]$ **do**
10:            forecast2[T,t]:=ANN.predict(inputs[T,t])
11:         measuredError[T] := realization[T]
           - forecast2[T]
12:         E:=E.add(measuredError[T])

---

3. Case study

3.1. System description

The NEST building at Empa in Switzerland (Richner et al., 2018), shown in Figure 2, is used as a case study. The building contains multiple individual units with different uses that can be added and removed from the building, in addition to office and meeting rooms that are permanently installed. At different times during the period covered by our data, two, three or four of these

individual units were in operation. One of the units was added during the training set (August 2017) and one during the test set (February 2018) of the case study. The configuration resembles a district heating system because different units have different use patterns, some of which resemble those of residential buildings while others those of office spaces. Moreover, all of the individual units are equipped with local controllers, whose details are unknown to the forecasting system. The demand forecast is therefore made for a collection of individual closed-loop energy systems. The units are connected via substations to a central heating system with a supply temperature of 35 °C and a return temperature of 25 °C. The heating system is served by a central heat pump with a maximum thermal capacity of 100 kW. A simplified schematic of the system can be seen in Figure 3. The heating demand that our algorithm is designed to forecast is the energy flow difference between supply line and the return line. For more information on the complete system see Lydon et al. (2017).

The building is set-up as a test bed for district heating and cooling and building energy experiments. It integrates approximately 540 sensors, whose data is stored in an SQL database with a temporal granularity of one minute, and approximately 200 controllable actuators. The measurement data used in this study contain 411 days (01.04.2017 to 16.05.2018) and include the measurements for ambient temperature and total heat consumption. The measured heating demand is shown in Figure 4. Figure 5 shows the ambient air temperature measured at the roof top of the building. Though, strictly speaking, weather forecasts should be used in demand forecasting, in this study we use the actual ambient air temperature directly as an input. This is done to avoid the additional uncertainty in the inputs and to allow direct comparison of the different forecasting methods.

3.2. Artificial Neural Networks

We use feed-forward ANN with the Python package Keras (Chollet, 2018), which is a higher-level API to TensorFlow (Abadi et al., 2016). We refrain from fundamental introduction to the concept here and refer the reader to other sources (for example Shanmuganathan (2016); Basheer and Hajmeer (2000)). Other topologies, such as Recurrent Neural Networks (RNN) and Long-Short-Term Memory Networks (LSTMN) were investigated in preliminary studies but showed weaker performance. To fit the ANN to the training data, we use the optimizer Adam (Yun et al., 2018) with standard parameters. This includes the learning rate of 0.001 for both offline and online learning. As a loss function, the
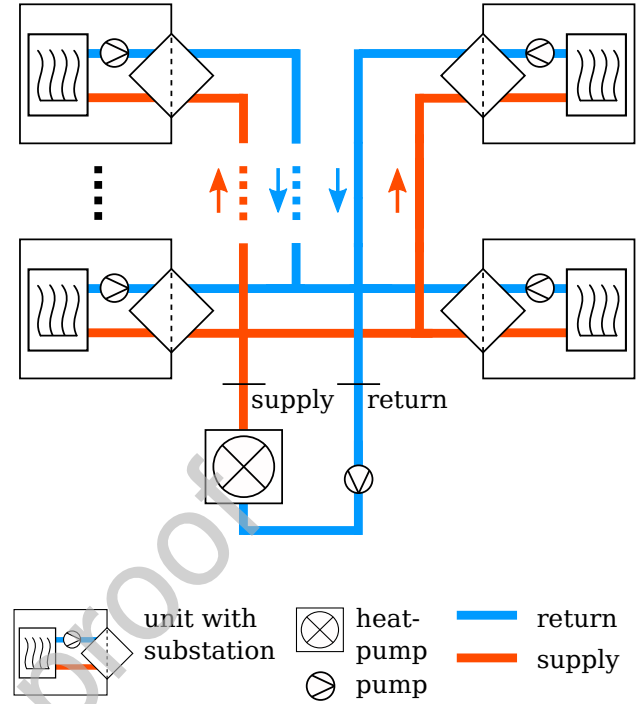

Figure 3: NEST heating system schematic

mean squared error and as activation functions, Rectified Linear Units (ReLu) are used.

3.2.1. Structure

As we want to forecast the heating demand in a day-ahead fashion, we forecast 96 data points at once. (One value every 15 minutes for one day.) For the structure of the ANN this gives two possibilities: 1) The network is built with 96 outputs and the forecast for the day is performed in one single step. 2) The network is built with one output and the forecast is performed with 96 different input vectors to forecast 96 single output values.

Option 1 has one major advantage: As the heating demand (forecast target) is highly correlated for small time delays, the accuracy of the prediction is very precise for the first few forecast data points at the beginning of the day. This is because the network makes use of the measured heating demands at the end of the previous day. In option 2 this is not possible as, for example, we cannot use the demand of $t - 4$ as an input for the network because for forecasts $t = 5, ..., 96$ this input has not yet been measured at the time the forecast is made. As an alternative, the forecasts of earlier time steps could be used as inputs, but this would potentially lead to error propagation. However, in case of option 1 the network
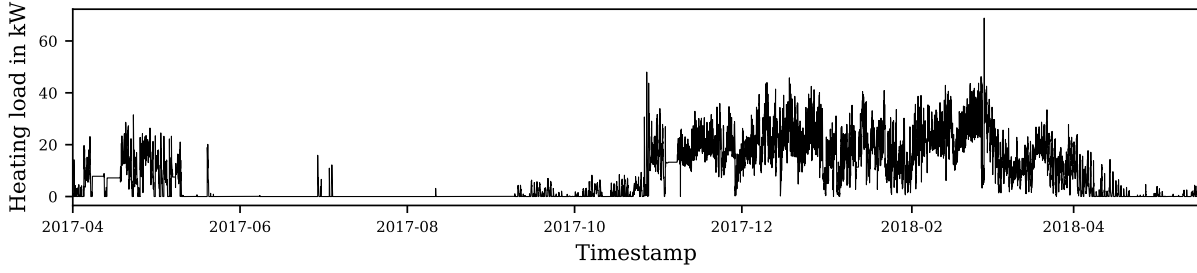
5

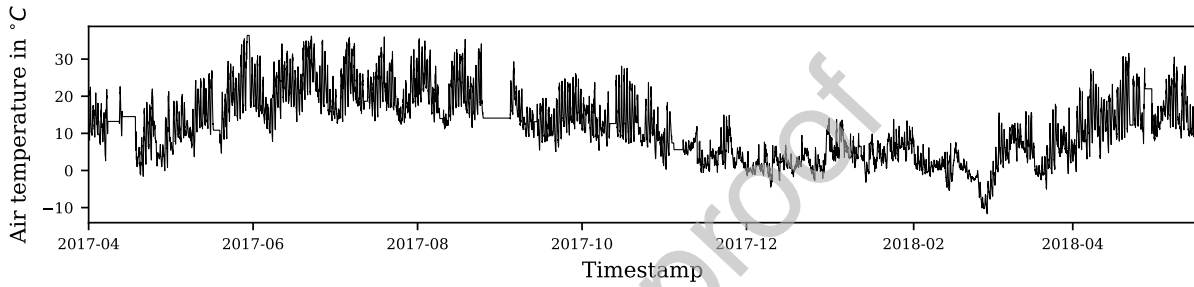Figure 4: Measured total heating load of the NEST building

Figure 5: Measured ambient temperature at the roof of NEST building

needs a much higher complexity in terms of number of layers and nodes, while the training set is not increased when compared to option 2. Preliminary forecast runs have shown that this results in reduced prediction performance. For option 2, the disadvantage of not being able to use the last measured heating demands of the previous day as inputs is compensated by the error-autocorrelation based correction method that was introduced in Section 2.2. We therefore use option 2 in this study.

3.2.2. Feature selection

Although NEST also has measurements for wind speed and solar radiation, the ambient temperature was selected as the only weather-related feature[3], because preliminary experiments suggested that the other measurements do not substantially improve forecasting accuracy. This might appear surprising as generally solar radiation has indeed an effect on the heating demand of a building, however, with a Pearson correlation coefficient of 0.54 between ambient temperature and solar

radiation in the considered data set, it appears that the ambient temperature incorporates much of the information of the solar radiation already.

With the assumption that the demand of the system shows daily and weekly patterns, as the units are used for living and as office spaces, two more features were added: the total heat consumption with a time delay of one day (again following the assumption "tomorrow will be like today") and the total heat consumption with a time delay of one week (motivated by the assumption that weekly routines can expected). This allows the prediction of the energy demand today based on the demand at the corresponding time yesterday and last week. Finally, the time stamp of all features was implemented as two inputs: hour of the day and weekday/weekend[4]. Hour of the day was one-hot encoded (Aggarwal, 2018), meaning that instead of one continuous input 24 binary inputs are used. Preliminary experiments suggested that this improves the forecast accuracy. We assume that the occupancy related internal gains are indirectly captured by the time-related inputs and do not use a separate occupancy model.

---

[3]The difference between "feature" and "input" is the level of detail in the description: E.g. "daytime" can be a feature and the corresponding inputs could be "hour of the day, encoded continuously" and "minute of the hour, encoded continuously".

[4]Weekday/weekend was chosen instead of working day/non-working day as the difference is marginal and the implementation effort substantially lower.

Table 1: ANN parameters

Parameter	Value
Number of hidden layers	2
Number of nodes per layer	8
Number of Epochs	10
Input scaling method	Linear [-1,1]

The resulting network has 28 inputs, of which three are continuous and 25 binary, and one continuous output.

3.2.3. Parameter selection

There is no proven rule on how to best determine the parameters of ANN. These values are commonly set based on heuristics and an iterative approach that involves a substantial amount of trial and error. Generally, a trade-off between overfitting and underfitting of the training data has to be found.

A preliminary sensitivity analysis was conducted to decide on the number of hidden layers, number of nodes per layer, number of training epochs (number of times the network's weights are updated based on each individual sample in the training set), and on the input scaling method. The results for the suggested optimal parameter values are given in Table 1. They were used in the case study.

3.3. Forecasting methods for benchmarking

To assess the performance of the ANN with the introduced correction methods for forecasting the thermal demand of NEST, the prediction accuracy is compared to other state of the art prediction methods: These are grey-box models in the form of resistor-capacitor building models, and black-box models in the form of other regression-based or machine-learning based approaches.

3.3.1. Grey-box method (R-C models)

For the grey-box approach, resistor-capacitor models are used. We note that while resistor-capacitor and other grey-box models are often used as optimization models for selecting control inputs for building level MPC, they are also used for forecasting (De Coninck et al., 2015; Zhou et al., 2008). Moreover, grey-box and white-box models are also commonly used for building performance simulation (which is essentially forecasting without using the forecasts online), see for example (Zhao and Magoulès (2012)).

Four different topologies were considered as shown in Figure 6, inspired by the models used in (De Coninck et al., 2015). In this approach, the thermal capacity of wall materials, floor materials and air volumes is represented by capacitors, whereas the thermal resistance of walls and floors is represented by resistors. The parameters of all resistors and capacitors are estimated to fit a heating demand measurement curve. As the R-C models are purely thermal models, the ambient temperature is the only used input.

The models were implemented in the modelling language Modelica (Mattsson et al., 1998) and simulated in Dymola (Brück et al., 2002). To estimate the parameters, a CMA-ES (Hansen et al., 2003) optimizer was used in Python.

In models 1R1C and 2R2C, the parameters were estimated such that the mean squared error between $\dot{Q}$ and the target heating demand is minimized. In model 1R1C, $\dot{Q}$ was set equal to the heat flux through R_{wall} and in model 2R2C $\dot{Q}$ was set equal to the heat flux through R_{wall2}, which keeps the temperature of C_{zone} at a constant 20 °C.

As an additional capacitor for the building core C_{core} was added for models 4R3C and 5R3C, the mapping of the target heating demand is more complex. The optimization was set up to minimize the mean squared error between the target heating and $\dot{Q}_{total}$,

$$\dot{Q}_{total} = \dot{Q}_1 + \dot{Q}_2, \tag{3}$$

which is the sum of the heat distributed via air conditioning to C_{zone} ($\dot{Q}_1$) and the heat distributed via floor heating to C_{core} ($\dot{Q}_2$). Moreover, as the direction of heat loss is not unique in these models (losses can go from C_{zone} to C_{core} and/or to T_{ground}), a P-controller was set-up to keep the temperature of C_{zone} at a constant $T_{set} = 20$ °C:

$$\dot{Q}_{total} = k_p \times (T_{set} - T_{zone}). \tag{4}$$

A P-controller is chosen because buildings are often controlled by a thermostat, which is a proportional controller, and the closed-loop demand response to the ambient temperature is assumed to be dominated by the building dynamics.[5] Additionally a fraction coefficient c_f was introduced to distribute the heat between air conditioning and floor heating, such that

$$\dot{Q}_1 = c_f \times \dot{Q}_{total} \tag{5}$$

and

$$\dot{Q}_2 = (1 - c_f) \times \dot{Q}_{total}. \tag{6}$$

---

[5]This is the case for any well-regulated building.

7

Model 1R1C
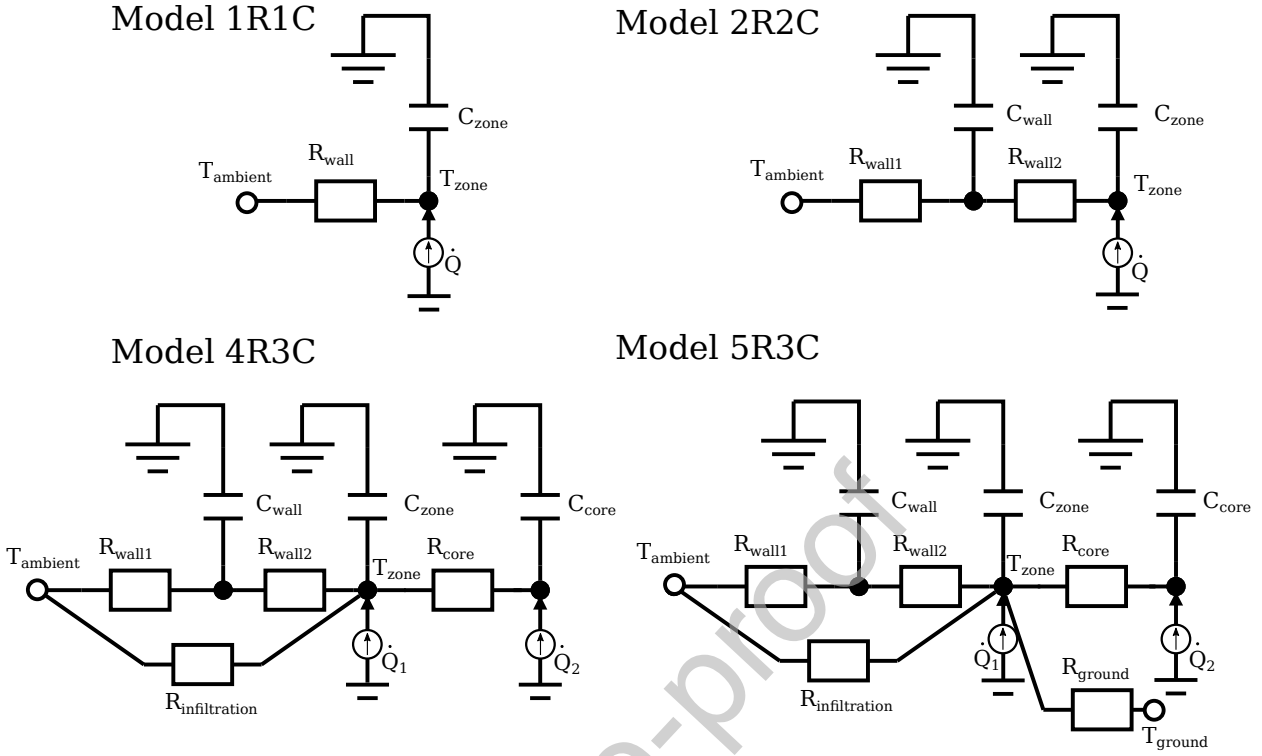
Model 2R2C

Model 4R3C

Model 5R3C

Figure 6: R-C building models

The values of k_p and c_f are also determined in the fitting process.

3.3.2. Black-box method (regression-based methods)

For the black-box methods that are used for benchmarking, we use Python's scikit-learn package (Pedregosa et al., 2011). The package features a variety of different regression or machine learning methods. In the case study, the following methods are used:

1. Least squares Linear Regression
2. Support Vector Machine
3. Huber Regressor
4. Orthogonal Matching Pursuit
5. SGD Regressor
6. Decision Tree Regression
7. Random Forest

The data sets used for these models were identical to that used for the ANN method, including the scaling of inputs and outputs and the one-hot encoding of categorical features. For all methods, a grid search regarding model tuning parameters was performed to reach a confident result. These parameters are in particular *kernel*, *gamma*, *C* and *degree* for Support Vector Machine,

epsilon and *alpha* for Huber Regressor, *nonzero coefficients* for Orthogonal Matching Pursuit, *alpha* and *L1 ratio* for SGD Regressor, *depth* for Decision Tree Regression, and *number of estimators* and *depth* for Random Forests. We refer the reader to the API reference of scikit-learn for details on the tuning parameters.

4. Results and discussion

For the case study, the measurement data from NEST is divided into a training set and a testing set. The training set consists of 70% of the whole data set, which corresponds to 287 days (27552 data points). The testing set consists of 124 days (11904 data points).

The data was post-processed with the help of the pandas package (Mckinney, 2011) in Python 3. Missing data points were first linearly interpolated or extrapolated from neighbouring data points. This excludes data points that have already been handled by the NEST database (e.g. the constant values around September 2017 in Figure 5). The data set was then re-sampled to a 15-minute interval using the mean value of all relevant data points.

8

To benchmark the prediction performance, the coefficient of determination,

$$R^2 = 1 - \frac{\sum_{i \in N} (y_i - f(x_i))^2}{\sum_{i \in N} (y_i - \bar{y}_N)^2}, \qquad (7)$$

is used. It becomes *zero* if the forecast $f(x_i)$ is as good as taking the average $\bar{y}_N$ of the data in the considered set N as a forecast, and becomes *one* if the forecast exactly matches the validation data y_i. In principle, R^2 may also become negative if the forecast is worse than taking the average, though this was not observed with any of the methods tested here. In the following, if more than one network is used, R^2 describes the average of all R^2 achieved, unless otherwise stated.

We train and forecast with 100 individual ANN as the initial input weights of each node in the network are initialized with a random value. This randomness leads to different prediction performance for each network as the training process involves the solution of a non-convex optimisation problem that generally converges to a local minimum. By simulating 100 networks, a more confident statement about R^2 can be made and the quartiles can be studied.

4.1. Prediction accuracy, quartiles and improvements through forecast corrections

Figure 7 shows the box plot of the coefficient of determination in the test set of 100 different networks for the uncorrected ANN, ANN with error-autocorrelation correction, ANN with online learning and ANN with error-autocorrelation + online learning for full day-ahead forecasts. First, it should be noted that the variance in performance for uncorrected ANN is indeed very high, as discussed in Section 1. The highest performing network reaches $R^2 = 0.868$, while the lowest performing network reaches $R^2 = 0.732$. It can be seen that the quartiles regarding prediction accuracy amongst different network instances are significantly reduced when correction methods are used. While the interquartile range (IQR) is 0.038 in the uncorrected case, it reduces to 0.023 in the case with error-autocorrelation correction, 0.009 in the case with online learning and 0.008 in the case where both correction methods are used. With both corrections in place all R^2 lie between 0.872 and 0.898. The average achieved R^2 are 0.818, 0.860, 0.878 and 0.885 subsequently. Figures A.14 - A.16 in the Appendix show the results of a second experiment with 100 runs for the variance of three different KPI, Mean Squared Error (MSE), Mean Absolute Error (MAE) and Coefficient of Variation of the Root-Mean Squared Error (CV RMSE). The trend is identical to the one observed for R^2 in the first experiment. For clarity, the analysis is therefore limited to R^2 in the following.

Figure 8 shows the coefficient of determination for a 2 hour ahead prediction and Figure 9 for a 24 hour ahead prediction, again for 100 networks. First, it can be noticed that the performance in terms of median and IQR is similar for the uncorrected forecasts. However, in case of the corrected forecasts, the performance is better in the 2h ahead case, because the forecast is closer to the model update and to the last measured error, on which the error autocorrelation correction is based on. Moreover, while the correction based on online learning shows a better performance in terms of median for the full forecast (Fig. 7) and 24h ahead forecast, the correction based on error autocorrelation is more effective in the 2h ahead case. However, the variance is significantly reduced by both correction methods in all cases.

These results suggest that the corrections introduced in Section 2 can to a large extent alleviate a main disadvantage of ANN, namely the dependence of the prediction performance on randomly initialized node input weights. Moreover, the average prediction performance is significantly improved.

4.2. Influence of training set on prediction performance

The training set and the online application of forecasts can often fall into different seasons for real-life applications, which results in different probability distributions of samples in training and application (or in case of this study in the testing set). To investigate the influence of different distributions, we have trained 100 networks based on randomly sub-sampled sets, with random length and random location in the original set, from the training set described in Section 4. The networks are validated on the same testing set for comparable R^2.

Figure 10 shows the variance of the coefficient of determination. It can be seen that median and IQR has significantly worsened for the uncorrected networks compared to Figure 7. However, in the case where both correction methods are applied, both median (0.868) and IQR (0.0183) are close to the ones achieved with the full training set.

4.3. Prediction error analysis

To further evaluate the impact of the correction terms introduced in Section 2 we analysed the statistics of the forecast error with and without the corrections.[6]

---

[6]The absolute error is used here instead of the relative error, as the latter is undefined for many of the data points where demand is close to zero.
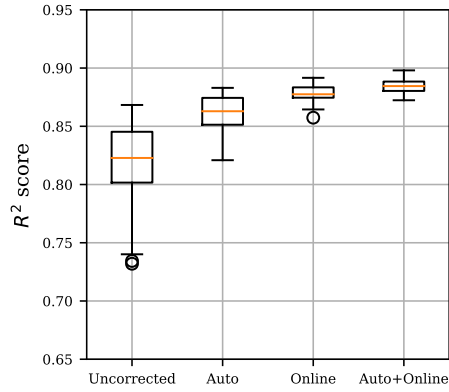
Figure 7: Coefficient of determination for ANN and forecast correction methods for 100 individual networks (with boxes describing the interquartile range, whiskers of 1.5 times the interquartile range, circles indicating outliers and the orange line indicating the median)
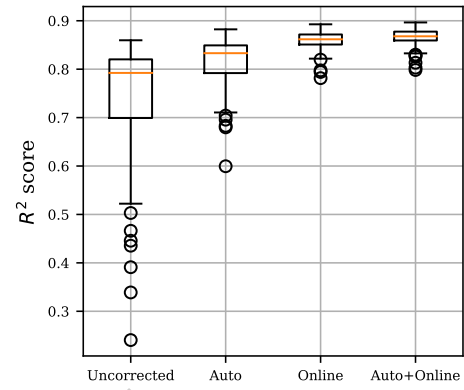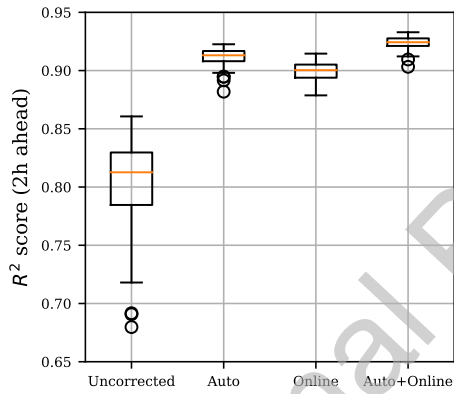

Figure 8: Coefficient of determination of 2h ahead prediction of 100 ANN and forecast correction methods

Figure 9: Coefficient of determination of 24h ahead prediction of 100 ANN and forecast correction methods

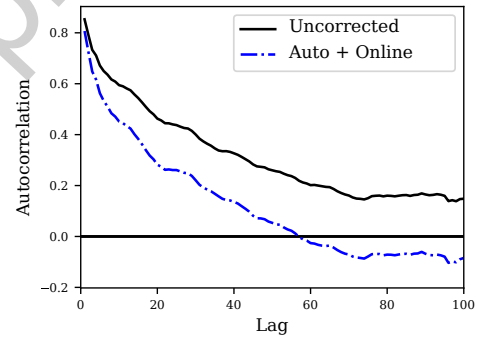Figure 10: Coefficient of determination of 100 ANN and forecast correction methods with randomized training period

Figure 11: Residual autocorrelation for uncorrected ANN and ANN with error-autocorrelation correction + online learning

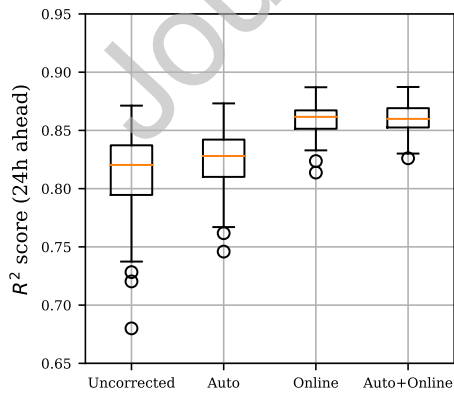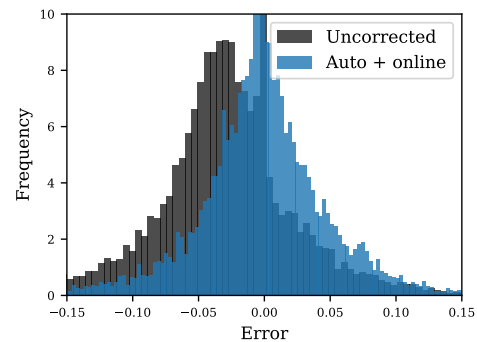Figure 12: Forecasting error histogram for uncorrected ANN and ANN with error-autocorrelation correction + online learning (Standard deviations are indicated by the vertical lines)

Figure 11 shows the autocorrelation of the error for one instance of an uncorrected ANN forecast (with a relatively high R^2 of 0.841) and the remaining error after correction for one instance of a corrected ANN with error-autocorrelation and online learning (with R^2 of 0.887). It can be seen that the signal for the corrected ANN is less correlated than the uncorrected signal. However, the residual is still not "white", as the autocorrelation is still substantial for many values of the lag. We conjecture that this residual autocorrelation is an intrinsic artefact of the day-ahead forecast: Disturbances that occur during the forecasting period will introduce a systematic forecasting error for the rest of the period, as there is no chance to correct the forecast until the next forecasting period begins.[7] Moreover, one can see that there is no peak at $t = 96$, i.e. after one day. This suggests that the network captured all regular daily time-dependencies of the demand and has no systematic error regarding this aspect.

Figure 12 compares the forecasting error histogram of one instance of ANN without correction to the one using both error correction methods.[8] The plot is cut at a frequency of 10 for the sake of readability, because by far most errors are 0 for both cases. The interquartile range reduces from 0.054 in the uncorrected case to 0.033 in the corrected case. Furthermore, the median improves from -0.025 to 0, which shows that the bias in the forecast is effectively eliminated by the forecast corrections.

4.4. Forecasting trajectory examples

Figure 13 shows examples of the forecast heating demand trajectory at different temporal resolutions for an ANN that includes the error-autocorrelation and online learning corrections. The figure also shows the empirical confidence bounds for the forecast computed with a posteriori analysis of the absolute error distribution of the testing set. The dark-grey background depicts a confidence bound of 68% (one standard deviation) and the light-grey background depicts a confidence bound of 95% (two standard deviations). Such confidence intervals could be useful if one uses the forecast for predictive building energy management based on robust or stochastic model predictive control. For certain control tasks, such as temperature control within a building, these confidence bounds might appear large. However, for high level energy management tasks, methods such as affine decision rules (Warrington et al., 2012, 2014) can be deployed to introduce recourse in receding horizon control schemes that significantly reduce the impact of forecast uncertainty on the control performance.

The top graph of Figure 13 shows the prediction and the actual heating demand for the complete testing set. The general trend is captured well. We note in passing that the model seems to cope well with another unit being added in February (as described in Section 3.1), as the general trend is matched equally well before and after February. The middle graph shows a more detailed view of five days with a medium heating demand in March. At this scale the capabilities and limitations of the forecast correction methods can be observed. At the beginning of the 18th of March and 19th of March, the forecast signal is corrected towards the true demand, because an error was observed at the end of the previous days. The time of correction is indicated by the vertical grey lines. On the other hand, in the middle of 17th of March the forecast deviates from the actual demand for several hours. However, this error can not be corrected, as the correction is applied only at the beginning of each 24 hour forecasting interval. In the bottom graph, a system shut-down or failure occurs in the afternoon of the 26th of February. The forecast is not corrected until the start of the next day. At the beginning of the 28th of February and 1st of March a correction of the forecast towards the real demand at the beginning of the day can be observed again.

4.5. Comparison of different forecasting methods

Table 2 compares the coefficient of determination in the testing set for all ANN and the benchmark forecasting methods for a full 24 hour forecast.[9] It can be seen that the ANN in all three corrected cases outperform the other forecasting methods.

Decision Tree Regression and Random Forest reach a high R^2 without any additional corrections. It is possible that with appropriate correction methods (similar to the ones introduced in this study) the performance of our corrected ANN can be reached or exceeded. With respect to the correction methods introduced here, we note that online learning is not a viable option for regression trees and random forests, because of the way

---

[7]To use the example of an opened window again: If the window is opened at day T-1, we can use the resulting forecast error in day T-1 in our forecast corrections for day T. However, if the window is opened at day T, for example at t=5, this will introduce a systematic error in the forecast for t=[6...96]. This error will show up in the autocorrelation plot and is impossible to avoid in this forecasting setting.

[8]The number of bins is set according to the Freedman-Diaconis rule.

[9]N-step ahead error metrics are not analysed because the models are not updated or corrected and the inputs are real measurements and not forecasts.
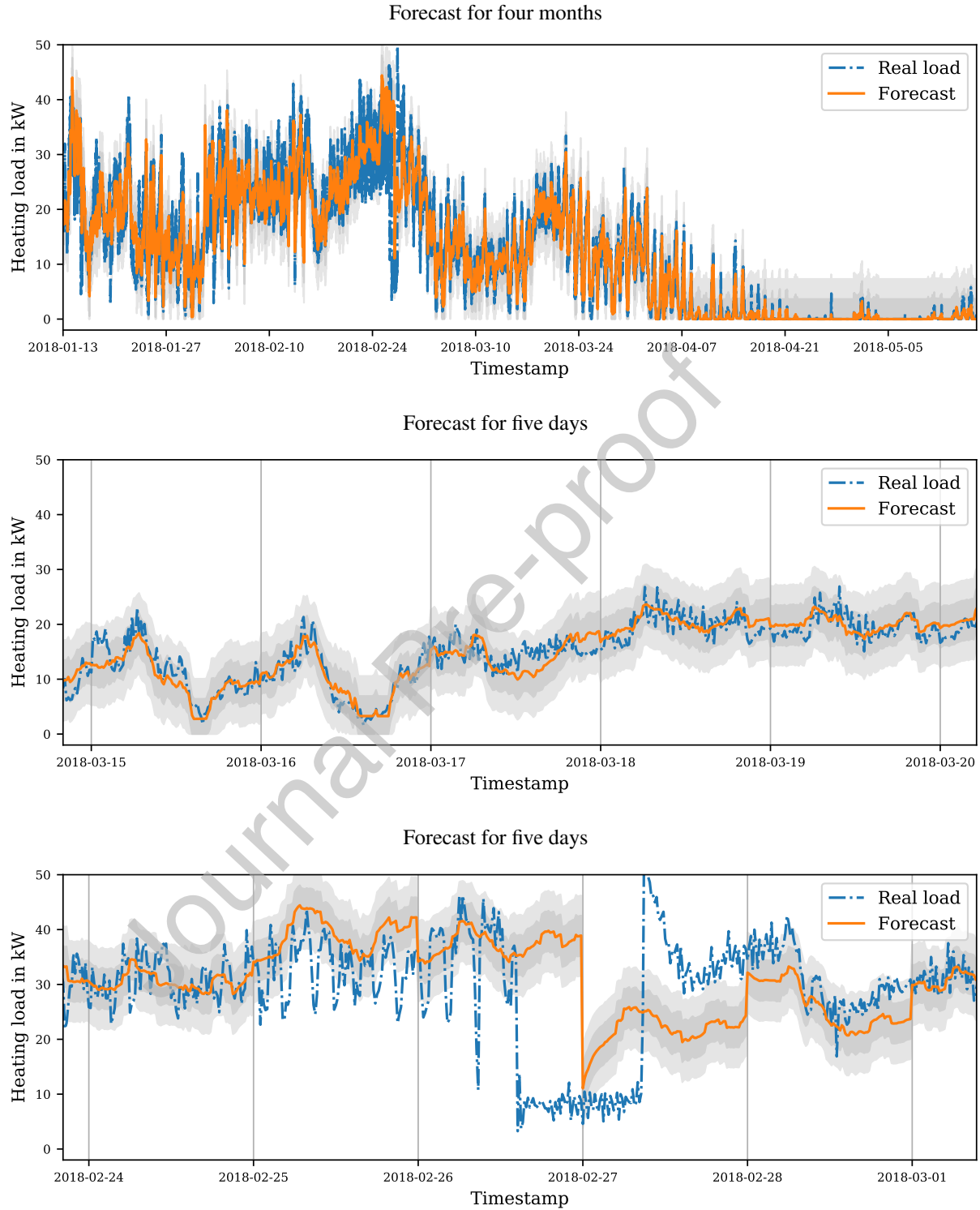
Figure 13: Forecasting examples with varying temporal resolution. The time of correction is indicated by the vertical grey lines. The grey shadings indicate confidence bounds of 68% and 95%.

12

Table 2: Coefficient of determination for all forecasting methods

Method	R^2
ANN online + autocorr.	0.885
Huber Regressor	0.768
Orthogonal Matching Pursuit	0.770
Least squares Linear Regression	0.770
SGD Regressor	0.775
Support Vector Machine	0.811
Decision Tree Regression	0.844
Random Forest	0.860
Model 1R1C	0.691
Model 2R2C	0.702
Model 4R3C	0.739
Model 5R3C	0.761

these are constructed. Both methods are based on decision trees in which the input variables are split in partitions and the output is approximated with a constant value in these partitions. The splitting variables and splitting points are chosen heuristically. In online learning with ANN, the network weights are updated with the measured input-output samples every day. The importance of the new samples can be adjusted with a learning rate. Such an adjustment of splitting variable and point is not possible in trees, as the whole tree would have to be rebuilt from the changed splitting point down. Thus, regression trees or random forests always have to be trained on the complete data set and cannot be updated with single samples. Retraining on a combined set of historical data and daily measurements is in theory possible, but it would not bias the forecast towards the recent behaviour of the building. Corrections based on error-autocorrelation could, however, be used to correct forecasts made by any forecasting method.

The resistor-capacitor based models perform worse than the majority of the regression based methods in this case study. This could be due to the fact that these models can only make use of the ambient temperature as an input, making it impossible to capture time-dependent phenomena, such as occupancy for example. This shortcoming could be improved by an occupancy model. Moreover, these models involve only a small number of parameters, in the form of resistor and capacitor values. Increasing the model complexity further could improve forecasting, as suggested by the fact that the R^2 of the R-C models in our results continue to increase as more parameters are added. The addition of more R-C

components of course implies increased modelling effort. Finally, R-C models are often used in this context for modelling building thermal dynamics (Sturzenegger et al., 2014) and not for demand forecasting. This study suggest that they are not equally suitable for the latter.

4.6. Limitations

As evident from the discussion in the previous section, the study has certain limitations. The scope of the study is to find correction methods that solve the problem of prediction accuracy variance of ANN and to increase their prediction performance. As KPIs of prediction performance parameters are not comparable between different studies (as they heavily depend on the predicted trajectory), ANN combined with online correction methods were benchmarked against other forecasting methods. As the scope of this study is not a comprehensive review and comparison of forecasting methods, the benchmarking methods have certain limitations and could potentially be improved in further studies. For example, besides the already mentioned measures, the prediction performance of the grey-box models could benefit from re-initializing the model states with temperature measurements from the building at the beginning of each forecasting period. However, there are also arguments that speak against this. In particular, temperature measurements of individual buildings are generally not available in an energy-hub or a different higher level control context.

4.7. Generalization to other buildings

To strengthen the confidence in the correction methods, the case study was extended to three other buildings: An eleven-storey office building at ETH Zrich with measurement data of 36 months, a three-storey laboratory building at Empa Dbendorf with data of 24 months and a three-storey office building at Empa with data of 24 months. The data quality of the last differs from the other buildings, as the quantization error in the heating demand measurement is at least three times higher than that of the other case studies.

For all cases, the data was divided into 70% for the training set and 30% for the testing set. The ANN used are identical to those in the NEST case studies. Again, 100 instances of ANN were trained and tested for each case in order to compensate for the variance in prediction performance.

Table 3 shows the average R^2 and the interquartile range of R^2 for each building with the uncorrected ANN and both correction methods applied. The IQR ist reduced by a factor between 3 and 20 for the different

Table 3: Performance of correction methods for other

Building	$\bar{R}^2$ uncorrected	$\bar{R}^2$ auto+online	R^2 IQR uncorrected	R^2 IQR auto+online
Office ETH	0.886	0.936	0.0329	0.0053
Laboratory Empa	0.674	0.809	0.0600	0.0033
Office Empa	0.856	0.862	0.0097	0.0032

cases. Moreover, it can be seen that the average coefficient of determination also increases for each case when both correction methods are applied. The relative improvement in case of the office building at Empa is small compared to the other buildings. The fact that the measurement data quality of this building is inferior to the other ones suggests that the methods are sensitive to data quality. A quantitative statement regarding this issue is the topic of current research. Although these results are by no means a theoretical proof, they give confidence that the methods generalize to different types of buildings.

5. Conclusion

Artificial Neural Networks can be used for demand forecasting in the building and district energy domain. However, their prediction accuracy has high variance and depends on network parameters that are commonly randomly initialized. Here, we have introduced two simple forecast correction methods to significantly reduce this variance without using ensemble methods. The correction methods are based on the autocorrelation of the forecasting error and on online learning.

With the help of a case study of a complex building with district energy system character, we have demonstrated that the methods significantly reduce variance in the prediction performance and also improve prediction accuracy and model bias for a sub-hourly day-ahead heat demand forecasting task. Furthermore, we demonstrated that ANN with both correction methods outperform other grey-box and black-box forecasting methods in the case study. The results were confirmed with case studies of three additional buildings.

We are currently investigating the use of the forecasting approach in a variety of Model Predictive Control and Data Predictive Control settings for building and district energy systems. We intend to use the forecasts as an input for control schemes to provide thermal reserves in low temperature district heating and cooling networks or secondary frequency control in electrical grids based on demand-side management in buildings. Moreover, different modelling approaches, such as ARMAX, are under investigation.

Conflict of interest

None.

Appendix A. Additional KPI

See Figures A.14 - A.16.

References

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., Zheng, X., Brain, G., 2016. TensorFlow: A System for Large-Scale Machine Learning. In: Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16). pp. 265–283. URL https://tensorflow.org.

Aggarwal, C. C., 2018. Text Sequence Modeling and Deep Learning. In: Machine Learning for Text. Springer International Publishing, Cham, pp. 305–360. URL http://link.springer.com/10.1007/978-3-319-73531-3{_}10

Ahmad, T., Chen, H., Guo, Y., Wang, J., apr 2018. A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand: A review. Energy and Buildings 165, 301–320. URL https://www.sciencedirect.com/science/article/pii/S0378778817329225

Amasyali, K., El-Gohary, N. M., jan 2018. A review of data-driven building energy consumption prediction studies. Renewable and Sustainable Energy Reviews 81 (1), 1192–1205. URL https://linkinghub.elsevier.com/retrieve/pii/S1364032117306093
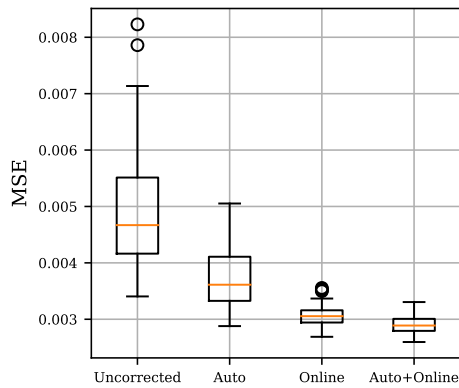
Figure A.14: Mean Squared Error (MSE) of 100 ANN and forecast correction methods
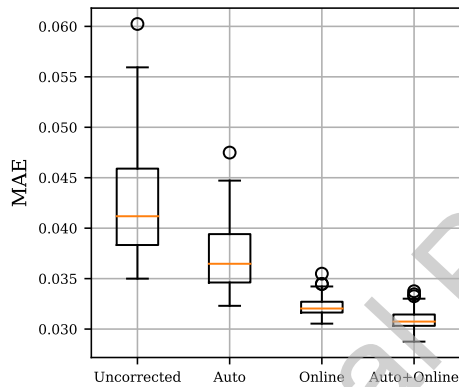

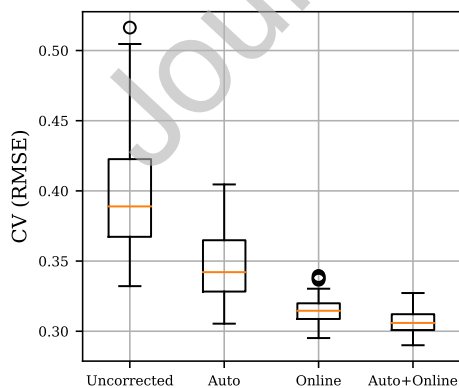Figure A.15: Mean Absolute Error (MAE) of 100 ANN and forecast correction methods

Figure A.16: Coefficient of Variation of Root-Mean Squared Error (CV RMSE) of 100 ANN and forecast correction methods

Arnold, M., Andersson, G., 2011. Model Predictive Control of Energy Storage Including Uncertain Forecasts. Proceedings of the 17th Power Systems Computation Conference (PSCC) 22 (1), 60–71.
URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.722.6663{&}rep=rep1{&}type=pdf

Arnold, M., Negenborn, R. R., Andersson, G., De Schutter, B., jul 2009. Model-based predictive control applied to multi-carrier energy systems. In: 2009 IEEE Power and Energy Society General Meeting, PES '09. IEEE, pp. 1–8.
URL http://ieeexplore.ieee.org/document/5275230/

Bagnasco, A., Fresi, F., Saviozzi, M., Silvestro, F., Vinci, A., 2015. Electrical consumption forecasting in hospital facilities: An application case. Energy and Buildings 103, 261–270.
URL http://dx.doi.org/10.1016/j.enbuild.2015.05.056

Basheer, I., Hajmeer, M., dec 2000. Artificial neural networks: fundamentals, computing, design, and application. Journal of Microbiological Methods 43 (1), 3–31.
URL https://www.sciencedirect.com/science/article/pii/S0167701200002013

Bengio, Y., 2012. Practical Recommendations for Gradient-Based Training of Deep Architectures. In: Montavon, G., Orr, G., Müller, K.-R. (Eds.), Neural Networks: Tricks of the Trade, 2nd Edition. Springer Berlin Heidelberg, pp. 437–-478.

BFE, 2014. Energieverbrauch in der Schweiz und weltweit. Tech. Rep. 4, Bundesamt für Energie BFE.
URL http://www.bfe.admin.ch/php/modules/publikationen/stream.php?extlang=de{&}name=de{_}802079548.pdf

Brück, D., Elmqvist, H., Erik Mattsson, S., Olsson, H., 2002. Dymola for Multi-Engineering Modeling and Simulation. In: 2nd International Modelica Conference. Vol. 55. pp. 1–8.
URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.217.1782{&}rep=rep1{&}type=pdf

Bünning, F., Sangi, R., Müller, D., may 2017. A Modelica library for the agent-based control of building energy systems. Applied Energy 193, 52–59.
URL https://www.sciencedirect.com/science/article/pii/S0306261917300612

Bünning, F., Wetter, M., Fuchs, M., Müller, D., 2018. Bidirectional low temperature district energy systems with agent-based control: Performance comparison and operation optimization. Applied Energy.

Chae, Y. T., Horesh, R., Hwang, Y., Lee, Y. M., 2016. Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings. Energy and Buildings 111, 184–194.
URL http://dx.doi.org/10.1016/j.enbuild.2015.11.045

Chollet, F., 2018. Keras: The Python Deep Learning library. Astrophysics Source Code Library.
URL http://adsabs.harvard.edu/abs/2018ascl.soft06022C

Darivianakis, G., Georghiou, A., Smith, R. S., Lygeros, J., 2017. The Power of Diversity: Data-Driven Robust Predictive Control for Energy-Efficient Buildings and Districts. IEEE Transactions on Control Systems Technology 99, 1–14.
URL http://ieeexplore.ieee.org/document/8100652/

De Coninck, R., Magnusson, F., Åkesson, J., Helsen, L., 2015. Toolbox for development and validation of grey-box building models for forecasting and control. Journal of Building Performance Simulation 9 (3), 1–16.
URL http://www.tandfonline.com/doi/full/10.1080/

19401493.2015.1046933

De Felice, M., Yao, X., aug 2011. Short-term load forecasting with neural network ensembles: A comparative study. IEEE Computational Intelligence Magazine 6 (3), 47–56.

Escrivá-Escrivá, G., Álvarez-Bel, C., Roldán-Blay, C., Alcázar-Ortega, M., 2011. New artificial neural network prediction method for electrical consumption forecasting based on building end-uses. Energy and Buildings 43, 3112–3119.
URL https://ac.els-cdn.com/S0378778811003422/1-s2.0-S0378778811003422-main.pdf?{_}tid=29d377cd-f9fa-405a-bef8-e504638b79e3{&}acdnat=1540288745{_}92407d2f0a7d8804ed29f6951c0068c9

Foucquier, A., Robert, S., Suard, F., Stéphan, L., Jay, A., 2013. State of the art in building modelling and energy performances prediction: A review. Renewable and Sustainable Energy Reviews 23, 272–288.
URL http://dx.doi.org/10.1016/j.rser.2013.03.004

Geidl, M., Koeppel, G., Favre-Perrod, P., 2007. The Energy HubA powerful concept for future energy systems. In: Third annual Carnegie mellon conference on the electricity industry. No. March.
URL https://www.ece.cmu.edu/{~}tanddconf{_}2004/2007/2007ConfPapers/AnderssonPaperfinal.pdf

Hameed Shaikh, P., Bin Mohd Nor, N., Nallagownden, P., Elamvazuthi, I., Ibrahim, T., 2014. A review on optimized control systems for building energy and comfort management of smart sustainable buildings. Renewable and Sustainable Energy Reviews 34, 409–429.
URL http://dx.doi.org/10.1016/j.rser.2014.03.027

Hansen, N., Müller, S. D., Koumoutsakos, P., mar 2003. Reducing the Time Complexity of the Derandomized Evolution Strategy with Covariance Matrix Adaptation (CMA-ES). Evolutionary Computation 11 (1), 1–18.
URL http://www.mitpressjournals.org/doi/10.1162/106365603321828970

Harish, V. S., Kumar, A., 2016. A review on modeling and simulation of building energy systems. Renewable and Sustainable Energy Reviews 56, 1272–1292.
URL http://dx.doi.org/10.1016/j.rser.2015.12.040

Henderson, P., Islam, R., Bachman, P., Pineau, J., Precup, D., Meger, D., sep 2018. Deep Reinforcement Learning that Matters. In: Thirthy-Second AAAI Conference On Artificial Intelligence (AAAI).
URL http://arxiv.org/abs/1709.06560

Huang, J., Smola, A. J., Gretton, A., Borgwardt, K. M., Schölkopf, B., 2006. Correcting Sample Selection Bias by Unlabeled Data. In: NIPS'06 Proceedings of the 19th International Conference on Neural Information Processing Systems. pp. 601–608.
URL http://papers.nips.cc/paper/3075-correcting-sample-selection-bias-by-unlabeled-data.pdf

Jamieson, K., Talwalkar, A., 2016. Non-stochastic Best Arm Identification and Hyperparameter Optimization. In: 19th International Conference on Artificial Intelligence and Statistics (AISTATS). pp. 240–248.
URL https://arxiv.org/pdf/1502.07943.pdfhttp://arxiv.org/abs/1502.07943

Jetcheva, J. G., Majidpour, M., Chen, W. P., 2014. Neural network model ensembles for building-level electricity load forecasts. Energy and Buildings 84, 214–223.

Johansson, C., Bergkvist, M., Geysen, D., Somer, O. D., Lavesson, N., Vanhoudt, D., jun 2017. Operational Demand Forecasting In District Heating Systems Using Ensembles Of Online Machine Learning Algorithms. Energy Procedia 116, 208–216.
URL https://linkinghub.elsevier.com/retrieve/pii/S1876610217322750

Jovanović, R., Sretenović, A. A., Živković, B. D., jun 2015. Ensemble

of various neural networks for prediction of heating energy consumption. Energy and Buildings 94, 189–199.

Kamaev, V. A., Shcherbakov, M. V., Panchenko, D. P., Shcherbakova, N. L., Brebels, A., 2012. Using connectionist systems for electric energy consumption forecasting in shopping centers. Automation and Remote Control 73 (6), 1075–1084.
URL https://link.springer.com/content/pdf/10.1134{%}2FS0005117912060124.pdfhttp://link.springer.com/10.1134/S0005117912060124

Kato, K., Sakawa, M., Keiichi Ishimaru, Ushiro, S., Toshihiro Shibano, oct 2008. Heat load prediction through recurrent neural network in district heating and cooling systems. In: 2008 IEEE International Conference on Systems, Man and Cybernetics. IEEE, pp. 1401–1406.
URL http://ieeexplore.ieee.org/document/4811482/

Kondoh, J., Lu, N., Hammerstrom, D. J., aug 2011. An evaluation of the water heater load potential for providing regulation service. IEEE Transactions on Power Systems 26 (3), 1309–1316.

Kwok, S. S., Lee, E. W., jul 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.
URL https://www.sciencedirect.com/science/article/pii/S0196890411000744?via{%}3Dihub

Leung, M., Tse, N. C., Lai, L., Chow, T., dec 2012. The use of occupancy space electrical power demand in building cooling load prediction. Energy and Buildings 55, 151–163.
URL https://www.sciencedirect.com/science/article/pii/S0378778812004367?via{%}3Dihub

Lydon, G. P., Hofer, J., Svetozarevic, B., Nagy, Z., Schlueter, A., mar 2017. Coupling energy systems with lightweight structures for a net plus energy building. Applied Energy 189, 310–326.
URL https://www.sciencedirect.com/science/article/pii/S030626191631741X{#}f0010

Mat Daut, M. A., Hassan, M. Y., Abdullah, H., Rahman, H. A., Abdullah, M. P., Hussin, F., apr 2017. Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review.
URL https://linkinghub.elsevier.com/retrieve/pii/S1364032116310619

Mattsson, S. E., Elmqvist, H., Otter, M., apr 1998. Physical system modeling with Modelica. Control Engineering Practice 6 (4), 501–510.
URL https://www.sciencedirect.com/science/article/pii/S0967066198000471

Mckinney, W., 2011. pandas: a Foundational Python Library for Data Analysis and Statistics. In: Python for High Performance and Scientific Computing. pp. 1–9.
URL http://pandas.sf.net

Mestekemper, T., Kauermann, G., Smith, M. S., jan 2013. A comparison of periodic autoregressive and dynamic factor models in intraday energy demand forecasting. International Journal of Forecasting 29 (1), 1–12.
URL https://www.sciencedirect.com/science/article/pii/S0169207012000520

Oldewurtel, F., Parisio, A., Jones, C., Morari, M., Gyalistras, D., Gwerder, M., Stauch, V., Lehmann, B., Wirth, K., 2010. Energy Efficient Building Climate Control using Stochastic Model Predictive Control and Weather Predictions. In: Proceedings of the 2010 American Control Conference. Ieee Service Center, 445 Hoes Lane, Po Box 1331, Piscataway, Nj 08855-1331 Usa, pp. 5100–5105.
URL https://infoscience.epfl.ch/record/169733

Ölz, S., Sims, R., Kirchner, N., 2007. CONTRIBUTION OF RENEWABLES TO ENERGY SECURITY. Tech. rep., International Energy Agency.

16

URL https://www.iea.org/publications/freepublications/publication/so{_}contribution.pdf

Park, T. C., Kim, U. S., Kim, L.-H., Jo, B. W., Yeo, Y. K., jul 2010. Heat consumption forecasting using partial least squares, artificial neural network and support vector regression techniques in district heating systems. Korean Journal of Chemical Engineering 27 (4), 1063–1071.
URL http://link.springer.com/10.1007/s11814-010-0220-9

Paudel, S., Elmtiri, M., Kling, W. L., Corre, O. L., Lacarrière, B., feb 2014. Pseudo dynamic transitional modeling of building heating energy demand using artificial neural network. Energy and Buildings 70, 81–93.
URL https://www.sciencedirect.com/science/article/pii/S0378778813007585?via{%}3Dihub

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, É., 2011. Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research 12 (Oct), 2825–2830.
URL http://www.jmlr.org/papers/v12/pedregosa11a.html

Recht, B., 2018. Updates on Policy Gradients arg min blog.
URL http://www.argmin.net/2018/03/13/pg-saga/

Richner, P., Heer, P., Largo, R., Marchesi, E., Zimmermann, M., Zimmermann, M., jan 2018. NEST una plataforma para acelerar la innovación en edificios. Informes de la Construcción 69 (548), 222.
URL http://informesdelaconstruccion.revistas.csic.es/index.php/informesdelaconstruccion/article/view/5879

Saloux, E., Candanedo, J. A., 2018. Forecasting District Heating Demand using Machine Learning Algorithms. In: Energy Procedia.

Shanmuganathan, S., 2016. Artificial neural network modelling: An introduction. In: Studies in Computational Intelligence. Vol. 628. pp. 1–14.
URL http://link.springer.com/10.1007/978-3-319-28495-8{_}1

Silver, D., Schrittwieser, J., Simonyan, K., Antonoglou, I., Huang, A., Guez, A., Hubert, T., Baker, L., Lai, M., Bolton, A., Chen, Y., Lillicrap, T., Hui, F., Sifre, L., Van Den Driessche, G., Graepel, T., Hassabis, D., oct 2017. Mastering the game of Go without human knowledge. Nature 550 (7676), 354–359.
URL http://www.nature.com/doifinder/10.1038/nature24270

Široký, J., Oldewurtel, F., Cigler, J., Prívara, S., sep 2011. Experimental analysis of model predictive control for an energy efficient building heating system. Applied Energy 88 (9), 3079–3087.
URL https://www.sciencedirect.com/science/article/pii/S0306261911001668

Sturzenegger, D., Gyalistras, D., Morari, M., Smith, R. S., jan 2016. Model Predictive Climate Control of a Swiss Office Building: Implementation, Results, and CostBenefit Analysis. IEEE Transactions on Control Systems Technology 24 (1), 1–12.
URL http://ieeexplore.ieee.org/document/7087366/

Sturzenegger, D., Gyalistras, D., Semeraro, V., Morari, M., Smith, R. S., jun 2014. BRCM Matlab Toolbox: Model generation for model predictive building control. In: 2014 American Control Conference. IEEE, pp. 1063–1069.
URL http://ieeexplore.ieee.org/document/6858967/

Suganthi, L., Samuel, A. A., feb 2012. Energy models for demand forecastingA review. Renewable and Sustainable Energy Reviews 16 (2), 1223–1240.

URL https://www.sciencedirect.com/science/article/pii/S1364032111004242

Suryanarayana, G., Lago, J., Geysen, D., Aleksiejuk, P., Johansson, C., 2018. Thermal load forecasting in district heating networks using deep learning and advanced feature selection methods. Energy.

Tobin, J., Fong, R., Ray, A., Schneider, J., Zaremba, W., Abbeel, P., mar 2017. Domain randomization for transferring deep neural networks from simulation to the real world. In: IEEE International Conference on Intelligent Robots and Systems. Vol. 2017-Septe. pp. 23–30.
URL http://arxiv.org/abs/1703.06907

Wang, Z., Srinivasan, R. S., aug 2017. A review of artificial intelligence based building energy use prediction: Contrasting the capabilities of single and ensemble prediction models. Renewable and Sustainable Energy Reviews 75, 796–808.
URL https://linkinghub.elsevier.com/retrieve/pii/S1364032116307420

Warrington, J., Goulart, P. J., Mariethoz, S., Morari, M., dec 2012. Robust reserve operation in power systems using affine policies. In: 2012 IEEE 51st IEEE Conference on Decision and Control (CDC). IEEE, pp. 1111–1117.
URL http://ieeexplore.ieee.org/document/6425913/

Warrington, J., Hohl, C., Goulart, P. J., Morari, M., jun 2014. Optimal unit commitment accounting for robust affine reserve policies. In: 2014 American Control Conference. IEEE, pp. 5049–5055.
URL http://ieeexplore.ieee.org/document/6858800/

Wojna, Z., Gorban, A. N., Lee, D. S., Murphy, K., Yu, Q., Li, Y., Ibarz, J., apr 2018. Attention-Based Extraction of Structured Information from Street View Imagery. In: Proceedings of the International Conference on Document Analysis and Recognition, ICDAR. Vol. 1. pp. 844–850.
URL http://arxiv.org/abs/1704.03549

Yudong Ma, Borrelli, F., Hencey, B., Coffey, B., Bengea, S., Haves, P., may 2012. Model Predictive Control for the Operation of Building Cooling Systems. IEEE Transactions on Control Systems Technology 20 (3), 796–803.
URL http://ieeexplore.ieee.org/document/5739562/

Yun, D., Lee, H., Choi, S. H., dec 2018. A deep learning-based approach to non-intrusive objective speech intelligibility estimation. IEICE Transactions on Information and Systems E101D (4), 1207–1208.
URL http://arxiv.org/abs/1412.6980

Zhao, H.-X., Magoulès, F., 2012. A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews 16, 3586–3592.
URL https://ac.els-cdn.com/S1364032112001438/1-s2.0-S1364032112001438-main.pdf?{_}tid=d24ff951-1882-4611-b89e-5ee08d2c0510{&}acdnat=1540308093{_}c9e893f0b56cb633368b99a0f42d87bc

Zhou, Q., Wang, S., Xu, X., Xiao, F., 2008. A grey-box model of next-day building thermal load prediction for energy-efficient control. INTERNATIONAL JOURNAL OF ENERGY RESEARCH Int. J. Energy Res 32, 1418–1431.
URL www.interscience.wiley.com

Felix Bünning: Conceptualization, Methodology, Software, Investigation, Writing - Original Draft, Visualization

Philipp Heer: Conceptualization, Resources, Writing - Review & Editing, Supervision, Project administration

Roy S. Smith: Conceptualization, Resources, Writing - Review & Editing, Supervision, Project administration

John Lygeros: Conceptualization, Resources, Writing - Review & Editing, Supervision, Project administration

DECLARATION OF INTEREST

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

We confirm that the manuscript has been read and approved by all named authors and that there are no other persons who satisfied the criteria for authorship but are not listed. We further confirm that the order of authors listed in the manuscript has been approved by all of us.

We confirm that we have given due consideration to the protection of intellectual property associated with this work and that there are no impediments to publication, including the timing of publication, with respect to intellectual property. In so doing we confirm that we have followed the regulations of our institutions concerning intellectual property.
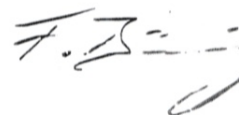
We understand that the Corresponding Author is the sole contact for the Editorial process (including Editorial Manager and direct communications with the office). He/she is responsible for communicating with the other authors about progress, submissions of revisions and final approval of proofs. We confirm that we have provided a current, correct email address which is accessible by the Corresponding Author and which has been configured to accept email from *felix.buenning@empa.ch*.

Signed by all authors as follows:

Felix Bünning   Date: 13.05.2019   Signature:

Andrew Bollinger   Date: 14.05.2019   Signature:

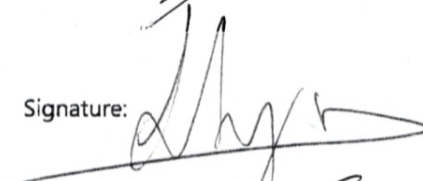Philipp Heer   Date: 15. 5. '19   Signature:

John Lygeros   Date: 20/5/2019   Signature:

Roy S. Smith   Date: 22-05-2019   Signature: