Machine Learning Recommendations for Control of Complex Building Systems Using Weather Forecasts

Paul Westermann[1], Nigel David[2], Ralph Evins[1]
[1]Energy Systems and Sustainable Cities group, University of Victoria
[2]SES Consulting Inc., Victoria

Abstract: *We present a machine learning model used to provide recommendations on chiller operation based on the prediction of cooling demand using a weather forecast. A long short term memory (LSTM) formulation was used, and achieved favourable results compared to a standard approach. The model captured the data to a reasonable extent (R^2 = 0.70), but was unable to predict very high loads at unexpected times. The model is intended to be used as an aid to a human operator, not as a replacement, and it is likely that many of these unexpected events could be overridden by the operator. Overall, the predictive model reduced the number of occasions in which a chiller was operating unnecessarily by 80.5%, or 469 hours. This demonstrates the power of data-driven predictive control to assist in the efficient operation of complex building systems, saving money, energy and operator time.*

Keywords: *Energy management system, human-in-the-loop control, Machine Learning*

Introduction

Non-residential, commercial and institutional buildings consume large amounts of energy. In Canada, they account for more than 10% of the end-use energy consumption (NRCan, 2017). Furthermore, in 2009 their energy intensity exceeded the ones of average Canadian households by almost 40% (NRCan, 2012).
Due to the sheer size of the buildings a share of 20% of energy in the total operating costs depicts a large amount of absolute energy payments. Hence, a strong driver to implement building retrofit measures exists. Regarding the complexity of heating and cooling systems of those large buildings, it is common that a specialised energy manager supervises the operation of all systems. Based on many factors like weather, occupant behaviour and other disturbances he decides which systems are switched on and how the temperature set-points are selected.

Recent advances in software development as well as increasing amounts of available data are promoting the development of methods to support or even automate this human-based energy management of complex building systems. One approach which received a lot of attention in building control is model predictive control (MPC) as shown by Oldewurtel et al. (2012). MPC uses a physical model to determine optimal system inputs for the upcoming hours or days. At each hour, the system inputs are optimised following certain objectives (e.g. lowering cost or energy consumption) subject to human comfort constraints. The approach takes inputs like weather conditions, electricity prices and occupancy patterns into account. However, wide application is yet to come, especially, because deriving a physical model of the building is work intensive. It requires to transfer all architectural and building system information of into a model. Furthermore, the modeller has to make assumptions about typical building operation patterns which may lead to modelling errors.

Hand in hand with the wide spread rise of machine learning (ML), building scientists have explored ways to exploit sensor data in buildings to train models capable of predicting future building behaviour and its energy consumption. Accurate forecasts can be used in a similar fashion as proposed by MPC, i.e. to optimise heating and cooling system inputs. Wei et al. (2018) explain the fundamentals and give a broad review on existing data driven models. A good example on the application of the most common ML model, artificial neural network (ANN), is given by Jetcheva et al. (2014), who generated an ensemble of multiple ANN models. Massana et al. (2015) address which kind of data is required to forecast building energy demand accurately. Especially, they highlight the importance of occupancy data.

Both physical model-based and ML-based optimisation of the building energy consumption struggle with the stochastics laying in building energy consumption due to multiple factors like weather forecast, or occupant behaviour uncertainty. One pathway in MPC research is the application of stochastic MPC Oldewurtel et al. (2010), where uncertainties and constraint violations are considered probabilistically to allow acceptable levels of risk in optimal control. In data driven research, methods which provide probabilistic forecasts, e.g. Gaussian

ISBN 978-2-921145-88-6

Processes, attract a lot of attention (Gray and Schmidt, 2016). Another option is the use of human domain knowledge as it is suggested for example in the field of crowd sourcing where human intelligence is used in cases when machine intelligence is less effective (Kamar et al., 2012). In case of buildings, one could exploit the knowledge of an on-site energy manager to remove some uncertainty the ML model fails to take into account. As an example, an energy manager of an office building may know of extraordinary events like meetings or conferences and thus, would know that heating demand will be higher than predicted by a data-driven model. Incidents like this are not straight forward to incorporate into a ML model.

We complement the existing literature with an application of ML based human-in-the-loop control. We use ML to generate point forecasts of the energy demand of a building and give recommendations to an energy manager on-site to switch individual systems on and off. The goal is to reduce the number of hours, when more systems run than required, which causes overall efficiency losses. We aim for a solution that avoids complex feature acquisition and extraction but only uses weather forecasts provided by an API as well as time information. Therefore, our approach offers a plug-and-play solution being highly generalizable as the model features do not need to be customized to changing available sensor data on different buildings.

Case study building

This study uses ML to optimise the operation of a complex cooling system of a large building in British Columbia, Canada. The system is equipped with three chillers to guarantee comfort within the building. Currently, the control of the cooling system is determined by an energy manager who decides on the use of the three different chillers. The chillers are all of different types with different efficiencies as shown in Table 1 and Figure 2. A simple operation routine for the three chillers is chosen by the building operator:[1]

$$\begin{aligned} 0t \leq D < 1100t \quad & \textit{Base operation}\text{: only chiller A} \\ 1100t \leq D < 3000t \quad & \textit{Co-operation}\text{: chiller A \& B} \\ 3000t \leq D \quad & \text{All chillers} \end{aligned} \tag{1}$$

As chiller C is supposed to act as a back-up chiller, which is switched on only in rare events like maintenance, outages of other chillers or heat waves, its operation is excluded in the following study. The given routine represents the most efficient operation strategy if the co-operation of both chillers runs optimally (see Figure 2). At the threshold of $1100t$ the co-operation becomes more efficient than the use of chiller A only. The optimisation of the simultaneous operation of both chillers is not part of this study, but we focus on the prediction of whether the *co-operation* of both chillers is necessary.

[1]The unit used is refrigeration ton t (also: RT). It originates from refrigeration with natural ice. $1t$ corresponds to the refrigeration supplied by melting one short ton of water over 24 hours(Avallone et al., 2006).

Name	System	Properties
Ch. A	Centrifugal chiller with variable speed drive (VSD) and condenser relief	*most efficient, good for changing loads*
Ch. B	Centrifugal chiller with with condenser relief	*less efficient worse for changing loads*
Ch. C	Back-up chiller (absorption chiller)	*least efficient, back-up or during heat waves*

Table 1: Overview of chiller systems of case study building.

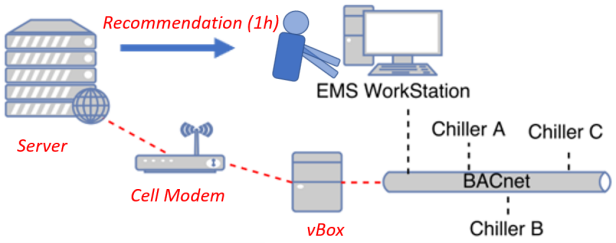

Figure 1: System architecture as introduced in Evins (2017).

Based on the considered data set (8760 hours), we observe that the actual operation did not always follow the presented routine which may have lead to significantly higher operating costs and energy use. It is found that both chillers run during 1998 hours however only during 1484 both chillers were actually needed, i.e. in 583 hours two chillers were running although only one chiller was required. Furthermore, ”cold” start of the second chiller occurred, leading to demand peaks which generates significant demand charges for the building operator (up to CAD$ 14k were reported).

The scope of the problem is therefore using demand forecasts to give recommendations to the energy manager to reduce

1. the number of hours of unnecessary *co-operation*,
2. peak demands.

The latter can be avoided by slow-starting the second chiller prior to the peak demand event (if the high peak demand is forecasted by the model) or by shifting the additional starting load to low demand hours (*load shifting*).

Human-in-the-loop control

Instead of developing a fully independent and automated control algorithm, this study proposes the use of ML based recommendations to support the building energy manager. Based on the recommendation, he decides to either switch

ISBN 978-2-921145-88-6

a chiller on or off. Once the system is switched on, it is controlled autonomously. As the energy manager fundamentally takes part in that control scheme we classify it as a human-in-the-loop approach.

The architecture of the intended control loop is shown in Figure 1. The ML model training and predictions are conducted on an external server connected to the internet. The server is also used to collect energy demand data and information on which chillers are running. The data is gathered via a mini computer running Volttron software, which is designed to access the BACnet of the building and retrieve data.[2] The transmission of the data happens over a cell modem.
Based on the information of which chiller is running the ML model recommends each hour of the day if chiller A or chiller B needs to be switched on or off. In case the building energy manager takes action, he accesses the energy management system (EMS) to control the individual chiller units. For practicability reasons this could also be done using wearables like a smart watch.

Having a human in a control loop, who is adjusting supervisory control parameters, is well understood (Stankovic, 2014). In those applications, the control loop is running autonomously with a human only intervening when it is necessary. An example for human-in-the-loop control in the building context may be found in Mirebrahim et al. (2017).

It has to be noted, that the provided solution is appropriate for large commercial and institutional buildings like hospitals, malls, offices and others. Due to the complexity of the heating and cooling system as well as the large uncertainty introduced by human activity, an interplay of machine intelligence and an energy manager seems promising in this case. In comparison, in buildings with low uncertainty in demand(e.g. data centres) pure machine learning methods have proven great success in lowering the energy demand. The company DeepMind reported a reduction of 40% in cooling demand using ML control.[3]

Energy demand forecasting

Methodologies in energy demand forecasting have a broad application including the optimisation of power systems or regarding buildings, either early building design optimisation or system control optimisation as found in this paper.
In this study we derived three models and compare them. In addition to two neural network models, a feed-forward neural network and a long short term memory (LSTM) model, we also included a common linear regression model. The models were programmed using Keras (Chollet et al., 2015) and Tensorflow (Abadi et al., 2015).
All models map weather data as well as time information (inputs) to cooling demand (outputs).[4] We received hourly chiller use data for 1 year (Dec. 2015 to Nov. 2016). For the same period weather data is available online via an API (www.vancouver.weatherstats.com) or was computed (sun position). Due to availability, we used weather measurements instead of forecast data. Thus, some uncertainty given by the inputs is ignored in this study and needs to be addressed in future. The heat map in Figure 3 shows the correlation of all used weather inputs to cooling demand.
The non-recursive structure of our models enables to choose the prediction horizon. For each input of hourly weather data the corresponding hourly cooling demand is given as an output of the model. Furthermore, no sensor data on the building state (e.g. temperature, internal gains, etc.) or on chiller A and B is used for predictions.

Traditional feed-forward neural networks are one of the most common ML algorithms. Each network consists of multiple layers with multiple cells (neurons). In comparison to LSTM networks they do not feature the capability to change an internal state depending on previous model outputs. Nonetheless, they have shown to be suitable for building energy demand predictions. Wei et al. (2018) give more explanations and a review on existing studies.

LSTM neural networks (Hochreiter and Schmidhuber, 1997) are part of the group of recurrent neural networks (RNN). RNN received a lot of attention in the time-series forecasting domain because they can store previous model outcomes. RNN are usually trained using back-propagation algorithm in real-time recurrent learning. Those algorithms are prone to vanishing or exploding gradients. The LSTM algorithm overcomes this problem and thus, it is popular in the ML world. A LSTM network consists of multiple cells whose outputs are looped back to the cell. Current model outputs can be stored in the cell via an internal state. This internal state is overwritten, erased or read at each model evaluation step.
LSTM cells can be stacked in a multi-layer architecture. Here, the final model architecture was determined by using a simple grid search. We explored different number of cells per layer (10 or 50), different number of layers (1,2) and different sizes of training batches (50 hours to 320 hours). The latter is specifically interesting as it determines on which size of individual batches the LSTM model is trained. The impact of the batch size had the most significant impact on the training performance. The final model consists of two layers with 50 cells each and was trained during 200 epochs on batches of 200 hours each.

[2] https://volttron.org/
[3] https://deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-40/
[4] For the specific building we also provide information if and to which extent chiller C is used to control for special events like maintenance.

ISBN 978-2-921145-88-6

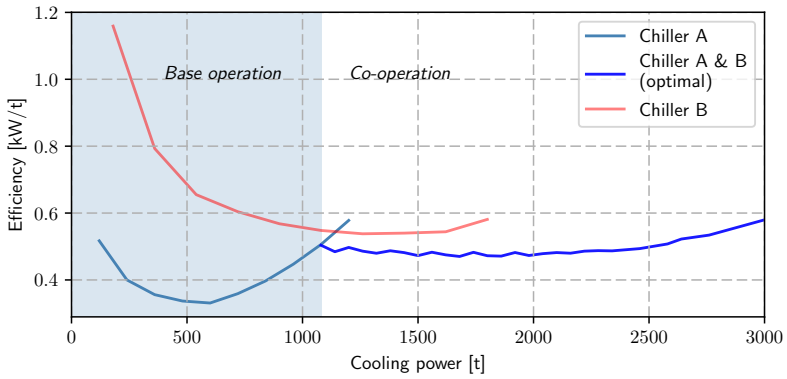

Figure 2: Overview of efficiency of Chiller A and B as a function of cooling power. Efficiency is quantified by the amount of electricity required to provide one refrigeration ton [kW/t].

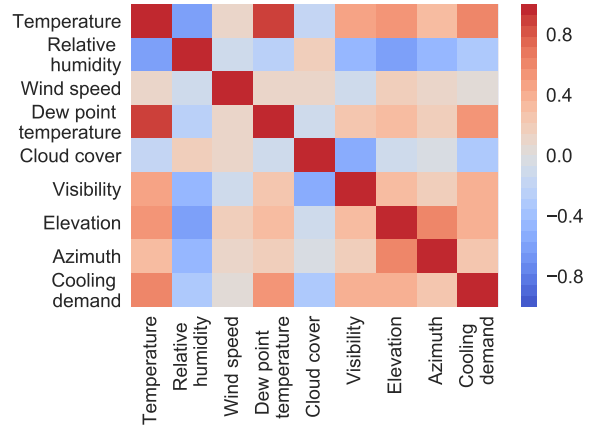

Figure 3: Heat map of correlation between weather inputs and cooling demand.

To generate outcomes with the model, the whole set of previous inputs has to be considered as the model predictions depend on the internal state of the LSTM, which depends on previous model outputs.

Results

The models were trained on a sequential block of data (70% of the data) and tested on the last block of the data set(30%). Last block testing is necessary for time-series data to receive an accurate estimation of the model performance (Bergmeir and Benitez, 2012). The training set consists of data from December to August and the test set from August to the end of November. The period of the test set covers late summer, autumn and the beginning of winter as such a large variety of cooling operation is included. Figure 4 shows that in the beginning of the test set consistently a high demand for cooling whereas later only occasional peaks occur.

	LR	**ANN**	**LSTM**
R^2_{train}	0.55	0.72	0.72
R^2_{test}	0.43	0.64	0.70

Table 2: Performance of applied models. The one of the LSTM surpassed the ones of the LR and ANN model. It is the most accurate one with the lowest degree of overfitting. The mean absolute error of the LSTM predictions on the test set is 24.2%.

The LSTM model surpassed the performance of all other models (see Table 2). The linear regression model had the worst performance which confirms that there are non-linearities in the relationship of cooling demand and weather data. The feed-forward network structure enables to capture any non-linearity and performed significantly better than a linear model. The best-performing (based on test data performance) neural network of the candidates in our grid search results in some overfitting as performance on the test data is significantly worse than on the training data. Similar overfitting behaviour was observed in the linear regression model. The LSTM model only showed weak overfitting. It seems that a network with memory for previous outputs is specifically well suited to our problem as no building state data (e.g. temperature) is given as an input to the model.

The performance of the LSTM model on the test data is shown in Figure 4. We find that in the beginning the model follows the real values well. However, at the end of the data two demand spikes occur which the model does not capture well. Looking into the data, it was observed that weather data does not indicate those peaks, i.e. temperature and solar gains do not show any abnormally high values. The model failure in those cases shows that eventually further input data on internal heat gains of the building would be required unless the building manager can compensate these inaccu-

ISBN 978-2-921145-88-6

racies. Overall, a mean absolute error of 24.2% was found for the LSTM model on the test data. If the demand peaks are not taken into account this value would be significantly better.

Another perspective on the model behaviour is given in Figure 5 where the real data is on the X-axis and predictions on the Y-axis. All perfect predictions lay on the 45° line. We see the strong correlation of predictions to the real data. However, two distinct areas are determined where the model struggles to give accurate predictions (see red dashed ellipses). The vertical ellipse shows the case when cooling demand was low and the chillers were close to idling operation, but the model predicts significant cooling. The horizontal ellipse indicates the opposite, i.e. the model predicts idling chiller operation but in reality, significant cooling was supplied. Again, this behaviour may be caused by a lack of information on occupant behaviour in the building. Either internal gains are higher than expected (horizontal ellipse) or lower than expected (vertical ellipse). Furthermore, a change in user comfort (i.e. changing temperature set points in multiple rooms) may cause modelling errors.

Recommendations for chiller operation

The predictions were processed to recommend if either chiller A or both chillers need to be switched on. Here, we programmed this process corresponding to the control routine defined above: any prediction of a cooling demand above 1100t means that the model recommends *co-operation* of both chillers. The resulting recommendations are shown in Figure 6 and emphasize that a ML approach to the problem is very suitable.

The three columns in the Figure represent which operation would be optimal according to the defined control routine (left), which operation was observed in the data (center) and which operation our prediction model suggests. We used boolean operators to filter the following cases:

1. The control routine suggests chiller A (*base operation*) is running AND only chiller A is observed/recommended to run (light blue).
2. The control routine suggests chiller A (*base operation*) is running AND both chillers are observed/recommended to run (green).
3. The control routine suggests both chillers (*co-operation*) are running AND both chillers are observed/recommended to run (white).
4. The control routine suggests both chillers (*co-operation*) are running AND only chiller A is observed/recommended to run (red).

The second case represents the erroneous switching on of two chillers during low demand hours. It was observed that in 92.9% only one chiller was running when the actual cooling demand was lower than 1100t (light blue area) and our prediction model recommended in 98.4% of the low demand hours to only use one chiller. In 7.1% of the cases it was found in the observed data that both chillers were running. The prediction model suggested only in 1.6% of the low demand hours to use two chillers (compare green area of center and right bar). With the prediction model 400 hours of unnecessary running of two chillers could have been avoided.

The third and fourth case addresses the prediction of the cooling operation during high demand hours (white and red area). In case the demand was between 1100t and 1200t (maximum capacity of chiller A), it was observed that only chiller A was running. For simplicity, we omit those samples (27) from this analysis, as the efficiency losses are small compared to the other cases. Therefore, these samples are not shown in Figure 6 and the white area of the left and the center bar are equally large. However, in the right bar a large red area is shown. This area shows that the prediction model recommends for 48.7% of the

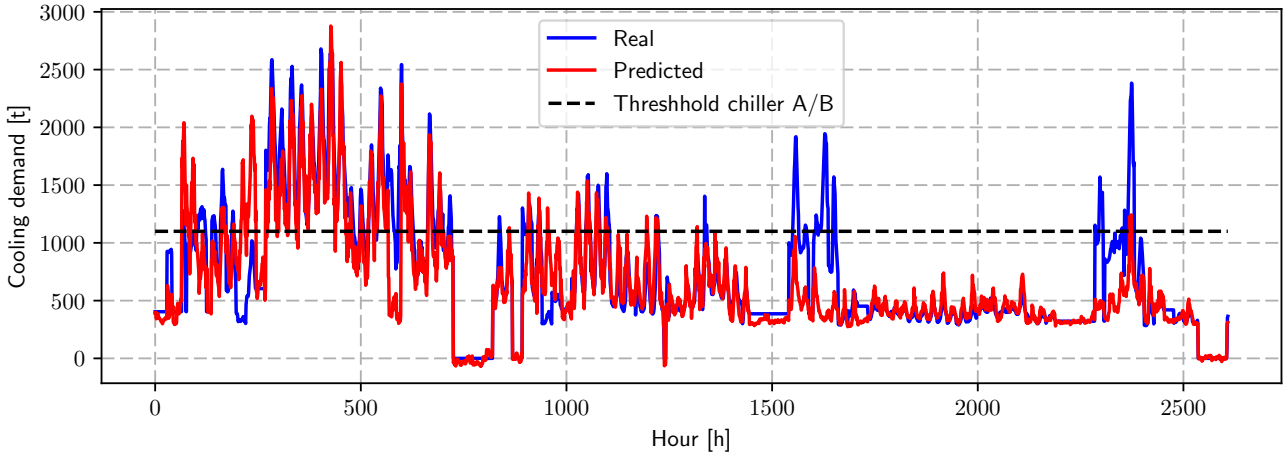

Figure 4: Sequential display of predicted and observed cooling demand of the test data set.

ISBN 978-2-921145-88-6

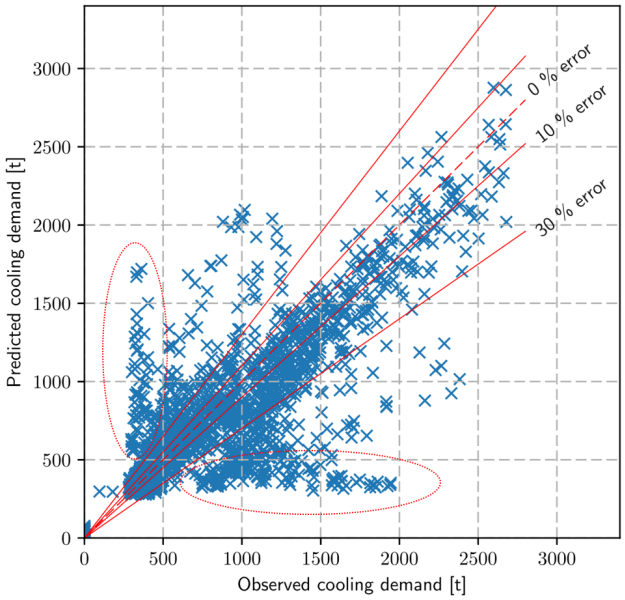

Figure 5: Predicted and observed cooling demand for test data set. The predictions which match the observed cooling demand lie on the 45° line. Ellipses indicate two distinct cases when the model fails to predict demand accurately. The horizontal ellipse highlights the area where the model predicts an idling chiller, but significant cooling is provided; the vertical ellipse corresponds to the opposite case.

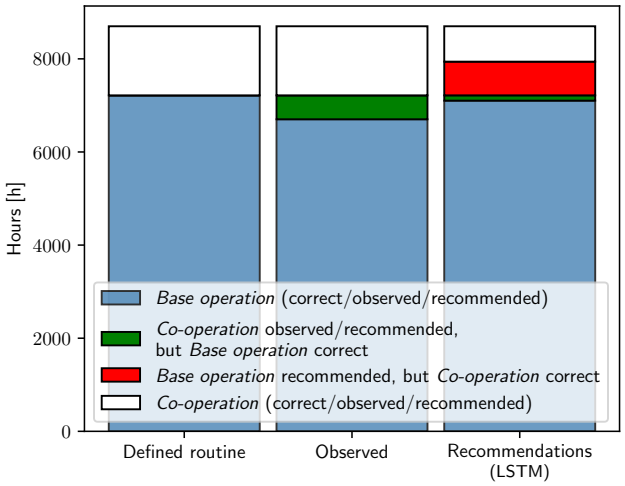

Figure 6: Performance of ML-based recommendations. Observed controls and recommendations are both compared to the optimum operation given by the previously defined control routine (see Eq. 1).
The unnecessary co-operation of two chillers (green section) is reduced if one follows the recommendations of the LSTM model. The red section shows that the LSTM recommendations sometimes fails to predict that two chillers are required. This emphasizes that even by using ML predictions, still unexpected cold starts of the second chiller may occur.

ISBN 978-2-921145-88-6

high demand hours that only one chiller is necessary. This possibly leads to cold starts of the second chiller and electricity demand spikes. In 51.3% of the time, the model predicts high demand well and prepares the energy manager for the upcoming need for the second chiller. This enables to slowly ramp up Chiller B or shift the incurred starting load to low demand hours. Nonetheless, without the prediction model no recommendation for the future are given and hence, the correct prediction in 51.3% of the time represents a benefit for the building.

Conclusions and Future work

The ultimate goal in building control research is optimal automated control of buildings. Although there are promising paradigms in development there is still significant work to be done.

We offer a simple approach to improve HVAC control by exploiting a combination of building operator's domain knowledge and the power of current ML algorithms. In this human-in-the-loop control scheme, a LSTM model is trained to predict energy demand of the building based only on weather forecast data. The demand predictions are used to give control recommendations to the building operator who can override recommendations if he receives additional information, which are not considered by the ML model as for example extraordinary occupant behaviour like conferences or meetings.

In this study, we applied that scheme to optimise chiller control. The cooling demand of a large commercial building was predicted with a LSTM model trained on one year of data. The model gives recommendations whether one or two chillers should be switched on. We compared the recommendations of the model to the observed decisions taken by the building operator: It was found that the model would have reduced the unnecessary use of the second chiller by 469 hours within one year (80.5% reduction). Furthermore, it correctly predicted the need for the second chiller in 51.3% (760 hours) of the cases it was required.

However, the model failed to predict the need for the second chiller in 48.7% of the cases. We investigated those errors and found that cooling was abnormal considering the associated weather inputs. This points out that further information on special happenings in the building are required which a building operator has. In those cases he should override model recommendations.

We aimed for a solution which can be generalized to a large variety of buildings. On the one side, the suggested method only uses weather forecast data and thus, no elaborate studying and selection of available sensor as model inputs has to be conducted. On the other, our approach requires good knowledge of the individual building system set-up. Furthermore, only buildings which are supervised by an energy manager are suitable to fully benefit.

The next step in research is to deploy hardware in the building (see Evins, 2017) to provide actual recommendations to the building operator. Recommendations of the model and control decisions of the operator need to be recorded. This will give understanding how well the collaboration of model and human works. It will be interesting to see if the accuracy of the model, which only relies on weather forecasts as inputs, is sufficient such that recommendations are accepted and in periods when the ML model is inaccurate, if the energy manager overrides the recommendations.

References

Abadi, M. et al. (2015). *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org.

Avallone, E., I. Baumeister, and A. Sadegh (2006). *Marks' Standard Handbook for Mechanical Engineers. 10*. New York: McGraw-Hill.

Bergmeir, C. and J. M. Benitez (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences 191*, 192–213.

Chollet, F. et al. (2015). Keras. `https://github.com/keras-team/keras`.

Evins, R.; David, N. (2017). Using simple predictive models to improve control of complex building systems. In *ACM International Conference on Systems for Energy-Efficient Built Environments (BuildSys), Delft, The Netherlands.*

Gray, F. M. and M. Schmidt (2016). Thermal building modelling using gaussian processes. *Energy and Buildings 119*, 119–128.

Hochreiter, S. and J. Schmidhuber (1997). Long short-term memory. *Neural computation 9*(8), 1735–1780.

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

Kamar, E., S. Hacker, and E. Horvitz (2012). Combining human and machine intelligence in large-scale crowdsourcing. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems-Volume 1*, pp. 467–474. International Foundation for Autonomous Agents and Multiagent Systems.

Massana, J., C. Pous, L. Burgas, J. Melendez, and J. Colomer (2015). Short-term load forecasting in a non-residential building contrasting models and attributes. *Energy and Buildings 92*, 322–330.

ISBN 978-2-921145-88-6

Mirebrahim, S. H., M. Shokoohi-Yekta, U. Kurup, T. Welfonder, and M. Shah (2017). A clustering-based rule-mining approach for monitoring long-term energy use and understanding system behavior. In *ACM International Conference on Systems for Energy-Efficient Built Environments (BuildSys), Delft, The Netherlands.*

NRCan (2012). Survey of commercial and institutional energy use – buildings 2009, *Detailed Statistical Report*. Technical report, Natural Resources Canada.

NRCan (2017). Energy fact book 2016-2017. Technical report, Natural Resources Canada.

Oldewurtel, F., A. Parisio, C. N. Jones, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and M. Morari (2012). Use of model predictive control and weather forecasts for energy efficient building climate control. *Energy and Buildings 45*, 15–27.

Oldewurtel, F., A. Parisio, C. N. Jones, M. Morari, D. Gyalistras, M. Gwerder, V. Stauch, B. Lehmann, and K. Wirth (2010). Energy efficient building climate control using stochastic model predictive control and weather predictions. In *American control conference (ACC), 2010*, pp. 5100–5105. IEEE.

Stankovic, J. A. (2014). Research directions for the internet of things. *IEEE Internet of Things Journal 1*(1), 3–9.

Wei, Y., X. Zhang, Y. Shi, L. Xia, S. Pan, J. Wu, M. Han, and X. Zhao (2018). A review of data-driven approaches for prediction and classification of building energy consumption. *Renewable and Sustainable Energy Reviews 82*, 1027–1047.

ISBN 978-2-921145-88-6