MODEL IDENTIFICATION FOR THE CONTROL OF NATURALLY VENTILATED BUILDINGS

Joshua S. Sykes¹, E. Abigail Hathway¹, Peter Rockett²

¹Department of Civil and Structural Engineering
²Department of Electrical Engineering
University of Sheffield, Sheffield, United Kingdom

ABSTRACT

In this paper, predictive models are developed to enable the application of model predictive control (MPC) to naturally ventilated buildings. The essential component of an MPC strategy is the predictive model of the building’s thermal dynamics, which is the focus of this study. An empirical approach is taken using multilayer perceptron (MLP) neural network models. The models presented were generated using data gathered from real buildings during operation and building simulation data generated using EnergyPlus. The resulting models were able to accurately predict internal conditions such as zone temperature. The problem of insufficient input excitation is highlighted and an identification procedure to overcome it is presented.

INTRODUCTION

Energy costs, climate change, mounting political and social pressure are examples of some of the drivers for the increasing attempts to reduce energy consumption. Buildings account for around 40% of total final energy consumption in developed countries, (Perez-Lombard et al., 2008), and in European countries around 76% of the energy consumed by buildings is used for comfort control, i.e. heating, ventilation and air conditioning (HVAC) (International Energy Agency, 2008). Reducing the amount of energy required by HVAC systems can be approached in a number of ways, for example, increasing airtightness, better insulation, increasing appliance efficiency, passive ventilation techniques, etc. In addition to energy concerns, there has been a growing awareness of the impact of indoor environmental quality (IEQ) upon occupants’ wellbeing (ASHRAE, 2013). IEQ refers to the quality of a building’s environment in relation to the health and wellbeing of those who occupy the space (CDC, 2013). There is a number of factors which contribute to IEQ including: air quality, temperature, lighting, contaminants etc.

Natural ventilation is the process of supplying and removing air to/from an indoor space without the aid of mechanical systems. Natural Ventilation is driven by pressure differences caused by wind, or temperature gradients (Awbi, 2003). As natural ventilation is affected by a number of factors, (such as external temperature, wind speed, wind direction and internal temperatures), it can be hard to predict the consequence of opening a window or vent, making control of naturally ventilated spaces more challenging than mechanically ventilated or air-conditioned spaces (Thomas, 2006). In this paper, we consider a control method which has the potential to reduce energy consumption and optimise occupant comfort in naturally ventilated spaces. Model Predictive Control (MPC) is a control method which originated in the process industries (Camacho and Bordons, 2007). MPC utilises a system model to optimise future outputs based upon possible inputs over a finite receding time horizon. At each time step, a minimisation of some objective function is carried out to determine the optimal control signals over a finite horizon. At each iteration, only the first step of the control strategy is then implemented. The control horizon is then shifted one step forward and the process repeated ad infinitum (Camacho and Bordons, 2007).

MODELLING

Modelling Strategy

Previous studies have investigated the potential to apply MPC techniques to HVAC systems. Existing work has focussed predominantly on applying MPC to mechanically ventilated buildings with a limited number of studies on mixed-mode spaces. In the current paper, the application of MPC to naturally ventilated spaces is investigated. Application of MPC to naturally ventilated spaces is likely to be more difficult compared with a mechanically ventilated scenario. With mechanical ventilation there is always some measure of how much cooling is being delivered in a space (e.g. fan power). However, with natural ventilation the cooling is provided by opening windows. In this scenario, the effect of the control action is highly changeable due to the number of variables that influence flow rate, such as temperature differences and wind speed. While the basic modelling procedure demonstrated in this paper is similar to previous studies on mechanical systems, a more complex identification procedure is carried out in order to incorporate the effect of opening windows.

In order for MPC to be successful, an accurate model of the system is required. The model should be as
simple as possible and have good prediction characteristics over the control horizon (Shook et al., 2002), (Lauri et al., 2010). There are two main approaches to system modelling which can be taken when applying MPC to HVAC systems. One approach is the use of first-principles models. These models are based upon our knowledge of the physical processes taking place within the building. In early applications, the first-principles models used were relatively simple linear models. Initial studies applied first-principles models to individual components in HVAC systems and then to simplified single-zone buildings and HVAC systems (Wang and Jin, 2000), (Yuan and Perez, 2006).

Recently there has been an increasing use of building energy modelling tools, typically multizone-network models such as EnergyPlus, TRNSYS etc. (Zhao et al., 2013), (May-Ostendorp et al., 2011). As with simple linear models, these models are based upon our knowledge of the physical processes taking place within the building. However, the use of building energy modelling tools allows for more complex building geometry and system modelling.

The alternative to the first-principles models is the use of ‘black-box’ data-driven models. These models are typically less computationally intensive to use, and once a suitable workflow has been devised, relatively simple to create. Empirical models have the advantage of modelling the processes which are actually happening within a space without including the assumptions which are necessary with a first-principles model. For example, with a simulation tool, such as EnergyPlus, it is possible to include detailed occupancy and activity schedules but it will be hard to fully capture the stochastic manner in which occupants interact with the building and their effect upon the building’s thermal environment. Empirical models can allow the relationship between variables to be mapped without calculating further, hard to predict variables. For example, in naturally ventilated buildings, it can be difficult to determine air flow rates; by using an empirical approach this becomes unnecessary as the effect of the weather conditions and window controls on indoor temperature can be modelled without the need for further information. Additionally, as we move towards “smart buildings”, there are increasing amounts of data available about how buildings are actually running, which have the potential to enable a data-driven approach.

A black-box approach to system modelling has a number of advantages over first-principles models. However, there is one significant disadvantage of the black-box approach. Often the inputs are insufficiently excited, and thus data collected from buildings during normal operation can fail to capture some important physical properties, resulting in a model inappropriate for MPC (Cigler and Privara, 2010). The need to carry out a specific identification experiment, whereby the inputs are excited, is often given as a reason to discount black-box modelling of HVAC systems (Cigler and Privara, 2010).

In this study, the initial models were generated using data collected from buildings during normal operation. Upon analysing the resulting models, lack of input excitation was found to be a problem. Although the resulting models were able to accurately predict future temperatures, the models did not capture the effect of the control input (window actuators). To investigate how this may be overcome, an identification experiment was carried out using EnergyPlus.

Neural Network Modelling

In this paper, we take a data-driven approach using multilayer perceptron (MLP) neural networks to predict zone temperatures in naturally ventilated spaces. Neural Networks have been used in previous studies for control of HVAC systems (Kusiak and Xu, 2012) and automated window blinds (Chen et al., 2009). According to Haykin (1998), neural networks are perhaps the most well-known class of nonlinear models. A multilayer neural network model is shown in Figure 1. An MLP neural network consists of multiple layers of nodes, where each layer is fully connected to the next. With the exception of the input nodes, each node has an associated non-linear processing function, in this case a sigmoid function (“S” shaped mathematical function (Bishop, 2006)), and a weight and bias parameter. As the neurons are nonlinear functions, the output of the network is a nonlinear function of the parameters (Nowak, 2002).

In this paper, we are using neural networks to model zone temperatures. As the current zone temperature will be related to previous temperatures, we can consider previous values as inputs. Hence, the model structure is essentially a non-linear autoregressive with exogenous external inputs (NARX) model. The defining equation for a NARX model is given by:

\[ \hat{y}(t) = f(y(t-1), y(t-2), ..., y(t-n), x(t-1), x(t-2), ..., x(t-n)) \]  

After comparing five different data-mining approaches, Kusiak et al. (2011) found that MLP neural networks gave the best prediction performance when predicting energy consumption in a mechanically ventilated space. In this study, the Neural Network Toolbox within MATLAB was used to train and test the networks.
The structure of neural network models is in some respects determined by the system being modelled (number of input and output nodes), however it is up to the user to determine the optimum number of hidden layers and hidden nodes contained within them. Although there is some guidance in the literature, this can often be contradictory (Blum, 1992), (Swingler, 1996), (Boger and Guterman, 1997). Therefore, determining the optimum structure for a particular problem and set of data is largely a process of trial and error. In this study, in addition to training networks with a range of architectures, a number of combinations of inputs and input delays were also tested.

Real Building Data

The building data used in this project comes from two sources: a recently-built school, and an office building in the north of England. Both are naturally ventilated and have a range of single-sided, cross-ventilated and buoyancy-ventilated spaces. The windows are a combination of occupant-controlled manual windows, and automated windows and vents.

![Figure 2 Workflow when using real building data.](image)

The workflow process is shown in Figure 2; the initial step was to collect the building data using the building management system (BMS). Data are available for the opening position of the automated windows in both buildings, however due to the lack of sensors on the manual occupant-operated windows; there is no information available on these. For this reason, the manual windows were treated as a disturbance. A total of eight zones within each building were studied. Data were collected for a number of variables (shown in Table 1) in 16 zones, for a full year of operation and sampled at ten-minute intervals.

In this study, there were two distinct phases in the pre-processing. First, was the processing carried out to extract and clean the data recorded by the BMS. This included linearly interpolating to replace missing data points and removing any obvious outliers. Outlier removal was carried out by calculating the standard score for each variable and then removing all values that fell outside of an expected range. The standard score of a variable is given by:

\[ z = \frac{x - \mu}{\sigma} \]  

(2)

Upon calculating the standard score, the results were used to determine the number of standard deviations away from the mean for which an observation could be considered an outlier (Howell, 1998). This was carried out on a case-by-case basis for each variable (typically three standard deviations from the mean was used to define outliers).

The quality of the data differed between variables. The model target (zone temperature) showed no obvious outliers in any of the 16 zones. However, input variables such as zone humidity, zone CO2 and wind speed had a relatively high number of erroneous observations. For example, in Figure 3 we can see the standard score plot for relative humidity in one of the classrooms within the school. There are clearly erroneous observations in these data as a standard score of around sixty equates to a relative humidity of almost 1000%. However, by removing observations which were outside 3 standard deviations of the mean, most of the sudden jumps, which were likely caused by errors in the measurement or recording equipment, were removed.

![Figure 3 Standard Score plot for one of the classrooms within the school.](image)

<table>
<thead>
<tr>
<th>VARIABLE</th>
<th>TYPE</th>
<th>NOTES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zone Temperature</td>
<td>Target</td>
<td>Only available for automated windows</td>
</tr>
<tr>
<td>Zone CO2</td>
<td>Input</td>
<td></td>
</tr>
<tr>
<td>Zone Humidity</td>
<td>Input</td>
<td></td>
</tr>
<tr>
<td>Outdoor Temperature</td>
<td>Input</td>
<td></td>
</tr>
<tr>
<td>Wind Speed</td>
<td>Input</td>
<td></td>
</tr>
<tr>
<td>Wind Direction</td>
<td>Input</td>
<td></td>
</tr>
<tr>
<td>Window Opening Percentage</td>
<td>Input</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 Variables recorded by BMS for use in system identification.
The second phase of pre-processing was carried out to improve network training. This included normalization to prevent saturation of the sigmoid transfer units in the network and to adjust the magnitudes of the various inputs. Typically, it is beneficial for network performance if inputs have a similar magnitude, unless there is intentional weighting being applied.

Following the initial data cleaning and pre-processing, the data were divided into three subsets using three contiguous blocks of the original data set. The first set was used for model training, the second for validation (this set was used to prevent over-fitting) and the final set was used as an unseen test set.

System Identification using Building Simulation

As previously mentioned, lack of input excitation can be a problem when using data collected from a building during occupation. Typically, buildings operate within a tight range of pre-specified temperatures and the standard input signals used to control actuators are insufficient to develop models with a suitable prediction capability. To overcome this, an identification experiment can be carried out, whereby the system is persistently excited. In this study, it was not possible to carry out an identification experiment on a real building. Therefore, building simulation was used to test the identification procedure using the workflow shown in Figure 4.

Simulation was used for the identification experiment as it allows you to test the building to extremes. Had a real building been used in this initial identification experiment there would have been significant disruption and unsuitable internal conditions for the building occupants. In this project, an open-loop system identification procedure is demonstrated. This resulted in a much greater range of internal temperatures than would be tolerated by building occupants. As a first step, open-loop system identification is the logical choice as it is more likely to sufficiently excite the system and result in models that capture the underlying dynamics. Given the large range of temperatures which resulted from the open-loop identification using simulation; in an occupied building, feedback control and a closed-loop identification may be necessary.

The model used in this study, was based upon one wing of the school building previously discussed (see Figure 5). For the purposes of the identification experiment, only one zone within the space was considered. EnergyPlus was used as the primary simulation program, with DesignBuilder used to generate the initial model geometry. To enable more complex control strategies to be tested, the Energy Management System functionality within EnergyPlus was then used to define temperature sensors and actuators for controlling the automated windows.

To implement the identification experiment, MLE+ was used. MLE+ is an open-source Matlab/Simulink toolbox for co-simulation with EnergyPlus (Bernal et al., 2012). By utilising MLE+, we can take advantage of the features available within the Matlab System Identification Toolbox.

In order to identify the parameters of a system model, the input signal used in the identification must be sufficiently rich. In this study, Gaussian white noise (GWN) was used as excitation to the system. GWN is often used as excitation in identification experiments (Nowak, 2002). If a system is subjected to a GWN stimulus over a sufficiently long enough time, there is a finite probability that any given stimulus waveform will be approximately represented by some sample of the GWN signal. Essentially, the system is being tested with every possible stimulus, or at least a large variety depending on the period over which the experiment is being carried out (Marmarelis and Marmarelis, 1978).

Using MLE+, the simulation was carried out with the window actuators being persistently excited. The resulting zone conditions along with the input signals were used to generate neural network models using the procedure previously described.

RESULTS

Real Building Data
To analyse the model performance, the model outputs were compared with the observed values for the unseen test data set. Alongside visual comparisons...
(see Figure 6), the following four metrics were used to measure the prediction accuracy of the model: the mean absolute error (MAE), the standard deviation of absolute error (StdAE), the mean absolute percentage error (MAPE), and the standard deviation of the absolute percentage error (StdAPE):

\[
AE = |y_i - \hat{y}_i|
\]

\[
MAE = \frac{\sum_{i=1}^{n} AE_i}{N}
\]

\[
APE = \frac{|\hat{y}_i - y_i|}{y_i}
\]

\[
MAPE = \frac{\sum_{i=1}^{n} APE_i}{N}
\]

\[
StdAE = \sqrt{\frac{\sum_{i=1}^{n} (AE_i - MAE)^2}{N - 1}}
\]

\[
StdAPE = \sqrt{\frac{\sum_{i=1}^{n} (APE_i - MAPE)^2}{N - 1}}
\]

The models developed in this paper were found to perform well with the unseen test data. The first models generated were for one-step-ahead prediction. As can be seen in Figure 7 the one-step-ahead model almost perfectly tracks the target temperatures and performs well in all of the evaluation criteria shown in Table 2. The multi-step-ahead models were also found to perform well. When predicting at ten and twenty-steps-ahead (n=10, i.e. 100mins in the future and n=20, i.e. 200 mins in the future) the error increased but the predictions still tracked the observed data reasonably well (see Figure 8).

Table 2 Temperature prediction performance of the neural network models generated using real building data for one-step-ahead prediction, n=10 and n=20.

<table>
<thead>
<tr>
<th>n</th>
<th>MAE</th>
<th>STD_AE</th>
<th>MAPE (%)</th>
<th>STD_APE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.195</td>
<td>0.164</td>
<td>0.0334</td>
<td>0.040</td>
</tr>
<tr>
<td>10</td>
<td>0.630</td>
<td>0.468</td>
<td>0.111</td>
<td>0.122</td>
</tr>
<tr>
<td>20</td>
<td>1.027</td>
<td>0.808</td>
<td>0.170</td>
<td>0.165</td>
</tr>
</tbody>
</table>

Upon closer inspection, it was observed that model performance was poorer during unoccupied periods. It was found that at the end of the week and during the nights, the predictions stray further from the target temperatures. This seems to indicate that occupancy can have a high impact upon the models. Potentially this could be overcome by creating two models for each zone, one for occupied periods and one for unoccupied. This is likely to improve accuracy; however, the degree to which this would impact upon the control performance may not justify the extra complexity.

While the initial results appeared very promising, there were clear inadequacies with the models developed. During training, a number of different combinations of the inputs shown in Table 1 were used. The addition of further information to the model did not improve performance over a purely autoregressive model based upon previous values of zone temperature alone. This suggests that previous values for zone temperature are a good enough predictor without additional data. While being able to discard weather inputs could have potential benefits in reducing model complexity, it is essential that the influence of control inputs (window opening positions) are captured by the model. It was confirmed by carrying out a sensitivity analysis that the window position had no impact upon the output temperature for both the single and multi-step-ahead models. This would prevent the models from being suitable for an MPC application.

System Identification: Building Simulation Data

The models developed using the data generated using EnergyPlus show a similar performance to those generated using the real data. When comparing the different performance criteria, it can be seen in Table 3 that the MAE for the models generated using the simulation data is slightly larger than that for the models generated using real building data, while the MAPE is actually smaller for the models generated using simulation data. This is because the input signal, used to regulate the window openings in the system identification experiment, is causing the building to operate over a larger range of temperatures. Hence, the percentage error is actually smaller while the mean is larger. The greater range of temperatures caused by the system identification experiment can be seen in Figures 9 and 10.

Table 3 Temperature prediction performance of the neural network models generated using simulation data for one-step-ahead prediction, n=10 and n=20.

<table>
<thead>
<tr>
<th>n</th>
<th>MAE</th>
<th>STD_AE</th>
<th>MAPE (%)</th>
<th>STD_APE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.064</td>
<td>0.131</td>
<td>0.31</td>
<td>0.63</td>
</tr>
<tr>
<td>10</td>
<td>0.150</td>
<td>0.207</td>
<td>0.72</td>
<td>0.99</td>
</tr>
<tr>
<td>20</td>
<td>0.261</td>
<td>0.221</td>
<td>1.1</td>
<td>1.07</td>
</tr>
</tbody>
</table>

Figure 6 Model outputs for windows fully open and windows fully closed for a week in the May.

As with the models generated using real building data, a sensitivity analysis was carried out. In this case, the window opening percentage was indeed having an influence on the model output. Figure 6 shows the
output of the model for two scenarios: windows fully open and windows fully closed. In both of these cases, the model outputs seem reasonable; with higher zone temperatures predicted when the windows are left closed and cooler predictions when the windows are left fully open.

DISCUSSION
The models developed using the real building data were able to predict internal temperature over a reasonable prediction horizon. In this study, results are presented for up to 200mins into the future. This should be sufficient for a receding horizon control strategy. However, a thorough study of is required to determine the optimum prediction horizon, which will likely vary between buildings.

The effect of occupancy seems to have been modelled well by the neural network models. However, the real building models did not capture the effect of the window opening. This would make them unsuitable for the MPC approach to ventilation control. The inability of the models to capture the effect of the control input is most likely due to lack of sufficient input excitation and is one of the common drawbacks when using data driven models (Shook et al., 2002), (Lauri et al., 2010). Buildings are typically operated within a tight range and the input is not persistently excited (Privara et al., 2011), (Cigler and Privara, 2010). This can lead to models which, while providing reasonable prediction capability, fail to capture underlying dynamics in essential physical relationships.

Although the models developed using real building data are unsuitable for the purpose of MPC, there are other potential uses for accurate data driven models such as those developed in this project. Previous studies have used empirical models for fault diagnosis (Lee et al., 2004), (Katipamula and Brambley, 2005) and to investigate potential overheating (Iddon et al., 2015). There could also be potential to incorporate a future temperature prediction within a traditional rule based control strategy.

By examining the input signals from the real building data set, it was found that the median position for all of the automated windows in the zones monitored is ‘closed’. In addition, the average time the windows were open was less than 6% during the observed period. While the windows being open for such a small percentage of time may have had an impact upon the indoor air quality it appears to have had an insufficient effect upon temperature to be captured by the models. Alongside the analysis of the neural network models developed, this suggests that if an empirical approach to modelling the thermodynamics of a naturally ventilated building is being taken, then collecting building data during normal operation is insufficient. In order for the models to capture the effect of inputs, an identification experiment such as the one demonstrated in this study must be carried out.

The identification procedure presented in this paper was successful. The resulting neural network models both gave accurate predictions over a reasonable horizon and captured the effects of window opening. However, the identification procedure used is relatively simple and would need refinement before being applied in a real building. The range of temperatures which resulted from the system identification experiment would be unacceptable in a real building. Further work is required with more refined proposals for the identification procedure. In particular, closed-loop system identification procedures may be more suitable for application in a real building. However, the degree to which the system is excited during a closed-loop experiment may result in models which are less informative.

CONCLUSIONS
The models developed using real building data gave a reasonable prediction for internal temperature. However, they did not capture the effect of window opening and as such, were unsuitable for MPC.

The identification procedure demonstrated using EnergyPlus shows that, with proper input excitation empirical neural network models can be created to model the thermal dynamics occurring within a naturally ventilated space. These models were able to give an accurate prediction of internal temperature over a reasonable prediction horizon and captured the effect of the control input successfully.

While the need to carry out an identification experiment is a disadvantage of empirical modelling, there are a number of advantages over using a simple first-principles or building simulation model: (1) once familiar with the techniques involved, creating the models is significantly less time intensive than building a dynamic thermal model; (2) empirical models do not require assumptions to be made about the building fabric or occupancy and are more likely to model the actuality within the building; (3) once developed, the models can output predictions much quicker than multi-zone building simulation programs as they require less computational effort.

Further Work
Having shown that empirical modelling techniques can be used for naturally ventilated spaces, there are two logical progressions. Firstly, further refinement and investigation of the identification procedure for application in a real building. Secondly, demonstrating a MPC approach in a naturally ventilated space.

NOMENCLATURE

\[
\begin{align*}
\mu &= \text{Mean of the population} \\
\sigma &= \text{Standard deviation of the population} \\
AE &= \text{Absolute Error} \\
APE &= \text{Absolute Percentage Error} \\
MAE &= \text{Mean Absolute Error}
\end{align*}
\]
MAPE = Mean Absolute Percentage Error
N = Number of observations
stdAE = Standard Deviation of Absolute Error
stdAPE = Standard Deviation of Absolute Percentage Error
x = Input variable
y = Measured target output
ŷ = Predicted output from the model
z = Standard Score

ACKNOWLEDGEMENT
This work was funded in part by Schneider Electric and by the EPSRC. The building data used in this paper was provided by Dr. C.R. Iddon of S.E. Controls. Thanks are also extended to Dr. Iddon for his useful advice and input throughout this project.

REFERENCES


Boger, Z., Guterman, H. 1997. Knowledge extraction from artificial neural network models, Man and Cybernetics Conference, Orlando Florida USA.


Figure 7 Comparison of one-step-ahead model output and observed temperatures for models developed with real building data.

Figure 8 Comparison of model output for n=10 and n=20, and observed temperatures for models developed with real building data.

Figure 9 Comparison of one-step-ahead model output and observed temperatures for models developed using data from system identification simulation.

Figure 10 Comparison of model output for n=10 and n=20, and observed temperatures for models developed using data from system identification simulation.