
This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Forecasting indoor temperatures during heatwaves: Do more complex models provide better predictions?

PLEASE CITE THE PUBLISHED VERSION

<http://www.ibpsa.org/building-simulation-2019>

PUBLISHER

International Building Performance Simulation Association (IBPSA)

VERSION

VoR (Version of Record)

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Gustin, Matej, Rob McLeod, and Kevin Lomas. 2021. "Forecasting Indoor Temperatures During Heatwaves: Do More Complex Models Provide Better Predictions?". Loughborough University.
<https://hdl.handle.net/2134/14812692.v1>.

Forecasting Indoor Temperatures During Heatwaves: Do More Complex Models Provide Better Predictions?

Matej Gustin^{1,2,*}, Rob S. McLeod^{1,2}, Kevin J. Lomas^{1,2}

¹School of Architecture, Building and Civil Engineering,
Loughborough University, LE11 3TU, UK

²London-Loughborough EPSRC Centre for Doctoral Training in Energy Demand,
Loughborough University, LE11 3TU, UK

*Correspondence: M.Gustin@lboro.ac.uk

Abstract

A novel application of semi-parametric Generalized Additive Models (GAMs) was developed to forecast elevated indoor temperatures. GAM models were compared to AutoRegressive models with eXogenous inputs (ARX) and validated against monitored data from two case study dwellings, located near to Loughborough in the UK, during the 2013 heatwave. Input variables were selected using backward stepwise regressions based on minimisation of the Akaike Information Criterion (AIC) and Mean Absolute Error (MAE), for the ARX and GAM models respectively. Comparison of the models showed that GAMs are capable of slightly improving the forecasting accuracy, but only at short horizons (3-6 hours ahead).

Introduction

Overheating in residential buildings is increasingly acknowledged as an emerging global health risk (NHBC, 2012; ZCH, 2016; Lomas and Porritt, 2017). Climate change projections indicate that the world's most populated regions will experience more frequent and intense heatwave periods over the coming decades (Meehl and Tebaldi, 2004; IPCC, 2014). The likelihood of events such as the 2003 heat wave (which was responsible for over 30,000 premature deaths across Europe (De Bono et al., 2004) recurring is projected to increase 100-fold by 2050 (Stott et al., 2004).

Understanding how individual buildings are likely to respond to extreme climatic events in the future is critical to mitigating their potentially life-threatening impacts. The complexity of this problem originates in the unique time-varying nature of the thermal behaviour of any given building, which is influenced both by its physical characteristics and the unique way in which it is occupied and operated.

Fully parametrised *Dynamic Thermal Simulation* (DTS) models have been widely used to assess current and future overheating risks (Porritt et al., 2012; McLeod et al., 2013; Mavrogianni et al., 2017; Symonds et al., 2017), however, the results of such studies often reveal a significant gap (De Wilde, 2014) between the empirically measured and modelled overheating performance of dwellings (Mantese et al., 2017). This 'modelling-gap' has led some researchers to question the applicability of using white-box DTS models for forecasting overheating (Lomas and Porritt, 2017). In contrast, the availability of

data from large monitoring studies (Beizaee et al., 2013; Lomas and Kane, 2013; Firth et al., 2016; Mavrogianni et al., 2017; Symonds et al., 2017; Buswell et al., 2018) offers the potential to develop empirical models of existing buildings which are capable of making predictions based on the data alone (i.e. machine learning) (Foucquier et al., 2013). In statistical black-box models (Amara et al., 2015), the time-varying responses of the building fabric, ventilation, etc., are all embedded in the past internal temperature data, obviating the need to make assumptions relating to the building's thermo-physical characteristics. Such models could be usefully deployed to provide tailored information to occupants (or their carers) and/or facilities managers of the impending risks of overheating in specific spaces, and potentially advise them on the level of preventative action needed to mitigate heat-related risks (Gustin et al., 2018).

In a previous study, the present authors (Gustin et al., 2018) have shown that linear *AutoRegressive models with eXogenous inputs* (ARX) can forecast indoor temperatures during heatwaves up to 72 hours in advance, with reasonable accuracy. However, it was postulated that the accuracy of such models might be improved by adopting non-linear models.

Several studies (Mechaqrane and Zouak, 2004; Thomas and Soleimani-Mohseni, 2007; Mustafaraj et al., 2011) have shown that *non-linear Artificial Neural Networks* (ANNs) such as *Non-linear ARX* (NARX) models outperform linear ARX models for forecasting indoor temperatures. Some researchers (Mechaqrane and Zouak, 2004; Mustafaraj et al., 2011) have posited that the higher forecasting accuracy of NARX models is attributable to their ability to capture the non-linear relationships that govern indoor temperatures. In contrast, Thomas and Soleimani-Mohseni (2007), showed that the differences between non-linear NARX and linear ARX models were minimal and Ferracuti et al. (2017) observed more accurate predictions with linear ARX models, both in summer and in winter. Whether or not non-linear models are a better choice than linear models, appears to depend on several factors, including: the period of testing, the structure of the models, and forecasting horizon. ANNs are also inherently limited by their lack of interpretability (Foucquier et al., 2013), which has been referred to as "the Achilles' heel of deep neural networks" (Zhang and Zhu, 2018).

In contrast to ANNs, semi-parametric models, also known as *Generalized Additive Models (GAMs)*, offer transparent interpretability of their results (Larsen, 2015) and for some problems, e.g. short-term forecasting of electricity demand (Fan and Hyndman, 2012), they have significantly outperformed ANNs. Because semi-parametric additive models allow non-linear and non-parametric terms to be included within the regression framework, they can readily capture complex non-linear relationships (Fan and Hyndman, 2012).

The monitored data set

To stress-test the predictive and generalisation capabilities of a model for overheating forecasting, it is important that it is tested and validated during a period in which external temperatures exceed those experienced during the training period. For this purpose, two rooms from two dwellings, located near to the town of Loughborough in the English Midlands (and monitored as part of the *LEEDR Smart Home dataset* (Buswell et al., 2018) were selected. These rooms were chosen because of the completeness of the data, their markedly different temperature response profiles during the 2013 heatwave¹. This UK-wide heatwave reached a peak temperature of 33.5°C and lasted from 3 to 23 July 2013 (Met Office, 2013), making it the second warmest July recorded in the UK, since 1910, in terms of both the mean and mean daily maximum temperatures (Met Office, 2018b).

To capture the most pronounced overheating, the internal temperatures (T_{int}) were logged at one-minute intervals, in the upstairs bedrooms. The weather data, consisting of the external air temperatures (T_{ext}) and Global Horizontal solar Irradiance (GHI), was recorded at the nearby Sutton Bonington meteorological station at hourly intervals. To achieve a compatible timestep, the data that was recorded in the dwellings was down-sampled by averaging the sub-hourly values to obtain hourly mean values (centred on each hour).

Outdoor air temperatures during spring and early summer 2013 were considerably below average. The external air temperature started to rise on the 3rd of July, resulting in a continuous hot spell that lasted until thunderstorms on the 22nd and 23rd of July broke the heatwave. During this extended hot spell, the indoor temperatures (recorded in the bedrooms) were noticeably elevated in both dwellings on 6-7 and 13-19 July. Although indoor temperatures in the two dwellings were very similar on some days, dwelling A warmed up considerably less than dwelling B on most days, with the most pronounced temperature difference (of 6.9°C) occurring on the 8th July.

Methods

In previous work (Gustin et al., 2018), overheating was forecasted using ARX models based on the lagged effects

of the internal temperature (T_{int}), external air temperature (T_{ext}) and Global Horizontal solar Irradiance (GHI). Here, additional predictor variables were considered for inclusion in both the newly developed GAMs and ARX models alongside those adopted in the previous study. These new additional variables were chosen based on inputs adopted by Fan and Hyndman (2010; 2012) and included: hour of the day (H), indoor temperature at the same time on the previous day ($T_{int}(t - 24)$), minimum and maximum indoor temperatures in the past 24 hours (T_{int}^- and T_{int}^+), and the 24-hour means of the indoor temperature ($\bar{T}_{int(24h)}$), outdoor temperatures, ($\bar{T}_{ext(24h)}$) and Global Horizontal solar Irradiance ($\overline{GHI}_{(24h)}$). These additional inputs were iteratively recalculated at every time step.

In GAMs the relationships between the dependent (output) and independent (input) and variables are represented by two-dimensional smooth functions². The only exception is the hour of the day, which was modelled as a cyclic cubic regression spline, which is a smooth function with a constrained relationship at either extreme (i.e. the first and last hours of the day, 00 and 23, adopt the same value). The hour of the day cannot be discretised as a single variable in a linear ARX model, because the relationship would be fixed as a constant for every hour. To perform the forecasts at a specific time-step (t) and forecasting horizon (h), the models are first fitted on the training data, a process which estimates the relationships (parametric for the ARX model and semi-parametric for the GAM) between the independent and dependent variables.

The general equation of the ARX model can be written in the form shown in equation (1).

$$T_{int}(t+h) = c + \sum_{i=1}^n p_{\Phi,i} T_{int}(t+h-i) + p_{\Phi,24} T_{int}(t+h-24) + \sum_{j=0}^n p_{\alpha,j} T_{ext}(t+h-j) + p_{\beta,j} GHI(t+h-j) + p^- T_{int(24h)}^- + p^+ T_{int(24h)}^+ + p_{\mu,1} \bar{T}_{int(24h)} + p_{\mu,2} \bar{T}_{ext(24h)} + p_{\mu,3} \overline{GHI}_{(24h)} + e(t+h) \quad (1)$$

where:

$T_{int}(t+h)$ forecasted hourly internal temperature at the time step t for the forecasting horizon h (°C)
 t hourly time step (h)

¹ According to the UK Met Office, based on the World Meteorological Organization definition, a heatwave is defined as, "A marked unusual hot weather (Max, Min and daily average) over a region persisting at least two consecutive days during the hot period of the year based on local climatological conditions, with thermal conditions recorded above given thresholds" (WMO, 2016; Met Office, 2018a).

² Non-parametric functions, are those where the shape of predictor variables (i.e. relationships between dependent and independent variables) are entirely determined by the data (Larsen, 2015).

h	forecasting horizon, hourly time steps ($h = 1, \dots, 72$) (h)
c	intercept ($^{\circ}\text{C}$)
n	maximum lag (previous n time steps) of the input variables that are being considered in the model
i	lag count (1-5) for autoregressive inputs (i.e. previous time steps of the output variable)
j	lag count (0-5) for exogenous inputs, where count 0 is weather data at the forecasted time step
$T_{\text{int}}(t+h-i)$	observed or forecasted hourly internal air temperature at lag i before the forecasting horizon h ($^{\circ}\text{C}$)
$p_{\Phi,i}$	parametric coefficients of the lagged (previous n) T_{int}
$T_{\text{int}}(t+h-24)$	observed or forecasted hourly internal air temperature 24 hours before the forecasting horizon h ($^{\circ}\text{C}$)
$p_{\Phi,24}$	parametric coeff. of the T_{int} on the previous day at the same hour (t-24)
$T_{\text{ext}}(t+h-j)$	observed or forecasted hourly external air temperature at lag j before the forecasting horizon h ($^{\circ}\text{C}$)
$p_{\alpha,j}$	parametric coefficients of the lagged (previous n) T_{ext}
$\text{GHI}(t+h-j)$	observed or forecasted Global Horizontal Irradiance at lag j before the forecasting horizon h (W/m^2)
$p_{\beta,j}$	parametric coefficients of the lagged (previous n) GHI
$T_{\text{int}}^{-}(24\text{h})$	minimum internal air temperature in the past 24 hours ($^{\circ}\text{C}$)
p^{-}	parametric coefficient of the minimum T_{int} in the past 24 hours
$T_{\text{int}}^{+}(24\text{h})$	maximum internal air temperature in the past 24 hours ($^{\circ}\text{C}$)
p^{+}	parametric coefficient of the maximum T_{int} in the past 24 hours
$\bar{T}_{\text{int}}(24\text{h})$	mean internal air temperature in the past 24 hours ($^{\circ}\text{C}$)
$\bar{T}_{\text{ext}}(24\text{h})$	mean external air temperature in the past 24 hours ($^{\circ}\text{C}$)
$\overline{\text{GHI}}(24\text{h})$	mean Global Horizontal Irradiance in the past 24 hours (W/m^2)
$p_{\mu,1}, p_{\mu,2}, p_{\mu,3}$	parametric coefficients of the mean values in the past 24 hours of T_{int} , T_{ext} and GHI respectively
$e(t+h)$	forecasting error: hourly difference between the forecasted and observed temperatures at the time step t ($^{\circ}\text{C}$)

The general equation of the GAM can be written in the form shown in equation (2).

$$g(T_{\text{int}}(t+h)) = c + \sum_{i=1}^n s_{\Phi,i} T_{\text{int}}(t+h-i) + s_{\Phi,24} T_{\text{int}}(t+h-24) + \sum_{j=0}^n s_{\alpha,j} T_{\text{ext}}(t+h-j) + s_{\beta,j} \text{GHI}(t+h-j) + s^{-} T_{\text{int}}^{-}(24\text{h}) + s^{+} T_{\text{int}}^{+}(24\text{h}) + s_{\mu,1} \bar{T}_{\text{int}}(24\text{h}) + s_{\mu,2} \bar{T}_{\text{ext}}(24\text{h}) + s_{\mu,3} \overline{\text{GHI}}(24\text{h}) + s_{cc} H(t+h) + e(t+h) \quad (2)$$

where:

g	gaussian (default) link function for GAM models
$T_{\text{int}}(t+h)$	forecasted hourly internal temperature at the time step t for the forecasting horizon h ($^{\circ}\text{C}$)
t	hourly time step (h)
h	forecasting horizon in hourly time steps ($h = 1, \dots, 72$) (h)
c	intercept ($^{\circ}\text{C}$)
n	maximum lag (previous n time steps) of the input variables that are being considered in the model
i	lag count (1-5) for autoregressive inputs (i.e. previous time steps of the output variable)
j	lag count (0-5) for exogenous inputs, where count 0 is weather data at the forecasted time step
$T_{\text{int}}(t+h-i)$	observed or forecasted hourly internal air temperature at lag i before the forecasting horizon h ($^{\circ}\text{C}$)
$s_{\Phi,i}$	smooth functions of the lagged (previous n) T_{int}
$T_{\text{int}}(t+h-24)$	observed or forecasted hourly internal air temperature 24 hours before the forecasting horizon h ($^{\circ}\text{C}$)
$s_{\Phi,24}$	smooth function of the T_{int} on the previous day at the same hour (t-24)
$T_{\text{ext}}(t+h-j)$	observed or forecasted hourly external air temperature at lag j before the forecasting horizon h ($^{\circ}\text{C}$)
$s_{\alpha,j}$	smooth functions of the lagged (previous n) T_{ext}
$\text{GHI}(t+h-j)$	observed or forecasted Global Horizontal Irradiance at lag j before the forecasting horizon h (W/m^2)
$s_{\beta,j}$	smooth functions of the lagged (previous n) GHI
$T_{\text{int}}^{-}(24\text{h})$	minimum internal air temperature in the past 24 hours ($^{\circ}\text{C}$)

s^-	smooth function of the minimum T_{int} in the past 24 hours
$T_{int(24h)}^+$	maximum internal air temperature in the past 24 hours ($^{\circ}C$)
s^+	smooth function of the maximum T_{int} in the past 24 hours
$\bar{T}_{int(24h)}$	mean internal air temperature in the past 24 hours ($^{\circ}C$)
$\bar{T}_{ext(24h)}$	mean external air temperature in the past 24 hours ($^{\circ}C$)
$\overline{GHI}_{(24h)}$	mean Global Horizontal Irradiance in the past 24 hours (W/m^2)
$s_{\mu,1}, s_{\mu,2}, s_{\mu,3}$	smooth functions of the mean values in the past 24 hours of T_{int} , T_{ext} and GHI respectively
H	Hour of the day (00-23)
s_{cc}	cyclic penalized cubic regression spline smooth function of H
$e(t+h)$	forecasting error: hourly difference between the forecasted and observed temperatures at the time step t ($^{\circ}C$)

To constrain the complexity of the models and thus the computational time, which is considerably longer for GAM than ARX models, the maximum lag (n), of the AutoRegressive (T_{int}) and eXogenous inputs (T_{ext} and GHI) was limited. As in previous work (Hyndman and Fan, 2010; Gustin et al., 2018), temperature variables (used as exogenous inputs) were set to a maximum lag n of 5 previous time steps.

For one-step-ahead forecasts, the models require only the observed past internal temperatures (T_{int}) as autoregressive inputs, while for multi-step-ahead forecasts, the model adopts partially (when $1 < h \leq n$) or exclusively (when $h > n$) the forecasted internal temperature estimates (generated at previous time steps). Similarly, with exogenous inputs, one-step-ahead forecasts require only the observed past weather data (T_{ext} and GHI) and the forecasted weather data for that specific time step ($t+1$). For multi-step-ahead forecasts, the model adopts the forecasted weather data partially (when $1 < h \leq n$) or exclusively (when $h > n$).

The developed models were coded in *R* (R Core Team, 2017) and the GAMs were implemented using the 'Mixed GAM Computation Vehicle with Automatic Smoothness Estimation' (*mgcv*) package (Wood, 2011; Wood, 2018b).

The accuracy of a forecasting model can only be evaluated based on how well it performs in relation to 'new' data (Hyndman and Athanasopoulos, 2018), and not in comparison to the 'past' data to which it was exposed during the training period. In this study, the initial training period spans from the 13th April 2013 to the 30th June at 23:00, during which there was a marked increase in the external air temperature and the heating was turned off. The forecasting period then starts immediately after this, on the 1st July at 00:00 (initial

forecasting origin). However, due to the 72-h forecasting window, it is not possible to evaluate the forecasting accuracy for the first three days, from 1st July at 00:00 to 3rd July at 23:00 for all forecasting horizons (h). The forecasting accuracy was evaluated at different forecasting horizons ($h = 1, 2, 3, 4, 5, 6, 12, 24, 36, 48, 60, 72$), using the *Mean Absolute Error* (MAE).

Rolling origin forecasts (i.e. sliding training and forecasting windows) were performed from the 1st July at 00:00 to 26th July at 23:00. However, because of the constraints imposed by using a 72-h forecasting window (as the longest forecasting horizon) a full comparison of the forecasting accuracy between the various forecasting horizons is only possible during the 19-day period from 4th July at 00:00 to 22nd July at 23:00, when complete forecasts are available for each forecasting horizon (h).

For the identification of the optimal linear ARX model, as in the previous study (Gustin et al., 2018), model selection was based on the minimisation of the *Akaike Information Criterion* (AIC). However, the consideration of additional input variables compared to the previous study (Gustin et al., 2018) leads to an increase in the number of viable model combinations from 131,072 (Gustin et al., 2018) to 8.4 million and 16.8 million for the ARX model and GAM respectively. This exponential increase in model combinations would render the testing of every possible combination computationally excessive. Therefore, to converge quickly on a near-optimal model, a backward stepwise regression (Hyndman and Athanasopoulos, 2018) selection procedure was adopted.

For the linear ARX model, the model selection algorithm begins by including all of the considered input variables in the calculation of the AIC. The algorithm then excludes one variable at a time, re-computing the AIC after each exclusion. The excluded parameter that decreases the AIC value the most is then permanently removed, and the improved model adopted as a reference for further parameter exclusions. The selection algorithm continues removing input variables iteratively until no further decrease in the AIC is observed, after which the final reference model is selected. This model selection procedure defines the structure of the model and is performed only once during the initial training period.

Model identification is more challenging for GAMs, due to their more complex structures. According to Wood (2018a), automatic model selection procedures for complex models that consider all of the possible inputs are often unsuccessful. Since the selection procedure (described above) based on the minimisation of the AIC did not show satisfactory results, a backward stepwise regression, based on minimisation of the out-of-sample predictive accuracy (as defined by the MAE) was adopted. This approach was demonstrated by Fan and Hyndman (Fan and Hyndman, 2012) to provide good results, for semi-parametric model selection. During this selection process, only the first part of the training period of the linear ARX model (75% of the data spanning from 13 April 2013 at 00:00 to 11 June 2013 at 23:00) was used to fit the models and the remaining 19 days (25% of the

data spanning from 12 June 2013 at 00:00 to 30 June 2013 at 23:00) were used to test the forecasting accuracy, as part of the backward stepwise selection process. As for the ARX models, the model selection procedure is performed only once during the initial training period.

In 'real-world' applications such a model would require forecasted weather data from one or more (Gustin et al., 2017) nearby meteorological station(s) as an input. Since the uncertainty of weather forecasts increases in proportion to the length of the forecasting horizon, their reliability several days ahead (particularly in a maritime climate) is questionable (Gustin et al., 2017); as a result, forecasting overheating risks at periods well beyond the forecasting origin is unlikely to be reliable. According to the UK Met Office, short-range (1-3 days ahead) weather forecasts, use data that is updated several times per day and are considered to be extremely accurate (Met Office, 2016). On the other hand, medium-range (3-10 days ahead) weather forecasts provide only a general synopsis on a day-to-day basis. For this reason, the developed models were constrained to forecasting indoor temperatures up to 72 hours (3-days) ahead. As in the previous study (Gustin et al., 2018), multi-step-ahead forecasts were performed by adopting a recursive strategy based on a rolling forecasting origin (i.e. utilising a sliding training and forecasting windows). This means that after each forecast, the model's training window moves forward by one time-step (i.e. 1 hour), before recalibrating the relationships of the previously selected predictors and then recalculating the subsequent forecasts. The model automatically stops forecasting when the sliding forecasting window (of 1-72 hours) reaches the end of the validation period. Once rolling origin forecasts have been completed for the entire validation period, it is then possible to assess the forecasting accuracy.

Results

To automatically select near-optimal models, backward stepwise regressions, based on the minimisation of the AIC and MAE were adopted for the ARX and GAM models respectively. During the model identification process, a number of the inputs (including T_{int} , T_{ext} , and/or GHI) were discarded from both the GAM and ARX models at some of the previous time steps. The internal temperature that was recorded at the same time on the previous day ($T_{int}(t-24)$), as well as the minimum and maximum internal temperature in the past 24 hours ($T_{int(24h)}^-$ and $T_{int(24h)}^+$), and the mean GHI in the past 24h ($\overline{GHI}_{(24h)}$) were selected in 3 out of the 4 models. Conversely, terms describing the mean internal and external temperatures in the past 24h ($\bar{T}_{int(24h)}$ and $\bar{T}_{ext(24h)}$) were never selected. Although the hour of the day (H) was included in the GAM models as a non-linear smooth function it was omitted by the selection algorithm for dwelling B.

Examining the fitting of the GAM provides a useful means of understanding how optimal relationships are

attributed to the various variables. It is evident from this analysis that the autoregressed variables of T_{int} assume the most dominant weights, and the nearer they are temporally located to the value that is being forecasted, the higher their weighting. Moreover, the final result is the sum of positive and negative effects, which in the ARX models is always linear, whereas in the semi-parametric GAM models might be non-linear. The exogenous inputs have considerably lower weights than the autoregressed variables and therefore act as a tuning effect on the predicted dependent variable.

The forecasts suggest that GAMs are capable of producing more accurate forecasts (Table 1) for $h \leq 6h$; whilst for $h = 12h$, the forecasting accuracy of the two models is very similar; however when $h \geq 24h$, the ARX models achieved consistently better predictions.

Whereas there is a comparable forecasting accuracy between the GAM and ARX models for $h \leq 12h$, for dwelling B, a localised disruption in the GAM forecast occurred on the 7th of July (Fig. 1). This is because when forecasting temperatures close to or above the maximum temperatures experienced during the training period some of the predictor variables contain estimates of the relationships which encompass a significant uncertainty. Therefore, until the model has been exposed to such hot conditions, the out of range values predicted by these terms remain highly uncertain. The recursive strategy used by GAMs for multi-step-ahead forecasts means that such errors compound exponentially. Thus, whilst the local over-prediction (seen in Fig. 1 on 7 July), is not unduly pronounced at short forecasting horizons ($h \leq 6$) it degenerates quickly as the forecasting horizon (h) increases (Table 1). This local disruption is evident in the MAE for $h \geq 24h$ (Table 1). For the ARX model, the errors are much smaller thereby avoiding the local disruptions that were observed with the GAM (Fig. 1). This is because ARX models only incorporate linear relationships, using the same regression coefficients, throughout the whole range of temperatures.

Table 1: Forecasting accuracy of GAM vs. ARX models in two dwellings during the 2013 heatwave.

Forecasting horizon h (hours)	Dwelling A		Dwelling B	
	ARX	GAM	ARX	GAM
	MAE (°C)	MAE (°C)	MAE (°C)	MAE (°C)
1	0.13	0.13	0.12	0.10
2	0.25	0.24	0.21	0.18
3	0.35	0.33	0.28	0.24
4	0.44	0.41	0.33	0.29
5	0.50	0.48	0.37	0.33
6	0.57	0.54	0.41	0.37
12	0.81	0.78	0.59	0.53
24	0.92	0.98	0.79	0.99
36	0.92	1.03	0.88	2.23
48	0.93	1.06	0.94	5.76
60	0.94	1.11	0.98	17.51
72	0.95	1.21	1.01	59.33

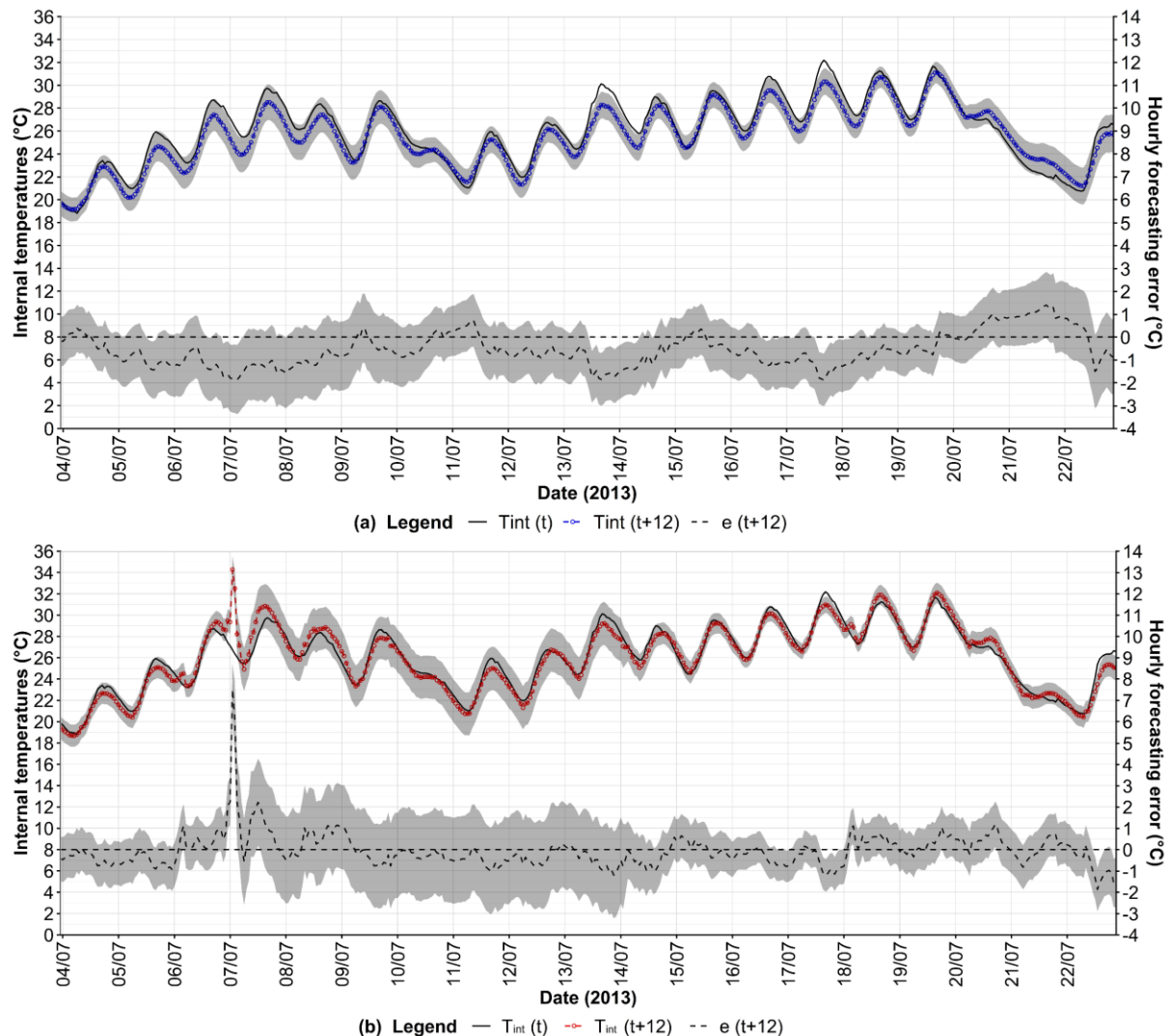


Figure 1: Dwelling B: observed - $T_{int}(t)$ vs. predicted - $T_{int}(t+h)$ hourly internal temperatures with hourly forecasting errors - $e(t+h)$ and 95% prediction intervals (grey band) for 12h forecasting horizons (h) with ARX (a) and GAM (b).

Following the first warm period, the non-linear relationships in the GAMs are recalculated, and as a result, the error in subsequent forecasts of impending high indoor temperatures are greatly reduced (Fig. 1). However, in terms of reliability in a ‘real-world’ application, it is concerning that a non-linear model might fail temporarily when rapidly approaching a considerably warmer period for the first time.

Discussion

The results demonstrate that the inclusion of substantially more input variables to the ARX models than in the authors’ previous study (Gustin et al., 2018) did not improve their accuracy at shorter forecasting horizons. For example, the 6h forecasts produced MAEs of 0.57 and 0.41°C for dwellings A and B respectively (Table 1) compared to MAEs of 0.21, 0.51 and 0.55°C in (Gustin et al., 2018). Over longer forecasting horizons, such as 72h, ARX models produced an MAE of 0.95 and 1.01°C, for dwellings A and B respectively which is higher than the MAEs of 0.49, 0.63 and 0.69°C recorded in the previous study (Gustin et al., 2018). However, the lower forecasting accuracy, reported here, should not be

attributed to poorer model performance but rather to the extended period over which it was evaluated. In the previous study (Gustin et al., 2018), the forecasting accuracy was computed for only one week of data where the day of, and the day after the two-day heatwave produced the largest forecasting errors. The intensive and long-lasting nature of the 2013 heatwave used in this study enabled errors to be computed over a 19-day period, during which there were several pronounced drops in the outdoor and indoor temperatures. The mean zonal indoor temperatures were also approximately 6.5°C (dwelling A) and 7.3°C (dwelling B) above the corresponding indoor temperatures during the initial training period. Considering these forecasting challenges, the ARX model can be considered to have performed well with good generalisation ability.

In the absence of previous results from the literature, the forecasting accuracy of the semi-parametric GAMs can be best assessed by comparison with the forecasts of the linear ARX models. The GAMs produced statistically better forecasts than the ARX models (at the 90% level) for horizons up to 6h ahead (with MAEs of 0.54 and

0.37°C for dwellings A and B respectively at 6h, Table 1). For forecasting horizons beyond 12h, the GAMs were not significantly better than ARX models. Moreover, the GAMs were vulnerable to disruptions when rapidly approaching a considerably warmer period for the first time, rendering them highly unstable, and difficult to control at longer forecasting horizons. While training the models on historical data from past heatwaves (if available) could potentially obviate this issue, any changes to the building fabric or occupancy in the interim would invalidate the previously established relationships embedded in the model. Considering these factors, linear ARX models appear to be a more reliable choice in the context of overheating forecasting. When computational time is considered, ARX models are also favoured due to their minimal fitting times. In contrast, GAMs require much longer to fit the models, an issue which increases exponentially with the number of predictor variables. For forecasts at shorter time horizons, however, when computational time is less relevant, the potentially higher forecasting accuracy of GAMs might be advantageous. The forecasting accuracies presented in this study are in line with previous studies involving the prediction of internal temperatures; although most previous research has focused on offices with mechanical cooling and with higher data resolutions. Mustafaraj et al. (2011) observed MAEs of 0.27-0.38°C for an ARX model predicting 1.5h ahead; cf. MAEs of 0.25 and 0.21°C for dwellings A and B at $h=2$ (Table 1). Forecasts by Mustafaraj et al. using a NARX model (Mustafaraj et al., 2011) were considerably better, with MAEs of 0.23-0.27°C at 2h ahead, which is very close to the MAEs achieved with the GAMs for $h=2$, 0.24°C and 0.18°C for dwellings A and B (Table 1). However, these results must be viewed in relation to the validation data used to test the models. Notably, the forecasts performed here took place in free-running dwellings with considerably higher indoor temperature variability than that observed in the study by Mustafaraj et al. (2011).

Conclusions

The ability of linear ARX models and semi-parametric GAMs to forecast indoor temperatures over the intense and long-lasting UK heatwave of 2013 was investigated using hourly data from two bedrooms, in two houses, located near to the town of Loughborough in the UK Midlands. A backward stepwise regression based on minimisation of the AIC (for ARX models) and MAE (for GAMs) was adopted for the model selection process. Recursive multi-step-ahead forecasts were produced by both the models using a rolling forecasting origin for the entire duration of the heatwave. Forecasts were made for time horizons of 1-6, 12, 24, 36, 48, 60 and 72 hours ahead, including the 95% prediction intervals, to provide a credible interval for the forecasted temperatures. The accuracy of the predictions was evaluated using the MAE to assess out-of-sample accuracy.

Comparisons between the ARX models and GAMs showed that although the GAMs were capable of slightly improved forecasting accuracy, the improvements were

only statistically significant up to 3-6 hours ahead. For longer forecasting horizons, ARX models provided an accuracy that was either equal to, or greater than the GAMs, with an MAE (up to 72 hours ahead) that was typically below 1°C for the entire heatwave. Considering the potential uncertainty associated with the non-linear GAMs relationships when exposed to higher temperature ranges for the first time, the subsequent risk of instability at longer forecasting horizons, higher computational time requirements, lower accuracy at longer forecasting horizons and marginal improvement of the predictive accuracy at shorter horizons; the adoption of such models appears unjustified for forecasting elevated internal temperatures in free-running buildings.

Overall this work suggests that more complex non-linear models do not necessarily produce better forecasts and that particular attention should be given to the use of GAMs when predicting out-of-range. By definition, there will always be limited data at the lower and upper ranges of the independent variables, which engenders increasing uncertainty when forecasting beyond the ranges for which the models were originally trained, with errors that are likely to amplify at longer forecasting horizons.

Future work will involve longitudinal testing of the prototyped forecasting models using larger datasets to quantify the reliability of predictions for different room, dwelling and household configurations across a wide range of geo-social contexts.

Acknowledgements

This research was made possible by the Engineering and Physical Sciences Research Council (EPSRC) support for the 'London-Loughborough (LoLo) Centre for Doctoral Training in Energy Demand' (grant EP/L01517X/1). Monitored data, indispensable to this study, was made available by the open access LEEDR project home energy dataset (Buswell et al., 2018), which was funded by the EPSRC: 'LEEDR: Low Effort Energy Demand Reduction' (grant EP/I000267/1).

References

- Amara, F., Agbossou, K., Cardenas, A., Dubé, Y., et al. (2015). Comparison and Simulation of Building Thermal Models for Effective Energy Management. *Smart Grid and Renewable Energy* 6(April), pp. 95–112.
- Beizaee, A., Lomas, K.J. and Firth, S.K. (2013). National survey of summertime temperatures and overheating risk in English homes. *Building and Environment* 65, pp. 1–17.
- De Bono, A., Giuliani, G., Kluster, S. and Peduzzi, P. (2004). Impacts of summer 2003 heat wave in Europe. *Environment Alert Bulletin UNEP* (August), p. 4.
- Buswell, R., Webb, L., Cosar-Jorda, P., Marini, D., et al. (2018). LEEDR project home energy dataset.
- Fan, S. and Hyndman, R.J. (2012). Short-term load forecasting based on a semi-parametric additive model. *IEEE Transactions on Power Systems* 27(1), pp. 134–141.

- Ferracuti, F., Fonti, A., Ciabattoni, L., Pizzuti, S., et al. (2017). Data-driven models for short-term thermal behaviour prediction in real buildings. *Applied Energy* 204, pp. 1375–1387.
- Firth, S., Kane, T., Dimitriou, V., Hassan, T., et al. (2016). REFIT Smart Home dataset.
- Fouquier, A., Robert, S., Suard, F., Stéphan, L., et al. (2013). State of the art in building modelling and energy performances prediction: A review. *Renewable and Sustainable Energy Reviews* 23, pp. 272–288.
- Gustin, M., Oraopoulos, A., McLeod, R.S. and Lomas, K.J. (2017). A new empirical model incorporating spatial interpolation of meteorological data for the prediction of overheating risks in UK dwellings. In: *Proceedings of PLEA 2017: the 33rd Passive and Low Energy Architecture International Conference*. Edinburgh, UK, pp. 3786–3793.
- Gustin, M., McLeod, R.S. and Lomas, K.J. (2018). Forecasting indoor temperatures during heatwaves using time series models. *Building and Environment* 143(October), pp. 727–739.
- Hyndman, R.J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. 2nd ed. OTexts.
- Hyndman, R.J. and Fan, S. (2010). Density Forecasting for Long-Term Peak Electricity Demand. *IEEE Transactions on Power Systems* 25(2), pp. 1142–1153.
- IPCC (2014). *Climate Change 2014: Synthesis Report*.
- Larsen, K. (2015). GAM: The Predictive Modeling Silver Bullet. *MultiThreaded*, pp. 1–27.
- Lomas, K. and Kane, T. (2013). Summertime temperatures and thermal comfort in UK homes. *Building Research and Information* 41(3), pp. 259–280.
- Lomas, K.J. and Porritt, S.M. (2017). Overheating in buildings: lessons from research. *Building Research & Information*, pp. 1–18.
- Mantese, E., Hopfe, C.J., Cook, M.J., Glass, J., et al. (2017). The modelling gap: Quantifying the discrepancy in the representation of thermal mass in building simulation. *Building and Environment* 131(August 2017), pp. 74–98.
- Mavrogianni, A., Pathan, A., Oikonomou, E., Biddulph, P., et al. (2016). Inhabitant actions and summer overheating risk in London dwellings. *Building Research & Information* 0(0), pp. 1–24.
- McLeod, R., Hopfe, C. and Kwan, A. (2013). An investigation into future performance and overheating risks in Passivhaus dwellings. *Building and Environment* 70, pp. 189–209.
- Mechaqrane, A. and Zouak, M. (2004). A comparison of linear and neural network ARX models applied to a prediction of the indoor temperature of a building. *Neural Computing and Applications* 13(1), pp. 32–37.
- Meehl, G.A. and Tebaldi, C. (2004). More Intense, More Frequent, and Longer Lasting Heat Waves in the 21st Century. *Science* 305(August), pp. 994–997.
- Met Office (2015). July 2013 heat wave - Met Office.
- Met Office (2016). 10 day weather forecast - Met Office.
- Met Office (2018a). Heatwave.
- Met Office (2018b). UK and regional series - Met Office.
- Mustafaraj, G., Lowry, G. and Chen, J. (2011). Prediction of room temperature and relative humidity by autoregressive linear and nonlinear neural network models for an open office. *Energy and Buildings* 43(6), pp. 1452–1460.
- NHBC (2012). *Overheating in new homes; A review of the evidence*.
- Porritt, S.M., Cropper, P.C., Shao, L. and Goodier, C.I. (2012). Ranking of interventions to reduce dwelling overheating during heat waves. *Energy and Buildings* 55, pp. 16–27.
- R Core Team (2017). R: A language and environment for statistical computing.
- Stott, P.A., Stone, D.A. and Allen, M.R. (2004). Human contribution to the European heatwave of 2003. *Nature* 432, pp. 610–614.
- Symonds, P., Taylor, J., Mavrogianni, A., Davies, M., et al. (2016). Overheating in English dwellings: comparing modelled and monitored large-scale datasets. *Building Research and Information* 45(1–2), pp. 195–208.
- Thomas, B. and Soleimani-Mohseni, M. (2007). Artificial neural network models for indoor temperature prediction: Investigations in two buildings. *Neural Computing and Applications* 16(1), pp. 81–89.
- De Wilde, P. (2014). The gap between predicted and measured energy performance of buildings: A framework for investigation. *Automation in Construction* 41, pp. 40–49.
- WMO (2016). Guidelines on the Definition and Monitoring of Extreme Weather and Climate Events. *World Meteorological Organization* (December 2015), p. 62.
- Wood, S.N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society*, pp. 3–36.
- Wood, S.N. (2018a). ‘mgcv’ package: Generalized Additive Model Selection.
- Wood, S.N. (2018b). ‘mgcv’ package for R (version 1.8-24): Mixed GAM Computation Vehicle with Automatic Smoothness Estimation.
- ZCH (2016). *Overheating in homes - The big picture*.
- Zhang, Q. and Zhu, S.-C. (2018). Visual Interpretability for Deep Learning: a Survey. *Frontiers of Information Technology & Electronic Engineering* 19(1423305), pp. 27–39.