



# Accurately forecasting temperatures in smart buildings using fewer sensors

Bruce Spencer<sup>1</sup> · Feras Al-Obeidat<sup>2</sup> · Omar Alfandi<sup>2</sup>

Received: 10 September 2017 / Accepted: 28 November 2017 / Published online: 16 December 2017  
© Springer-Verlag London Ltd., part of Springer Nature 2017

## Abstract

Forecasts of temperature in a “smart” building, i.e. one that is outfitted with sensors, are computed from data gathered by these sensors. Model predictive controllers can use accurate temperature forecasts to save energy by optimally using heating, ventilation and air conditioners while achieving comfort. We report on experiments from such a house. We select different sets of sensors, build a temperature model from each set, and compare the accuracy of these models. While a primary goal of this research area is to reduce energy consumption, in this paper, besides the cost of energy, we consider the cost of data collection and management. Our approach informs the selection of an optimal set of sensors for any model predictive controller to reduce overall costs, using any forecasting methodology. We use lasso regression with lagged observations, which compares favourably to previous methods using the same data.

**Keywords** Energy efficiency · Sensor networks · Model predictive control · Temperature forecast · Feature selection · Internet of things

## 1 Introduction

According to recent studies, about 40% of energy produced worldwide is consumed by buildings, and more than half of this is used by heating, ventilation and air conditioning (HVAC) systems [3, 8, 10]. Pan et al. [9] point out that, due to thermal inertia, it is more efficient to maintain temperature in a room or building than to raise or lower the temperature.

Accurate temperature forecasts can help reduce energy usage in buildings by using future values of temperature when deciding whether or not to activate the HVAC [22]. Moreno et al. [7] achieve estimated energy savings of 20% in a realistic situation based on the presence of persons in a room. Yuan et al. [20] achieve 20% savings while exploiting thermal inertia when assigning rooms for meetings by scheduling contiguous meetings in the same room. Prívarva et al. [11] report 17–24% savings for a large university building, by using a model predictive controller.

Model predictive controllers (MPC) produce a control signal for HVAC systems, and minimize a cost function based on energy consumption. The cost function takes into account a prediction horizon and a control horizon [1]. The prediction horizon used in practise depends on how much data is needed by the HVAC controller to achieve acceptable comfort while reducing energy consumption.

While the costs savings may be significant, the overhead and operational costs associated with MPC may discourage adoption. These costs include the installation and maintenance of the sensing devices, a wireless sensor network, and the computational cost of modelling temperature as a function of the data generated by the sensors. To encourage wider adoption of MPC, in this paper we seek to reduce these associated costs. Specifically, we identify sensor data with little influence on forecast accuracy.

In the remainder of this paper, we review the sensor data related to temperature forecasting reported from a smart house. We discuss the nature of the search, provide a best-first search procedure to select sensors, and compare the outcomes as we vary the history horizon, forecast horizon and the error metric. We report on related work on this data, and conclude with recommendations for using our results in both new and existing installations.

---

✉ Bruce Spencer  
bspencer@unb.ca

<sup>1</sup> University of New Brunswick, Fredericton, Canada

<sup>2</sup> Zayed Univesity, Abu Dhabi, United Arab Emirates

This paper is a full exposition of initial work [16]. This presentation adds to the previous work in two main ways: It covers longer sensor histories and applies the technique to forecasts of temperature differences.

## 2 Background

### 2.1 Data from a smart home

The SML House [22] competed in the Solar Decathlon 2012 competition [19], using 88 sensors and 49 actuators. In this paper and in our previous work [14, 15], we use a publicly available subset of this data [18], reporting some of the collected data, specifically values collected during March and April 2012 from 18 sensors every quarter-hour.

The sensors reported are as follows:

1.  $W_i$  – wind speed
2.  $T_w$  – twilight indicator
3.  $TP$  – predicted temperature
4.  $TL$  – living room temperature
5.  $TD$  – dining room temperature
6.  $T$  – external temperature
7.  $SW$  – sun on the west wall
8.  $SS$  – sun on the south wall
9.  $SE$  – sun on the east wall
10.  $Pcp$  – precipitation
11.  $P$  – sun irradiance measured by a pyranometer
12.  $LL$  – lights in the living room
13.  $LD$  – lights in the dining room
14.  $HL$  – humidity in the living room
15.  $HD$  – humidity in the dining room
16.  $H$  – external humidity
17.  $CL$  – carbon dioxide sensor in the living room
18.  $CD$  – carbon dioxide sensor in the dining room

### 2.2 Linear and lasso regression

Our forecasting methods are based on linear regression defined as follows. Given a set of independent variables  $x_1, \dots, x_n$  and a dependent variable  $y$  of interest that we want to forecast, we seek parameters  $\beta_0, \dots, \beta_n$  so that  $\hat{y}_i = \beta_0 + \sum_{i=1}^n \beta_i x_i$  is a good approximation of  $y$ . When presented with a set of  $m$  instances of each  $x_i$ , called  $x_{i,j}$  and the corresponding instances  $y_j$ , we select the  $\beta_i$  parameters to minimize the residual sum of squares (RSS):

$$\sum_{j=1}^m (\beta_0 + \sum_{i=1}^n \beta_i x_{i,j} - y_j)^2$$

Lasso regression [17] minimizes  $RSS + \lambda \sum_{j=1}^m |\beta_j|$  where  $\lambda$  is a tuning parameter that balances the emphasis between reducing error and using small  $\beta$  coefficients. Some  $\beta$  may reduce to zero, which deselects that variable  $x$ , thus

endowing lasso regression with feature selection. For lasso regression, we use the R library `glmnet` [4, 5, 13].

(We avoid using the usual vector notation shorthand for linear regression so we can better relate this background with Section 3, which requires requires us to adjust the indexed of summation for a precise description.)

### 2.3 Feature selection

Having too many independent variables, or features, can confound a forecast model. Irrelevant details overwhelm the modelling technique, which prevents it from computing an accurate forecast. Feature selection, the process of selecting specific features from which to build a model, is roughly divided into wrapper techniques, filter methods and embedded methods. Wrapper techniques enumerate various combinations of features and measure the accuracy of the resulting models, selecting that combination that exhibits the best error. Filter techniques measure the usefulness of features using computationally fast metrics. Embedded techniques identify useful features during the modelling process as a by-product. Lasso regression is an example.

In this paper, we focus on wrapper techniques that are guided by best-first provided by the R library `FSelector` [12] and embedded techniques, using lasso regression.

## 3 Models using lagged sensor readings

When creating a model from which to forecast temperatures, we provide multiple historical readings from each sensor. Given a history of  $b$  time periods, where readings are taken every quarter-hour, we provide  $b + 1$  lagged readings from each of  $s$  sensors, which includes the current period at lag 0. Let  $x_{k,t}$  be the  $t^{\text{th}}$  observation for sensor  $k$  counting from the first observation at time  $t = 1$ , as it appears in the training data. Let  $y_t$  be the internal temperature the house at time  $t$ . We are given observations over the  $m$  time periods in the training data. We create a linear a model for each future period  $f$ . We define the RSS as

$$RSS(f) = \sum_{t=b+1}^m (\beta_{f,0} + \sum_{g=0}^b \sum_{k=1}^s \beta_{f,k,g} x_{k,t-g} - y_{f+t})^2$$

In this equation,  $t$  starts at  $b+1$  because there are no observations for the lagged readings for the first  $b$  data points. Using lasso regression, we choose values for the coefficients  $\beta_f = \{\beta_{f,0}\} \cup \{\beta_{f,k,g} \mid g = 0, \dots, b, k = 1, \dots, s\}$  where  $g$  identifies the lag and  $k$  identifies the sensor. The coefficients in  $\beta_f$  specify a model for each future interval  $f$ . We use two different forecast horizons;  $h$  is either 12 or 48 future time periods, i.e. 3 or 12 h.

The coefficients are computed on the training data which is the first 2/3 of the data. Once they are computed, we switch over to using test data, which is the final 1/3 of

the data. Thus  $x$  and  $y$  below refer to observations in the test data and  $m$  to the number of observations in the test data. We report the RMSE for each future interval  $f$ . In our experiments  $f = 1, \dots, 12$  for forecasts 3 h into the future, and  $f = 1, \dots, 48$  for forecasts to 12 h.

$$\text{RMSE}(f) = \sqrt{1/(m-b) \sum_{t=b+1}^m (\beta_{f,0} + \sum_{g=0}^b \sum_{k=1}^s \beta_{f,k,g} x_{k,t-g} - y_{f+t})^2}$$

We report error metrics on all forecasts  $f$  over the forecast horizon  $h$ , including Mean RMSE  $= 1/h \sum_{f=1}^h \text{RMSE}(f)$  and Maximal RMSE  $= \max_f \text{RMSE}(f)$ .

### 4 Useful and confounding sensors

In our experiments we consider various sets of sensors, and from each, we measure the error from a forecast model based on the data from those sensors. To measure error, with the exception of the selected set of sensors, we hold all other factors fixed, including the training and test data, the size  $h$  of the forecast horizon and the number  $b$  of back observations. Thus, the error from the model is a function only of the set of sensors.

It will often occur that one sensor in a set of sensors is useful in that it provides predictive power. Let  $S$  be a set of sensors and let  $a$  and  $b$  be individual sensors. We say  $a$  is *useful* in  $S$  when the error from  $S \setminus \{a\}$  is greater than the error from  $S$ . If  $a$  is useful in  $S$  then  $a \in S$ . It may also occur that two sensors each provide that same predictive power, for instance when they report similar information. In this case we can use either one. More precisely, we say  $a$  and  $b$  are *interchangeable* in  $S$  when  $a$  and  $b$  are useful in  $S$  and the error from  $S \setminus \{a\}$  is the same as the error from  $S \setminus \{b\}$ .

The definitions in this paper are relative to some tolerance, below which forecast error is insignificant. We do not define this tolerance here, but note that it will be determined by the model predictive controller as follows: If an increase in the forecast error does not affect the controller’s ability to save energy, then that increase is below the tolerance. In this paper, we speak informally and understand an error to be greater than another when the difference exceeds this tolerance and likewise say that two errors are the same when their difference falls below this tolerance. In these experiments, since we are not measuring the performance of a controller, we take the tolerance to be 0.

Note that useful and interchangeable are defined with respect to a set of sensors. We may find that while  $a$  is useful in  $S \setminus \{b\}$ ,  $a$  is not useful in  $S \cup \{b\}$ . For instance, this will happen when  $a$  and  $b$  are interchangeable in  $S$ .

We may also observe that including a sensor in a model gives rise to a higher error. This can happen when the

sensor leads us “down the garden path,” for instance, when it appears to be correlated to the observed temperature in the training data, but oppositely correlated in the test data. We say that  $a$  *confounds*  $S$  when the error from  $S \setminus \{a\}$  is lower than the error from  $S \cup \{a\}$ .

We may also observe sensors that together increase accuracy but individually do not. This can happen when the model uses an interaction between the sensors. Suppose the laundry is always done on Saturday and no other day, and starts when someone enters the laundry room on Saturday. Suppose one of the sensors reports the day of the week and another reports motion in the laundry room. Then the modeller may recognize a heating event—for the room heats up when the laundry is done—occurs when both sensors are activated. In this case, if the modeller associated a heating event just based on motion in the laundry room, regardless of the day of the week, it would be misled on the non-Saturdays, and the model’s error would increase. Likewise, it would be misled by associating a heating event with Saturday for those weeks where no laundry was done. Thus, the laundry room motion sensor and the day of week sensor each individually confounds the model. However, together they improve the model. We say that two sensors  $a$  and  $b$  are *co-dependent* in  $S$  if individually each of  $a$  and  $b$  confound  $S \setminus \{a, b\}$ , but the error of  $S \cup \{a, b\}$  is smaller than the error of  $S \setminus \{a, b\}$ .

We seek a set  $S^*$  of sensors that has minimal error among the power set of sensors. This implies all sensors in  $S^*$  are useful, and that all sensors not in  $S^*$  confound  $S^*$ . We say that an ordered set of sensors *gracefully degrades* if we can remove one sensor at a time in that ordering, such that the error always increases. Given an ordered set of sensors that gracefully degrades, its *gracefully degrading* sequence of sets of sensors is that sequence that arises from removing sensors according to the ordering.

To guide the cost-benefit analysis, given a gracefully degrading set of sensors, we advocate computing two costs for each set of sensors in the gracefully degrading sequence. One cost is the measured energy costs which is higher when the error is higher. This always occurs with fewer sensors in the gracefully degrading sequence. The other cost arises from purchasing, installing, and maintaining the sensors, so it is lower when the number of sensors is lower.

### 5 Ordering sensors by influence

Our goal is to identify sensors that should be included in the model. Because we use lagged data in our model, each sensor provides many predictors in the regression, one for each quarter-hour of historical observations. For a given sensor, we may consider whether to include all

of the predictors arising from this sensor, some of them, or none of them. This leads to a large search space. For instance, given one hour of lagged observations (plus the current observation) for each of 18 sensors, gives  $5 \times 18 = 90$  predictors in the model. This gives rise to  $2^{90}$  sets of predictors, which is clearly infeasible to search entirely. We also want to consider 2 h of readings per sensor, but to avoid searching a space of  $2^{162}$  sets of sensors, which would take us almost  $10^{42}$  years to search if we could consider one set each second.

We rely on lasso regression, which selects features among the predictors in the regression. Since lasso feature selection is in place, we need only consider sensor selection so the search space is reduced from  $2^{90}$  to  $2^{18}$ , and is independent of the number of lagged readings per sensor. A complete search would take about 3 days if we could consider one set per second. We simplify it further by employing best-first search, which is a variant of bottom-up search that limits non-deterministic choices and is guided by a heuristic. Our heuristic prefers lower forecast error. This search procedure and the heuristic are provided by *FSelector*. Since best-first search is well-known, we describe it here at a high level.

*Available* is the set of sets of sensors that may be considered during the running of the Algorithm 1.

*Visited* is the set of sets of sensors that is accumulated during the running of the algorithm. It contains all of the sets of sensors that were explored.

**Algorithm 1 best-first search** Initially *Visited* is an empty set, and *Available* contains the empty set of sensors. The model for this empty set simply predicts the mean temperature. The search proceeds with a selection of  $S$  as the set from *Available* with lowest error.  $S$  is removed from *Available*. Nondeterministically, *FSelector* selects a new sensor  $a$  among the sensors in the SML house that do not occur in  $S$ . *FSelector* usually tries up to five different choices. For each of these choices for  $a$ , if  $a$  is useful in  $S \cup \{a\}$ , then  $S \cup \{a\}$  is added *Available* for future consideration.  $S \cup \{a\}$  is also added to *Visited*.

When there are no more sets in *Available*, the algorithm concludes. Among the sets in *Visited*, the set with minimal error,  $S'$ , is taken as the estimate of  $S^*$ .

Because the heuristic guides the search toward the most promising parts of the search space, good estimates of  $S^*$  are expected.  $S'$  is confounded by all sensors not in  $S'$ , so it is a local minimum. However, the non-deterministic choices made by *FSelector* do not consider all possible choices. The search space is not entirely explored and  $S'$  is not guaranteed to be a global minimum.

Given the sets that were visited by best-first search, we use a second algorithm to generate a sequence of these sets with gradually increasing error.

**Algorithm 2 construct the sensor sequence** Let  $S_1 = S'$ , which is the set in *Visited* with lowest error. Let  $i = 1$  and define  $S_{i+1}$  as the set with lowest error that is both a subset of  $S_i$  and a visited set. Proceed to increment  $i$  and compute the next set  $S$  until  $S_i$  is empty. Report the sequence of  $S$ 's and the sequence of set differences between them. In most cases the set differences will be individual sensors.

Because Algorithm 2 considers only visited sets, there is no error calculation required and Algorithm 2 is very efficient.

In the next section, we consider the effectiveness of this best-first search using the data of the SML house. There is no guarantee that Algorithm 1 will deliver the overall best set  $S^*$ . There is no guarantee that Algorithm 2 will generate the best sequence. However, Algorithm 1 and Algorithm 2 together a sequences that can be used to guide the cost-benefit analysis.

## 6 Experimental results

We investigate the sensor selection when forecasting temperature for different forecast horizons, different amounts of historical data and different error metrics. We also investigate sensor selection when forecasting temperature differences.

### 6.1 Forecasting temperature from one hour of sensor data

We ran experiments using four readings per sensor, shown in Tables 1, 2, 3 and 4. We varied the forecasting horizon to 3 and 12 h into the future.

Since we compute RMSE for each period, we have a series of RMSE values. We compute the mean and the maximal values of this series. Both maximal RMSE and mean RMSE are interesting here since the effectiveness of the model predictive controller can be affected in different ways. A large mean RMSE indicates errors in many of the quarter-hourly forecasts. As a result the MPC will consistently incur higher energy costs than necessary. A large maximal RMSE indicates at least one big forecast error. The inefficiency could be very great for that one period. Since each error metric indicates a separate kind of error, we report both.

Consider the example from Table 1, where we used four historical observations per sensor, generated 3-h forecasts, and measured maximal RMSE. Starting from the empty set, the search in Algorithm 1 considers sets up to about 10 sensors. Overall it visited 135 sets of sensors, which is a sharp reduction from the possible  $2^{18} = 262,144$  sets. The minimal error occurs with nine sensors: Wi, Tw, TL, TD, T, SW, SE, CL and CD, so this is our estimate of  $S^*$ . The other nine sensors, namely TP, SS, Pcp, P, LL, LD, HL, HD and H, confounded it.

**Table 1** Maximal RMSE for selected sensors, generating 3-h forecasts of temperature, from 4 historical readings

Maximal RMSE	Sensors	Remove Next
0.4589	Wi+Tw+TL+TD+T+SW+SE+CL+CD	CL
0.461	Wi+Tw+TL+TD+T+SW+SE+CD	CD
0.4638	Wi+Tw+TL+TD+T+SW+SE	Wi
0.4662	Tw+TL+TD+T+SW+SE	TD
0.4951	Tw+TL+T+SW+SE	Tw
0.5219	TL+T+SW+SE	SW
0.5403	TL+T+SE	T
0.5886	TL+SE	SE
0.68	TL	

Using Algorithm 2, we progressively remove sensors from  $S^*$  to increase the error gradually. The error increases only by 0.0021 °C if we remove CL, the carbon dioxide sensor in the living room. Another small increase, 0.0028 °C, occurs if we ignore the carbon dioxide sensor in the dining room.

We observe some trends in the results. Maximal RMSE is larger than mean RMSE, but within a factor of about two. Forecasting 3 h into the future has lower error than forecasting 12 h, usually within a factor of 5. This difference is always less than 2°, when comparing best models short forecasts to best models for long forecasts. From these two observations we conclude having longer forecast horizons and more sensitive error calculations increase the measured error, which is consistent with intuition.

All of our tabular results degrade gracefully. This suggests that the model does not need to consider interactions between sensors, as we described in Section 4. It also means that sensors that do not appear in the model have been found to be not useful, according to our definition of useful.

There is some consensus, across the different forecast horizons, about which sensors are useful. No best model

**Table 2** Mean RMSE for selected sensors, generating 3-h forecasts of temperature, from 4 historical readings

Mean RMSE	Sensors	Remove Next
0.2404	Tw+TP+TL+TD+T+SW+SE+CL+CD	CL
0.2429	Tw+TP+TL+TD+T+SW+SE+CD	CD
0.2434	Tw+TP+TL+TD+T+SW+SE	Tw
0.2488	TP+TL+TD+T+SW+SE	T
0.2672	TP+TL+TD+SW+SE	SW
0.2827	TP+TL+TD+SE	TD
0.284	TP+TL+SE	TP
0.299	TL+SE	SE
0.3407	TL	

**Table 3** Maximal RMSE for selected sensors, generating 12-h forecasts of temperature, from 4 historical readings

Maximal RMSE	Sensors	Remove Next
1.8598	Tw+TP+T+SW+LD	TP
1.9032	Tw+T+SW+LD	SW
1.975	Tw+T+LD	Tw
2.1399	T+LD	LD
2.3384	T	

made use of H, HD, HL or LL, which are, respectively, the humidity externally, in the dining room and in the living room, and the lighting in the living room. Based on this analysis, we would not recommend installing these sensors in this house for the purpose of forecasting temperature.

We cannot identify sensors that are always the most useful, but there is some consistency. We find that T, TD, TP, LD, SW, SE and Tw appear frequently in the smaller sets of sensors. These are, respectively, the external temperature, temperature in the dining room, and the predicted temperature, the light in the dining room, the sun on the west and east wall, and the pyranometer which measures sunlight intensity. There is also some influence from the  $CO_2$  sensors. Thus we can conclude that the future temperature results from a combination of human activities, ambient internal conditions, external weather conditions and time of day.

### 6.2 Forecasting temperature from longer histories

Starting from the models computing from four historical sensor observations, we ran the same models using eight historical observations per sensor, as shown in Tables 5 and 6. The picture that emerges is similar to when we used four historical observations. The errors are not significantly different; they are sometimes smaller and sometimes larger. This indicates the extra hour of observations is not particularly helpful to the lasso regression model. Again, the maximal RMSE is about twice as large as the mean RMSE

**Table 4** Mean RMSE for selected sensors, generating 12-h forecasts of temperature, from 4 historical readings

Mean RMSE	Sensors	Remove Next
1.1539	Tw+TP+TL+T+SW+LD+CL	TP
1.1566	Tw+TL+T+SW+LD+CL	CL
1.1598	Tw+TL+T+SW+LD	SW
1.1876	Tw+TL+T+LD	LD
1.2552	Tw+TL+T	T
1.3545	Tw+TL	Tw
1.4215	TL	

**Table 5** Mean RMSE for selected sensors, generating 3-h forecasts of temperature, from 8 historical readings

Mean RMSE	Sensors	Remove Next
0.2499	Tw+TP+TL+TD+T+SW+SE+Pcp+CL+CD	CD
0.2509	Tw+TP+TL+TD+T+SW+SE+Pcp+CL	CL
0.2517	Tw+TP+TL+TD+T+SW+SE+Pcp	Pcp
0.254	Tw+TP+TL+TD+T+SW+SE	SE
0.2738	Tw+TP+TL+TD+T+SW	T
0.2867	Tw+TP+TL+TD+SW	Tw
0.3027	TP+TL+TD+SW	SW
0.3142	TP+TL+TD	TD
0.322	TP+TL	TP
0.3486	TL	

and so we report only the mean error. The same observations apply with regard to longer forecast horizons increasing the measured error. The most useful sensors are approximately the same for both 4- and 8-history models, although there is some variation in the apparent importance of each.

The best set of sensors is smaller when considering forecasts for longer period into the future. This suggests some factors have influence over the temperature for a brief period, but are less useful at later times. Others are useful over the entire period. For instance, TW and SW appear quite often in the smaller sets of sensors, especially when there are eight historical values available. This suggests that knowing these longer histories for these two sensors, in particular, is especially useful for making longer predictions. Factors that usually do not appear in the best twelve-hour forecast models include the predicted temperature, the lights and the  $CO_2$  sensors. The observation that these sensors have short term value for predictions agrees with intuition. The temperature predictions are probably not as accurate for 12 h into the future as they are for shorter periods. The lighting and  $CO_2$  sensors also provide information about activities that are not likely to have a long term effect on temperature. These sensors report on activities and movements of building occupants that most likely do not occur according to any

**Table 6** Mean RMSE for selected sensors, generating 12-h forecasts of temperature, from 8 historical readings

Mean RMSE	Sensors	Remove Next
1.1303	Tw+TL+TD+SW+CL+CD	CD
1.1367	Tw+TL+TD+SW+CL	CL
1.1477	Tw+TL+TD+SW	TD
1.169	Tw+TL+SW	SW
1.2395	Tw+TL	Tw
1.3838	TL	

**Table 7** Mean RMSE for selected sensors, generating 3-h forecasts of temperature, from 4 historical readings

Mean RMSE	Sensors	Remove Next
0.0512	Tw+TL+T+SW+SE+Pcp+CL+CD	CL
0.0515	Tw+TL+T+SW+SE+Pcp+CD	CD
0.0516	Tw+TL+T+SW+SE+Pcp	Pcp
0.0519	Tw+TL+T+SW+SE	Tw
0.0535	TL+T+SW+SE	SW
0.0553	TL+T+SE	T
0.0599	TL+SE	SE
0.0631	TL	

schedule and have only a short term effect on the ambient temperature.

### 6.3 Forecasting temperature differences

Temperature controllers respond to changes in temperature, so some researchers investigate methods that forecast changes in temperature, such as the SML team [22].

Given a time sequence of temperatures  $y$  at each quarter-hour, we define  $z_i = y_i - y_{i-1}$  as the sequence of temperature differences over the previous quarter-hour. We set our goal to forecast  $\hat{z}_i$ , and otherwise follow the same method that we used when generating the forecasts  $\hat{y}_i$ . The results of selecting sensors for this forecasting problem are shown in this section.

Table 7 shows the errors when forecasting changes in temperature over the next 3 h, using one hour of historical observations, according to various sensors. Table 8 shows the same results but forecasting 12 h of temperature differences.

Our first observation is that forecasting temperature differences gives much smaller errors. When forecasting temperatures, we usually forecast numbers in the range of about  $20^\circ$ , whereas when forecasting temperature differences, we are forecasting numbers that are usually much less than  $1^\circ$ . Thus we should expect the errors to

**Table 8** Mean RMSE for selected sensors, generating 12-h forecasts of temperature, from 4 historical readings

Mean RMSE	Sensors	Remove next
0.0727	Tw+TP+TL+TD+SW+SE+Pcp+LD+CD	LD+CD
0.0728	Tw+TP+TL+TD+SW+SE+Pcp	TD
0.0731	Tw+TP+TL+SW+SE+Pcp	Pcp
0.0739	Tw+TP+TL+SW+SE	TP
0.0746	Tw+TL+SW+SE	SE
0.0752	Tw+TL+SW	SW
0.0775	Tw+TL	Tw
0.0863	TL	

**Table 9** Mean RMSE for selected sensors, generating 3-h forecasts of temperature changes, from 8 historical readings

Mean RMSE	Sensors	Remove Next
0.0512	Tw+TP+TL+T+SW+LD+HL	TP
0.0517	Tw+TL+T+SW+LD+HL	HL
0.052	Tw+TL+T+SW+LD	SW
0.0524	Tw+TL+T+LD	LD
0.0561	Tw+TL+T	T
0.0592	Tw+TL	Tw
0.0622	TL	

be much lower. For instance, it would not be informative in this house if we were to claim we could forecast that the temperature will not change by more than 1° over any quarter-hour, since it rarely does.

Now we consider slightly increasing the sensor history when forecasting differences in temperature. Table 9 shows the errors when forecasting changes in temperature over the next 3 h, using 2 h of historical observations, according to various sensors. Table 10 uses the same history length, but forecasts 12 h of temperature differences.

We notice that when forecasting temperature differences further into the future, our mean forecast error does not decrease as much as it does when forecasting temperature. Instead, the forecast errors are only about 40% greater when forecasting 12 h than they are when forecasting 3 h.

When forecasting temperature differences with more historical readings, that is, using 2 h instead of one hour, we see an effect similar to what we saw when forecasting temperatures. We see there is little improvement in mean RMSE gained by the extra available data.

### 6.4 Forecasting from much longer histories

Finally, we conducted two tests to further explore the effect of additional historical data readings from the sensors. In Section 6.2, we have shown that increasing from 1 to 2 h of sensor data did not improve forecast accuracy very much. In this section, we consider 12 h of sensor data.

**Table 10** Mean RMSE for selected sensors, generating 12-h forecasts of temperature changes, from 8 historical readings

Mean RMSE	Sensors	Remove Next
0.0704	Tw+TP+TL+SW+Pcp+LD	Pcp
0.0711	Tw+TP+TL+SW+LD	LD
0.0716	Tw+TP+TL+SW	TP
0.0722	Tw+TL+SW	SW
0.0735	Tw+TL	Tw
0.0854	TL	

**Table 11** Mean RMSE for selected sensors, generating 3-h forecasts of temperature, from 48 historical readings

Mean RMSE	Sensors	Remove Next
0.0469	Tw+TP+TL+TD+T+SW+SS	T
0.0493	Tw+TP+TL+TD+SW+SS	SS
0.0515	Tw+TP+TL+TD+SW	Tw
0.0528	TP+TL+TD+SW	TP
0.0529	TL+TD+SW	TL
0.0549	TD+SW	SW
0.0568	TD	

Table 11 shows the results of computing 3 h of temperature forecast from 12 h of sensor data. The largest set of sensors has a mean RMSE of 0.05 °C. The selected sensors included are a combination of external conditions and ambient temperatures, while no importance is placed on activities of the occupants. This error is significantly less than the error reported on Tables 2 and 5, both of which are about 0.25 °C.

Table 12 reports on the accuracy of computing changes in temperature over each 15-min interval in the next 48 h, using 12 h of historical readings. This forecast combines mostly environmental factors with some activities of the building occupants. Specifically it found that the twilight indicator, the sun on the walls, the atmospheric pressure and the lights in the dining room were most informative. For 12-h forecasts, the ambient room conditions were not found to be useful.

Table 12 shows an error that is surprisingly low for a 48 hour forecast. It is approximately the same error as seen in the 3-h forecasts in Table 7. It is smaller than error for the 12-h forecast in Table 8, even though the forecast horizon is increased by a factor of four.

## 7 Related work

The SML team reports [21] accuracy when forecasting temperature differences over future quarter-hour intervals,

**Table 12** Mean RMSE for selected sensors, generating 48-h forecasts of temperature changes from 48 historical readings

Mean RMSE	Sensors	Remove Next
0.0562	Tw+SW+SS+SE+P+LD	SS
0.0563	Tw+SW+SE+P+LD	SW
0.0565	Tw+SE+P+LD	LD
0.057	Tw+SE+P	SE
0.058	Tw+P	Tw
0.0639	P	

using data from among the 88 sensors they installed. They selected from among this set: internal temperature (TD and TL), irradiance (P), internal humidity (HD and HL) and precipitation (PCP). Based on their results for forecasting over 3 h, a selection of three sensors gave the lowest errors: internal temperature, solar irradiance, and a time-categorical variable. Using these sensors, forecasts for each quarter-hour over 3 h were generated using a combination of forecast models based on ANNs. They achieve error of about 0.11 °C Mean Absolute Error. We report forecast with about half of the error for the same problem. See Table 11.

The SML team, in later work [22] explored the selection of sensors for forecasting temperature differences over each quarter-hour with a forecast horizon of 48 h. They report a maximal MAE of about 1 °C, although the accuracy was often much smaller.

In our investigations of this same data and the same 48 hour forecasting problem, we experimented with a naive forecast, computed by always forecasting the mean temperature difference and looking at no sensors at all. This gives a mean RMSE of 0.12 °C. This result also can be compared to our Table 6 which is based on the same data. These forecasts have mean MAE of about 0.06 °C, which is an order of magnitude smaller than the previously published error.

Feature extraction shares some similarities with feature selection. Feature extraction is the process of defining new features from existing ones, by selecting those features with good predictive accuracy, and repackaging them into linear combinations that are considered new features. Partial least squares and principal component analysis are two feature extraction techniques [2, 6].

We used the same SML data for partial least squares and principle components [15]. Using four historical readings per sensor, we found the RMSE forecast error for both methods to be about 0.7 3-h forecasts whereas the comparable mean RMSE values in this paper range from 0.24 to 0.34. Likewise for twelve-hour forecasts, the RMSE for the feature extraction methods was about 1.7 for twelve-hour forecasts, and ranged from 1.15 to 1.42 in this paper. The results were similar for eight historical readings per sensor. Thus, lasso regression and best-first search exhibit better forecast accuracy than these feature extraction methods for temperature forecasting.

## 8 Conclusion

A model predictive controller can achieve significant savings by using an accurate temperature forecast when determining whether or not to engage HVAC systems. Temperature forecasts are informed by sensor data. We propose a cost-benefit analysis that balances the cost arising from installation, operating and computation against the

benefit of saving energy. A sensor's cost exceeds its benefit if it does not improve forecast accuracy by an amount sufficient to be useful to the controller.

The method we describe generates accurate temperature forecasts using lasso regression. It uses a best-first search technique to incrementally consider larger sets of sensors until no additional sensor improves the forecast accuracy. It then reduces this set by removing sensors incrementally and reporting the resulting sequence of forecast errors. If we assume that energy savings increase with forecast accuracy, this sequence of sets of sensors should help finding the optimal set of sensors.

Our system computes a gracefully degrading set of sensors for different situations, depending on the length of the forecast horizon, the number of historical observations, and whether the controller performs better with a lower mean error or a lower maximal error. Our findings indicate that the selection of sensors will be affected by these factors. In a new installation, we propose to temporarily install a large set of sensors and to collect readings from these sensors over several weeks. Then, it should be possible to determine which sensors to permanently install. Alternately, in an existing installation, the maintenance and computation costs may be reduced by removing sensors that are not providing benefit. The same gracefully degrading sequence can guide this selection.

Our experiments show accuracy increases as more data is available for forecasting. Shorter-term forecasts are more accurate than longer-term forecasts, and derive benefit from more sensors than longer-term forecasts.

We have used lasso regression over lagged data as the underlying modelling technology. While the search technique we employ for selecting sensors can be applied to any underlying modelling technology, lasso regression has shown good performance. In a comparison with previously published forecasts based on artificial neural nets, the lasso forecasts show considerable improvement.

In the future, we plan to apply this proof of concept to a set of small university buildings with 12 sensors, and a model predictive controller.

**Funding information** The authors gratefully acknowledge the financial support of their organizations, and Zayed University's RIF 17062 fund.

## References

1. Álvarez J, Redondo J, Camponogara E, Normey-Rico J, Berenguel M, Ortigosa P (2013) Optimizing building comfort temperature regulation via model predictive control. *Energ Buildings* 57:361–372



2. Bjørn-Helge MRW (2007) The `pls` package: principal component and partial least squares regression in R. *J Stat Softw* 18(2):1–24. <https://doi.org/10.18637/jss.v018.i02>
3. EIA (2017) Frequently asked questions how much energy is consumed in u.s. residential and commercial buildings? <https://www.eia.gov/tools/faqs/faq.php?id=86&t=1>
4. Friedman J, Hastie T, Tibshirani R (2010) Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 33(1):1–22. <http://www.jstatsoft.org/v33/i01/>
5. Friedman J, Hastie T, Simon N, Tibshirani R (2016) Package `glmnet`: lasso and elastic-net regularized generalized linear models ver 2.0-. <https://cran.r-project.org/web/packages/glmnet/glmnet.pdf>
6. Mevik BH, Wehrens R, Liland KH (2015) `pls`: Partial least squares and principal component regression. <https://CRAN.R-project.org/package=pls>, r package version 2.5-0
7. Moreno MV, Zamora MA, Skarmeta AF (2014) User-centric smart buildings for energy sustainable smart cities. *Transactions on emerging telecommunications technologies* 25(1):41–55. <https://doi.org/10.1002/ett.2771>
8. Morosan P, Bourdais R, Dumur D, Buisson J (2010) Building temperature regulation using a distributed model predictive control. *Energ Buildings* 42(9):1445–1452
9. Pan D, Yuan YDW, Xu X, Peng Y, Peng X, Wan PJ (2012) Thermal inertia: towards an energy conservation room management system. In: Greenberg A, Sohaby K (eds) *INFOCOM*. IEEE, pp 2606–2610
10. Pérez-Lombard L, Ortiz J, Pout C (2008) A review on buildings energy consumption information. *Energ Buildings* 40:394–398
11. Prívara S, Šíroký J, Ferkl L, Cigler J (2011) Model predictive control of a building heating system: The first experience. *Energ Buildings* 43(2):564–572. <https://doi.org/10.1016/j.enbuild.2010.10.022>. <http://www.sciencedirect.com/science/article/pii/S0378778810003749>
12. Romanski P, Kotthoff L (2016) Package ‘FSelector’ selecting attributes. R package version 0:21. <https://cran.r-project.org/web/packages/FSelector/FSelector.pdf>
13. Simon N, Friedman J, Hastie T, Tibshirani R (2011) Regularization paths for cox’s proportional hazards model via coordinate descent. *J Stat Softw* 39(5):1–13. <http://www.jstatsoft.org/v39/i05/>
14. Spencer B, Al-Obeidat F (2016) Temperature forecasts with stable accuracy in a smart home. In: *SEIT Sustainable Energy Information Technology Conference*. <https://doi.org/10.1016/j.procs.2016.04.160>
15. Spencer B, Al-Obeidat F, Alfandi O (2016) Short term forecasts of internal temperature with stable accuracy in smart homes. *Int J Thermal Environ Eng* 13(2):81–89
16. Spencer B, Al-Obeidat F, Alfandi O (2017) Selecting sensors when forecasting temperature in smart buildings, vol 109, pp 777–784. <https://doi.org/https://doi.org/10.1016/j.procs.2017.05.321>. <http://www.sciencedirect.com/science/article/pii/S1877050917309857>, 8th International Conference on Ambient Systems, Networks and Technologies, ANT-2017 and the 7th International Conference on Sustainable Energy Information Technology, SEIT 2017, 16–19 May 2017, Madeira, Portugal
17. Tibshirani R (1996) Regression shrinkage and selection via the lasso: a retrospective. *J R Stat Soc Ser B (Stat Methodol)* 58(1):267–288. <http://www.jstor.org/stable/2346178>
18. UCI (2010) `Sml2010` data set. <https://archive.ics.uci.edu/ml/datasets/SML2010>
19. United States Department of Energy (2012) Solar decathlon Europe competition. <http://www.solardecathlon.gov>
20. Yuan Y, Pan D, Wang D, Xu X, Peng Y, Peng X, Wan PJ (2013) A study towards applying thermal inertia for energy conservation in rooms. *ACM Trans Sen Netw* 10(1):7:1–7:25. <https://doi.org/10.1145/2529050>. <http://doi.acm.org.proxy.hil.unb.ca/10.1145/2529050>
21. Zamora-Martínez F, Romeu P, Botella-Rocamora P, Pardo J (2013) Towards energy efficiency: forecasting indoor temperature via multivariate analysis. *Energies* 6(9):4639. <https://doi.org/10.3390/en6094639>. <http://www.mdpi.com/1996-1073/6/9/4639>
22. Zamora-Martínez F, Romeu P, Botella-Rocamora P, Pardo J (2014) On-line learning of indoor temperature forecasting models towards energy efficiency. *Energ Buildings* 83:162–172