Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 126 (2018) 1903–1912

Procedia Computer Science

www.elsevier.com/locate/procedia

22nd International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

# An acquisition system of in-house parameters from wireless sensors for the identification of an environmental model

Giorgio Biagetti[a], Diego Coccia[a], Paolo Crippa[a,*], Laura Falaschetti[a], Claudio Turchetti[a]

[a]DII – Department of Information Engineering, Università Politecnica delle Marche, Via Brecce Bianche 12, I-60131, Ancona, Italy

**Abstract**

This paper presents a system for the acquisition of in-house parameters, such as temperature, pressure, humidity and so on, that can be used for the intelligent control of a building. The main objective of this work is to determine an environmental model of an in-house room using machine learning techniques. The system is based on a low data-rate network of sensing and control nodes to acquire the data, realized with a new protocol, called ToLHnet, that is able to employ both wired and wireless communication on different media. Several standard machine learning techniques, namely linear regression, classification and regression tree algorithm, support vector machine, have been used for the regression of the input-output thermal model. Additionally, a recently proposed new technique named particle-Bernstein polynomial has been successfully applied. Experimental results show that this technique outperforms the previous techniques, for both accuracy and computation time.



*Keywords:* Machine learning, regression, identification, environmental model

## 1. Introduction

Primary energy consumption in buildings represents about 40% of total energy produced world wide, and more than a half is used by heating ventilation and air conditioning (HVAC) systems [2]. Accurate internal temperature forecast can reduce energy usage in building firstly because it is more efficient to maintain temperature in a room than to heat or cool it and secondly because HVAC controller system is able to minimize a cost function based on energy consumption. The most challenging phase in predictive building controller designs is the control-oriented modeling of building thermal dynamics. Time-series prediction has demonstrated as one of the most powerful method to model indoor temperature behaviour [3, 16, 11]. Additionally the increasing availability and use of sensor devices enables to collect a large amount of data [15]. As a consequence, combining these two key aspects, machine learning techniques

* Corresponding author. Tel.: +39-071-220-4541 ; fax: +39-071-220-4464.
*E-mail address:* p.crippa@univpm.it

1877-0509 
10.1016/j.procs.2018.08.072

for the identification of input-output nonlinear systems from data, can be fruitfully adopted for internal temperature modeling [12, 13].

The aim of this paper is to present a system for the acquisition of in-house parameters, such as temperature, pressure, humidity and so on, that can be used for an intelligent control of building system. The key points of such a system are: *i)* low data-rate network of sensing and control nodes, *ii)* the machine learning techniques for the environmental modeling of an in-house room. As far as the first point is concerned a new protocol, called ToLHnet [5, 6, 1, 4], that is able to employ both wired and wireless communication on different media among thousands of nodes, has been adopted. With reference to the second point several standard machine learning techniques [9], namely linear regression (LR) [14], classification and regression tree algorithm (CART) [8], support vector machine (SVM) [10], have been used. Additionally a recently proposed new technique named particle-Bernstein polynomials (PBP) [7] has been used for the regression of input-output thermal modeling, that outperforms previous techniques.

The remainder of this paper is organized as follows. In Section 2 we describe the wireless in-house parameters acquisition system used for the experimentation. In Section 3 the environmental model identification problem is formalized and the machine learning techniques for regression are described. Section 4 is devoted to experimental results. Finally, Section 5 summarizes some conclusions.

## 2. Wireless in-house parameters acquisition system

### 2.1. ToLHnet protocol

ToLHnet is a simple communication protocol that was conceived for implementing low-data-rate networks of sensing and control nodes, using a tree-based routing scheme. The protocol is able to employ both wired and wireless communication over different media among thousands of nodes. The protocol was designed so as to allow a strongly asymmetrical implementation, moving most of the complexity out of the standard nodes and into a single special node that will be the master controller of the network. To this end, as the master controller takes care of computing routing tables, assigning addresses, configuring the network, it shall have larger computing and memory resources than those required by other nodes. The main tasks the master controller carries out are:

- finding the optimal tree;
- choosing the routers;
- building appropriate routing tables;
- dispatching the tables to the routers.

### 2.2. Sensor network

The sensor network used for the experiments includes 19 Bluetooth Low Energy (BLE) sensors, 3 nodes and a master controller equipped with a Linux operating system. With the firmware developed a maximum of 7 BLE sensors can be polled by each ToLHnet node, thus the architecture chosen in this work is as shown in Fig. 1.

The sensors are able to provide measures of ambient temperature, humidity and pressure, at a rate of 1 reading/minute by sending them to the linked node. The sensor network topography used for the experiments is reported in Fig. 2; as you can see the sensors have been deployed in a 5 m × 8.93 m room to form a uniformly spaced grid.

### 2.3. Client/server architecture

The acquisition system has been organized on the basis of a client/server architecture that, through a browser, allows the user to have access to the data that reside on a server. The architecture uses an HTTP communication protocol as shown in Fig. 3. Two databases have been used to storage the data: SQLite for the storage of the sensors data and MySQL for the management of the graphical interface (create sensors map, update sensors map, select node, change node, . . . ).

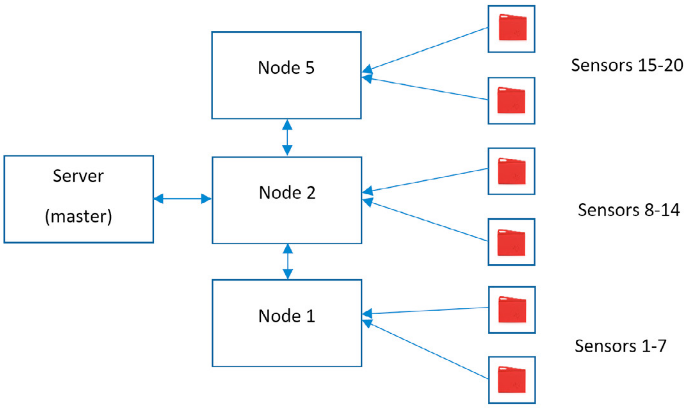


Fig. 1. Architecture of sensor network.

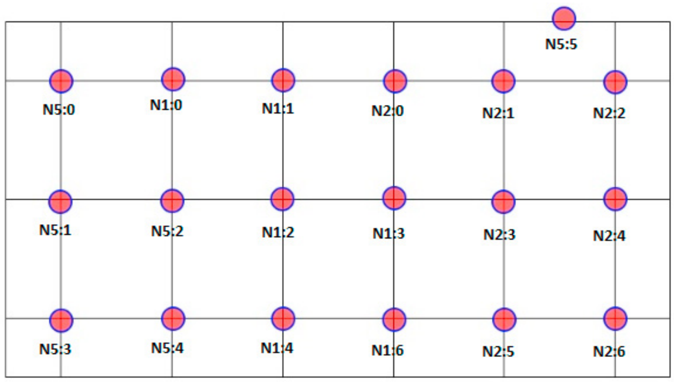


Fig. 2. The map of sensors.

## 3. Environmental model identification

The main objective of this work is to determine an environmental model of an in-house room using machine learning techniques. Mathematically this problem can be formalized as the identification of an input-output relationship

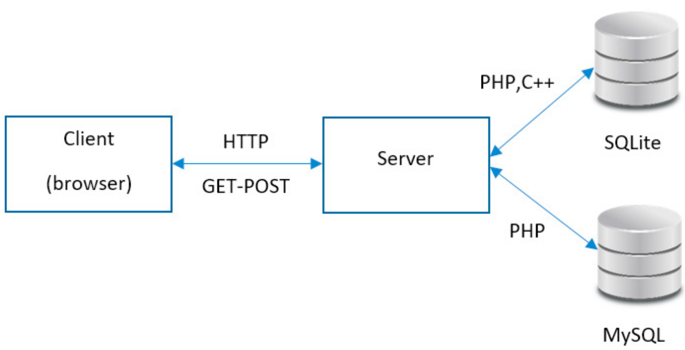


Fig. 3. Client/server architecture.

$$T_{in}(t) = h\left(T_{ext}(t)\right), \; t = 1, \dots, N \tag{1}$$

between internal temperature $T_{in}(t)$ and external temperature $T_{ext}(t)$, sampled at different time instants $t$. In this work we refer to the network of Fig. 1, thus $T_{in}(t)$ is a vector that contains the 18 temperatures acquired by the internal sensors $(N_{5:0}, N_{1:0}, \dots)$, while $T_{ext}(t)$ is the external temperature measured by the sensor $N_{5:5}$. Due to the thermal inertia the internal temperature $T_{in}(t)$ at time instant $t$, depends on present and past values of $T_{ext}(t)$, thus (1) can be rewritten as

$$T_{in}(t) = h\left(T_{ext}(t), T_{ext}(t-1), \dots, T_{ext}(t-p)\right) \tag{2}$$

where $T_{ext}(t-i)$, $i = 1, \dots, p$ are the lagged values of $T_{ext}(t)$. The identification of model (2) corresponds to the regression of the function $f(\cdot)$ through input-output data, thus can be viewed as a supervised learning problem. In particular the data matrices used for learning have the following form

$$X = \begin{bmatrix} x(t) & x(t-1) & x(t-2) & \cdots & x(t-p) \\ x(t+1) & x(t) & x(t-1) & \cdots & x(t-p+1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x(t+N) & \cdots & \cdots & \cdots & x(t-p+N) \end{bmatrix} \tag{3}$$

$$Y = \begin{bmatrix} y_1(t) & y_2(t) & \cdots & y_{18}(t) \\ y_1(t+1) & y_2(t+1) & \cdots & y_{18}(t+1) \\ \vdots & \vdots & \vdots & \vdots \\ y_1(t+N) & y_2(t+N) & \cdots & y_{18}(t+N) \end{bmatrix} \tag{4}$$

where $N$ is the number of the observations.

### *3.1. Regression model of thermal behaviour*

The regression of input-output relationship (2) given the data matrices (3) and (4) can be solved using a machine learning approach. In this paper some standard techniques, LR, CART, SVM, and a novel technique, PBP, have been used for comparison. As the standard techniques are well known we will briefly describe only the most recent PBP technique.

### *3.2. Machine learning regression based on particle-Bernstein polynomials*

This particularly effective algorithm has been recently proposed for the regression of input-output relationships from data, and is based on Bernstein polynomials. The $m-$degree Bernstein polynomials are defined by

$$b_k^m(x) = \binom{m}{k} x^k (1-x)^{m-k}, \quad k = 0, 1, \dots, m \tag{5}$$

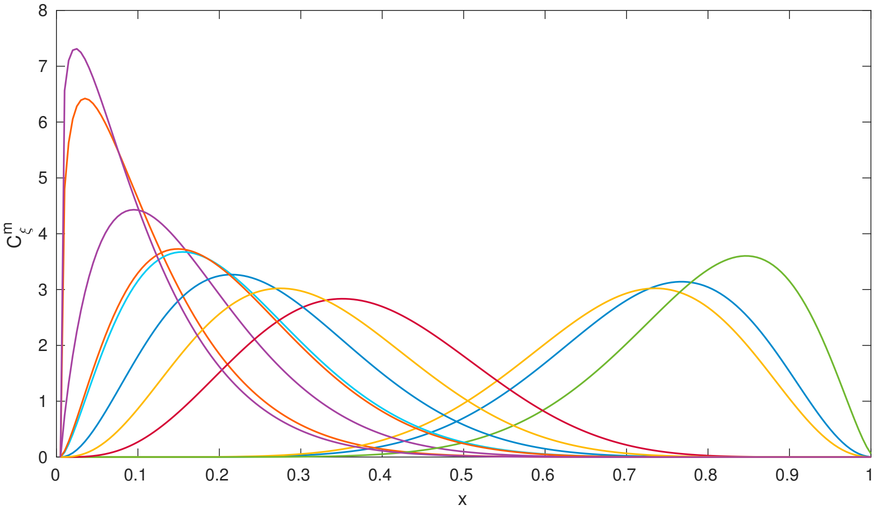


Fig. 4. Particle-Bernstein polynomials.

where $x \in [0, 1]$ and $\binom{m}{k} = \frac{m!}{(m-k)!\,k!}$. In (5) both the variables $m$ and $k$ are integer, as the binomial coefficients are defined for integer values. Relaxing this constraint assuming $k$ is real and denoting this value with $\xi$, a new set of functions called particle-Bernstein polynomials can be defined as follows

$$C_\xi^m(x) = \alpha_\xi^m\, x^\xi\, (1-x)^{m-\xi} = \alpha_\xi^m\, k_\xi^m(x), \quad \xi \in \mathbb{R}^1, \quad \xi \in [0, m] \quad , \tag{6}$$

where the coefficients $\alpha_\xi^m$ are chosen in such a way the integral constraint

$$\int_0^1 C_\xi^m(x)\, dx = 1, \tag{7}$$

holds. Several examples of functions defined by (6) for $m = 20$ and different values of $\xi$ are reported in Fig. 4.

## 4. Experimental results

The experimental data used for validating the proposed approach were collected by the sensor network of Fig. 2 over a period of six days. Temperature was sampled at a rate of 1 sample every minute, so as 8640 observations were gathered. Since an observation includes 18 values of internal temperature and 1 value of external temperature, a total of 164160 temperature reading were collected. The identification of the input-output relationship (2) was performed using several different algorithms, namely LR, CART, SVM, PBP. For each algorithm two models were derived using two different input/output data matrices:

1. Input data matrix $X_1(8620 \times 20)$: in this case the temperature at time instant $t$ depends on the $p = 19$ lagged values;
   Output data matrix $Y_1(8620 \times 18)$.
2. Input data matrix $X_2(8540 \times 100)$: in this case the temperature at time instant $t$ depends on the $p = 99$ lagged values;
   Output data matrix $Y_2(8540 \times 18)$.

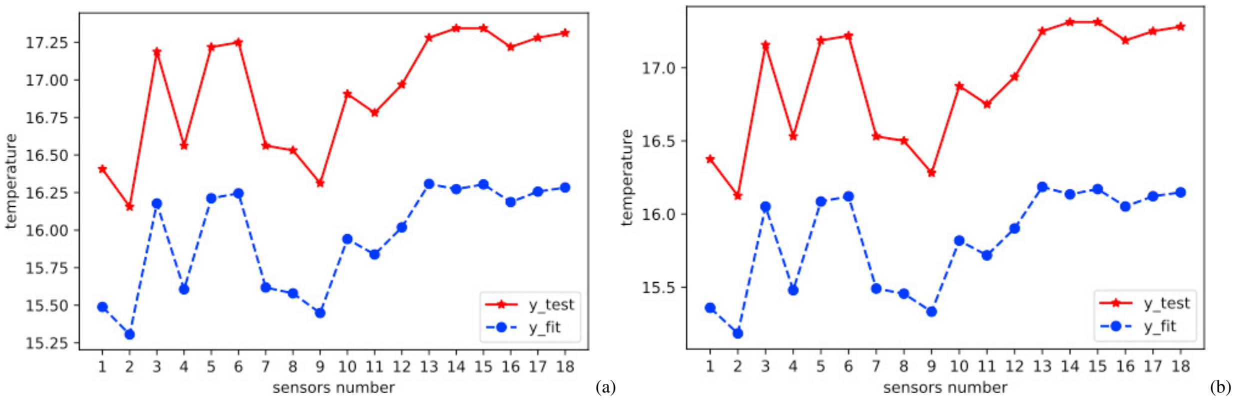


Fig. 5. LR: (a) Input data matrix $X_1$, frame 1000. (b) Input data matrix $X_2$, frame 1000.

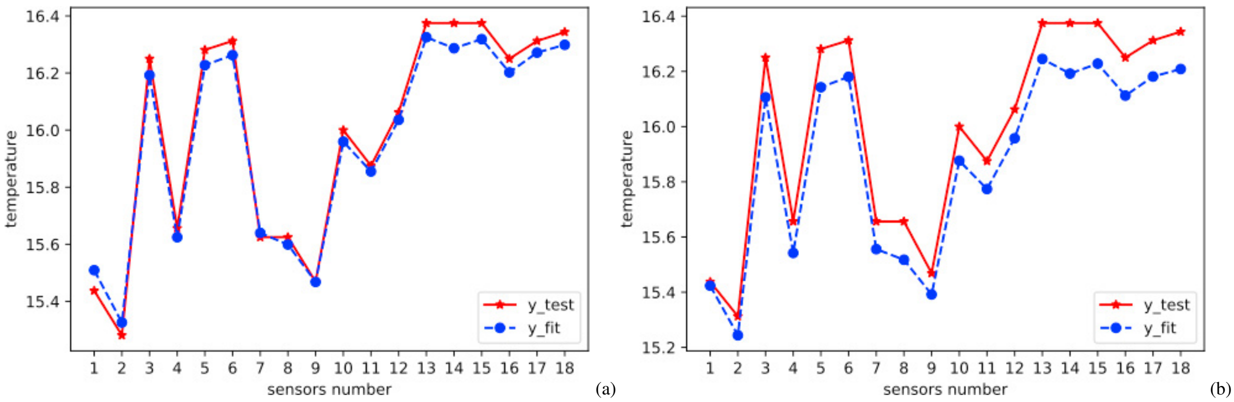


Fig. 6. LR: (a) Input data matrix $X_1$, frame 3000. (b) Input data matrix $X_2$, frame 3000.

Four testing vectors were used corresponding to the frames 1000, 3000, 6500 and 8621 for $p = 19$ and to the frames 1000, 3000, 6500 and 8541 for $p = 99$. For each test, the corresponding row and its corresponding $p$ following lines were removed from the data matrix used for training.

The experiments were conducted on a computer equipped with:

- CPU Intel Core i7-6800K (15M Cache, up to 3.60 GHz);
- Ram DIMM DDR4 32GB at 2666 MHz.

### 4.1. Linear regression (LR)

The results achieved with this algorithm for the frame 1000 and 3000 are reported in Fig. 5, 6.

### 4.2. Classification and regression trees algorithm (CART)

The results achieved with this algorithm for the frame 1000 and 3000 are reported in Fig. 7, 8.

### 4.3. Support vector machine algorithm (SVM)

The results achieved with this algorithm for the frame 1000 and 3000 are reported in Fig. 9, 10.

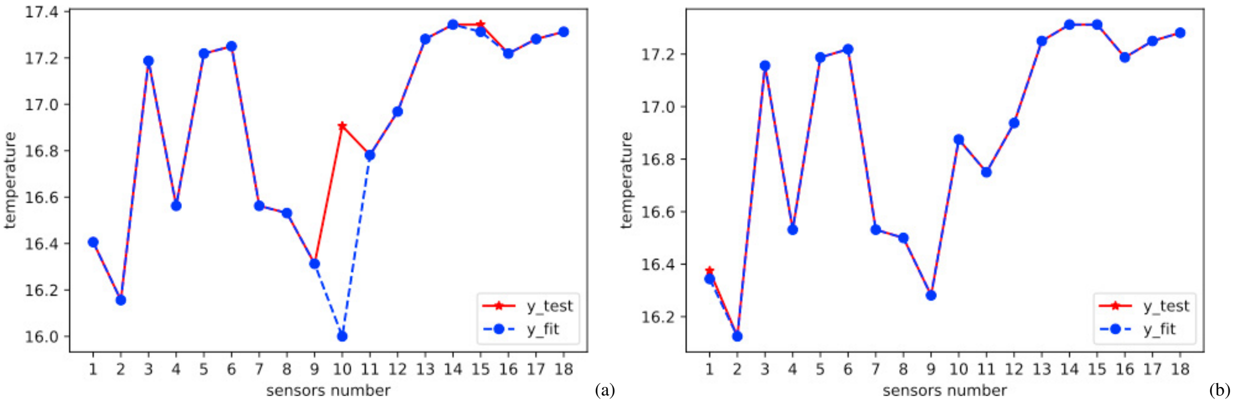


Fig. 7. CART: (a) Input data matrix $X_1$, frame 1000. (b) Input data matrix $X_2$, frame 1000.

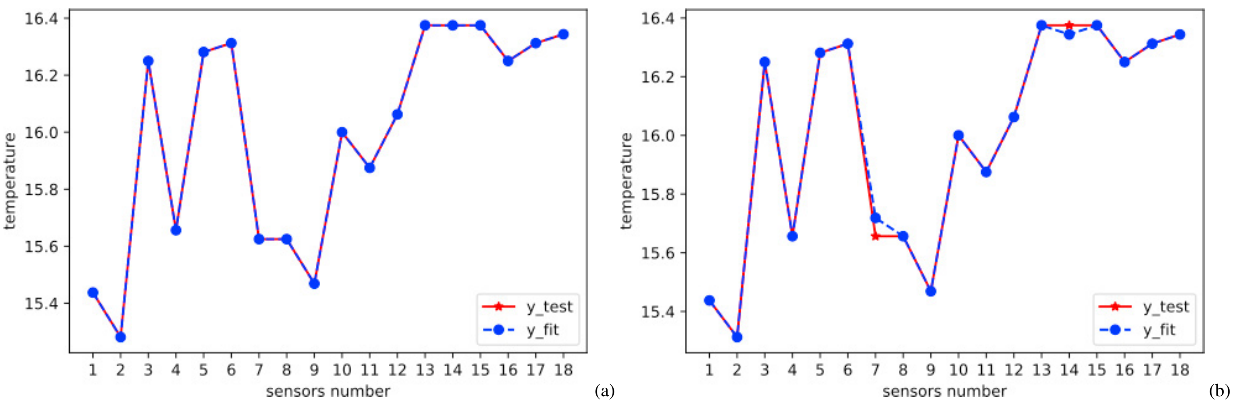


Fig. 8. CART: (a) Input data matrix $X_1$, frame 3000. (b) Input data matrix $X_2$, frame 3000.

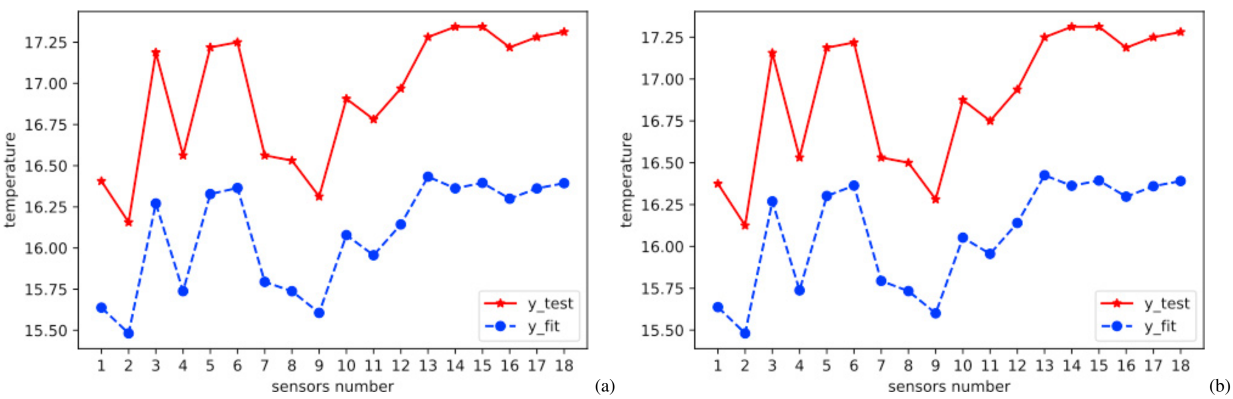


Fig. 9. SVM: (a) Input data matrix $X_1$, frame 1000. (b) Input data matrix $X_2$, frame 1000.

### *4.4. Particle-Bernstein polynomial algorithm (PBP)*

The results achieved with this algorithm for the frame 1000 and 3000 are reported in Fig. 11, 12.

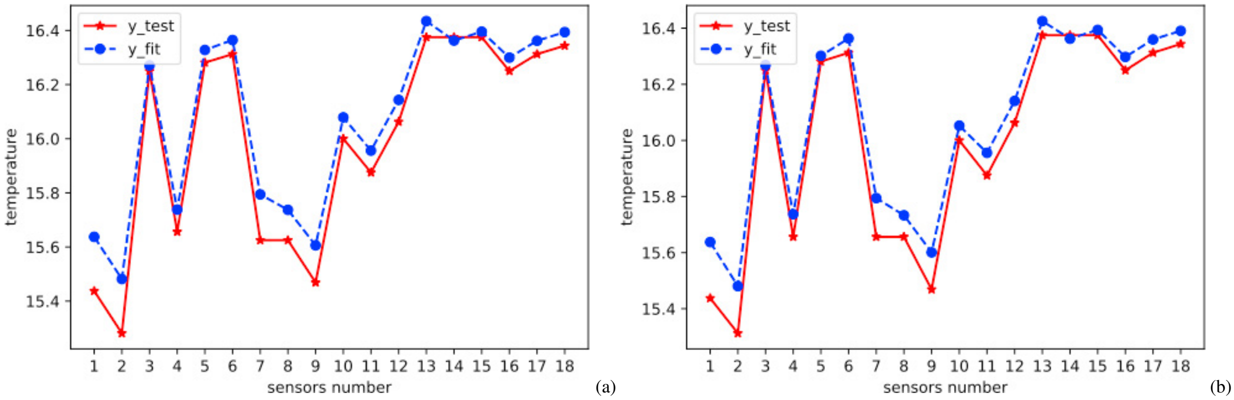


Fig. 10. SVM: (a) Input data matrix $X_1$, frame 3000. (b) Input data matrix $X_2$, frame 3000.

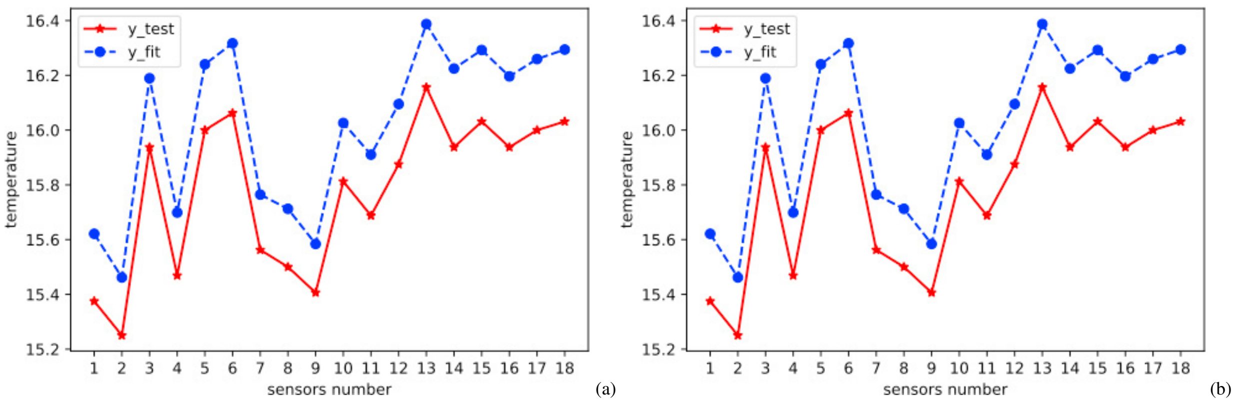


Fig. 11. PBP: (a) Input data matrix $X_1$, frame 1000. (b) Input data matrix $X_2$, frame 1000.

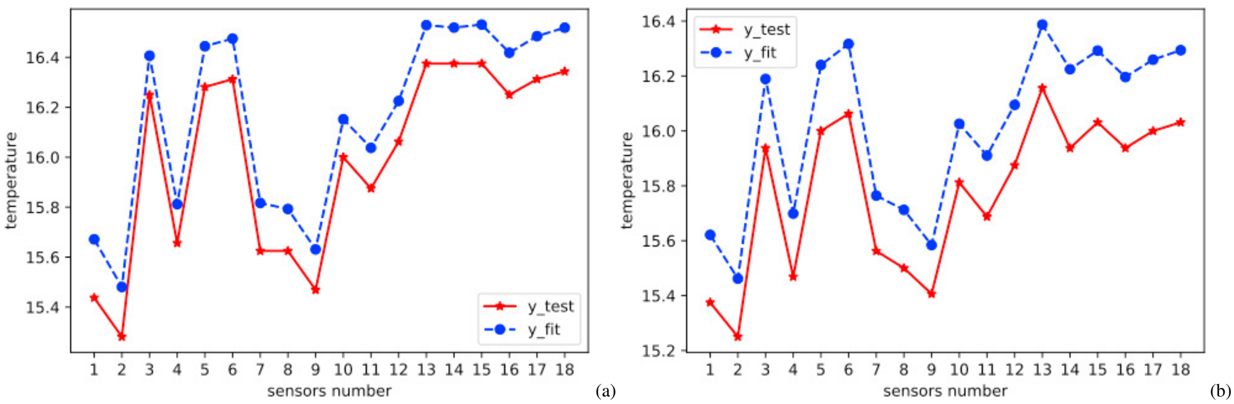


Fig. 12. PBP: (a) Input data matrix $X_1$, frame 3000. (b) Input data matrix $X_2$, frame 3000.

## 4.5. *Algorithm comparison*

The complete results are summarized in Table 1 for $p = 19$ and Table 2 for $p = 99$. These tables report the performances of each algorithm in terms of mean squared error regression loss (the best value is 0.0), explained variance regression score function (the best possible score is 1.0), and $R^2$ (coefficient of determination) regression score function (the best possible score is 1.0).

Table 1. MSE, Variance score and $R^2$ score for $p$ = 19.

| frame number | MSE | | | |
|---|---|---|---|---|
| | LR | CART | SVM | PBP |
| 1000 | 0.952207 | 0.952207 | 0.723615 | **0.056278** |
| 3000 | **0.002207** | **0.002207** | 0.010166 | 0.028946 |
| 6500 | 0.100648 | 0.100648 | 0.090766 | **0.051349** |
| 8621 | 2.950574 | 3.002713 | 4.678558 | **0.384642** |
| frame number | **Variance score** | | | |
| | LR | CART | SVM | PBP |
| 1000 | 0.979 | 0.979 | 0.957 | **0.991** |
| 3000 | 0.990 | 0.990 | 0.975 | **0.997** |
| 6500 | 0.101 | 0.990 | 0.956 | **0.991** |
| 8621 | 0.959 | -6.034 | 0.968 | **0.995** |
| frame number | **$R^2$ score** | | | |
| | LR | CART | SVM | PBP |
| 1000 | 0.979 | 0.979 | 0.957 | **0.991** |
| 3000 | 0.990 | 0.990 | 0.975 | **0.997** |
| 6500 | 0.101 | 0.990 | 0.956 | **0.991** |
| 8621 | 0.959 | -6.034 | 0.968 | **0.995** |

Table 2. MSE, Variance score and $R^2$ score for $p$ = 99.

| frame number | MSE | | | |
|---|---|---|---|---|
| | LR | CART | SVM | PBP |
| 1000 | 1.148228 | **0.000054** | 0.679614 | 0.056278 |
| 3000 | 0.015192 | **0.000271** | 0.008087 | 0.056278 |
| 6500 | 0.183171 | **0.001302** | 0.097049 | 0.056278 |
| 8541 | 2.700020 | 0.000217 | 4.657414 | **0.000006** |
| frame number | **Variance score** | | | |
| | LR | CART | SVM | PBP |
| 1000 | 0.975 | **1.000** | 0.957 | 0.991 |
| 3000 | 0.990 | **0.998** | 0.978 | 0.991 |
| 6500 | 0.974 | 0.984 | 0.936 | **0.991** |
| 8541 | 0.963 | 0.998 | 0.966 | **1.000** |
| frame number | **$R^2$ score** | | | |
| | LR | CART | SVM | PBP |
| 1000 | -6.441 | **1.000** | -3.404 | 0.247 |
| 3000 | 0.885 | **0.998** | 0.939 | 0.247 |
| 6500 | -1.572 | **0.982** | -0.363 | 0.247 |
| 8541 | -19.070 | 0.998 | -33.621 | **1.000** |

Table 1 shows that the better performances are obtained with the PBP algorithm, while, as shown in Table 2, with $p$ = 99, CART algorithm achieves the better accuracy. However, in terms of computation time, PBP achieves always the better results, as shown in Tables 3, 4.

## 5. Conclusion

In this paper, an acquisition system of in-house parameters from wireless sensors for the identification of an environmental model, has been presented. The system employs a low data-rate network of sensing and control nodes,

Table 3. Execution time for $p = 20$.

| frame number | Training & Testing time [s] | | | |
|---|---|---|---|---|
| | LR | CART | SVM | PBP |
| 1000 | 0.063287 | 1.414434 | 41.268678 | **0.014175** |
| 3000 | 0.059241 | 1.314986 | 41.318781 | **0.018037** |
| 6500 | 0.059966 | 1.384132 | 39.158068 | **0.013535** |
| 8621 | 0.057017 | 1.379291 | 41.202026 | **0.009359** |

Table 4. Execution time for $p = 100$.

| frame number | Training & Testing time [s] | | | |
|---|---|---|---|---|
| | LR | CART | SVM | PBP |
| 1000 | 0.506092 | 8.561662 | 94.922802 | **0.037444** |
| 3000 | 0.508766 | 8.133819 | 95.163664 | **0.038142** |
| 6500 | 0.507753 | 8.578798 | 90.216109 | **0.049462** |
| 8541 | 0.514602 | 8.525970 | 94.794116 | **0.036270** |

using a new protocol, called ToLHnet. The regression of the input-output thermal model has been realized using several standard machine learning techniques. Among this, a recently proposed new technique named particle-Bernstein polynomial has been successfully applied. Experimental results show that, in term of mean squared error regression loss, variance regression score and $R^2$ regression score, for $p = 19$, this technique achieves the best performance. Furthermore, in terms of computation time, this new technique achieves always the better results.

## References


[1] Alessandrini, M., Biagetti, G., Crippa, P., Falaschetti, L., Orcioni, S., Turchetti, C., 2016. Optimizing linear routing in the ToLHnet protocol to improve performance over long RS-485 buses. EURASIP Journal on Embedded Systems 2017, 7.

[2] Atam, E., Helsen, L., 2016. Control-oriented thermal modeling of multizone buildings: Methods and issues: Intelligent control of a building system. IEEE Control Systems 36, 86–111.

[3] Barbosa, M.R., Lopes, A.M., 2017. Temperature time series: Pattern analysis and forecasting, in: 2017 4th Experiment@International Conference (exp.at'17), pp. 226–231.

[4] Biagetti, G., et al., 2014. ToLHnet - Tree or Linear Hopping network. URL: http://www.tolhnet.org/.

[5] Biagetti, G., Crippa, P., Curzi, A., Orcioni, S., Turchetti, C., 2014. ToLHnet: A low-complexity protocol for mixed wired and wireless low-rate control networks, in: Proceedings of the 6th European Embedded Design in Education & Research Conference (EDERC 2014), Milan, Italy. pp. 177–181.

[6] Biagetti, G., Crippa, P., Falaschetti, L., Orcioni, S., Ortolani, N., Turchetti, C., 2015. Improvement of RS-485 performance over long distances using the ToLHnet protocol, in: Proceedings of the 2015 12th International Workshop on Intelligent Solutions in Embedded Systems (WISES 2015), Ancona, Italy. pp. 85–89.

[7] Biagetti, G., Crippa, P., Falaschetti, L., Turchetti, C., 2017. Machine learning regression based on particle Bernstein polynomials for nonlinear system identification, in: 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1–6.

[8] Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and Regression Trees. Wadsworth & BrooksCole Advanced Books & Software., Monterey, CA.

[9] Christopher, M.B., 2016. Patter Recognition and Machine Learning. Springer-Verlag New York.

[10] Cortes, C., Vapnik, V., 1995. Support-vector networks. Machine Learning 20, 273–297.

[11] Grigorievskiy, A., Miche, Y., Ventelä, A.M., Séverin, E., Lendasse, A., 2014. Long-term time series prediction using OP-ELM. Neural Networks 51, 50 – 56.

[12] Mateo, F., Carrasco, J.J., Sellami, A., Millán-Giraldo, M., Domínguez, M., Soria-Olivas, E., 2013a. Machine learning methods to forecast temperature in buildings. Expert Systems with Applications 40, 1061 – 1068.

[13] Mateo, F., Fernández, J.J.C., Millán-Giraldo, M., Sellami, A., Escandell-Montero, P., Martínez-Martínez, J.M., Soria-Olivas, E., 2013b. Temperature forecast in buildings using machine learning techniques, in: ESANN, pp. 357 – 362.

[14] Seber, G.A., Lee, A.J., 2012. Linear regression analysis. volume 329. John Wiley & Sons.

[15] Spencer, B., Alfandi, O., 2016. Forecasting internal temperature in a home with a sensor network. Procedia Computer Science 83, 1244 – 1249. The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016) / The 6th International Conference on Sustainable Energy Information Technology (SEIT-2016) / Affiliated Workshops.

[16] Zamora-Martínez, F., Romeu, P., Botella-Rocamora, P., Pardo, J., 2014. On-line learning of indoor temperature forecasting models towards energy efficiency. Energy and Buildings 83, 162 – 172.