A soft computing based method for detecting lifetime building thermal insulation failures

Javier Sedano[1], Leticia Curiel[2], Emilio Corchado[2], Enrique de la Cal[3], José R. Villar[3]
[1] *Electromechanic Engineering Department, University of Burgos, Spain*
[2] *Department of Civil Engineering, University of Burgos, Spain*
[3] *Computer Science Department, University of Oviedo, Spain*

Abstract. The detection of thermal insulation failures in buildings in operation responds to the challenge of improving building energy efficiency. This multidisciplinary study presents a novel four-step soft computing knowledge identification model called IKBIS to perform thermal insulation failure detection. It proposes the use of Exploratory Projection Pursuit methods to study the relation between input and output variables and data dimensionality reduction. It also applies system identification theory and neural networks for modeling the thermal dynamics of the building. Finally, the novel model is used to predict dynamic thermal biases, and two real cases of study as part of its empirical validation.

1. Introduction

Predicting the thermal dynamics of a building is a complex task. The dynamic thermal performance of a building has mainly been used to estimate its power requirements. As an example, the difficulties in obtaining a black-box model for a generic building are documented [72, 20]. Assessing thermal insulation is a well-known problem that has not as yet been fully resolved [31, 80]. Several different techniques are proposed in the literature. In [33], thermal insulation leaks are found by measuring thermal resistance and infrared (IR) thermography, while in [12, 59] only IR thermography is used to locate thermal insulation failures. As the main drawback of using IR thermography is the high cost of equipment, it is interesting to analyze alternatives using different technologies.

Local regulations generally stipulate how thermal insulation should be calculated in new buildings. In the case of Spain, building and heating system regulations are adapted to five winter climate zones and five summer climate zones across the entire country. Building materials, insulation thickness, materials, and so on, are calculated according to each climate zone.

In this research, a novel four-step methodology is presented, which aims to generate a model for estimating the behaviour of indoor temperature in a building of a specific configuration. This methodology is called IKBIS, which stands for Intelligent Knowledge-Based Identification System. Firstly, the dynamic thermal behaviour of a specific configuration is obatined. Then, a post-processing step should be carried out to obtain suitable datasets. In this methodology, the dataset should be analysed using several statistical methods, Exploratory Projection Pursuit (EPP) [24], Principal Component Analysis (PCA) [56, 38, 77, 76, 27, 78], Maximum Likehood Hebian Learning [16, 26] and Cooperative Maximum-Likelihood Hebbian Learning (CMLHL) [17]. This analysis extracts the main relationships between the variables. A model is then generated to estimate the indoor temperature at a specific configuration; this step is based on the application of system identification theory [39, 11, 40]. Finally, the thermal insulation failure is identified when the temperature error, measured as the difference between the indoor temperature and the model output temperature, rises above a pre-set threshold.

The remainder of this paper is organized as follows. The following Sub-Section 1.1 details the problem description. Section 2 introduces unsupervised connectionist techniques for analysing the datasets in order to extract their relevant internal structures. Section 3 presents the knowledge-based approach for system identification. Section 4 describes the multi-step procedure. In Section 5, the experiments and results are presented and commented on. Finally, the conclusions are set out and comments are made on future lines of research.

1.1. *Spanish Regulations and the Problem Description*

In 2007, several regulations on buildings and construction were approved in Spain dealing with energy efficiency [2], project development and specifications [6, 5, 3], the energy consumption limitation in buildings [6, 1], heating systems in buildings is the RITE (Reglamento de las Instalaciones Térmicas en los Edificios) [4], and certifying energy efficiency in new buildings is detailed in [7].

As established in [7], the energy efficiency of a new building that is being designed should be calculated using the CALENER software package [8]. The energy efficiency in the case of buildings in operation (during the lifetime of the building) is still an open issue, and the assumption is that it will be based on heat flux and conductivity measurement. Such a procedure would have to comply with the aforementioned Spanish Regulations by considering the different climate zones, the different building materials, etc..

2. Soft Computing for Feature Selection

Soft computing is [43, 57, 10, 45] a set of various technologies which are used to solve inexact and complex problems [82]. It investigates, simulates, and analyzes complex issues and phenomena in an attempt to solve real-world problems [73].

Feature Selection and extraction [30, 46] involve feature construction, space dimensionality reduction, and sparse representations among others. These are all commonly used pre-processing tools in soft computing that undertake pattern recognition. Our approach to feature selection is based on the dimensionality reduction issue. Initially, we apply the following three projection methods: PCA [56, 38], MLHL [16, 26] and CMLHL [17]. They are applied again in a second step to analyse the internal structure of a data set that is representative of a case of study. If after applying these models a clear internal structure may be identified, it means that the data are sufficiently informative. Otherwise, further data must be collected again. In a third step, these models are used to perform space dimensionality reduction in order to identify interesting dimensions or projections.

2.1. *Data structure analysis using connectionist techniques*

Principal Component Analysis (PCA) [56, 38] is a statistical method which aims to find the orthogonal basis which maximizes the projection variance of the data for a given basis dimensionality. It may be used as a dimension reduction technique which preserves as much information as possible in the remaining dimensions. If we consider only the largest eigenvalues corresponding to the principal components, we can also find those components that contain most information, which may provide insight into the structure of the data.

Exploratory Projection Pursuit (EPP) [24, 16] is a recent statistical method which centres on solving the difficult problem of how to identify structure in complex high dimensional data. There is an index that measures the “interestingness” of a given projection, and which then represents the data in terms of projections that maximise that index.

Maximum-Likelihood Hebbian Learning (MLHL) [16, 26] is an implementation of EPP. The MLHL-based method has been widely used in the field of pattern recognition [16, 26, 17, 18] as an extension of PCA. It identifies interestingness [24, 16, 26] by maximising the probability of the residuals using specific probability density functions that are non-Gaussian, which analyse the fourth-order statistic or the kurtosis. Cooperative Maximum-Likelihood Hebbian Learning model (CMLHL) [17] is based on MLHL [16, 26] but adds lateral connections [17, 18] which have been derived from the Rectified Gaussian Distribution [66]. Let consider an N-dimensional input vector (x), and an M-dimensional output vector (y), with W_{ij} being the weight (linking input j to output i). CMLHL is computed by using four equations: where Eq. (1) corresponds to the feed-forward step, Eq. (2) with the lateral activation passing, Eq. (3) represents the feedback step and Eq. (4) is the rule for updating the weights. In these equations, τ is the "strength" of the lateral connections, η is the learning rate, b the bias parameter, p is a parameter in order to choose a function to maximise the likelihood of the residuals under particular models of probability density functions and A is a symmetric matrix used to modify the response to the data according to the distances between the output neurons. Finally, $[\,]^{+}$ is necessary to ensure that the y-values remain within the positive quadrant.

$$y_i = \sum_{j=1}^{N} W_{ij} x_j, \forall i \tag{1}$$

$$y_i(t+1) = \left[y_i(t) + \tau(b - Ay)\right]^{+} \tag{2}$$

$$e_j = x_j - \sum_{i=1}^{M} W_{ij} y_i, \forall j \tag{3}$$

$$\Delta W_{ij} = \eta y_i \, sign(e_j) |e_j|^{p-1} \tag{4}$$

3. System Identification

System identification is concerned with obtaining a model that best suits a certain process behaviour (see Figure 1) [47]. Firstly, several are sampled from the process. The data is then analysed to obtain a model that estimates the desired process behaviour. The model is then used to optimize the process output. Finally, the process is modified to enhance its outcome. If more adjustments are needed the cycle is repeated.

The system identification procedure includes the experimental design, data visualisation and analysis, and the learning, testing, and validation of the model [47, 70, 65, 79, 52, 60].

When the data set is prepared, several tasks should be carried out: eliminating missing data and outliers [28, 29, 14, 21] scaling and normalizing the data [67], etc.

Selection of the model structure, its training and validation represents the core of the system identification. According to [47], several measures have been proposed in the literature to evaluate the goodness of a model, i.e., the one-step ahead prediction error (FIT1), the ten-step ahead prediction error (FIT10), or the simulation error, (FIT), the loss or the error function (V), the generalization error value (the numeric value of the normalized sum of squared errors, NSSE), or the final prediction error (FPE) among others.

Choosing the most suitable model structure requires a degree of expertise. Model structures range from classical black-box models – such as FIR (finite impuse response), OE (output error), autoregressive models ((ARMA, ARX, ARMAX, ARARX, ARARMAX, etc.) and the BJ (box-Jenkins) models- to fuzzy systems [81], neural networks [22] and support vector regressions [25], including the N4SID (numerical algorithm for subspace state space) [55], the MOESP (multivariable output error state-space model) [71], the Laguerre models [75], the nonlinear models from the Volterra series [42], the NARMAX model [15], and ANN models with a set of regressors NARX, NOE, NARMAX, NFIR, NSSIF [54, 64, 58].

The Artificial Neural Networks (ANN) are well-known universal approximators or predictors [19, 37]. A Multilayer Perceptron (MLP) network [69] with two layers is shown in Figure 2. Several well-known model structures are used when merging system identification with ANN. If the ARX model is used as the regression vector θ, the model structure is called NNARX (neural network for ARX model). Likewise, NNFIR, NNARMAX and NNOE structures, are also extensively used [54].

4. IKBIS: an intelligent knowledge-based identification method for detecting thermal insulation failures in buildings

IKBIS is a hybrid approach based on a knowledge-based tool applying system identification and Exploratory Pursuit methods. Its purpose is to detect thermal insulation failures in buildings in operation, that is, in a certain moment during the lifetime of the building.

IKBIS is presented in Figure 3. Firstly, the thermal evolution of the building should be obtained by means of simulation. Then the data is processed, the relevant relationships shall be found inducind a feature selection step. With the dimensional reduced data set the models for the temperature evolution in each space are generated. Finally, mesaurements of the selected features are carried out, the models estimate the temperatures in the spaces which are to be compared with the real temperature. Insulation failures are proposed when the bias are higher that expected. A description of each step is in the following subsections.

4.1. *Experiment design*

The required output of this first step is the thermal dynamics of the building, which should be recorded in the building under operating conditions. For this reason, the experiments should consider the specific building regulations, and the real metereological historical data on the climate zone in which the building is based.

A network of sensors would be necessary to measure and to gather data from a building in operation [13]. Nevertheles, it is not always feasible to install a sensor network, but also it is anoreous and no advantages are obtained with respect to IR thermography. Alternatively, simulation software allows to obtain realistic data from modelled buildings [44]. Realistic simulations should consider the different profiles, i.e. the occupancy and the lighting profiles, as close as possible to the currently used in the building. Also, the model should use exactly the same building materials and dimensions, the geometry, etc. Finally, the simulation should accomplish with the country specific building regulations, establishing the suitable climate zone, the same building topology, the use of historical metereological data for the same season and period of the year, etc.

The simulation sofware tool should accept all of this parameters and manage them so the calculations would be realistic. Also, the output data set should include information about heat flux, sun radiation, the state of the heating system and the indoor temperature for all the spaces. In this method we propose the use of HTB2 [44] as it accomplishes with all the requirements.

From the point of view of IKBIS, the process is represented by a data set in which each column contains values of a certain variable [48].

4.2. *Data pre-processing and analysing*

The data set gathered in the previous step includes information on a set of variables for each space in the building; consequently, data set dimensionality should be reduced. IKBIS completely defines the data pre-processing and analysing step as shown in Figure 4.; although it is not completely automated yet. Firstly, Data Transformation is carried out to generate a valid data set –elimination of missing data, etc. Then, dimensionality reduction is performed in the data analysis step. Finally, Data partitioning is considered in the case of a data set with a reduced number of examples. For sake of brevity, the different techniques that can be used are not enumerated in this research.

IKBIS performs the data analysis using Exploratory Pursuit methods, which are described as follows. Firstly, the internal structure analysis is conducted, after which a feature selection stage is performed.

4.2.1. *Internal structure analysis*

EPP models are applied to visualize the internal structure of a data set (Figure 5) helping to identify any clear structure or patterns, which are the sign of a robust data gathering process. Otherwise, the experiment to collect a representative data set should be performed again.

4.2.2. *Feature selection*

Once a satisfactory internal data set structure is identified, the underlying features are selected (Figure 5) by using a EPP model to perform dimensionality reduction.

4.3. *The system identification module*

This IKBIS step establishes the best model structure, its parameters and delays. It was implemented in Matlab and made use of several toolboxes: the System Identification Toolbox [48], the Neural network-based System Identification Toolbox [53], and the Control System Toolbox [50].

Thus, IKBIS performs as follows. Firstly, the user establishes the required criteria. Then, the IKBIS estimates the model structure and the learning method, although the user can also choose both manually.

Using the above mentioned toolboxes, a total of 76 methods are used, which include the following techniques:

- the frequency response analysis based on spectrum analysis and the Fourier Fast Transform (FFT),
- correlation analysis, finite impulse response method (FIR),
- the estate space analysis, with 6 different models and parameters, such as the N4SID algorithm by Van Overschee and de Moor, CVA algorithm by Larimore or the MOESP algorithm by Verhaegen,
- the black-box model analysis, with 31 different models and parameters, such as the least squares method, (QR factorization) ARX, or the recursive normalized gradient algorithm RARMAX,
- the ANN models, with 31 different models and training methods [23, 32, 34, 36],
- the residual analysis based on cross correlation between the residual $\widehat{R}_{\varepsilon}^{N}(\tau)$, between the residual and the input $\widehat{R}_{\varepsilon u}^{N}(\tau)$, and the non-linear residual correlation $\widehat{R}_{\varepsilon^2 u^2}^{N}(\tau)$.

Then, the IKBIS generates a search of the structural parameters of the model using the chosen criteria to analyse all possible model structures. IKBIS proceeds with the validation of the models and the most satisfactory models are visualised and their criteria values are shown. The user can choose from among them the most appropriate system model for the input data set.

4.4. *Detection of thermal insulation failures*

Once the model has been obtained, then it is used as a reference model to compare with the measurements sampled in the building in normal operation conditions. If building performance is similar to the reference model output then its thermal insulation is satisfactory; otherwise, a failure in the thermal insulation will have been detected. A soft computing model should be developed to automatically detect the insulations failure detection. Developing this model requires real experiments to validate the approach, and is now in research.

4.5. *Discussion on knowledge-based identification and KB maintenance*

The IKBIS is a KBS that searches in the model space trying to find the model that is best suited to a given problem. As stated by [61, 49, 51], data mining techniques can improve the performance of KBS, while also helping to manage the knowledge base.

The main problem in these approaches is the generation of comprehensive and exhaustive data sets to learn the classifiers. The classifiers represents the knowledge base, such that if a new problem needs to be solved, the classifiers assist in reducing the search space. This is the step where the advantages of the IKBIS become fully apparent, as it will incoporate typical signal processing measurements - mutual information, information gain, largest Lyapunov exponent, etc.

Moreover, IKBIS may also generate information on each of the models in the search process. This information, can then be used for training classifiers that would be incorporated in the knowledge base of a second release. Heuristics such as Case-Based Reasoning [9, 41, 74, 68], or Genetic Fuzzy rule Systems [35, 63, 62], appear to be valid for this purpose.

Finally, IKBIS has no graphic human machine interface (HMI) but is based on a sequence of line commands. This is due to the fact that IKBIS is at an early stage of development, and it is better to continue developing more reliable knowledge bases following the above mentioned ideas than to design an HMI that would probably have to be re-designed in the short term as the knowledge base evolves. The right time to design an HMI will be when the knowledge base is stabilized.

5. Cases of study and results

In this work, the method is applied in Spain. As seen in Sub-Section 1.1, Spanish building regulations establish several winter/summer zones, from E1 (more severe climate zones) to A3 (gentler climate zone). The typical values that each variable could take for a C winter climate zone of maximum severity in Spain -i.e. the cities of Bilbao, San Sebastian, and La Coruña, among others- are shown in Table 1 and are the following six variables: air temperature of the house, exterior air temperature, heater gain, small power and occupancy gain, lighting gain and ventilation gain.

Two case studies have been analysed in this research to illustrate the IKBIS procedures. Both correspond to Spanish cities in different climatic zones as defined by Spanish regulations. The first case of study is the city of Santander (in the north coast of Spain). The second case of study is the city of Avila, in central Spain, which experiences some of the lowest temperatures on the Meseta. Both cases were analysed in February over a simulation period of ten days. Real metereological data was used and the materials were arranged according to the Spanish regulations. Realistic profiles for occupancy, lighting and the operation of small power devices operation were employed. The HTB2 output dataset includes 14,400 features -the indoor temperature, the instantaneous heating power, the lightning power, and so on.

5.1. *Results*

5.1.1. *Case of study: the city of Santander*

The IKBIS procedure (see Figure 3) was applied to data gathered in the city of Santander. The HTB2 output data set was analysed in order to select the features that best describe the relationships with indoor temperature. PCA, MLHL and CMLHL were applied to analyse the data. The best results, which came from the application of CMLHL, are shown in Figure 5. It can be concluded that CMLHL identified two different clusters ordered by small power and occupancy. Inside each cluster there are further classifications by lighting and heater power output and the dataset may be said to have an interesting internal structure. The initial data set is then represented by these four variables. Finally, the system identification module is applied to find the most appropriate model.

Table 2 shows the results obtained for different model structures. It may be seen that the NNARX is the most appropriate model for monitoring the thermal dynamics of the building. As the heating process exhibits nonlinear behaviour between output and inputs, the linear modelling techniques do not behave properly except in the linear behaviour zones of the process. Consequently, ANN-based models are the most appropriate when using the IKBIS.

Figure 6 and Figure 7 show the time responses of the indoor temperature - $y_1(t)$- and of the estimated indoor temperature - $\hat{y}_1(t \mid m)$- for the NNARX model. The former corresponds to the training data set, while the latter corresponds to the validation data set. The X-axis shows the number of samples used in the estimation and validation of the model and the Y-axis represents the normalized indoor temperature of the house. The training and the validation data sets include 388 and 336 samples, respectively, and have a sampling rate of 1 sample/minute.

It can be concluded from Table 2 that the pruned network NNARX model is able to predict the behaviour of indoor temperature in the building. This model may not only be used to predict indoor temperature but can also determine the normal operating conditions of thermal insulation in buildings.

5.1.2. Case of study: the city of Avila

The same features chosen for the first case of study and the same model structure are used for this second case of study, so the validity of the method can be tested. This means that the corresponding HTB2 simulation was carried out and the same features used in the first case of study were chosen. Figure 8 and Figure 9 show the time responses of the indoor temperature -$y_1(t)$- and of the estimated indoor temperature -$\hat{y}_1(t \mid m)$- for the NNARX model. The former corresponds to the training data set, while the latter corresponds to the validation data set. The X-axis shows the number of samples used in the estimation and validation of the model and the Y-axis represents the normalized indoor temperature of the house. The training and the validation data sets include 2,000 and 1,126 samples, respectively, and have a sampling rate of 1 sample/minute.

The performance of the pruned network NNARX model is shown in Table 3. As in the previous case of study, this model is able to simulate and predict the behaviour of the indoor temperature of the house and it can be used to predict the indoor temperature and to determine the normal operating conditions of thermal insulation in buildings.

6. Conclusions and Future Work

Effective thermal insulation is an essential component of energy efficient heating systems in buildings. The more effective the insulation in the buildings, the lower the energy losses due to insulation failures. Thus, the possibility of improving the detection of thermal insulation failures represents a challenge in building energy management.

IKBIS represents a novel method for detecting thermal insulation failures, although its validation in a preliminar phase. This procedure makes use of several different techniques, such as Exploratory Pursuit methods, and neural networks modelling, among others.

Future work will cover finishing the validation of the method, but also modeling the heat flux crossing through the walls of the building – a regulatory test in the evaluation of the energy efficiency in buildings in Spain-, and the development of an automatic thermal insulation failure detection system, which will improve overall performance. Finally, this method would also be applied in other countries.

Acknowledgements

We would like to extend our thanks to Phd. Magnus Nørgaard for his freeware version of Matlab Neural Network Based System

Identification Toolbox. This research has been partially supported through projects of the Junta of Castilla and León (JCyL): [BU006A08] and Business intelligent for production, Technological Institute of Castilla y León (ITCL) and investment and services agency (ADE), project of the Spanish Ministry of Education and Innovation [CIT-020000-2008-2] and project of Spanish Ministry of Science and Technology [TIN2008-06681-C06-04]. The authors would also like to thank the vehicle interior manufacturer, Grupo Antolin Ingenieria, S.A., within the framework of project MAGNO2008 - 1028.- CENIT also funded by the same Government Ministry.

References

[1] CTE-HE ahorro de energía, aplicación a edificios de uso residencial vivienda-dav. Consejo Superior de Colegios de Arquitectos de España, España, 2006.

[2] Directive 2002/91/CE of the European Parliament and the Council of 16 December 2002 on the energy performance of buildings, Oficial Journal of the European Community, 2003.

[3] Orden VIV/984/2009, de 15 de Abril, por el que se modifican determinados documentos básicos del código técnico de edificación aprobados por Real Decreto 314/2006 y 1371/2007, BOE núm 99, España, 2009.

[4] Real Decreto 1027/2007, de 20 de Julio, por el que se aprueba el reglamento de instalaciones térmicas en edificios, BOE núm. 207, España, 2007.

[5] Real Decreto 1371/2007, de 19 de Octubre, por el que se aprueba el documento básico «DB-HR protección frente al ruido» del código técnico de la edificación y se modifica el Real Decreto

314/2006, de 17 de Marzo, por el que se aprueba el código técnico de la edificación. BOE núm 254, España, 2007.

[6] Real Decreto 314/2006, de 17 de Marzo, por el que se aprueba el código técnico de la edificación. BOE núm 74, España, 2006.

[7] Real Decreto 47/2007, de 19 de Enero, por el que se aprueba el procedimiento básico para la certificación de eficiencia energética de edificios de nueva construcción, BOE núm. 27, España, 2007.

[8] www.calener.es.

[9] A. Aamodt and E. Plaza. Case-based reasoning: Foundational issues, methodological variations, and system approaches. *Artificial Intelligence Communications-AICom*, 7(1):39–59, 1994.

[10] H. Adeli and S.L. Hung. *Machine Learning - Neural Networks, Genetic Algorithms, and Fuzzy Systems*. Wiley & Sons, New York, USA, 1995.

[11] H. Adeli and X. Jiang. Dynamic fuzzy wavelet neural network model for structural system identification. *Journal of Structural Engineering, ASCE*, 132(1):102–111, 2006.

[12] C. A. Balaras and A. A. Argiriou. Infrared thermography for building diagnostics. *Energy and Buildings*, 34(2):171–183.

[13] R. Banchs, H. Klie, A. Rodriguez, S.G. Thomas M.F., and Wheeler. A neural stochastic multiscale optimization framework for sensor-based parameter estimation. *Integrated Computer-Aided Engineering*, 14(3):213–223, 2007.

[14] G.E.A.P.A. Batista and M.C. Monard. An analysis of four missing data treatment methods for supervised learning. *Applied Artificial Intelligence*, 17(5):519–533, 2003.

[15] S. Chen and S.A. Billings. Representations of non-linear systems: The narmax model. *Int. J. Control*, 49(3):1013–1032, 1989.

[16] E. Corchado, D. MacDonald, and C. Fyfe. Maximum and minimum likelihood hebbian learning for exploratory projection pursuit. *Data Mining and Knowledge Discovery*, 8(3):203–225, 2004.

[17] E. Corchado and C. Fyfe. Connectionist techniques for the identification and suppression of interfering underlying factors. *International Journal of Pattern Recognition and Artificial Intelligence*, 17(8):1447–1466, 2003.

[18] E. Corchado and C. Fyfe. Orientation selection using Maximum Likelihood Hebbian learning. *International Journal of Knowledge-based Intelligent Engineering*, 7(2):11-30, 2003.

[19] G. Cybenko. Aproximation by superpositions of sigmoidal function. *Math. Control, Sygnals and System*, 2(4):303–314, 1989.

[20] E. de la Cal, J.R. Villar, and J. Sedano. A thermodynamical model study for an energy saving algorithm. In *Lecture Notes in Artificial Intelligence:Hybrid Artificial Intelligence Systems*, Vol. 5572, pp. 384–390. Springer, 2009.

[21] J. Deogun, W. Spaulding, B. Shuart, and D. Li. Towards missing data imputation: A study of fuzzy k-means clustering method. In *Lecture Notes in Computer Science: Proceedings of the 4th International Conference of Rough Sets and Current Trends in Computing (RSCTC'04)*, Vol. 3066, pp. 573–579, Uppsala (Sweden), 2004. Springer.

[22] T. Efendigil, S. Önüt, and C. Kahraman. A decision support system for demand forecasting with artificial neural networks and neuro-fuzzy models: a comparative analysis. *Expert Systems with Applications*, 36(3, Part 2):6697–6707, 2009.

[23] R. Fletcher. *Practical Methods of Optimization*. Wiley & Sons, Chichester, UK, 2nd edition, 1987.

[24] J.H. Freedman and J.W. Tukey. Projection pursuit algorithm for exploratory data-analysis. *IEEE Transactions on Computers*, 23(9):881–890, 1974.

[25] Y.Y. Fu, C.J. Wu, J.T. Jeng, and C.N. Ko. Identification of mimo systems using radial basis functions networks with hybrid learning algorithms. *Applied Mathematics and Computation*, 213:184–196, 2009.

[26] C. Fyfe and E. Corchado. Maximum likelihood hebbian rules. In *Proceedings of the 10th European Symposium on Artificial Neural Networks(ESANN 2002)*, pp. 143–148, Bruges, Belgium, 2002. D-side Publishers.

[27] S. Ghosh-Dastidar, , H. Adeli, and N. Dadmehr. Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. *IEEE Transactions on Biomedical Engineering*, 55(2):512–518, 2008.

[28] P.A. Gourraud, E. Ginin, and A. Cambon-Thomsen. Handling missing values in population data: Consequences for maximum likelihood estimation of haplotype frequencies. *European Journal of Human Genetics*, 12(10):805–812, 2004.

[29] J.W. Grzymala-Busse, L.K. Goodwin, W.J. Grzymala-Busse, and X. Zheng. Handling missing attribute values in preterm birth data sets. In *Lecture Notes in Computer Science: Proceedings of the 10th International Conference of Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC'05)*, Vol. 3642, pp. 342–351, Regina (Canada), 2005. Springer.

[30] I. Guyon and A. Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research, Special Issue on variable and Feature Selection*, 3:1157–1182, 2003.
[31] J. Han, L. Lua, and H. Yang. Investigation on the thermal performance of different lightweight roofing structures and its effect on space cooling load. *Applied Thermal Engineering*, 29(11–12):2491–2499, 2009.
[32] L.K. Hansen and M.W. Pedersen. Controlled growth of cascade correlation nets. In *International Conference on Artificial Neural Networks (ICANN1994)*, pp. 797–800, Italy, 1994. Eds. M. Marinaro and P.G. Morasso.
[33] D.A. Haralambopoulos and G.F. Paparsenos. Assessing the thermal insulation of old buildings. the need for in situ spot measurements of thermal resistance and planar infrared thermography. *Energy Conversion and Management*, 39(1-2):65 – 79, 1998.
[34] B. Hassibi and D.G. Stork. Second order derivatives for network pruning: Optimal brain surgeon. In *Advances in neural Information Processing System 5*, pp. 164–171, San Mateo, CA, USA, 1993. Eds. S.J. Hanson et al.
[35] F. Herrera. Genetic fuzzy systems: Taxonomy, current research trends and prospects. *Evolutionary Intelligence*, 1:27–46, 2008.
[36] J. Hertz, A. Krogh, and R.G. Palmer. *Introduction to the Theory of Neural Computation*, Vol. 1. Addison-Wesley, Santa Fe Institute Studies In The Sciences Of Complexity, 1991.
[37] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal aproximators. *Neural Networks*, 2(5):359–366, 1989.
[38] H. Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of Education Psychology*, 24:417–444, 1933.
[39] X. Jiang and H. Adeli. Dynamic wavelet neural network for nonlinear identification of highrise buildings. *Computer-Aided Civil and Infrastructure Engineering*, 20(5):316–330, 2005.
[40] X. Jiang, S. Mahadevan, and H. Adeli. Bayesian wavelet packet denoising for structural system identification. *Structural Control and Health Monitoring*, 14(2):333–356, 2007.
[41] A. Karim and H. Adeli. CBR model for freeway work zone traffic management. *Journal of Transportation Engineering, ASCE*, 129(2):134–145, 2003.
[42] T. Koh and E.J. Powers. Second-order volterra filtering and its application to nonlinear system identification. In *IEEE Transactions on*

acoustics, Speech and Signal Processing ASSP-33(6), pp. 1445–1455, 1985.
[43] H. Lee, E. Kim, and M. Park. A genetic feature weighting scheme for pattern recognition. *Integrated Computer-Aided Engineering*, 14(2):161–171, 2007.
[44] P.T. Lewis and P.K. Alexander. A flexible model for dynamic building simulation: HTB2. *Building and Environment*, 1:7–16, 1990.
[45] D.T. Lin and D.C. Pan. Integrating a mixed-feature model and multiclass support vector machine for facial expression recognition. *Integrated Computer-Aided Engineering*, 16(1):61–74, 2009.
[46] H. Liu and L. Yu. Toward integrating feature selection algorithms for classification and clustering. *IEEE Knowledge and Data Engineering, IEEE Transactions*, 17(4):491–502, 2005.
[47] L. Ljung. *System Identification. Theory for the User*. Prentice-Hall, Upper Saddle River, N.J., USA, 2nd edition.
[48] L. Ljung. System Identification Toolbox User´s Guide. The MathWorks, Inc., 2003.
[49] B. Majhi and G. Panda. Development of efficient identification scheme for nonlinear dynamic systems using swarm intelligence techniques. *Expert Systems with Applications: An International Journal*, 37(1), 2010.
[50] *MathWorks. Control System Toolbox User´s Guide. The MathWorks, Inc., 1998.*
[51] N.N. Nandola and S. Bhartiya. Hybrid system identification using a structural approach and its model based control: An experimental validation. *Nonlinear Analysis: Hybrid Systems*, 3:87–100, 2009.
[52] O. Nelles. *Nonlinear System Identification. From Classical Approaches to Neural Networks and Fuzzy Models*. Springer, Berlin, Germany, 2001.
[53] M. Nørgaard. Neural Network based system identification toolbox. Technical Report 00-E-891, Department of Automation, Technical University of Denmark, 2000.
[54] M. Nørgaard, O. Ravn, N.K. Poulsen, and L.K. Hansen. *Neural Networks for Modelling and Control of Dynamic Systems*. Springer-Verlag, London U.K., 2000.
[55] P.V. Overschee and B. Moor. N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30(1):75–93, 1994.
[56] K. Pearson. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2(6):559–572, 1901.

[57] K. Perusich. Using fuzzy cognitive maps to identify multiple causes in troubleshooting systems. *Integrated Computer-Aided Engineering*, 15(2):197–206, 2008.

[58] A. Poznyak and L. Lennart. On-line identification and adaptive trajectory tracking for nonlinear stochastic continuous time system using differential neural networks. *Automatica*, 37:1257–1268, 2001.

[59] S. Ribaric, D. Maroetic, and D. S. Vedrina. A knowledge-based system for the non-destructive diagnostics of façade isolation using the information fusion of visual and ir images. *Expert Systems with Applications*, 36(2):3812–3823, 2009.

[60] F. Rodriguez, J.L. Guzman, M. Berenguel, and M. Ruiz. Adaptive hierarchical control of greenhouse crop production. *International Journal of Adaptive Control and Signal Processing*, 22(2):180–197, 2007.

[61] S. Saitta, B. Raphael, and I.F.C. Smith. Data mining techniques for improving the reliability of system identification. *Advanced Engineering Informatics*, 19:289–298, 2005.

[62] K. Sarma and H. Adeli. Fuzzy discrete multicriteria cost optimization of steel structures. *Journal of Structural Engineering, ASCE*, 126(11):1339–1347, 2000.

[63] K. Sarma and H. Adeli. Fuzzy genetic algorithm for optimization of steel structures. *Journal of Structural Engineering, ASCE*, 126(5):596–604, 2000.

[64] J. Schoukens, J. Nemeth, P. Crama, Y. Rolain, and R. Pintelon. Fast approximate identification of nonlinear systems. *Automatica*, 39(7):1267–1274, 2003.

[65] J. Schoukens, Y. Rolain, and R. Pintelon. Improved approximate identification of nonlinear systems. In *21st IEEE Instrumentation and Measurement Technology Conference*, pp. 2183–2186, Como, Italy, 2004.

[66] H.S. Seung, N.D. Socci, and D. Lee. The rectified gaussian distribution. *Advances in Neural Information Processing Systems*, 10:350–356, 1998.

[67] L.A. Shalabi, Z. Shaaban, and B. Kasasbeh. Data mining: A preprocessing engine. *Journal of Computer Science*, 2(9):735–739, 2006.

[68] G.F. Sirca and H. Adeli. Case-based reasoning for converting working stress design-based bridge ratings to load factor design-based ratings. *Journal of Bridge Engineering*, 10(4):450–459, 2005.

[69] M.J. Smith, M.F. Goodchild, and P.A. Longley. *Geospatial Analysis. A Comprehensive Guide to Principles, Techniques and*

Software Tools -Third Edition-. SPLINT, UCL, 2009.
[70] T. Söderström and P.Stoica. *System identification*. Prentice Hall International, Englewood Cliffs, NJ, 1989.
[71] M. Verhegen. Identificacion of the deterministic part of mimo state space models. *Automatica*, 30(2):61–74, 1994.
[72] J.R. Villar, E. de la Cal, and J. Sedano. Minimizing energy consumption in heating systems under uncertainty. In *Lecture Notes in Artificial Intelligence: Hybrid Artificial Intelligence Systems*, Vol. 5271, pp. 583–590. Springer, 2008.
[73] J.R. Villar, E. de la Cal, and J. Sedano. A fuzzy logic based efficient energy saving approach for domestic heating systems. *Integrated Computer-Aided Engineering*, 16(2):151–163, 2009.
[74] A. Waheed and H. Adeli. Case-based reasoning in steel bridge engineering. *Knowledge–Based Systems*, 18(1):37–46, 2005.
[75] B. Wahlberg. System identification using Laguerre models. In *IEEE Transactions on Automatic Control*, Vol. 36(5), pp. 551–562, 1991.
[76] Q. Wu and J. Ben-Arie. View invariant head recognition by hybrid PCA based reconstruction. *Integrated Computer-Aided Engineering*, 15(2):97–108, 2008.
[77] J. Xu, A. Roy, and M.H. Chowdhury. Noise separation in analog integrated circuits using independent component analysis technique. *Integrated Computer-Aided Engineering*, 15(2):163–180, 2008.
[78] H. Yin and I. Hussain. Independent component analysis and non-gaussianity for blind image deconvolution deblurring. *Integrated Computer-Aided Engineering*, 15(3):219–288, 2009.
[79] P.C. Young, C.J. Taylor, W. Tych, D.J. Pedregal, and P.G. McKenna. *The Captain Toolbox*. Centre for Research on Environmental Systems and Statistics, Lancaster University, 2004.
[80] J. Yu, C. Yang, L. Tian, and D. Liao. Evaluation on energy and thermal performance for residential envelopes in hot summer and cold winter zone of china. *Applied Energy*, 86(10):1970–1985, 2009.
[81] W. Yu and X. Li. Online fuzzy modeling with structure and parameter learning. *Expert Systems with Applications*, 36(4):7484–7492, 2009.
[82] L.A. Zadeh. Soft computing and fuzzy logic. *IEEE Software*, 11(6):48–56, 1994.

Table 1 Typical values of each variable in a C winter climate zone city in Spain

Variable (Units)	**Range of values**	**Transmittance level (W/m^{2}K)**
Air temperature of the house (ºC), $y_1(t)$.	17 to 24	-External cavity wall: 0.68
Exterior air temperature in February (ºC), $u_1(t)$.	8 to 10	-Double glazing: 2.91
Heater gain (W), $u_2(t)$.	0 to 4,250	-Floor/ceiling: 1.96
Small power and occupancy gain (W), $u_3(t)$.	0 to 1,200	-Party wall between buildings: 0.96
Lighting gain (W), $u_4(t)$.	0 to 500	-Others party wall: 1.050
Ventilation gain (m^3/min), $u_5(t)$.	0.5 to 7.5	-Internal partition: 2.57

Table 2 Case of study: the city of Santander. The values of the quality indexes for the obtained models.

Model	Indexes
Black-box ARX model with $n_a=3$, $n_{b1}=1$, $n_{b2}=3$, $n_{b3}=2$, $n_{b4}=2$, $n_{b5}=1$, $n_{k1}=2$, $n_{k2}=10$, $n_{k3}=10$, $n_{k4}=10$, $n_{k5}=1$, [*3 1 3 2 2 1 2 10 10 10 1*]. The model is estimated using the least squares method, QR factorization, the degree of the model selection is carried out with the best Akaike information criterion (AIC) -the structure that minimizes AIC-.	FIT1:73.93% FIT10: 26.6% V: 0.15 FPE:0.165 NSSE:0.11
Black-box ARMAX model with $n_a=3$, $n_{b1}=1$, $n_{b2}=3$, $n_{b3}=2$, $n_{b4}=2$, $n_{b5}=1$, $n_C=3$, $n_{k1}=2$, $n_{k2}=10$, $n_{k3}=10$, $n_{k4}=10$, $n_{k5}=1$, [*3 1 3 2 2 1 3 2 10 10 10 1*]. The model is estimated using the prediction error method, the choice of the model order is Realized from the best AIC criterion of the ARX model.	FIT1:73.90% FIT10: 19.5% V: 0.163 FPE:0.177 NSSE:0.112

Table 2 (cont.) Case of study: the city of Santander. The values of the quality indexes for the obtained models.

Model	Indexes
ANN model for the heating process, NNARX regressor, the order of the polynomials of the initial fully connected structure are $n_a=3$, $n_{b1}=1$, $n_{b2}=3$, $n_{b3}=2$, $n_{b4}=2$, $n_{b5}=1$, $n_{k1}=2$, $n_{k2}=10$, $n_{k3}=10$, $n_{k4}=10$, $n_{k5}=1$, [*3 1 3 2 2 1 2 10 10 10 1*]. The model was obtained using the regularized criterion. This model was optimised by CMLHL analysis, residual analysis and the pruned network, using optimal brain surgeon (OBS). The model structure has 10 hidden hyperbolic tangent units and 1 linear output unit. The network is estimated using the Lenvenberg-Marquardt method, and the model order is decided on the basis of the best AIC criterion of the ARX model.	FIT1:92.23% V: 0.022 FPE:0.14 NSSE:0.01
ANN model for the heating process, NNARMAX regressor, the order of polynomials of the initial fully connected structure are $n_a=3$, $n_{b1}=1$, $n_{b2}=3$, $n_{b3}=2$, $n_{b4}=2$, $n_{b5}=1$, $n_c=3$, $n_{k1}=2$, $n_{k2}=10$, $n_{k3}=10$, $n_{k4}=10$, $n_{k5}=1$, [*3 1 3 2 2 1 3 2 10 10 10 1*]. The model was obtained using the regularized criterion and was optimised by CMLHL analysis, residual analysis and the pruned network, using OBS. The model structure has 10 hidden hyperbolic tangent units and 1 linear output unit. The network is estimated using the Lenvenberg-Marquardt method, and the model order is decided on the basis of the best AIC criterion of the ARX model.	FIT1:84.2% V: 0.041 FPE:0.142 NSSE:0.043

Table 3 Case of study: the city of Ávila. The values of the quality indexes for the NNARX model.

Model	Indexes
ANN model for the heating process, NNARX regressor, the order of the polynomials of the initial fully connected structure are $n_a=3$, $n_{b1}=2$, $n_{b2}=3$, $n_{b3}=3$, $n_{b4}=3$, $n_{b5}=2$, $n_{k1}=5$, $n_{k2}=5$, $n_{k3}=5$, $n_{k4}=5$, $n_{k5}=2$, [*3 2 3 3 3 2 5 5 5 5 2*]. The model was obtained using the regularized criterion. This model was optimised by CMLHL analysis, residual analysis and the pruned network, using OBS. The model structure has 10 hidden hyperbolic tangent units and 1 linear output unit. The network is estimated using the Lenvenberg-Marquardt method, and the model order is decided on the basis of the best AIC criterion of the ARX model.	FIT1:93.42% V: 0.0056 FPE:0.102, NSSE:0.0036

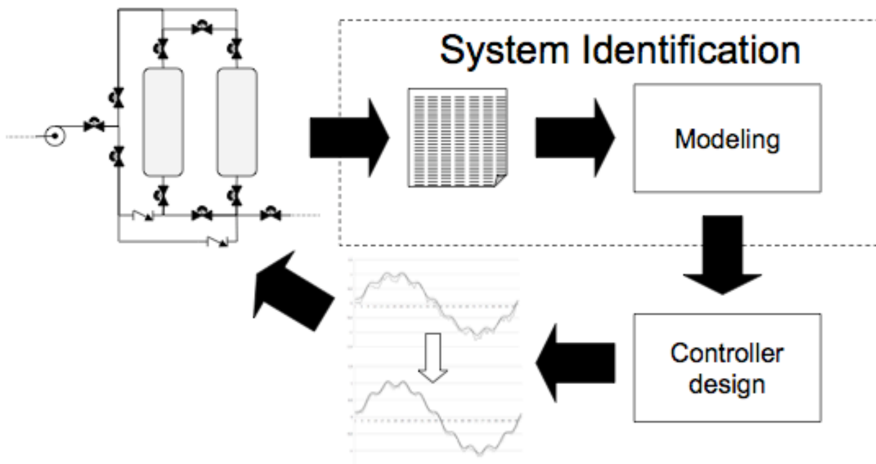

Figure 1. The data source cycle: the data is gathered through the sensors from a process in operation –the set of valves and the tanks in the figure. This data is then processed and a better controller is found. The controller is then used. Whenever the behaviour of the system may be improved the cycle is repeated.

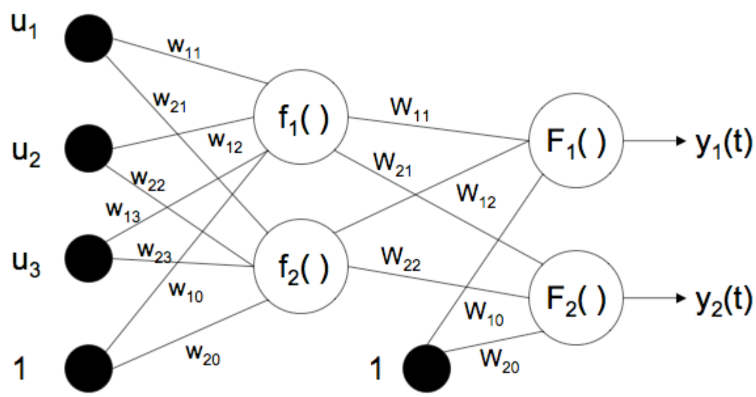

Figure 2. A MLP network with two layers, with two nodes per layer, and three inputs. W_{ij} is the weight matrix between the hidden and output layer, while w_{ij} is the weight matrix between the inputs and the hidden layer. The network has two bias nodes with value 1.

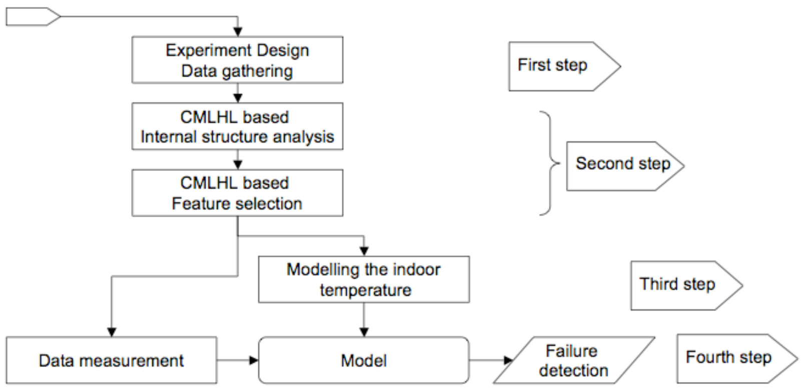

Figure 3. The schema of the IKBIS to detect thermal insulation failures. The relevance of CMLHL is apparent in two ways as it allows the relevant features to be selected i.e., to reduce the variables used to train and to analyse the internal structure of the data set.

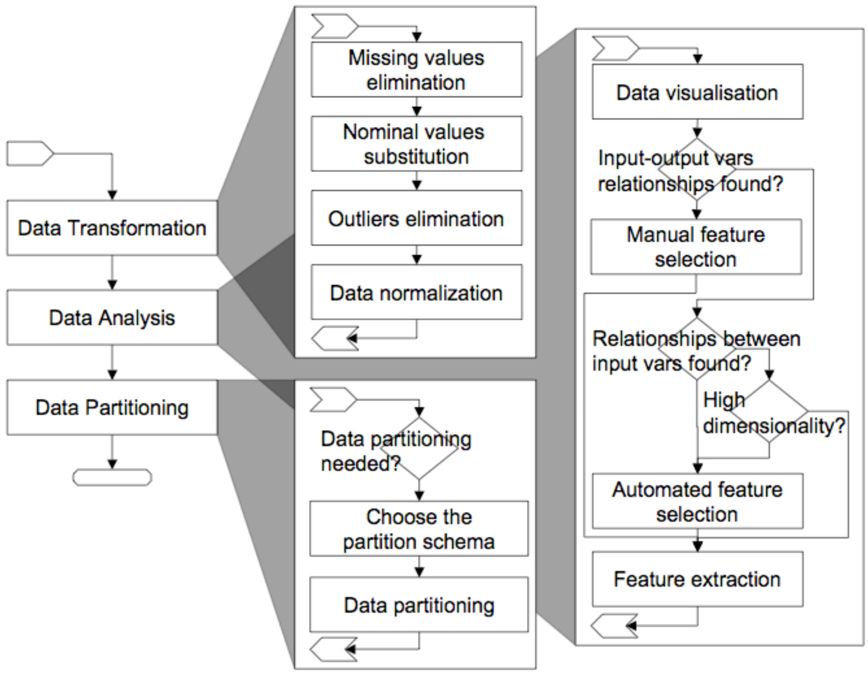

Figure 4 The data pre-processing and analysis step flowchart. On the left, the main algorithm; each box in the main algorithm is unfolded on the right.

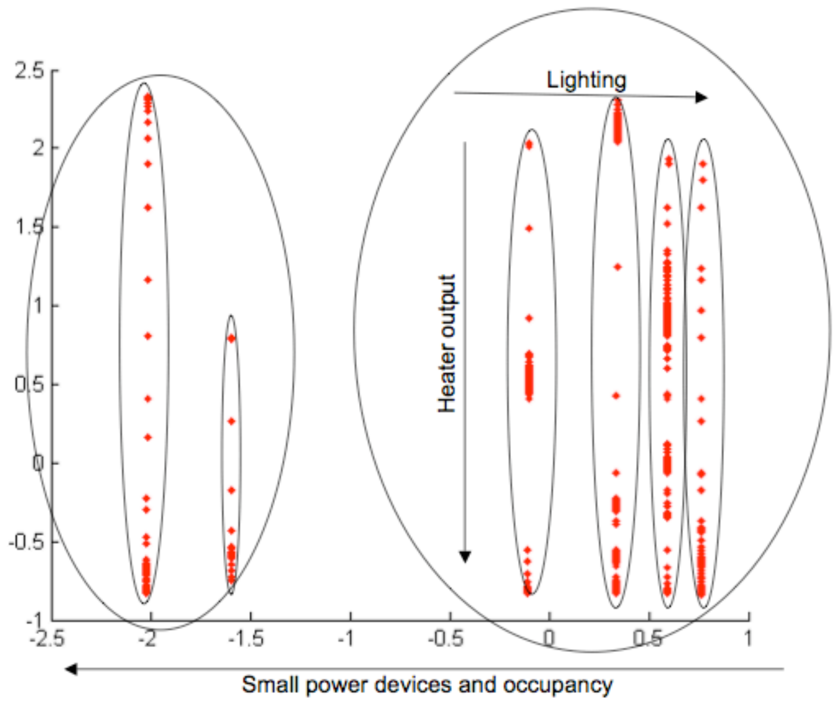

Figure 5. The CMLHL projection. As can be seen, there are four relevant features that allow the indoor temperature to be modelled: the occupancy, the small power devices total power, the lighting electrical power and the heating system power.

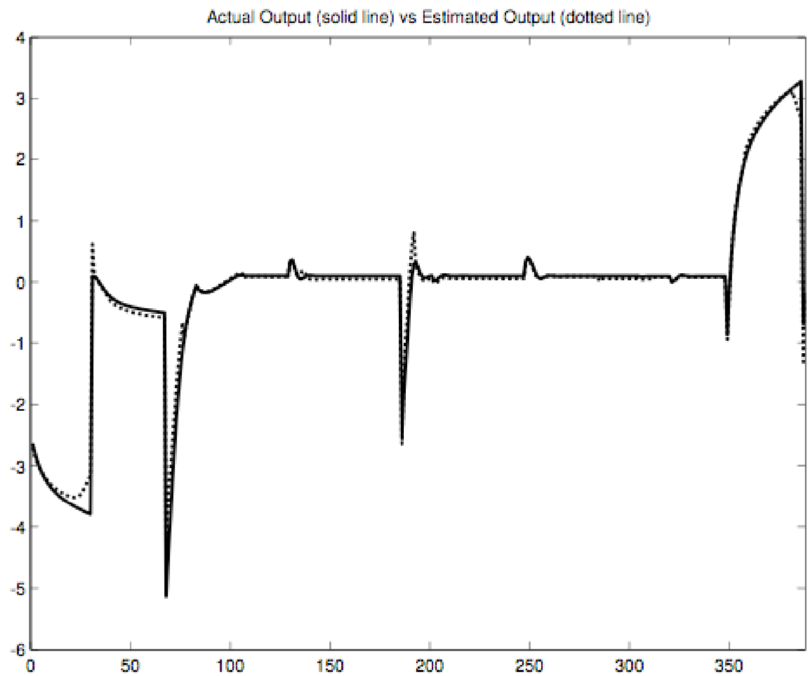

Figure 6. Case of study: the city of Santander. Output response of NNARX model for the training data set. The actual output (solid line) is graphically presented with one-step-ahead prediction (dotted line). The X-axis represents the time steps, the Y-axis represents the signal value.

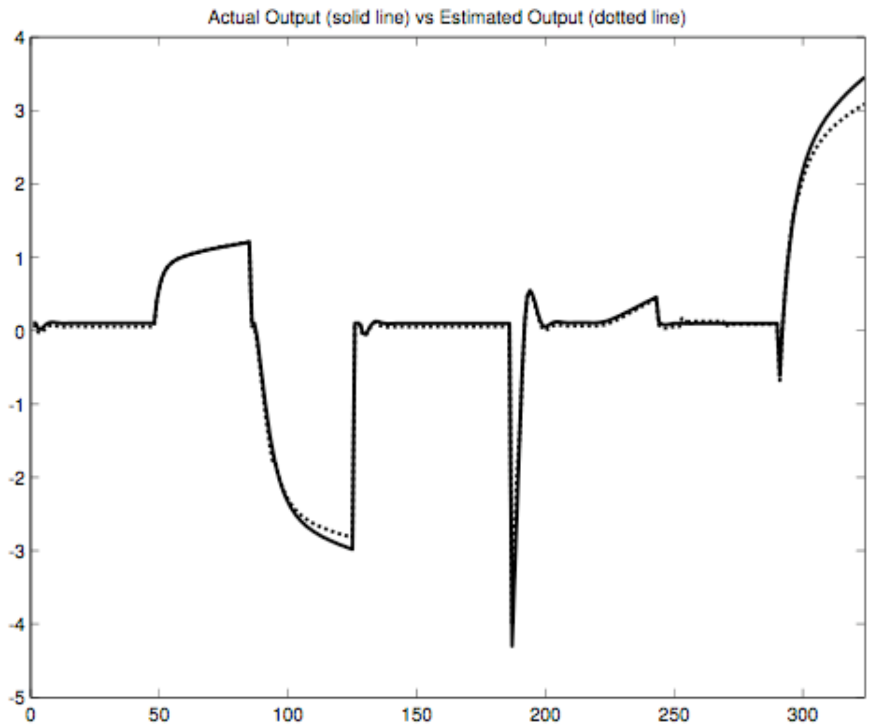

Figure 7. Case of study: the city of Santander. Output response of NNARX model for the validation data set. The actual output (solid line) is graphically presented with one-step-ahead prediction (dotted line). The X-axis represents the time steps, the Y-axis represents the signal value.

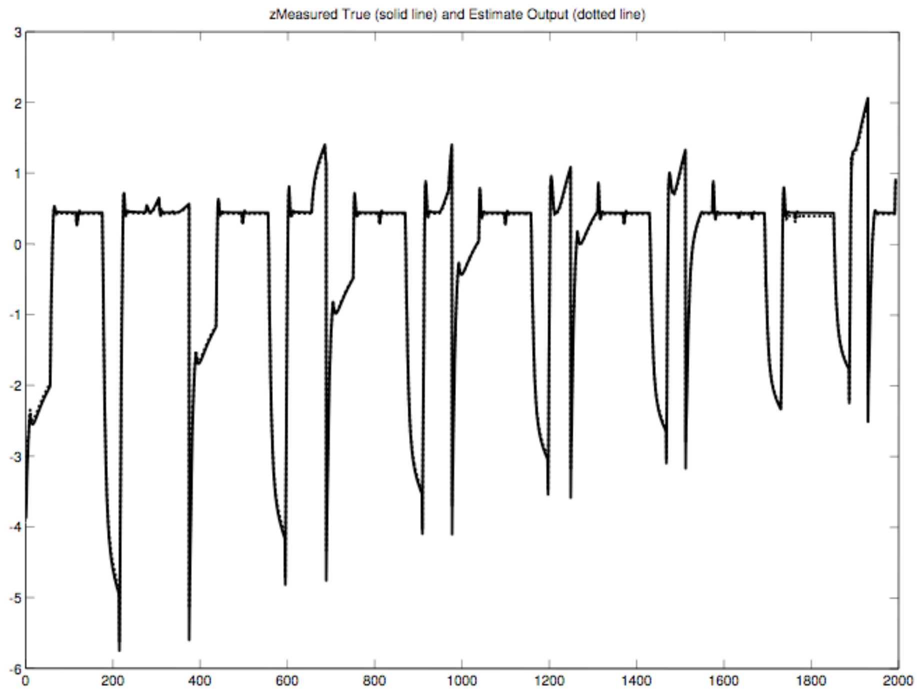

Figure 8. Case of study: the city of Ávila. Output response of NNARX model for the training data set. The actual output (solid line) is graphically presented with one-step-ahead prediction (dotted line). The X-axis represents the time steps, the Y-axis represents the signal value.

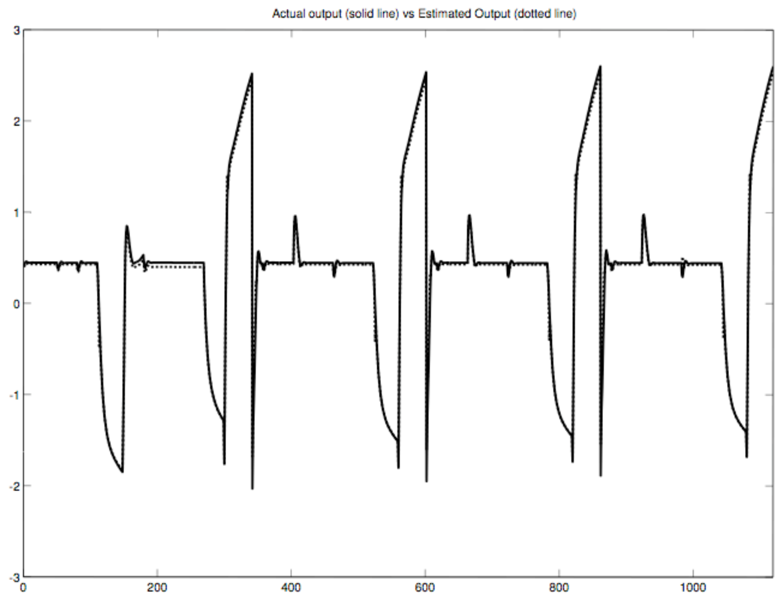

Figure 9. Case of study: the city of Ávila. Output response of NNARX model for the validation data set. The actual output (solid line) is graphically presented with one-step-ahead prediction (dotted line). The X-axis represents the time steps, the Y-axis represents the signal value.