

Clustering of a Greenhouse Fuzzy Model

Paulo Salgado^a, Paulo A. Afonso^b and J. Boaventura Cunha^a

^a CETAV- Universidade de Trás-os-Montes e Alto Douro, 5000-911 Vila Real, Portugal, { psal@utad.pt; jboavent@utad.pt }

^b ESTGA – Escola Superior de Tecnologia e Gestão de Águeda, 3750-127 Águeda, Portugal, { paulo@estga.ua.pt }

Abstract

This paper describes the identification of greenhouse climate processes with multiple fuzzy models, derived from an organization of information of one global (flat) fuzzy model. This concept is called Separation of Linguistic Information Methodology - SLIM, and it is based on a new concept of relevance, which has been proposed to measure the relative importance of sets of rules. It allows the automatic organization of the sets of fuzzy IF ... THEN rules of one fuzzy system into a multimodel Hierarchical Structure. This task is accomplished by a new Fuzzy Clustering of Fuzzy Rules Algorithm (FCFRA), which is a generalization of the well-known Fuzzy *c*-means, applied to a cluster of fuzzy rules instead of data points. This organizational structure makes the fuzzy greenhouse climate model interpretable, as in the case of the physical model. This new methodology was tested to split the inside greenhouse air temperature and humidity flat fuzzy models into fuzzy sub-models, which have alike counterparts on the physical sub-models.

Key words: Hierarchical fuzzy model; Fuzzy clustering; Greenhouse climate modelling.

1 Introduction

Many processes related to the climate of the greenhouse are difficult to describe mathematically, especially when the structure of the system and the relationship between variables is unknown or too complex, giving rise to intractable mathematical expressions. Other strategies called intelligent could also be valuable alternatives, as the Fuzzy modelling. Combining the well-established learning techniques with the proposed methodology enables the ability to learn from real world observation and describes naturally the behaviour as a series of understandable linguistically human rules. Despite this effort, it is not possible to guarantee a direct relation between the linguistic information and the physical processes.

A fuzzy system establishes a fuzzy map, which assigns for each input fuzzy set in U an output fuzzy set in V , through relations expressed by a set of fuzzy rules on the product space $S = U \times V$. The organization of the information of the fuzzy model could be obtained by partition of this product space, which may be achieved through the decomposition of these relationships. In the context of the modelling of a fuzzy system, the organization of the information is considered as a methodology capable of making the fuzzy system more interpretable and legible, preventing losses of information. These processes are repetitive and reversible.

This work addresses this fundamental goal of fuzzy modelling by using an algorithm that implements Fuzzy Clustering of Fuzzy Rules. The proposed algorithm allows the decomposition of the fuzzy relation into sub-relations, through a process of unfolding the fuzzy rules. The obtained sub-rules are then grouped into c subgroups (clusters), by similarity association. Both tasks are implemented simultaneously by the proposed algorithm, which is a generalization of the Probabilistic Clustering Algorithm, traditionally known as the fuzzy *c*-means algorithm. The clusters obtained with this Fuzzy Clustering of Fuzzy Rules Algorithm, FCFRA (Salgado and Cunha, 2004), are used to compose the new fuzzy sub-models. As a result, the original fuzzy relation is layered through the c levels of the hierarchical fuzzy system. In other

words, the application of the FCFRA to the clustering of a flat fuzzy system leads to the distribution of the information carried by the system among various layers of a hierarchical collaborative structure (HCS).

The practical goal of this work is to model the greenhouse air temperature and humidity under process control. As expected, the fuzzy identification procedure for this kind of process models leads to an unreadable model, with a very high number of rules. To solve this problem, the flat fuzzy system information is organized into subsystems by using the FCFRA algorithm. The resulting HCS structures improve the model readability without loss of valuable information. In this final process, each fuzzy sub-model reflects the behaviour of one real physical sub-model (air leakage, ventilation, etc.) (Salgado and Cunha, 2004).

The paper is organized as follows. Section 2 gives the basic description of the Mandani Fuzzy System. A new algorithm that implements Fuzzy Clustering of Fuzzy Rules is given in section 3, where the automatic design method of the hierarchical Fuzzy System is also presented. In section 4, a flat fuzzy system is obtained to model the greenhouse climate system. The proposed Fuzzy Clustering of Fuzzy Rules algorithm is used to obtain a hierarchical collaborative structure, HCS, with physical meaning. Some simulation results and a short discussion on the fuzzy system identification and its clustering process are presented in section 5. Finally, section 6 contains the main conclusions and future work.

2 The fuzzy model

In this section, it is assumed that fuzzy systems are multi-input single-output systems $y:U \mapsto V$, where $U = U_1 \times \dots \times U_n$ is the input space and V is the output space. A multi-output system can always be separated into a group of single-output systems. In this paper the fuzzy system is composed by four main components: fuzzifier, fuzzy rule base, fuzzy inference engine, and defuzzifier.

Consider a system $y = f(x)$, where y is the output variable and $x = [x_1, x_2, \dots, x_n]^T \in R^n$ is the input vector. Let $U = [\alpha_1, \beta_1] \times \dots \times [\alpha_n, \beta_n]$ be the domain of the input vector $x = [x_1, x_2, \dots, x_n]^T$. In fuzzy systems modelling, this relationship is represented by a collection $\mathfrak{F}$ of fuzzy IF-THEN rules:

$$R_l : \text{IF } x_1 \text{ is } A_{l,1} \text{ and } x_2 \text{ is } A_{l,2} \text{ and } \dots \text{ and } x_n \text{ is } A_{l,n} \text{ THEN } y \text{ is } C_l \quad (1)$$

where $A_{l,j}$ in U_j and C_l in V are linguistic terms characterized by fuzzy membership functions $A_{l,j}(x_j)$ and $C_l(y)$, respectively.

The fuzzifier is the most commonly used singleton fuzzifier. By using the product inference engine and the center-average defuzzifier, the overall output of the fuzzy system can be obtained by (Zeng *et al.*, 1994):

$$f(x) = \frac{\sum_{l=1}^M A_l(x) \theta_l}{\sum_{l=1}^M A_l(x)} \quad (2)$$

where θ_l is the point in V at which $C_l(y)$ achieves its maximum value and l is the index of rule ($l=1, \dots, M$). In this paper $C_l(y)$ is a normal and centroid fuzzy set, $C_l(y = \theta_l) = 1$. Then, the antecedent of the fuzzy rule R_l in eq. (1) can be viewed as one aggregated fuzzy set in the product space U with membership function $A_l(x) = T_j^n A_{l,j}(x_j) \in U$, where T is a t-norm operation (usually the min or product T-norm operator).

For each input j , ($j=1,2,\dots,n$), N_j fuzzy sets are defined in $U_j = [\alpha_j, \beta_j]$, using the following triangular membership functions:

$$A_1^j(x_j) = \mu(x_j; \bar{x}_j^1, \bar{x}_j^1, \bar{x}_j^2); A_{r-1}^j(x_j) = \mu(x_j; \bar{x}_j^{N_j-1}, \bar{x}_j^{N_j}, \bar{x}_j^{N_j}); A_r^j(x_j) = \mu(x_j; \bar{x}_j^{r-1}, \bar{x}_j^r, \bar{x}_j^{r+1}) \quad (3)$$

for $r=2,3,\dots,N_{j-1}$, $\alpha_j = \bar{x}_j^1 < \bar{x}_j^2 < \dots < \bar{x}_j^{N_j} = \beta_j$ and $\mu(x; a, b, c)$ is the triangular membership, where a , b and c are the values of the ordinate left, centre and right borders of the base of triangle, respectively.

At the end of this task, the domain space is partitioned by a grid of triangular memberships, where the extreme of the memberships, $\bar{x}_j^r$, are met with the left and the right center of the adjacent membership function, respectively $\bar{x}_j^{r-1}$ and $\bar{x}_j^{r+1}$. After this process has been performed for all N fuzzy partitions of the input space, the rule base is represented by a table of indexes of input memberships containing at least one point of the training data. The consequences of rules are represented by the θ_i parameters.

The fuzzy system of eq. (2) can be represented by a linear combination of its fuzzy basis functions (FBF's) as follows:

$$y_k = \mathbf{p}^T(\mathbf{x}_k) \cdot \boldsymbol{\theta} + e_k \quad (4)$$

where $\mathbf{p}(\mathbf{x}) = [p_1(\mathbf{x}), \dots, p_l(\mathbf{x}), \dots, p_M(\mathbf{x})]^T$ and $\boldsymbol{\theta} = [\theta_1, \dots, \theta_l, \dots, \theta_M]^T$ are the vectors.

The fuzzy basis functions (FBF) of the fuzzy system are given by $p_l(\mathbf{x}) = A_l(\mathbf{x}) / \sum_{l=1}^M A_l(\mathbf{x})$. Equation (4) can be viewed as a special case of the linear regression model (Lawson and Hanson, 1974) where $\mathbf{p}(\mathbf{x})$ is the regressor vector, $\boldsymbol{\theta}$ are the parameters, and e_k is an error signal, which is assumed to be uncorrelated with the regressors $\mathbf{p}(\mathbf{x})$. For all input-output pairs $\{x_k, y_k\}$, $k=1,2,\dots,n_p$, the parameter vector $\boldsymbol{\theta}$ in eq. (4) can be determined by minimizing the Euclidean norm of the residual vector that characterizes the LSE problem (Golub and Van Loan, 1989).

3 The greenhouse model

The greenhouse climate model describes the dynamic behaviour of the stated variables by the use of differential equations for the air temperature, humidity and CO₂ concentrations. These differential equations are results of combination of the various physical processes involving heat and mass transfer taking place in the greenhouse and from the greenhouse to the outside air.

In this paper, only the temperature variables presented in Figure 1 are considered. The presented physical model has as the main objective of validating the proposed diffuse model and the performance of the FCFRA algorithm. The model considers a greenhouse divided into two control system devices (heating and ventilation), the covering material, the air inside, the outside space and the soil beneath the greenhouse. These components are characterized by the heating pipe water temperature, T_{heat} , for the heating system; inside air temperature, T_{in} ; outside air temperature, T_{out} and the soil temperature, T_{soil} . Other fundamental measured variables of the process are the solar radiation, Rad [W/m²], and two main control signals (control variables): heating ($0 \leq u_{heat} \leq 1$) and ventilation ($0 \leq u_{vent} \leq 1$), corresponding to a range from 0 to 100% of the actuator nominal power.

The energy balance to the greenhouse air is affected by the energy supplied by the heating system, $Q_{T,heat}$, the energy losses to the outside air due to the conduction through the greenhouse cover and the forced ventilation exchange, $Q_{T,out}$ [Wm⁻²], the energy exchange with the soil $Q_{T,soil}$ [Wm⁻²] and the heat supplied by radiation from the sun, $Q_{T,rad}$ [Wm⁻²]. These fluxes obey to the following relations:

- Sunlight radiant energy,

$$Q_{T,Rad} = c_{rad} \cdot Rad \quad (5)$$

where the coefficient c_{rad} denotes the usual fraction of the solar radiation transformed into heat;

- Energy losses to the outside air,

$$Q_{T,out} = (c_{vent} \phi_{vent} + c_{leak,roof}) \cdot (T_{in} - T_{out}) \quad (6)$$

where $c_{leak,roof}$ is the heat transfer coefficient through greenhouse cover. The ventilation flux, ϕ_{vent} , is the sum of natural ventilation ($\phi_{leak,vent}$) with the forced ventilation ($c_{force,vent} \cdot u_{vent}$), i.e., $\phi_{vent} = c_{force,vent} \cdot u_{vent} + \phi_{leak,vent}$ and c_{vent} is the heat capacity per unit volume of air.

- Energy transferred from the soil,

$$Q_{T,soil} = c_{soil} \cdot (T_{soil} - T_{in}) \quad (7)$$

where c_{soil} is the heat transfer coefficient.

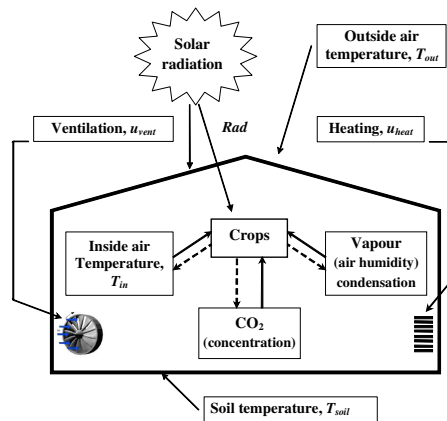


Figure 1 - Greenhouse Climate Model Structure

In the inside volume, a heat balance equation can be written and solved taking into account the different heat fluxes present here. From this analysis, the following differential equation results:

$$\frac{dT_{in}}{dt} = \frac{1}{C_{cap,T}} (Q_{T,heat} - Q_{T,out} + Q_{T,soil} + Q_{T,rad}) \quad (8)$$

where $C_{cap,T}$ is the heat capacity of the greenhouse air. The computed parameters of the greenhouse climate model are showed in Table 1. Theses parameters were obtained by employing a nonlinear optimization strategy to a set of real measured data. The training data are a collection of data recorded between January 15 and February 4 of 1998, with a sampling interval of 1 minute, in a greenhouse located at the university campus in the north of Portugal.

Here, the required task is to develop the fuzzy systems that can match all the N pairs of collected data to a given level of accuracy. The air temperature model is broken in two parts: with and without forced ventilation sub-models, LDT - Low Dynamical Temperature and HDT - High Dynamical Temperature, respectively. The equation that describes the LDT air temperature sub-model is:

$$\left. \frac{dT_{in}}{dt} \right|_{u_{vent}=0} = f_{Temp}(\Delta T, Rad, Q_{heat}, Q_{soil}) \quad (9)$$

where $Q_{heat}(t) = (T_{heat}(t) - T_{in}(t)) \cdot u_{heat}(t)$ is the heat flux from the heating system, $\Delta T(t) = T_{in}(t) - T_{out}(t)$ is the difference between inside and outside air temperature and $Q_{soil}(t) = (T_{soil}(t) - T_{in}(t))$ is the heat exchange between greenhouse air and soil.

The HDT air temperature sub-model (valid for $u_{vent} > 0$) is described by:

$$\left. \frac{dT_{in}}{dt} \right|_{u_{vent} > 0} = f_{Temp,vent}(Q_{vent}) \quad (10)$$

where $Q_{vent}(t) = (T_{out}(t) - T_{in}(t)) \cdot u_{vent}(t)$ is the heat flux exchange from outside to inside air. The dynamic model of the inside temperature is the sum of the contributions of (9) and (10).

Table 1: Parameters of the greenhouse climate model.

Parameter	Value	Dimension
$C_{cap,T}$	21208	$J m^{-2} ^\circ C^{-1}$
c_{heat}	1.148	$W m^{-2} ^\circ C^{-1}$
c_{soil}	4.554	$W m^{-2} ^\circ C^{-1}$
c_{rad}	0.5	-
c_{vent}	1290	$J m^{-3} ^\circ C^{-1}$
$c_{leak,roof}$	1.374	$W m^{-2} ^\circ C^{-1}$
$c_{force,vent}$	0.010	$m s^{-1}$
$\phi_{leak,vent}$	0.0014	$m s^{-1}$

The identification process was performed using triangular membership functions with $N_i = 4$ fuzzy sets. A set of 158 and 6 fuzzy rules were generated for the LDT and HDT models, respectively. The simulated and measured air temperatures are compared during a test period. These curves show a good agreement between measured data and the models output, as shown in Figure 2.

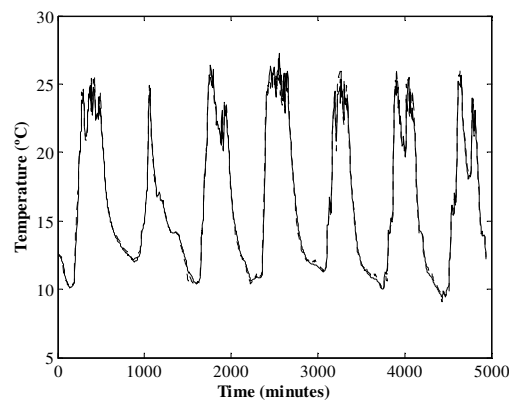


Figure 2 - Air Temperature: measured (dotted line) and simulated with fuzzy model (full line).

4 The clustering algorithm

Cluster analysis is a technique that is used to seek out natural clusters corresponding to natural classes in the data, dividing all objects (samples) into smaller sub groups, and classifying them according to the similarities among them. Clustering techniques belong to the classes of unsupervised (learning) methods,

since they do not use prior class identifiers. In general, the objective of clustering methods is to perform a partition of the collection of elements into c datasets with respect to a given criterion, where c is defined (given) as the number of clusters. In the generality of the situations, these elements are vectors of points in a space with a norm and distance measured and its results is a set of clusters of discrete points. In this work, this idea is generalized to the clustering of fuzzy numbers and fuzzy rules, where the result is the partition of a set of fuzzy belong to a fuzzy system by a set of clusters of fuzzy rules.

The objective of clustering is to partition a data set Z into c clusters, where the available identification data, $Z = [X^T, y]^T$ is formed by concatenating the regression data matrix X and the output vector y , $X^T = [x_1, \dots, x_N]$, $y = [y_1, \dots, y_N]$. This means that each observation consists of $n+1$ variables, grouped into an $(n+1)$ dimensional line vector $z_k = [x_{1,k}, \dots, x_{n,k}, y_k]$. Through clustering, the fuzzy partition matrix $U = [u_{ik}]_{c \times M}$ is obtained, whose element u_{ik} represents the degree of membership of the observation z_k in the cluster, $v_i, i = 1, \dots, c$. The fuzzy c-mean algorithm is based on the minimization of the sum of the weighted squared distances between the data points, z_k , and the cluster centers, $v_i, i = 1, \dots, c$:

$$J(Z, U, V) = \sum_{i=1}^c \sum_{k=1}^N u_{i,k}^m D_{i,k}^2 \quad (11)$$

where $m \in [1, \infty)$ is a weighting exponent that determines the fuzziness of the resulting clusters (usually chosen as $m=2$). The minimum of eq. (11) is usually solved by the alternating optimization (AO) algorithm.

The main objective of clustering the fuzzy rules is to decompose the fuzzy rules relations and the associated fuzzy sets in order to organize the information of the fuzzy system. This strategy of separation results from the unfolded process on the antecedent set A and on the consequent set B , that lead, necessarily, to the decomposition of the early relation $R_{A \rightarrow B}^l(x, y)$ (fuzzy rule) into sub-relations (in others nested fuzzy rules), on the same product space. This task is performed by a new algorithm that implements FCFR.

Consider now $Z = [X^T, y]^T$, identified by a fuzzy model $y = f(x)$ with an adequate accuracy, through a set fuzzy rule base, with M rules, and its engine mechanism. Contrarily to what happens in the classical fuzzy clustering, here the partition of the input-output space is made through the clustering of the fuzzy rules of f . A fuzzy clustering partition has the goal of separating a set of fuzzy rules $\mathcal{F} = \{R_1, R_2, \dots, R_M\}$ in c clusters, according to a similarity criterion. The result is a fuzzy partition matrix $U = [u_{il}]_{c \times M}$, whose element $u_{il} \in [0, 1]$ represents the degree of membership of rule l in cluster i . Clusters of different shapes can be obtained by using an appropriate definition of cluster prototypes (e.g., linear varieties) or by using different distance measures. The setup prototype for the clusters is given by $v = (v_1, v_2, \dots, v_c)$. Moreover, in the context of fuzzy modelling, a measure of the relevance of fuzzy rules is necessary to do the clustering of the fuzzy rules (Salgado, 2005). The relevance of a fuzzy rule l on the point, z_k , of the space work of the fuzzy system, measures the relative contribution of rule l to modelling that region and is characterized by a value in the unit interval, $0 \leq \mathfrak{R}_l(z_k) \leq 1$. For a vector $z_k \in Z$, a set of fuzzy rules is

activated and the constraint problem results in $\sum_{i=1}^c \sum_{l=1}^M \mathfrak{R}_l(z_k) \cdot u_{il} = 1, \forall z_k \in S$.

The partition should find a vector of clusters centers, V , and a partition matrix, U . Each value u_{ik} of U represents the membership degree of the k^{th} rule, R_k , to the i^{th} cluster, A_i . The FCFRA is based on the minimization of

$$J(U, V) = \sum_{k=1}^n \left[\sum_{i=1}^c \sum_{l=1}^M (u_{il} \mathfrak{R}_l(x_k))^m \|x_k, v_i\|^2 \right] \quad (12)$$

where the minimum is found by the alternating optimization (AO) algorithm given below:

Step 1 – For a set of points $\mathbf{Z}$, with $\mathbf{z}_k \in S$, and a set of rules $\mathcal{F} = \{R_1, R_2, \dots, R_M\}$, with relevance $\mathfrak{R}_l(\mathbf{z}_k)$, $k = 1, \dots, M$, keep c , $2 \leq c < n$, and initialize $U(0) \in M_{cf}$.

Step 2 – On the r^{th} iteration, with $r = 0, 1, 2, \dots$, compute the c mean vectors.

$$v_i^{(r+1)} = \frac{\sum_{l=1}^M \left(u_{il}^{(r)} \right)^m \cdot \sum_{k=1}^{np} \left(\mathfrak{R}_l(\mathbf{z}_k) \right)^m \cdot \mathbf{z}_k}{\sum_{l=1}^M \left(u_{il}^{(r)} \right)^m \cdot \sum_{k=1}^{np} \left(\mathfrak{R}_l(\mathbf{z}_k) \right)^m} \quad \text{where } \left[u_{il}^{(r)} \right] = U^{(r)}, i = 1, 2, \dots, c.$$

Step 3 – Compute the new partition matrix $U(r+1)$ using the expression:

$$u_{il}^{(r+1)} = 1 / \sum_{j=1}^c \left(\frac{\sum_{k=1}^{np} \left(\mathfrak{R}_l(\mathbf{z}_k) \right)^m \cdot d_{ik}}{\sum_{k=1}^{np} \left(\mathfrak{R}_j(\mathbf{z}_k) \right)^m \cdot d_{ik}} \right)^{\frac{2}{m-1}} \quad \text{with } 1 \leq i \leq c, 1 \leq l \leq M.$$

Step 4 – Compare $U(r)$ to $U(r+1)$: If $\|U(r+1) - U(r)\| < \varepsilon$ then the process ends. Otherwise let $r=r+1$ and go to step 2. A small real positive constant is denoted by ε .

The application of the FCFRA to a fuzzy system results in a HCS fuzzy system, where the i^{th} fuzzy sub-model is

$$f_i(\mathbf{x}) = \sum_{l=1}^M \left(A^l(\mathbf{x}) \cdot u_{il} \right) \cdot \theta^l / \sum_{l=1}^M \left(A^l(\mathbf{x}) \cdot u_{il} \right) \quad (13)$$

The previous flat fuzzy models have been organized in HCS structures by FCFRA into c clusters (sub-models). This strategy will be used in next section to break the greenhouse fuzzy climate temperature model into different parts, with a correspondent meaning to the analogous physical sub-model.

5 Experimental results

In this section the results achieved with the proposed greenhouse climate model strategies are presented: physical sub-models and the clustered fuzzy models. The identification of both models was done by using input–output data previously collected. The physical model was built by adding the contributions from the different heat mechanisms involved in the greenhouse climate behaviour. This model is used as a comparison to the clustered fuzzy models.

In Figure 3, the heat fluxes computed with the physical and the clustered fuzzy sub-models are plotted. The natural heat leakage responses, $Q_{out,leak} = (c_{vent} \phi_{leak,vent} + c_{leak,roof}) (T_{in} - T_{out})$ (without forced ventilation) is plotted in (a); the heat input due to the sunlight radiation, $Q_{T,rad}$ in (b); the heat supplied by the heating system, $Q_{T,heat}$ in (c) and the heat exchanges between the inside air and the soil, $Q_{T,soil}$, in (d). These results show a good agreement between the clustered fuzzy model and its physical sub-models counterparts.

6 Conclusion

Fuzzy models, especially if acquired from data, often contain a large number of fuzzy sets. A closer study of the fuzzy rule base may reveal that the relationship between the input and output fuzzy sets is organized in a way that does not reflect the internal process dynamics. Despite the success of the linguistic interpretation by the fuzzy model, its utility in describing linguistically non-additive relationships has been neglected.

This work has shown that the clustering strategy is capable of organising the fuzzy models in order to decompose its structure into a hierarchical and collaborative way. The result is a set of fuzzy sub-models that can reflect some type of individual relationships. This task is carried out without degradation of the global quality of the original fuzzy model. The proposed method can be successfully applied to real problems such as modelling of physical and biological processes. In this paper, an application of fuzzy modelling the air temperature of an agricultural greenhouse has been shown. The clustering of the fuzzy model has been tested to split the inside greenhouse air temperature flat fuzzy model into fuzzy sub-models. These sub-models have alike counterparts on the physical modelling that represent the contributions of the process mechanisms involved in the global system dynamics.

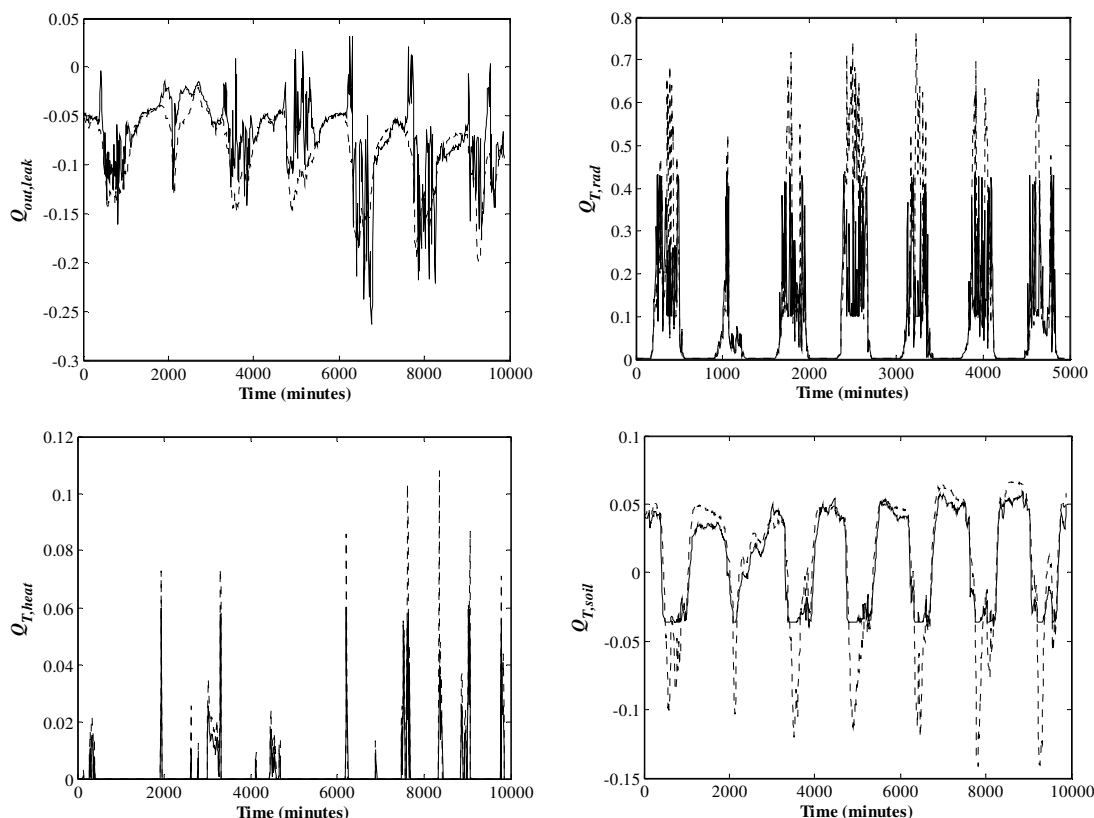


Figure 3 - Heat fluxes computed with physical model (dotted line) and HCS fuzzy sub-model (full line).

7 References

- Golub, G.H. and Van Loan, C.F., Matrix Computations, John Hopkins, Baltimore, 1989.
- Lawson, C.F. and Hanson, R.J., Solving Least Squares Problems, Prentice-Hall, Englewood Cliffs, NJ, 1974.
- Salgado, P., "Clustering and hierarchization of fuzzy systems", Soft Computing Journal, Springer-Verlag, in-press, 2005.
- Salgado, P. and Cunha, J.B., "Greenhouse Climate Hierarchical Fuzzy Modelling", Control Engineering Practice, Elsevier, Vol 13/5, pp 613-628, 2004.
- Zeng, X.J. and Singh, M.G., "Approximation theory of fuzzy systems – SISO case," IEEE Trans. Fuzzy Syst., vol. 2, pp. 162-176, 1994.

Acknowledgements

This work was supported by Fundação para a Ciência e Tecnologia (FCT) – Proj. POSI/SRI/41975/2001.