Fuzzy reference model for daily outdoor air temperature

Eugenio Fco. Sánchez Úbeda, Ana Berzosa Muñoz
Instituto de Investigación Tecnológica
Universidad Pontificia Comillas
28015 Madrid
Eugenio.Sanchez@iit.upco.es, Ana.Berzosa@iit.upco.es

Abstract

In this paper it is proposed a new methodology for modeling daily outdoor air temperature based on the Linear Hinges Model (LHM). In particular, the proposed model tries to capture the seasonal evolution of temperature throughout the year and it consists of two terms: the deterministic component describing the expected temperature, and the stochastic component describing the variance. Using a fuzzy representation of the LHM, a set of humanly understandable fuzzy rules are obtained. These rules are useful to describe qualitatively the evolution of temperature throughout the year. Several application examples are shown and salient features of the model are also discussed.

1. Introduction

The outdoor temperature time series can be described by a process fluctuating randomly around a seasonal long-run mean temperature. Figure 1 shows the daily evolution of the maximum and minimum outdoor air temperature recorded in the airports of Barcelona and Madrid from October 1st 1999 to September 31st 2003. Note that departures from the "usual" behavior are so frequent and significant that one-year ahead punctual predictions of that variations are useless. However, the obvious annual periodicity of temperature time series allows us to obtain good daily estimates of both the mean temperature and a measure of the volatility around that reference temperature.

The outdoor air temperature has a direct impact on many industrial and economic processes. Thus, many studies have been carried out to understand such relationships.

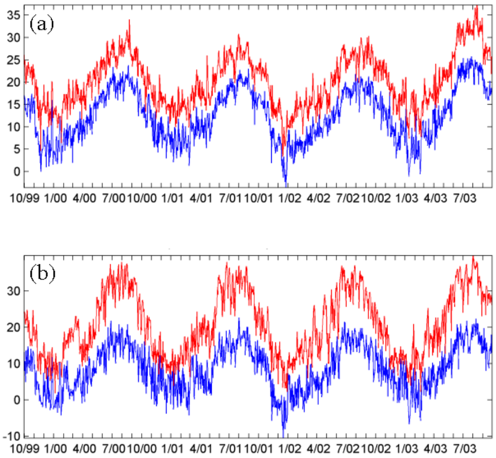

Figure 1 daily evolution of the maximum and minimum outdoor air temperature at (a) the Barcelona airport and (b) the Madrid airport

For example, the influence of temperature in energy consumption has been widely analyzed in the past, see e.g. [1][2][3][4]. This influence is usually captured by using the so-called heating degree days (HDD) and the cooling degree days (CDD). They are defined as the difference between the mean temperature for a day and some reference temperature:

$$HDD(t) = \max(T_R - T(t), 0) \qquad (1)$$
$$CDD(t) = \max(T(t) - T_R, 0)$$

where $T(t)$ is the mean outdoor air temperature at day t and T_R is a reference temperature (e.g. 65ºF in the USA, 15ºC in Spain). If $T(t)$ is above T_R, there are no HDD that day. If the $T(t)$ is less than T_R, there are no CDD that day. Thus, these indexes try to measure the intensity and duration of cold and heat in winter and summer days, respectively. Although these degree days can help to explain, for example, the daily variations of electricity and natural gas consumption due to temperature, they have severe limitations.

Actas del III Taller Nacional de Minería de Datos y Aprendizaje, TAMIDA2005, pp.271-278
ISBN: 84-9732-449-8

In this paper we propose a fuzzy reference model for daily outdoor air temperatures. This model can be used to improve and simplify many applications where temperature models are used (e.g. demand forecasting).

The rest of the paper is organized as follows. Section 2 introduces the LHM as a general-purpose automatic learning model. This is the basis of the proposed approach. Section 3 describes how to apply the LHM to obtain a useful temperature reference model. Section 4 provides the fuzzy interpretation of the LHM. Section 5 shows illustrative example. Finally, conclusions are pointed out in Section 6.

2. The Linear Hinges Model

The LHM has been proposed in [5]. It is an efficient approach to flexible and robust one-dimensional curve fitting under stringent high noise conditions. Given a set of N points

$$\left(x_e, y_e = f(x_e) + \varepsilon_e\right)_{e=1,N} \quad (2)$$

the objective in the curve fitting from scatterplot data is to find a simple enough function *f* such that the following equation holds

$$y_e = \hat{f}(x_e) + \hat{\varepsilon}_e\,, \quad (3)$$

with small enough error estimates ε_e, as measured empirically by the overall mean squared error (MSE):

$$MSE = \frac{1}{N}\sum_{e=1,N}\hat{\varepsilon}_e^2 \quad (4)$$

Next we introduce the LHM, summarizing its main characteristics as well as its learning algorithm.

2.1. Model definition

The LHM is a piecewise linear model defined by K knots, the points specifying the pieces (Fig. 2a). In particular, it can be expressed mathematically as:

$$\boldsymbol{H}(x) \equiv \left\{H_j(x) \quad : k_{j-1} \le x \le k_{j+1}\right\}_{1\le j\le K} \quad (5)$$

where one hinge is defined by the current knot plus two straight-line segments:

$$H_j(x) \equiv 1\left[k_{j-1}, k_j\right] H_{l,j}(x) + 1\left(k_j, k_{j+1}\right] H_{r,j}(x) \quad (6)$$

where

$$H_{l,j}(x) = h_{j-1} + \frac{\Delta h_{l,j}}{\Delta k_{l,j}}\left(x - k_{j-1}\right) \quad (7)$$

$$H_{r,j}(x) = h_{j+1} + \frac{\Delta h_{r,j}}{\Delta k_{r,j}}\left(x - k_{j+1}\right) \quad (8)$$

$$\begin{aligned}\Delta h_{l,j} &= h_{j-1} - h_j \;;\quad \Delta h_{r,j} = h_j - h_{j+1}\\ \Delta k_{l,j} &= k_{j-1} - k_j \;;\quad \Delta k_{r,j} = k_j - k_{j+1}\end{aligned} \quad (9)$$

and 1[•] and 1(•] are the indicator functions of the left and right intervals. Note that the subscripts *l* and *r* denote the left- and right-hand side of the hinge.

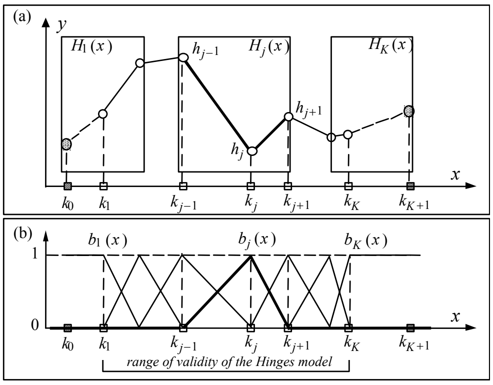

Figure 2 (a) LHM definition; (b) basis functions for it

2.2. Features of the LHM

The main general characteristics of the LHM are:

- One of the salient features is its flexibility. This approach is able to produce adequate models in many different situations. The model is able to adapt its complexity to the quality and availability of the data.
- Another very important advantage of the LHM is its accuracy. According to [6], the LHM seems to be in general less complex than other automatic learning models such as the Supersmoother [7] or the Hermite polynomial model.
- The computational efficiency is another characteristic of the LHM. It is extremely fast. For problems with a few thousand learning

examples the computing time at the learning stage is in the order of magnitudes of milliseconds, (see examples below). On the other hand, in terms of testing CPU times, it is ultra fast (a few milliseconds).

- Finally, simplicity is another feature of the LHM. Because it is generally desirable to interpret the obtained result, models in closed form approximations of maximal simplicity are in demand. Indeed, this characteristic has been used to develop the ORTHO model, a multidimensional model based on the LHM which is fully interpretable [6][8].

2.3. Learning algorithm

The learning algorithm to fit the LHM combines a greedy divide-and-conquer strategy with a computationally efficient pruning approach and special updating formulas. In particular, it consists of four main stages: smoothing, growing, pruning and refitting. The first three steps are used to identify automatically the complexity required for a given problem, (i.e. the number of parameters of the model). The refitting stage then improves the accuracy of the model. A detailed description, including analytical formulas of the learning algorithm, can be found in [5] or [6].

Alternatively, this learning algorithm can be described as a particular implementation of the "backfitting" algorithm (see [6]). Friedman and Stuetzle initially proposed this term in the context of projection-pursuit regression [9]. In [10] authors make intensive use of this general alternating optimization strategy, providing justifications for its use.

Basically, in this generic algorithm the parameters of the model are grouped such that the solution for those in each group is straightforward given fixed values for those outside the group. A solution is obtained for the parameters in a group, using these solution values as current values for the parameters in that group. This process is repeated for each group, one by one. The overall process is repeated several times until it agrees with the termination criteria used.

This powerful algorithm enables us to fit any additive model, but its iterative nature has a cost. In some circumstances, according to [11], it can converge slowly. However, if the optimization process required for each group of parameters is not only straightforward but also fast enough, then the computer time for the overall process can be really acceptable. This is the situation of our particular implementation for the LHM. Its parameters are grouped into hinges, being the solution for each hinge straightforward given fixed values for the parameters of the rest of hinges. Figure 3 shows the hierarchical grouping used implicitly during the adjustment of the LHM.

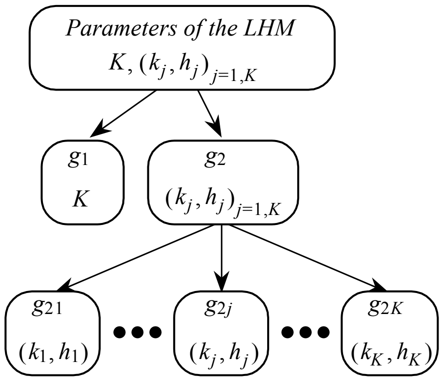

Figure 3 Hierarchy used to fit the LHM, via the backfitting algorithm

2.4. Fuzzy interpretation of the LHM

In this section we relate the LHM to fuzzy rule-based systems, where a collection of (fuzzy) rules specifies an input-output mapping by associating fuzzy sets in the input space with output fuzzy sets. The goal is to show that such fuzzy representation of the LHM is straightforward as well as useful for model interpretation.

This fuzzy interpretation of the LHM allows coding the dynamic temperature behavior with a natural language syntax with which people are comfortable.

First we describe the LHM in terms of a set of basis functions. According to this decomposition the connection of this model with some existing neurofuzzy approaches (e.g. [12][11]) becomes obvious. Next we discuss two alternative specifications of the rule outputs. Finally, an illustrative example is presented.

2.5. The LHM in terms of a set of basis functions

Although no basis functions have been used to adjust the LHM, this model can be expressed as the linear combination of basis functions:

$$H(x) = \sum_{j=1,K} h_j b_j(x), \qquad (10)$$

where the triangular-shape basis functions are given by (Fig 2b):

$$b_1(x) = 1(k_0, k_1] + 1(k_1, k_2]\frac{x-k_2}{k_1-k_2}$$

$$b_j(x)\Big|_{j=2,K-1} = 1[k_{j-1}, k_j]\frac{x-k_{j-1}}{k_j-k_{j-1}} + 1(k_j, k_{j+1}]\frac{x-k_{j+1}}{k_j-k_{j+1}}$$

$$b_K(x) = 1[k_{K-1}, k_K]\frac{x-k_{K-1}}{k_K-k_{K-1}} + 1(k_K, k_{K+1})$$

(11)

Thus, according to this formulation of the LHM, it is straightforward to represent the model as a neural network, similar to a RBFN. Figure 4 shows this representation, where the input layer merely sends the input value to the hidden layer. This hidden layer consists of K units, where each unit has a triangular-shape activation function (i.e. a basis function). Finally, the output layer carries out the weighted linear combination of (10). Note that the weights feeding the hidden layer are fixed to one.

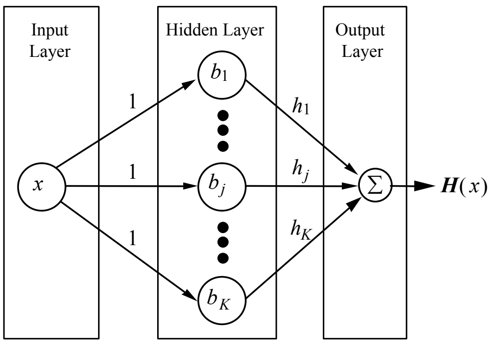

Figure 4 The LHM as a simple neural network

2.6. Automatic fuzzy-partition of the input space

A fuzzy system is a model whose behavior is described by a set of rules of the form "*if x is A, then y is B*" connecting antecedents (x is A) with consequences (y is B). The linguistic terms A and B are specified by their membership functions $\mu_A(x)$ and $\mu_B(y)$, respectively. For example, the fuzzy membership function $\mu_{small}(x)$ represents the grade of membership of a quantity value x belonging to the set *small*.

The membership functions have values in the interval [0, 1]. The value 0 means that the variable x is not a member of the set small, whereas $\mu_{small}(x) = 1$ indicates that x entirely belongs to the fuzzy set small. For further reading see e.g. [13][14].

In a traditional fuzzy system these input and output membership functions as well as the fuzzy rules are typically provided by application domain experts. On the other hand, more recent neurofuzzy systems try to estimate the input fuzzy sets from experimental data [11][12][13]. It is well-known that the modeling capabilities of a neurofuzzy model are mainly determined by the number, shape and distribution of the fuzzy input membership functions. In this respect, the LHM produces automatically a fuzzy partition of the input space as it finds the number, shape and distribution of the fuzzy input membership functions. Thus, the user only needs to decide a fuzzy label for each membership function to obtain the input fuzzy sets.

Each basis function in (11) can be viewed as a membership function describing a particular abstract concept (fuzzy set):

$$\mu_{A,j}(x) = b_j(x) \quad j = 1, \cdots, K \qquad (12)$$

where A_j is a fuzzy label or linguistic term like 'small' or 'large' describing the variable x. This label is the adjective selected by the user to refer to the concept defined by the membership function (12). Obviously, to obtain a truly transparent description these linguistic terms should be appropriate adjectives.

2.7. Takagi-Sugeno fuzzy rules (crisp output)

Using the fuzzy partition of the input space provided by the LHM one can derive the following handful of fuzzy rules:

If x is A_j then y = h_j , $\quad j = 1, \cdots, K$

where A_j is the fuzzy label defined by $\mu_{A,j}(x)$ (known in this context as rule strength) and *hj* is the vertical position of the hinge *j*. This type of rules are known as (zero-order) Takagi-Sugeno fuzzy rules [13] and they simply set the position of the hinges in both input and output spaces.

Although in our approach these rules are used to describe the input-output mapping provided by the LHM in terms of humanly understandable statements, they could be used to infer the output for a given input. In particular, a standard fuzzy system with singleton fuzzification, the product as fuzzy implication operator for rule evaluation and centroid defuzzification will produce the same output than the LHM, (see e.g. [13][11] or [12] for further details). Note that these prescriptions are commonly used in practice.

2.8. Zadeh-Mamdani fuzzy rules (fuzzy output)

Another alternative is to use rules where the output is also fuzzy (known as Zadeh-Mamdani rules [13]). Indeed, because the issue of this fuzzy representation is the linguistic interpretation of the LHM, this last approach seems to be particularly interesting: one can obtain rules of different complexity describing the model by using different fuzzy output partitions (i.e. quantizations). For example, a coarse partition of the output can consist of two fuzzy sets (e.g. 'small' and 'large'), whereas a fine partition has a larger number of output fuzzy sets.

These fuzzy rules can be formally defined as:

$$\textit{If } x \textit{ is } A_j \textit{ then } y \textit{ is } B_i\ (c_{ij}) \quad j = 1,\cdots,K; i = 1,\cdots,W$$

where A_j and B_i are the fuzzy labels defined on the input and output variables, respectively. The input fuzzy sets are specified by the LHM, whereas the W output ones are provided by the user. Each rule has a rule confidence c_{ij}, representing the confidence in the rule being true (0 means the rule does not contribute to the output). In particular, these rule confidences are given by:

$$c_{ij} = \mu_{B,i}(h_j), \tag{13}$$

where h_j is the vertical position of the hinge j and $\mu_{B,i}(x)$ is the membership function specified by the user to precisely define the vague fuzzy label B_i in the present context.

Note that these rules will produce the same output than the LHM when one uses some typical prescriptions for fuzzification, rule evaluation and defuzzification, outlined in §2.7. However, we recall that these rules have been generated for linguistic interpretation of the LHM. To obtain the output one should use the more compact and computationally efficient mathematical form of (5).

3. Modeling temperatures via the LHM

In this section we describe the proposed fuzzy reference model for daily outdoor air temperature. In particular, the proposed model tries to capture the seasonal evolution of temperature throughout the year and it consists of two terms:

- The deterministic component describes for each day of the year the (mean) expected temperature.
- The stochastic component provides interval predictions based on confidence intervals around the mean value.

To build this fuzzy model in terms of a LHM, first we need to form the set of learning examples describing the temperature behavior. This learning set is obtained by overlapping years of 365 days (discarding all February 29th).

In order to obtain the determinist component, a LHM is built based on this set by applying the learning algorithm described above. Next, the stochastic component is obtained. This component is also modeled as a LHM, but a new scatterplot is used as learning set. Given the previous learning set and the deterministic component, the new set is built by computing for each day the square root of the difference between the real temperature value and the deterministic component. Figure 5 sketches these two steps.

4. An illustrative example

In order to illustrate the proposed approach, we have studied the minimum outdoor air temperature at the Madrid airport. In particular, we have used 23 years (from 1980 to 2003, i.e. 8395 points) to build the model shown in Figure 6. This model, with only 7 hinges, summarizes the functional dependence of the outdoor air temperature on the day of the year (1 to 365). The Mean Absolute Error, estimated using the 8395 samples, is 2.6367ºC.

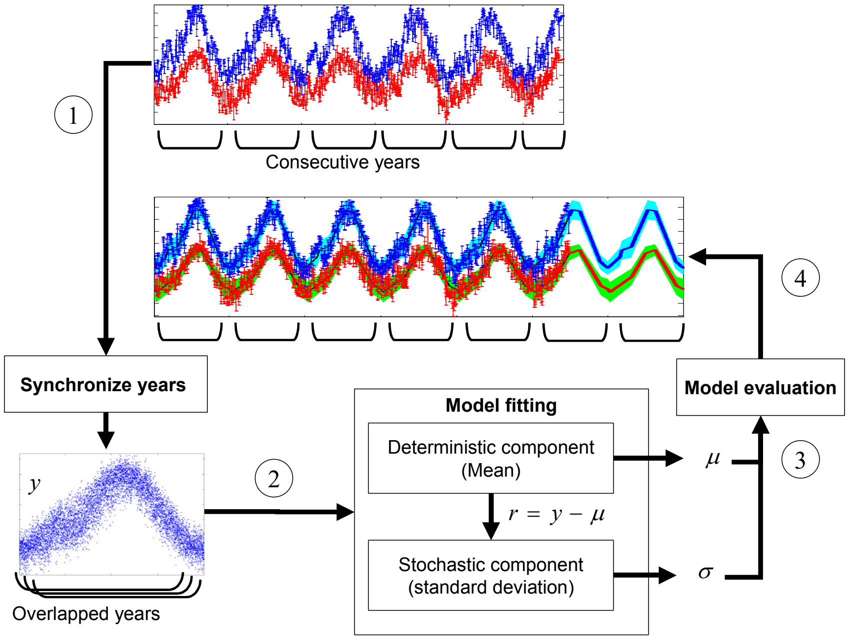

Figure 5 General scheme showing the overall process proposed to built the fuzzy reference model

The previous LHM can be described in terms of the following set of Takagi-Sugeno fuzzy rules (see §2.7):

if the day is WTR then the temp is –1ºC
if the day is SPG then the temp is 6ºC
if the day is SPG-SMR then the temp is 16ºC
if the day is SMR then the temp is 17ºC
if the day is ATN then the temp is 2ºC
if the day is ATN-WTR then the temp is 1ºC

where the fuzzy labels WTR (winter), SPG (spring), SPG-SMR (spring-summer), SMR (summer), ATN (autumn) and ATN-WTR (autumn-winter) describing the variable "minimum temperature at the Madrid airport" have been precisely defined by the membership functions of Figure 6.

On the other hand, according to §2.8, it is possible to use fuzzy sets for encoding the output variable. Following our expectation about real daily minimum temperatures in Madrid, it could be enough to distinguish between L (low), M (medium) and H (high) temperatures.

According to Figure 7, the LHM of Figure 6 can be described in terms of the following set of Zadeh-Mamdani fuzzy rules:

if the day is WTR then the temp is L (1.0)
if the day is SPG then the temp is L (0.8)
if the day is SPG then the temp is M (0.2)
if the day is SPG-SMR then the temp H (1.0)
if the day is SMR then the temp H (1.0)
if the day is ATN then the temp is L (1.0)
if the day is ATN-WTR then the temp is L (1.0)

where the number between brackets is the rule confidence.

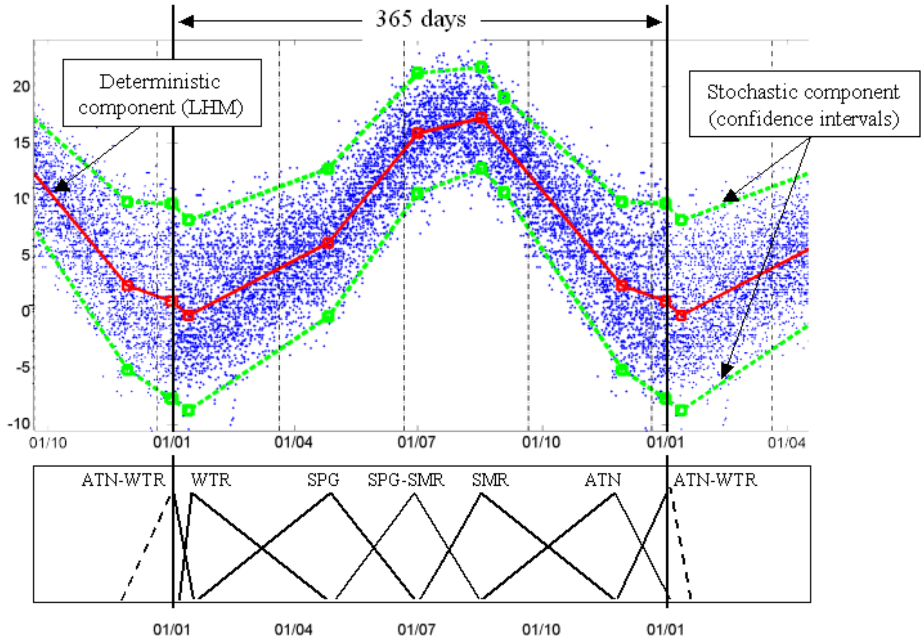

Figure 6 (a) Obtained LHM for minimum outdoor air temperature and input fuzzy sets describing the variable "temperature". The letters above the sets are linguistic labels (see main text)

It is clear that the previous definition of fuzzy sets is very context-dependent, i.e. these sets may seem inappropriate to the reader. This is due to the nature of the language. However, using our approach the user only needs to label the abstract concepts (i.e. basis functions) identified automatically by the LHM.

Finally, Figure 8 shows the real daily outdoor air temperature recorded in the Madrid airport from July 1st 2001 to June 31st 2002, as well as the obtained fuzzy reference model. Using this reference model it is possible to identify the temperature drop during December 2001.

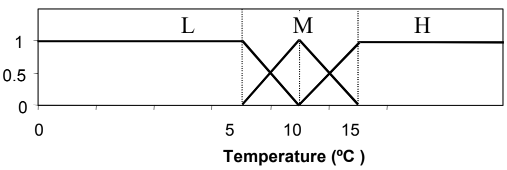

Figure 7 A group of possible fuzzy sets defining the output variable "temperature"

5. Conclusions

In this paper we have proposed a novel approach to fuzzy modeling daily outdoor air temperature. It exploits a general-purpose automatic learning model, the so-called Linear Hinges model (LHM), to extract simple and at the same time reliable models of daily outdoor air temperature series. The LHM provides relevant interpretable information, which may receive various interesting uses. In this respect several applications have been illustrated, showing very promising results and in particular the ability of the model to capture the interesting information as well as to describe in terms of a set of humanly understandable fuzzy rules the input-output mapping.

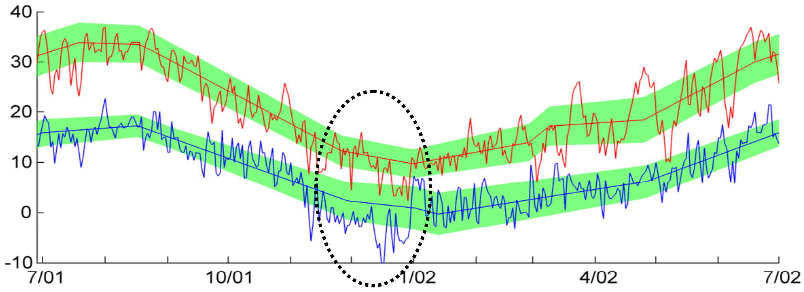

Figure 8 Maximum and minimum daily outdoor air temperature recorded in the Madrid airport, and obtained fuzzy reference model for both temperatures

Furthermore, such invertible relationship between the mathematical form of the LHM and its qualitative form in terms of fuzzy-rules can allow us to exploit and combine both expert knowledge and empirical data. In particular, this approach provides two main advantages:

- Due to the ability of the LHM to capture the relevant information, it can be useful within some analytical tool to examine how temperatures are working.
- The fuzzy interpretation of the LHM is particularly beneficial in applications where some subsequent human decision-making must be carried out.

References

[1] D.J. Sailor, J.R. Muñoz, "Sensitivity of electricity and natural gas consumption to climate in the USA-methodology and results for eight states", Energy 22, 987-998, 1997.

[2] S. Ayyash, M. Salman, N. Al-Hafi, "Modelling the impact of temperature on summer electricity consumption in Kuwait", Energy 10, 941-949, 1985.

[3] A. Pardo, V. Meneu, E. Valor, "Temperature and seasonality influences on Spanish electricity load", Energy Economics 24, 55-70, 2002.

[4] E. Valor, V. Meneu, V. Caselles, "Daily air temperature and electricity load in Spain", Journal of Applied Meteorology, vol. 40, no. 8, Aug. 2001, pp. 1413-21, 2001.

[5] E.F. Sánchez-Úbeda and L. Wehenkel, "The Hinges model: A one-dimensional continuous piecewise linear model", Proc. of IPMU'98, pp. 878-885, Paris, July 1998.

[6] E.F. Sánchez-Úbeda, Models for data analysis: contributions to automatic learning, Ph.D. thesis, Universidad Pontificia Comillas of Madrid, 1999.

[7] J.H. Friedman, "A Variable Span Smoother", Tech. Report LCS 05, Stanford University, November 1984.

[8] E.F. Sánchez-Úbeda and L. Wehenkel, "Automatic fuzzy-rules induction by using the ORTHO model", Proc. of IPMU'2000, invited paper, Madrid, July 2000.

[9] J.H. Friedman and W. Stuetzle, "Projection Pursuit Regression", Journal of the American Statistical Association, Vol. 76, No. 376, December 1981.

[10] T.J. Hastie and R.J. Tibshirani, Generalized Additive Models, Chapman and Hall, London, 1990.

[11] K.M. Bossley, Neurofuzzy Modelling Approaches in System Identification, thesis, Univ. of Southampton, 1997.

[12] E. Uchino, T. Yamakawa, "Neo-Fuzzy-Neuron Based New Approach to System Modeling, with Applicatioin to Actual System", Proc. of the 6th International Conf. on tools with artificial intelligence, New Orleans, Nov. 1995.

[13] V. Cherkassky and F. Mulier, Learning from data: concepts, theory and methods, John Wiley, N.Y., 1998.

[14] G.J. Klir and B. Yuan, Fuzzy sets and fuzzy logic: theory and applications, Prentice Hall, New Jersey, 1995.