A Neural Network model for the estimation of bioclimatic indexes

F. Patania[1], A. Gagliano[1], R. Caponetto[2], F. Nocera[1] & A. Galesi[1]
[1]*Department of Industrial and Mechanics Engineering,
Catania University, Italy*
[2]*Department of DIEES, Catania University, Italy*

Abstract

Many researches have highlighted the influence of climate on mortality, showing a high increase in mortality in the summer time during "heat waves", periods of very high temperature and humidity levels.

The bioclimatic indexes are used in urban climate studies to describe the level of thermal sensation that a person experiences due to the modified climatic conditions of an urban area.

The index provides a meaningful and realistic indicator that can not only be used as an information as to how hot it feels, but also as a readily identifiable warning for individuals subject to the physiological dangers of heat exposure.

The authors have developed a methodology that, by means of the Neural Network (NN), permits one to predict the values of meteorological data and then the calculation of the bioclimatic indexes.

The meteorological data required for the calculation of the bioclimatic index concerning hourly values of air temperature, relative humidity, wind speed have been used according to the records of the meteorological station of the Pergusa Lake (EN), for the year from 2003 to 2006

NN-estimated, the bioclimatic indexes values were compared with coincident bioclimatic indexes values obtained from air temperature and relative humidity observations recorded at standard meteorological stations.

Statistical analysis showed a good agreement between the NN-estimated and the station-observed bioclimatic indexes values, with a root mean square error (RMSE) < 1%.

The proposed methodology demonstrates the potential of using the NN for defining the bioclimatic indexes and its feasibility as an indicator to pre-alert authority to heat related risks and dangers of heat waves

Keywords: heat-waves, bioclimatic index, Neural Network.

1 Introduction

The most direct effect of weather on human mortality in mid-latitudes are observed during and after summer heat waves that lead to a significant increase and intra-seasonal shifts in total mortality [1].

Recent blistering summer heat-waves have caused serious health and social problems in the WHO European Region as, during the summer of 2003, heat waves can be not only extremely uncomfortable, but even deadly.

During the summer season, or in very hot periods anyway, the human organism makes use of perspiration to maintain its temperature within proper physiological limits. The sweat, which is mostly made of water, evaporates (endothermal phenomenum) taking away heat so as to have a cooling effect on the skin. A high humidity level in the surrounding environment may obstruct this process limiting the evaporation. The human body therefore cannot eliminate the excessive heat (compared to its own physiological limits), receiving a sensation of a higher temperature. Overheating can be very dangerous; heat exhaustion occurs when the body loses fluid and salt through perspiration faster than can be replaced, causing dizziness. The elderly and ill are not the only people that heat kills. It also kills healthy young people, usually because they do not recognize the dangers of exercising in hot weather, especially hot, humid weather.

Exercising in hot weather can cause muscle cramps, especially in the legs, because of brief imbalances in body salts. People not used to exercising in heat may experience a quick drop in blood pressure that can lead to fainting and, in some cases, extreme heat can cause the body temperature to rise to 40,5 degrees or higher, causing a heat stroke, with confusion and unconsciousness.

When heat and humidity combine with the slow evaporation of sweat from the body, outdoor exercise becomes dangerous even for those in good shape.

Many researches have highlighted the influence of climate on mortality, showing a high increase in mortality in the summer time during "heat-waves", periods with very high temperature and humidity levels.

This relation seems to be stronger than those between mortality and other environmental factors, such as atmospheric pollution [6].

Dangers of heat include

- Heat cramps: exercising in hot weather can lead to muscle cramps, especially in the legs, because of brief imbalances in body salts. Cramps become less frequent as a person becomes used to the heat.
- Heat syncope or fainting: anyone not used to exercising in the heat can experience a quick drop in blood pressure that can lead to fainting. As with heat cramps, the cure is to take it easy.
- Heat exhaustion: Losing fluid and salt through perspiration or replacing them in an imbalanced way can lead to dizziness and weakness. Body

temperature might rise, but not above 39,0 degrees. In some cases victims, especially the elderly, should be hospitalized. Heat exhaustion is more likely after a few days of a heat wave than when one is just beginning.

- Heatstroke: In some cases extreme heat can upset the body's thermostat, causing body temperature to rise to 40,5 degrees or higher. Symptoms are lethargy, confusion and unconsciousness. Even a suspicion that someone might be suffering from heatstroke requires immediate medical aid. Heatstroke can kill.

The risk of premature death due to temperature increases associated with global warming is six times higher for Europeans who are already suffering from respiratory problems, according to the European Respiratory Society (ERS): for instance, a one degree Celsius increase in temperature produces a 1-3% increase in deaths in the general population, but this same temperature increase results in a 6% increase in deaths among people with respiratory problems.

2   Artificial Neural Networks

One attempt to emulate the imprecise and incomplete way of treating information that is typical of the human brain is represented by Neural Networks (NNs), which were born as structure models of the brain but are currently used as calculation paradigms for approximating non-linear functions and classification.

One of the characteristics of NNs is that of emulating the human brain, even though simplistically, by exploiting its main features, *i.e.*, the ability to learn from experience.

Thus, NNs emulate the structure and functioning of biological neurons, even if in fact no neural network is able to reach, except to a minimal extent, the elaborating capacity of the human brain, and therefore the analogy is a purely formal one.

Artificial NNs can thus be defined as a connected *computational system* able to *store* and *utilize knowledge* acquired through experimenting. The knowledge acquired is stored with the help of values of certain parameters, called *weights*, which connect the computational units, known as *nodes* or *neurons*, whose values are fixed during the training phase.

Every neuron is an entity which has access to several inputs and one output only. It receives the inputs from *neighboring* neurons, elaborates them and transmits the output to other neurons, appropriately weighting them by means of the connective values. The neuron model currently most used is reported in 1, where x_i represents the ith input (the ith component of the vector X), w_i is the relative weight at the ith input (the ith component of the vector W), and $f(W^t X)$ is a function, generally a non-linear one, known as the *activation function* . The value of the activation function, calculated at the weighted addition of the inputs, thus represents the *neuron output*.

The number of input and output neurons is known being related to the application while the *hidden structure* is chosen from one time to another and determines the *performance* of the net.
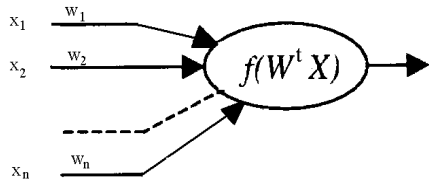
Figure 1:   Neuron model.

There are three different *operational phases* of the networks. The first is called the *learning or training phase*. During this phase a set of example are provided to the network. These data, using different learning algorithm, chosen according to the network structure and type, are used to update the weight values.

The second phase is called the *test phase* or *checking* phase and, by using a different data set, serves to verify that at the end of the learning the network is functioning correctly. In this phase, the weights of the net are "frozen" at the previously determined, learning phase, values. Once these two phases are completed, the net can be used for calculating the output in the face of *unknown input*.

The neural structure applied in the paper is the *Multilayer Perceptron* (MLP). As shown in Figure 2, the MLP is made up of a layer of input neurons that have the goal of transferring the signal to the next layer, giving it the appropriate weights from a certain number of internal layers, also known as *hidden layers*, and from a layer of output neurons. The neurons of each layer are all connected by the *weights* with those of the previous and successive layers. The MLP structure is thus characterized by the *number of hidden layers* and the *number of neurons* in each layer [3].
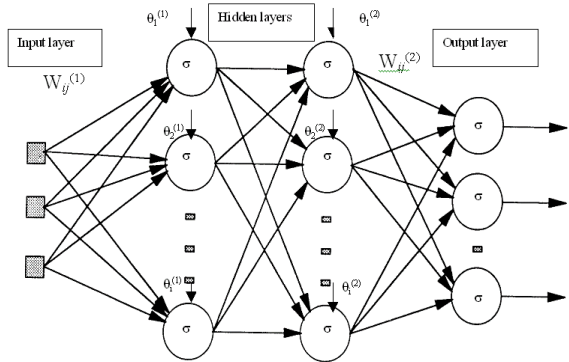

Figure 2:   Multilayer perceptron diagram with two hidden layers.

With regard to Figure 2, the following notations are used:
- *N1*: number of layers;
- *s*=1,...,N1: index of layers;

- $N^{(s)}$: number of neurons in the layer s (in particular $N^{(1)}$ indicates the input while $N^{(Nl)}$ indicates the number of output);
- $O_i^{(s)}$: i^{th} input of the layer s (for $s=1$ it represents the ith input of the network);
- $W_{ij}^{(s)}$: weight of the interconnection between the ith neuron of the layer s and the jth neuron of the layer s-1;
- $\theta_i^{(s)}$: the bias factor of the ith unit of the layer s, which corresponds to the weight of an interconnection with a neuron activated at the unitary value;
1. $\sigma i(\bullet)$: the activation function (generally, a continuous and monotone increasing derivable function, bounded both above and below).

The formulas which allow us to calculate the network outputs, once the inputs $O_i^{(1)}$ are known, are the following:

1. calculation of the activation for each neuron

$$x_i^{(s)} = \sum_{j=1}^{N_{s-1}} W_{ij}^{(s-1)} O_j^{(s-1)} + \theta_i^{(s)} \qquad \text{for } s=2,...,N_e \text{ e } i=1,...,N^{(s)} \qquad (1)$$

2. calculation of the neuron output with a logistic sigmoidal activation function

$$O_i^{(s)} = f(x_i^{(s)}) = \frac{1}{1+e^{-x_i^{(s)}}} \qquad (2)$$

The calculations that allow us, given the input values to the network, to calculate the outputs $O_i^{(Nl)}$ for $i = 1,...,N^{(Nl)}$ are generally called *forward propagation* phase.

The Stone-Weierstrass theorem which offers a rigorous demonstration of the *universal interpolation* property of an MLP with at least one layer of hidden neurons, provides the theoretical background for the NN application. In fact, it has been proved that an MLP-type artificial NN with at least three layers and a *sigmoidal activation* function for the neurons of the hidden layer is able, to *any degree of accuracy* that may be required, to interpolate a continuous non-linear function, or one with a finite number of discontinuities.

With regard to the *learning* of an MLP, it takes place by means of an *iterative algorithm* that updates the values of the network interconnections such that a total *square error functional* is optimized on a set of input/output data defined [1]. At this point, the gradient algorithm for *back-propagation* of the output layer error is applied until the first hidden layer, simultaneously updating the value of the weights, according to the *deepest-descent* gradient formula [4].

3 The bioclimatic index

The body attempts to maintain a constant internal temperature of 37°C at all times. In hot weather, the body produces sweat, which cools the body as it evaporates. As the humidity or the moisture content in the air increases, sweat does not evaporate as readily.

Sweat evaporation stops entirely when the relative humidity reaches about 90 percent. Under these circumstances, the body temperature rises and may cause illness.

The bioclimatic index is used to quantifying human discomfort due to excessive heat and humidity.

The bioclimatic indexes most commonly used in urban climate studies are the Humidex, the Discomfort Index (DI) of Thom, the Summer Simmer Index (SSI), the Apparent Temperature, etc.

The Humidex combines the temperature and humidity into one number to reflect the perceived temperature. As it takes into account the two most important factors that affect summer comfort, it can be a good measure of how stifling the air feels than either temperature or humidity alone.

The Humidex formula is based on the work of Masterton and Richardson [3].

$$\text{Humidex} = T + 5/9 * (e^{-10}) \qquad (3)$$

$e = (6.112*10^{\wedge(7.5*T/(237.7+T)})*H/100$ vapour pressure.
T= air temperature (degrees Celsius).
H= humidity (%).

Table 1 shows the correlation between the values of the Humidex and the correlated human discomfort.

Table 1:   Humidex and the correlated human discomfort.

Humidex Range	Degree of Comfort
20-29	comfortable
30-39	some discomfort
40-45	great discomfort; avoid exertion
above 45	dangerous
above 54	heat stroke imminent

The new SSI gives an indication of the dangers of exposure to various combinations of temperature and humidity, based on physiological studies.

$$\text{SSI} = 1.98 \times (T - (0.55 - 0.0055 \times (H)) \times (T - 58)) - 56.83 \qquad (4)$$

T = Air Temperature (°F);
H = Relative Humidity (%).

The results can be illustrated in tabular form as a function of temperature and humidity, or as a function of temperature and dew point.

Figure 3 show the values of SSI as a function of temperature and humidity.

Temp (° F)	Relative Humidity (%)										
	5	10	20	30	40	50	60	70	80	90	100
70		66	69	70	71	72	73	73	74	74	75
75		73	75	76	76	77	78	79	80	81	82
80		78	80	82	83	84	86	88	91	93	97
85		83	86	87	88	92	96	100	104	109	117
90		88	92	94	97	100	106	115	123	135	*
95		94	99	103	107	112	123	135	149		
100		100	104	112	117	128	137	*			
105	102	106	112	119	131	142	*				
110	108	112	119	131	144	*					
115	113	117	128	142	*						
120	118	124	138	*							

Figure 3:   The new Summer Simmer Index [6].

Seven zones are identified in figure 3.

For SSI values between 70 and 77(zone 1), most people are comfortable, but slightly cool.

For SSI values between 77 and 83 (zone 2), nearly everyone feels quite comfortable.

For SSI values between 83 and 91 (zone 3) most people are comfortable, but slightly warm.

For SSI values between 91 and 100 (zone 4), increasing discomfort is experienced (warm).

For SSI values between 100 and 112 (zone 5), a caution of sunstroke and heat exhaustion exists for prolonged exposure and activity, along with significant discomfort (extremely warm).

For SSI values between 112 and 125 (zone 6), virtually everyone is uncomfortable, a danger of heatstroke and great discomfort exists (hot).

For SSI values between 125 and 150 (zone 7) there is an extreme danger of heatstroke, especially for the weakened or elderly, and even young children, whosebody metabolism demands cooler effective temperatures than most adults. Maximum discomfort exists at these conditions (extremely hot).

Beyond 150, circulatory collapse is imminent for prolonged exposure.

Obviously the formulas used to calculate the discomfort indexes produce theoretic average data which may be highly influenced by a lot of human and environmental factors such as: height, weight and sex of the individual, clothes used, presence of shade or wind, carried out activity.

4   Methodology

The general form of a time-discrete NARMAX (Non-Linear Auto-Regressive Moving Average with Exogeneous Input) model is the following:

$$y(k) = F[y(k-1),........,y(k-n_y),u(k-1),.....,u(k-n_u)]$$

being u(k) and y(k) respectively the input and output of the system. In the study case y(k) and u(k) represent samples of the considered variables that are: air temperature, in the following indicated as (T), solar radiation (RS), relative humidity (UR), wind speed (VV) and wind direction (DV).

Daily data, from years 2003 to 2006, recorded by meteorological station of the Pergusa Lake (Enna, Italia) have been used to train and test neural networks. Training phase has been performed using years from 2003 to 2005, while 2006 has been used for the test phase. As previously introduced, the unknown function F is identified by means of artificial neural networks, as described in the following models. In particular, an MLP [1] with 20 hidden neurons has been adopted.

Several models with a different number of regression for each variable have been tested. An extensive trial-and-error phase has led to the models characterized by the following structures:

T(k)=F[T(k-2), RS(k-2), UR(k-2), DV(k-2), VV(k-2), T(k-1), RS(k-1), UR(k-1), DV(k-1), VV(k-1)];
UR(k)=F[UR(k-2), T(k-2), RS(k-2), DV(k-2), VV(k-2), UR(k-1), T(k-1), RS(k-1), DV(k-1), VV(k-1)];
VV(k)=F[VV(k-2), T(k-2), RS(k-2), UR(k-2), DV(k-2), VV(k-1), T(k-1), RS(k-1), UR(k-1), DV(k-1)].

5 Results

Figure 4 shows the comparison between the NN-estimated Humidex values and the coincident bioclimatic indexes values obtained from air temperature and relative humidity observations recorded at standard meteorological stations during the summer of 2006.

Statistical analysis showed a good agreement between the NN-estimated and the station-observed bioclimatic indexes values, with a root mean square error (RMSE) < 1%.

It is possible to notice that for more or less the 20% of cases (20 days) have been verified to be extremely dangerous conditions for elderly and persons with pre-existing cardiovascular and respiratory diseases.

Figure 5 shows the comparison between the NN-estimated SSI values and the coincident bioclimatic indexes values obtained from air temperature and relative humidity observations recorded at standard meteorological stations during the summer of 2006.

Figure 6 shows the percentage distribution of the Humidex (a) and the SSI (b) in function of the perceived human discomfort.

It has to be noted that for more or less in the 25% of cases (23 days) have been verified extremely dangerous conditions for elderly and persons with pre-existing cardiovascular and respiratory diseases.
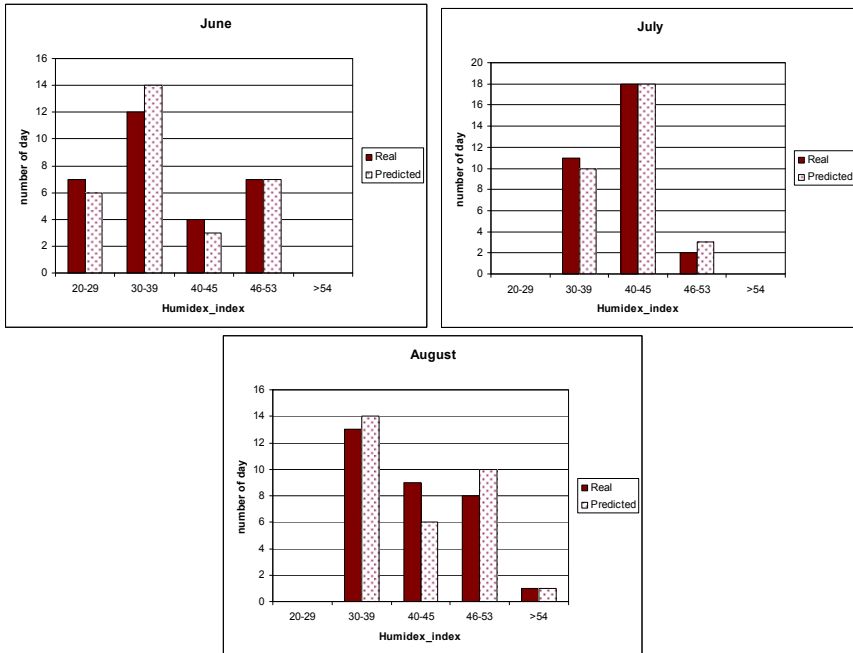
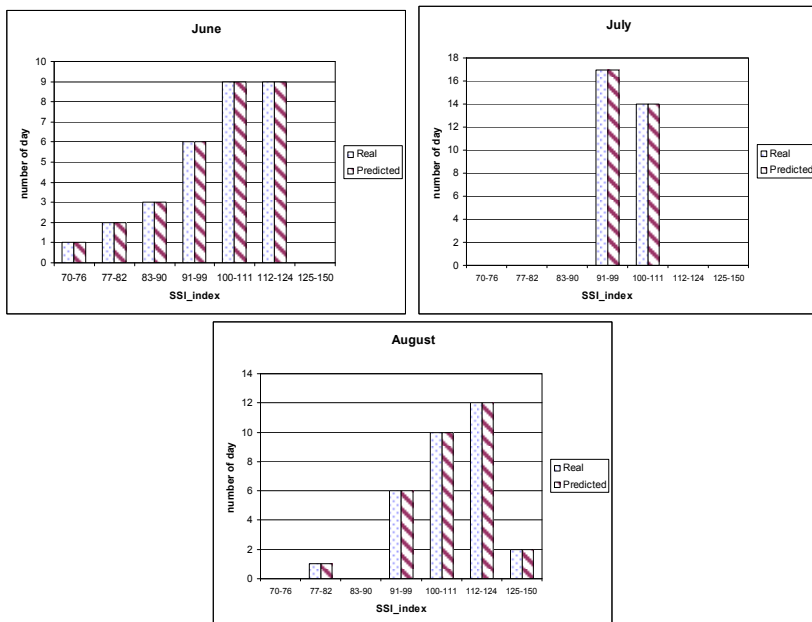Figure 4:   Predicted and "measured" value of Humidex index.

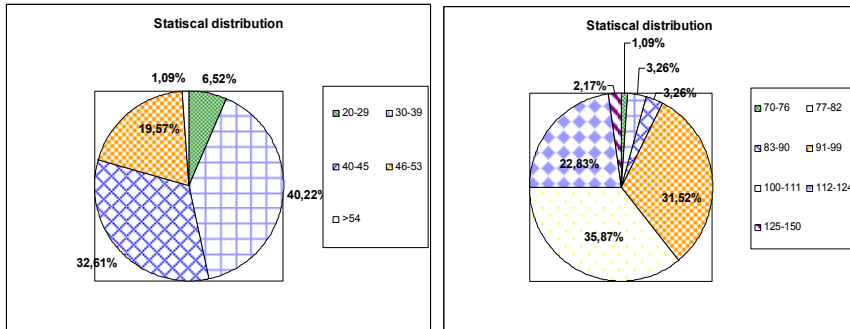Figure 5:   Predicted and "measured" value of SSI (June 2006).

Figure 6:   Statistical distribution of Humidex (a) and SSI (b) index during the summer 2006.

The analysis of results shows that no substantial differences have been found for the two different bioclimatic indexes considered.

6   Conclusions

In this study the authors have proposed a methodology that permits one to predict the daily value of bioclimatic indexes (Humidex and the SSI) utilising the data recorded by the meteorological station in the Urban Area of Enna (ITALY).

The bioclimatic index provides a meaningful and realistic equivalent temperature that can not only be used as a marker as to how hot it feels, but also as a readily identifiable warning for individuals subject to the physiological dangers of heat exposure.

The obtained results demonstrate a good correlation between predicted values of bioclimatic indexes, calculated by means of the NN, and real values of bioclimatic indexes

The elderly and persons with pre-existing cardiovascular and respiratory diseases are most at risk of death from ambient heat exposure; these risks can be prevented if we can provide warning systems and information for respiratory patients.

The proposed methodology to predict the value of the bioclimatic indexes could be used as a public awareness for elderly and persons with pre-existing cardiovascular and respiratory diseases that are most at risk of death from ambient heat exposure.

References

[1] Meehl, G. A., and C. Tebaldi, 2004: More intense, more frequent, and longer lasting heat waves in the 21st century. *Science*, **305**, 994-997.
[2] Hajat, S., B. J. Armstrong, M. Baccini, A. Biggeri, L. Bisanti, A. Russo, A. Paldy, B. Menne, and T. Kosatsky, 2006: Impact of high temperatures on mortality. *Epidemiology*, **17**, 632–638

[3] L. Fortuna, G. Rizzotto, M. Lavorgna, G. Nunnari, M.G. Xibilia, R. Caponetto, "Soft Computing", Advanced Textbooks in Control and Signal Processing, Springer, Berlino, 2001

[4] Jan Kysely, Bohumir Kriz "Decrease impacts of the 2003 heat waves on mortality in the Czech Republic: an improved response?, Journal of Biometeorology (2008) 52: 733-7452 DE Rumelhart, JL McClelland. "Parallel distributed processing: explorations in the microstructure of cognition: foundations". vol.1, MIT Press, Cambridge, MA, USA 1986

[5] J.M. Masterton and F.A. Richardson of Canada's Atmospheric Environment Service in 1979.

[6] Pepi W.J. 1987. The Summer Simmer Index, Weatherwise, Vol 40, No. 3, June