

Estimation of fungal spore concentrations associated to meteorological variables

Antonella Angelosante Bruno · Loretta Pace · Barbara Tomassetti ·
Erika Coppola · Marco Verdecchia · Giovanni Pacioni · Guido Visconti

Received: 9 May 2007 / Accepted: 8 June 2007
© Springer Science+Business Media B.V. 2007

Abstract A 1-year time series of fungal spore concentrations has been used to calibrate an artificial neural network for the estimation of *Alternaria* and *Pleospora* concentrations associated to observed meteorological variables. Analysis of the results revealed that the daily average values of these meteorological variables are suitable to predict with high confidence the number of fungal spores that are actually observed. The calibrated neural network has also been used randomizing each single input parameter in order to evaluate which meteorological variable contributes more to the formation and the

depletion of the selected fungal spores. Emphasis is given to the possibility of using the proposed model for operational activities, predicting the future spore concentrations on the basis of meteorological forecasts.

Keywords Aerobiological monitoring · *Alternaria* · Artificial neural network · Fungal spores · Meteorological parameters · *Pleospora*

1 Introduction

Spore concentration is an important cause of allergies and many other poor health conditions, the main symptoms of which are usually associated with respiratory problems. The spores seem to be responsible for clinical symptoms of oculorhinitis and bronchial asthma. Following house dust mite and grass pollen, *Alternaria* and *Cladosporium* spores are considered to be two of the most common causal factors eliciting symptoms of asthma, eczema and rhinitis (Tariq et al. 1996). The potential to predict spore growth under specific meteorological conditions could be a great advantage for the physician as a means to establish a preventive appropriate strategy for allergy sufferers.

The critical correlation between spore concentrations and meteorological conditions has been long recognized by many authors. Lyon et al. (1984) used a multiple-regression technique to determine the

A. Angelosante Bruno · L. Pace (✉) · G. Pacioni
Dipartimento di Scienze Ambientali, Università degli
Studi dell'Aquila, Via Vetoio, L'Aquila 67100 Coppito,
Italy
e-mail: loretta.pace@univaq.it

A. Angelosante Bruno · L. Pace
Centro di monitoraggio RIMA (AQ1), Università
dell'Aquila, L'Aquila, Italy

B. Tomassetti · E. Coppola · M. Verdecchia · G. Visconti
Centro di Eccellenza CETEMPS, Università dell'Aquila,
L'Aquila, Italy

E. Coppola
The Abdus Salam Centre for Theoretical Physics,
Trieste, Italy

M. Verdecchia · G. Visconti
Dipartimento di Fisica, Università dell'Aquila,
L'Aquila, Italy

possible interaction of several meteorological factors on the spore concentrations at 1.5 m above ground level, which is the height at which most people would inhale allergenic fungal spores. They found a direct correlation between *Cladosporium* and ascospore concentrations and a minimum wind velocity and atmospheric relative humidity, while an inverse correlation was found with maximum wind speed. A long time series of 10 years of data was used by Hjelmroos (1993) to investigate the relationship between the presence of *Cladosporium* and *Alternaria* spores and weather variables. In this case, a regression analysis technique was adopted to investigate the relationship, which revealed a good correlation with cloud cover, relative humidity and precipitation. Trouitt and Levetin (2001) showed that different combinations of factors can be predictors of the concentration of the spore types being examined, with atmospheric temperature seemingly the most significant predictor for *Cladosporium* and *Alternaria* spore concentration, while dew point and pressure possibly playing a secondary role. The variable role of temperature on conidia production and conidia release was also investigated by Kurkela (1997). All of these studies lead to the conclusion that the concentrations of fungal spores in the atmosphere largely depend on many different meteorological factors.

We report here our investigation on the possibility of mapping such a complex relationship using an artificial neural network (ANN) based on a numerical model. We demonstrate how the observed values of daily average minimum and maximum temperature, the daily average relative humidity and the daily total precipitation represent a complete set of predictors that can be used to efficiently estimate the concentration of *Alternaria* and *Pleospora* spores.

This article is organized such that we first provide a detailed description of spore concentration measurements obtained at the University of L'Aquila and describe the meteorological data sets used in our study. We then provide an explanation of the basic concepts of ANN theory and present the characteristics of specific applications. The results of our ANN model for the prediction of *Alternaria* and *Pleospora* concentrations are then reported and discussed. The last section focuses on the future developments and applications of the proposed neural approach.

2 Spore concentration and meteorological parameter measurements

Spore concentration and meteorological variable measurements were collected in L'Aquila (Italy) from January to October 2004. L'Aquila is located in central Italy at latitude 42.355° N and longitude 13.395° E and at about 700 m a.s.l. The site is inland and situated at the southwest foothills of the Gran Sasso peak (2912 m a.s.l.). The landscape is made up of oak wood (*Quercus pubescens*), pine wood (*Pinus nigra*) and a community of pastures. Typical communities along the rivers are willows (*Salix* sp.) and poplars (*Populus* sp.). Invasive plants are present in the surroundings of L'Aquila, especially in the uncultivated fields, road, path margins and dumps.

A time series of spore concentrations has been collected using a Lanzoni VPPS 2000 Hirst volumetric sampler (Hirst, 1952) that was installed 20 m above ground level. Melinex tape impregnated with a 2% silicon solution was used as trapping surface. Average daily spore concentration estimations (spores m^{-3}) were carried out from January to October 2004: the spores were identified under a Leica light microscope at 4000× magnification on the basis of morphological characteristics; the counting procedure of the conidia followed the model proposed by Mandrioli et al. (1998). The observed time series for different species of spores are reported in Fig. 1.

As clearly seen in Fig. 1, the concentration of *Alternaria* spores increased from early April, with the maximum peaks (around 80 m^{-3}) occurring during the summer months, while the maximum concentration of *Pleospora* spores (about 40 m^{-3}) occurred during April, with only a few spikes observed during summer months. All other fungal spores never reached concentrations higher than 5 m^{-3} . The peak period for *Epicoccum* spores was August; for *Stemphylium* spores, May; for *Torula* and *Helminthosporium* spores, June; for *Polythrincium*, April. Due to the seasonal variability of different taxa shown in Fig. 1, it is more interesting to correlate spore concentrations and meteorological conditions for *Alternaria* and *Pleospora*.

Data of meteorological variables for the same period were collected using different sensors with a 1-h time resolution. In order to reduce the complexity of the regression model discussed in the subsequent sections, we used the daily average for wind speed,

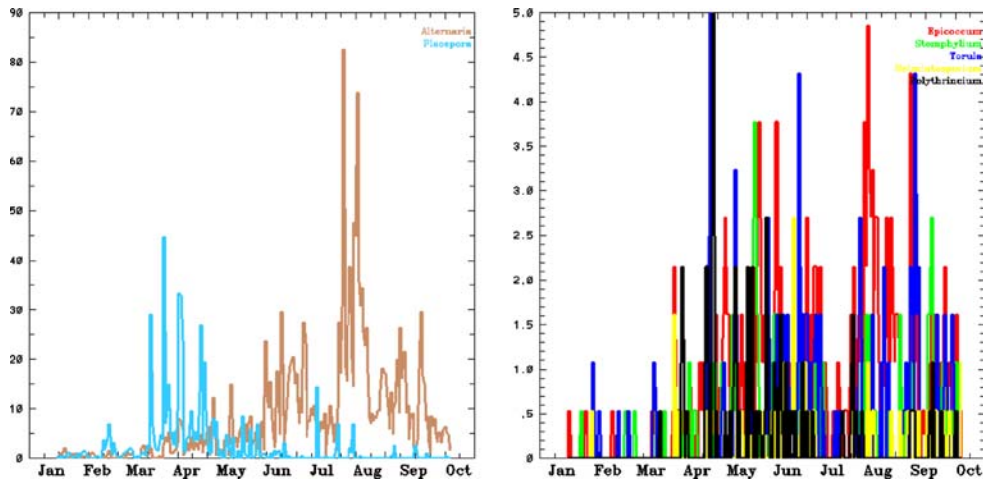


Fig. 1 Spore concentration (m^{-3}) measured in L'Aquila between January and October 2004. *Left panel*—The concentration of *Alternaria* (brown line) and *Pleospora* (cyan line) spores. *Right panel*—The observed time series for *Epicoccum* (red line), *Stemphylium* (green line), *Torula* (blue line), *Helminthosporium* (yellow line) and *Polythrincium* (black line)

precipitation and atmospheric relative humidity, and we also calculated the average, maximum and minimum temperatures on a daily basis.

The time sequences of the main meteorological variables used in the present work are shown in Fig. 2. Based on the data recorded in Fig. 2 (left

concentrations. The maximum *Alternaria* spore concentration (about 80 m^{-3}) was observed during the summer months, while the maximum peaks for *Pleospora* (about 40 m^{-3}) spore concentration occurred during April. All other species never reached concentrations higher than 5 m^{-3}

panel), most of the rainfall events in 2004 occurred between March and May, with typical values of precipitation being around 10 mm day^{-1} and only occasionally values greater than 20 mm day^{-1} being reached. A long period of drought with only few precipitation events are also observed from June until

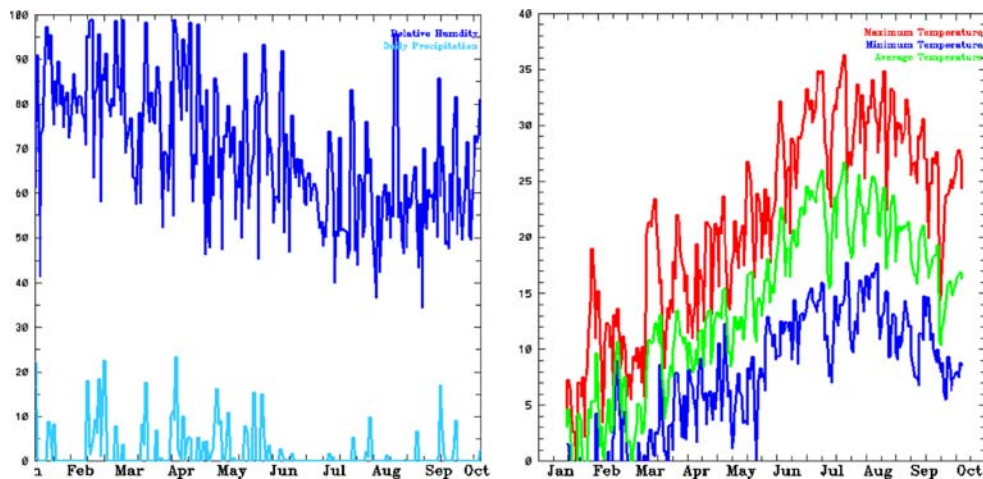


Fig. 2 *Left panel*—The observed sequences for percentage relative humidity (blue line) and daily total precipitation (cyan line) for the selected time interval in 2004. Most of the rainfall events occurred between March and May, while a long period of aridity with only few precipitation events were observed from June until the end of September. The average values of relative humidity decreased slightly from spring to summer.

Right panel—Time sequences for maximum (red line), minimum (blue line) and average (green line) temperature. The winter temperatures mainly ranged from 0 to 10°C , and strong fluctuations in the daily temperature occurred during the summer months, with maximum daily values of around 30°C or more and minimum values falling to around 15°C or lower

the end of September. The average values of relative humidity decrease slightly from spring to summer. The time sequences for maximum, minimum and average temperature are reported in Fig. 2 (right panel). Winter temperatures mainly ranged from 0 to 10°C. During the summer months, there were strong daily fluctuations in temperature, with the maximum temperature during the day commonly reaching 30°C or more and the minimum temperature at night commonly falling to around 15°C or lower.

3 The neural network model

ANN represents a powerful numerical device to map complex relationships between different sets of observed variables. In the last two decades a neural-based approach has been applied to a wide number of problems in different disciplines, including those in atmospheric physics and the natural Sciences (see, for example, Coppola et al. 2006, Grimes et al. 2003). A complete description of the ANN technique is beyond the scope of the present paper; instead, we limit our explanation to a very quick introduction to ANN theory, focusing the discussion on the basic concepts needed for the understanding of the proposed application. A complete and rigorous description of ANN theory can be found in Hect-Nielsen (1991).

The design of a formal artificial neuron is inspired by the physiology of neural cells. The status of a neuron is a linear function of several input stimuli:

$$O_I = \sum_j O_j w_{ij}$$

where O_I represents the status of the neuron I , and O_j are the different inputs received by the cell. The weights w_{ij} are equivalent to the synapses of biological neural network. For practical applications, the stimuli O_j consist of the status of the different other neurons to which the neuron is connected to.

An ANN consists of a collection of different neurons connected to each other. The presence of a connection between neuron I and j means that the status of neuron I is one of the inputs of another neuron, j , by means of a weight (synapse) w_{ij} . The single neurons are often called the nodes of the network. The architecture of an ANN is defined by establishing how the single neurons of the network are

connected to each other. One of the most popular types of ANN architecture, used in a wide range of applications, is called the ‘multilayer perceptron’. In this architecture, neurons are arranged in different layers, with the nodes in each layer receiving inputs only from the nodes in the preceding layer and subsequently passing on their outputs (status) only to the nodes of the next layer (feed forward multi-layer perceptron). An example of this architecture is given in Fig. 3. The first layer of the network, usually referred to as the ‘input layer’, comprises seven neurons, the second layer, usually called the “hidden layer”, comprises four neurons and the last layer (output layer) contains only one neuron. For our application, the use of this kind of network architecture is straightforward: a complete set of meteorological parameters is presented as the input of the network of Fig. 3 (first layer), which returns, as output, the corresponding spore concentration. It is essential that the weights (synapses) of the network are carefully determined in such a way that the network can reproduce the observed relationship between meteorological conditions and spore concentrations.

According to the Hect-Nielsen theorem (Hect-Nielsen 1991), any analytical application can be computed with high accuracy using a feed-forward three-layer network with a full connection between adjacent layers like the network shown in Fig. 3. Nevertheless, a more ambitious approach could be

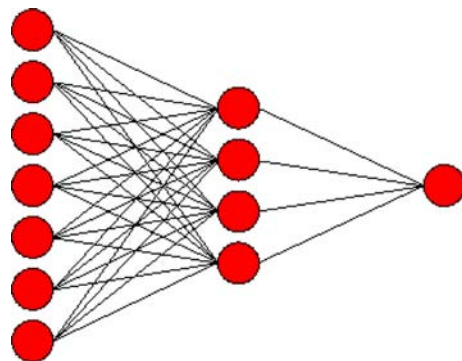


Fig. 3 An example of a ‘feed forward multi-layer perceptron’. The neurons are arranged in different layers, with the nodes in each layer receiving inputs only from the nodes in the preceding layer and subsequently passing on their outputs (status) only to the nodes of the next layer. In this example, the first layer of the network, usually referred to as the ‘input layer’, consists of seven neurons, the second layer, usually called ‘hidden layer’, consists of four neurons and the last layer (output layer) contains only one neuron

adopted for our application. The basic idea is that spore depletion and production are not only influenced by the meteorological conditions of a single day, but depend on the sequence of meteorological events (several days). In order to take this multi-day effect into account, we adopt a more complex architecture, usually referred to as the Elman neural network (Elman 1990). An example of this architecture is schematically reported in Fig. 4. In this case, the status of the neurons in the hidden layer is back-propagated to the input layer. In fact, the neurons in the second layer contain compressed information of the meteorological parameters at the previous time step, and the architecture shown in Fig. 4 implements a sort of dynamical memory of meteorological events. For a more complete description of Elman architecture, the reader is referred to Elman (1990).

It has to be mentioned that the number of the neurons in the hidden layer (four) has been chosen in order to limit the total number of weights to at least a few fold lower than the total number of training examples, thereby avoiding a critical problem known as overtraining of the network (Hect-Nielsen 1991).

Figure 4 also shows the list of input variables used for the application presented in the present paper. We chose to use as input all of the meteorological parameters that were measured with spore concentrations: the daily average temperature (T_{av}), the daily

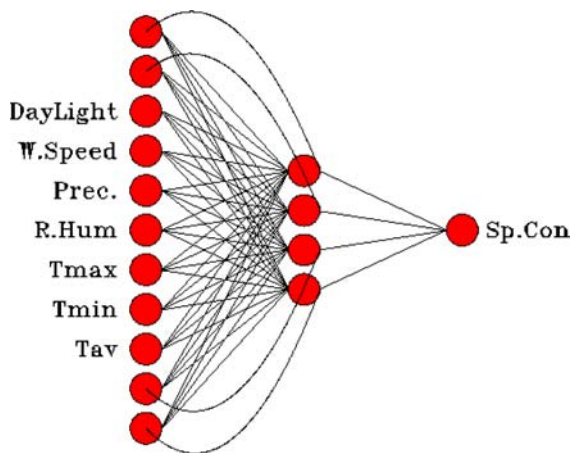


Fig. 4 The Elman neural network architecture proposed for the estimation of spore concentrations from meteorological parameters. In comparison to the feed forward neural network, the architecture of the Elman neural network implements a sort of dynamical memory that enables the network to take the sequence of meteorological events into account

minimum temperature (T_{min}), the daily maximum temperature (T_{max}), the daily average wind speed (Wind), the total daily precipitation (Prec) and the daily average relative humidity (RelH). In addition, we assume that spore concentrations also depend on other heterogeneous seasonal factors and, in order to take into account these factors, we also consider an additional input variable, Day Light, which is a sinusoidal function of Julian day assuming the value 1 at the summer equinox and 0 at the winter equinox.

As mentioned above the process of calibration, usually referred to as the ‘training’ of the ANN, consists of determining all of the weights (synapses) of the network using the observed input/output patterns. To this end, we used each input (meteorological) pattern and the corresponding output (spore concentrations) value in sequence; a single input/output pattern is usually called a ‘training example’: an atmospheric pattern used as input is presented to the first layer of the ANN, the output of the network (i.e. the status of the neuron in the last layer) is calculated and then this result is compared with the spore concentration observed with the given input pattern. The values of the synapses are then modified in order to reduce the difference between observed and predicted spore concentrations. Formally, for each, training pattern the weights W_{ij} are modified by a quantity ΔW_{ij} , according to:

$$\Delta W_{ij}(t) = -\alpha \frac{\partial E}{\partial W_{ij}} + \beta \Delta W_{ij}(t-1)$$

where α and β are two small coefficients and E is a measure of the error. The process is carried out for all the examples and is iterated until good performances for the ANN device are obtained. For a more complete description of the training techniques, see Hect-Nielsen (1991).

For the practical implementation of our model, a modified version of JETNET package developed by Lombladd et al. (1991) has been used. The modification of the software is needed to include Elman network architecture not implemented in the original package.

4 Results and discussion

The neural network described in the previous section (see Fig. 4 and related discussion) has been used to

built a regression model for evaluating the *Alternaria* and *Pleospora* fungal spore concentrations from meteorological parameters. In order to quantitatively evaluate the performances of the neural model, we used several statistical parameters. The bias is defined as

$$\sum_{l=1}^n (x_l - y_l)$$

where the vectors x_l and y_l represent the time series of the estimated and observed concentration, respectively, and the sum is carried out over all of the available observations. The bias value allows to highlight a systematic underestimation or overestimation of the predicted quantity. The root mean square error is defined as:

$$\sqrt{\frac{\sum_{l=1}^n (x_l - y_l)^2}{n}}$$

and is a measure of the average error affecting the estimation. Finally, the correlation coefficient of the two time series is calculated as:

$$\frac{\sum_{l=1}^n (x_l - \bar{x})(y_l - \bar{y})}{\sqrt{(\sigma_x^2 \sigma_y^2)}}$$

where σ_x and σ_y are the standard deviation of two time series. An optimal value of this coefficient is 100 (perfect correlation) and the worst value is -100 (the two series are totally uncorrelated). A graphical comparison of estimated and observed time series for *Alternaria* spore concentrations is given in Fig. 5.

In Fig. 5, the red line represents the observed time series for *Alternaria* spore concentrations (the same shown in the left panel of Fig. 1), and the blue line represents the time series rebuilt by the neural regression. As it can be seen, the occurrence of the main peaks during the summer season are localized very well and the intensities are also reasonable. The ability of the ANN to reproduce the critical events is of course of great importance for the future possible operational use of this numerical estimator. Nevertheless, a few peaks of concentration are clearly underestimated by approximately twofold during June and September. A number of factors could cause this underestimation: (1) Poor statistics of the

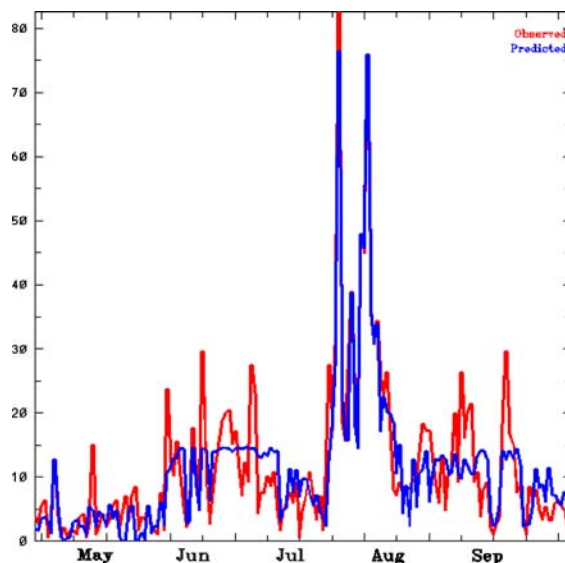


Fig. 5 Comparison between estimated (blue line) and observed (red line) *Alternaria* concentration time sequences

events (less than 1 year) that do not allow a totally adequate calibration of the numerical device; (2) seasonal factors other than those included in our meteorological-based model that affect spore concentration; an explanation supported by many authors (see Lyon et al., 1984; Abdel Hameed, 2005) who claim that other biological factors critically affect spore depletion and growth.

A comparison similar to what was reported in Fig. 5 is given in Fig. 6, but for *Pleospora* concentrations. In this case, the reconstruction of observed time series is very close to being perfect. The curves are in fact superimposed for a long range, and the occurrence and intensity of the main peaks are very well reproduced.

The results of the quantitative analyses using the statistical indices already defined are summarized in Table 1.

For both *Alternaria* and *Pleospora* taxa, the bias values are negligible, and the predicted time series are therefore not affected by systematic overestimation or underestimation. The root mean square error is good for both time series and very good (less than 2 m^{-3}) for *Pleospora*. Correlation coefficients are, in both cases, greater than 90%.

Based on these good results, we decided to carry out an addition test to check the different influences of meteorological parameters on our numerical

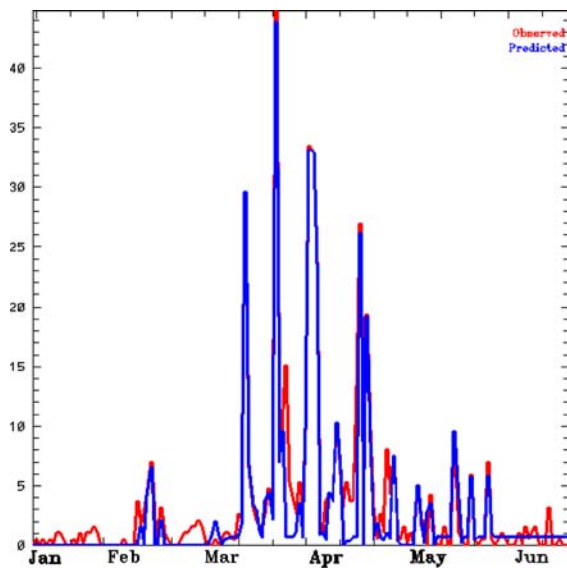


Fig. 6 The same comparison as in Fig. 5 but for *Pleospora* concentrations

Table 1 The statistical indices for *Alternaria* and *Pleospora* concentration estimations

	Bias	RMSQ	Correlation
<i>Alternaria</i>	0.2	4.1	92.0%
<i>Pleospora</i>	0.2	1.9	93.4%

RMSQ, Root mean square

models for spore growth and depletion. For this purpose, the same statistical indices of in Table 1 are calculated, but in this case the different input parameters are randomized in order to evaluate how the statistical indices are affected by the loss of one data set. The results are summarized in Table 2.

As expected, because of the random perturbation of input parameters, bias values were not significantly

affected. The RMSE always increased up to a value around 10 m^{-3} . The values of correlation coefficients for the perturbed configuration support the hypothesis that the concentration of *Alternaria* spores is strongly affected by the minimum and average temperatures, while that of the *Pleospora* spores is critically dependent on precipitation. It has to be specified that the test reported in Table 2 must be considered as indication only, and it is not an effective measurement of different influences of single input parameters to the network output. In fact, following the theory of ANNs, the neural box must be considered as a ‘distributed device’; this means that the output of the network is always built using the complete input pattern, and we cannot determine a particular relationship between the output and a single input parameter. From a different point of view, the test of Table 1 demonstrates how the neural device described in this paper is suitable to estimate – with a reasonable error (about 10 m^{-3}) – the spore concentrations also when the input pattern is noisy; for example, because one of the meteorological sensors is broken.

We also carried out a further analysis to assess if the neural model performs in a different way simulating the growth and the depletion of fungal spores. To check this point, we selected two subsets of the concentration time series: for pre- and post-peak occurrences, respectively, and we recalculated the statistical indices for these subsets. The selection was done as follows: we located all of the peaks of the time series and then considered the four points (days) before each peak as defining the *growth subset* and the four points following the peak as defining the *depletion subset*. In order to filter the high-frequency variability, we also considered only those peaks above a specified threshold (20 m^{-3}). As could be

Table 2 The same statistical indices as in Table 1 but obtained by randomizing a single input parameter

	<i>Alternaria</i>			<i>Pleospora</i>		
	Bias 0.2	RMSE 4.1	Correlation 91.9	Bias 0.2	RMSE 1.9	Correlation 93.4
Random T_{av}	3.0	12.5	15.3	−2.4	10.3	42.1
Random T_{min}	2.6	11.9	24.2	−1.8	9.1	35.3
Random T_{max}	−1.5	12.2	64.8	−1.6	8.0	48.5
Random RHum	0.5	10.0	54.4	−0.7	7.8	35.3
Random prec.	4.2	10.1	48.6	−12.2	21.8	1.1
Random wind	−1.0	11.2	48.2	−2.0	9.9	31.0

expected, for both partial time series, statistical parameters are quite similar to those obtained for the whole measurement set, with the differences being less than few percentage points. It should be noted here that the a *neural device* is able to map a class of relationship that is larger than the class mapped by the analytical functions (see Kolmogorov theorem in Hecht-Nielsen 1991). We can therefore imagine, as an example, that the ANN presented in this paper implements two mutually exclusive functions, the first mapping the growth of the concentration of the spores and the other one mapping the depletion of the spores. This explains why a neural approach usually does not present the type of asymmetry that is often observed using any kind of statistical regression.

5 Conclusion

A complex recurrent ANN has been described which has been calibrated to estimate the concentration of two different spore taxa from meteorological parameters. The statistical analysis of ANN performances shows the ability of the proposed technique to provide a good estimation of *Alternaria* spore concentration, with an average error of about 4 m^{-3} , and an even better performance of the estimation of *Pleospora* spore concentration. To actually realize this objective, at least 1 more year of observed data would be needed. Nevertheless, the excellent results obtained, especially in terms of localizing the major peaks of spore concentrations, demonstrate the potential of this the model with a longer time series. In addition to using the numerical model for a deeper investigation of the relationship between spore growth and depletion, there is the additional possibility of using the proposed model for predicting spore growth from meteorological forecasts. This application could be of great importance for both human health and agriculture activities. In terms of the former, the possibility of predicting spore growth from meteorological forecasts could be

of great usefulness to physicians, who would then be able to establish preventive appropriate strategies for allergy sufferers. An accurate estimation of spore concentrations over wide geographical areas could also facilitate local authorities in making their choice of the appropriate fungicide treatments during agricultural activities, thereby avoiding unsafe and useless pollution of the atmosphere, crops and fields.

References

- Abdel Hameed, A. A. (2005). Vegetation: A source of air fungal bio-contaminant. *Aerobiologia*, 21, 53–61.
- Coppola, E., Grimes, D. I. F., Verdecchia, M., & Visconti, G. (2006). Validation of improved TAMANN neural network for operational satellite-derived rainfall estimation in Africa. *Journal of Applied Meteorology*, 45, 1557–1572.
- Elman, L. J. (1990). Finding structure in time. *Cognitive Science*, 14, 179–211.
- Grimes, D. I. F., Coppola, E., Verdecchia, M., & Visconti, G. (2003). A neural network approach to real-time rainfall estimation for Africa using Satellite data. *Journal of Hydrometeorology*, 4, 1119–1133.
- Hecht-Nielsen, R. (1991). *Neurocomputing*. Reading, Mass.: Addison Wesley.
- Hirst, J. M. (1952). An automatic volumetric spore trap. *Annals of Applied Biology*, 39, 257–265.
- Hjelmroos, M. (1993). Relationship between airborne fungal spore presence and weather variables. *Grana*, 32, 40–47.
- Kurkela, T. (1997). The number of *Cladosporium* conidia and the air in different weather conditions. *Grana*, 36, 54–61.
- Lonbladd, L., Peterson, C., & R ngvaldsson, T. (1992). Pattern recognition in High Energy Physics with Artificial Neural Network – Jetnet 2.0. *Computer Physics Communications*, 167–182.
- Lyon, F., Kramer, C. L., & Eversmeyer, M. G. (1984). Variation of airspora in the atmosphere due to weather condition. *Grana*, 23, 177–181.
- Mandrioli, P., Comtois, P., & Levizzani, V. (1998). *Methods in aerobiology* (pp. 79–82). Bologna: Pitagora Editrice.
- Tariq, S. M., Matthews, S. M., Stevens, M., & Hakim, E. A. (1996). Sensitization to *Alternaria* and *Cladosporium* by the age of 4 years. *Clinical and Experimental Allergy*, 26, 794–798.
- Troutt, C., & Levetin, E. (2001). Correlation of spring spore concentrations and meteorological conditions in Tulsa, Oklahoma. *International Journal of Biometeorology*, 45, 64–74.