Turkish Journal of Botany

http://journals.tubitak.gov.tr/botany/

Review Article

Turk J Bot
(2019) 43: 703-709
© TÜBİTAK
doi:10.3906/bot-1909-35

Artificial neural models of concentrations of fungal spores in the air for aerobiological research

Agnieszka GRINN-GOFRON*
Institute of Botany, University of Szczecin, Szczecin, Poland

Received: 23.09.2019 • **Accepted/Published Online:** 06.11.2019 • **Final Version:** 21.12.2019

Abstract: This article describes in detail an advanced statistical method, the artificial neural network, and the possibilities for its application in aerobiological analyses. The study and models involve the concentration of fungal spores in the air and their relationship with various biological and environmental factors. The author hopes that this work will contribute to a wider use of this method not only in the study of spores but also the concentration of pollen grains.

Key words: Aerobiology, airborne fungal spores, artificial neural models, modeling, statistical techniques

1. Introduction

Fungi are one of the most important groups of organisms on the planet. This is easy to overlook, given their largely hidden, unseen actions and growth. Fungi are found almost everywhere in very large numbers and inhabit most of the water and land ecosystems in the world. Together with bacteria, fungi are responsible for breaking down organic matter and releasing chemical elements (among others: carbon, oxygen, nitrogen, and phosphorus) into the soil and the atmosphere. Fungi are essential in many household and industrial processes, such as the making of bread, beer, wine, and some types of cheeses. In systematics mushrooms are classified as a separate kingdom, while their participation in the construction and functioning of all ecosystems is indisputable. In every ecosystem on earth parasitic and saprophytic fungi exist, having a large impact on the plants and animals forming those ecosystems. For asexual reproduction, fungi produce huge amounts of spores, releasing them into the atmosphere. Fungal spores are found all over the world in most climate zones and ecosystem types. In addition, they can move very long distances, even crossing the oceans, by using air currents. Spores are found in the highest layers of the troposphere and even in outer space. Their ability to survive makes them almost perfect objects in many types of research. They may live as saprophytes, parasites, or symbionts of animals and plants in indoor as well as outdoor environments. Outdoor spore concentrations range from 230 to 10^6 spores/m^3 (Lacey, 1980; D'Amato and Spieksma, 1995). The atmospheric fungal spore concentration exceeds mean pollen concentration by 100–1000 times (Burge, 1988). The spore concentration in the air varies substantially depending on climatic factors such as air temperature, wind speed and direction, and moisture (in terms of relative humidity and precipitation). The majority of fungal species grow in outdoor environments. Examples are *Alternaria*, *Cladosporium*, *Epicoccum*, and *Ganoderma* (Levetin and Horner, 2002).

1.1. How can we consider the presence of spores in the air from an aerobiological point of view?

The three most important aspects of research on spores in aerobiology are their allergenic and pathogenic properties and their presence in the air as pollution.

1.1.1. As allergens

Research on changes in the concentrations of fungal spores causing allergies is carried out all over the world. Modern epidemiological studies from various countries indicate that currently 15%–20% of the average population suffers from allergic diseases (https://www.worldallergy.org/adrc/). In contrast to pollen, fungal spores and/or mycelial cells may cause many allergic symptoms and diseases, including allergic bronchopulmonary mycoses, allergic sinusitis, hypersensitivity pneumonitis, and atopic dermatitis. Sensitization to molds has been reported in up to 80% of asthmatic patients. Out of over 100,000 fungal species reported, more than 80 mold genera have been shown to induce type I allergies in susceptible persons, whereas allergenic proteins have been identified in 23 fungal genera (Simon-Nobbe et al., 2008). Most people

* Correspondence: agofr@univ.szczecin.pl

hypersensitive to this group of allergens present year-round symptoms, with periods of seasonal exacerbations. In contrast to other allergenic sources, fungi are very common in the environment; hence, exposure to airborne spores is almost constant throughout the year. A major difference to other sources, e.g., house dust mites or pollen, is that fungi may colonize the human body, and they may damage the airways by the production of toxins, proteases, and enzymes (Kauffman et al., 2000) and volatile organic compounds (Fischer et al., 1999). Thus, molds have a far greater impact on patients' immune systems than pollen or other allergenic sources. It is estimated that sensitization to pollen and fungal allergens impacts a growing number of people. The necessity of such research results from the observed sudden increase in the number of allergies to fungal spores in the last decades. This phenomenon is particularly intense in industrial areas and in large cities (Kurup, 2003).

1.1.2. As biopollutants

Fungal spores in some countries, including Canada, have been recognized as biopollutants and special standards for their acceptable levels in the air are being developed. Accordingly, it must be remembered that a significant portion of atmospheric aerosol is of biological origin. The consideration of fungal spores as biopollutants is closely related to their allergenic properties and to their relationships with sources of air pollution like nitrogen dioxide (NO_2) and trioxide (NO_3), sulfur dioxide (SO_2), ozone (O_3), and particulate matter (PM). Therefore, in aerobiological analyses and statistical modeling, these two aspects are most often combined. Because fungal spores are an important component of bioaerosols and are also considered to act as indicators of the level of atmospheric biopollution, better understanding of these phenomena demands a detailed survey of airborne particles (Ščevková, 2019).

1.1.3. As pathogens

The third important point about fungi is that they are plant pathogens. Many of them cause damage to crops with serious economic effects. The losses that they can cause in various types of crops are so significant that preventive actions have become a very important issue. Aerial dispersal of spores over short or long distances affects the epidemiology of many fungal plant pathogens and long-distance dispersal is an important strategy for a number of pathogens, which may lead to invasions in new areas or the spread of aggressive pathogens on a global scale (Brown and Hovmøller, 2002). The most important parameter in the infection process is dispersal. Dispersal mechanisms can be grouped into two types: one is passive dispersal by wind, water, or animals (Inglod, 1953) and the other is active dispersal, such as by shooting ascospores through the boundary layer of air surrounding the fruiting body by forcible discharge (Trail, 2007). Similarly, spores can be grouped into two types according to their motility. In fungi nonmotile spores include sexual spores such as ascospores, rust urediniospores, sclerotia, and conidiospores, while nonmotile oomycete spores include oospores, sporangiospores, and conidia. Motile spores with flagella, called zoospores, are ubiquitous among oomycetes and are also found in chytrid fungi (James et. al., 2006). Pathogens have evolved several mechanisms that include structural and/or enzymatic components in order to enter their plant hosts (Agrios, 2005). Many fungi develop appressoria to directly penetrate plant cuticles (Hardham, 2001). Aerobiological monitoring fits perfectly with many types of preventive measures. If high levels of spores of a particular pathogen appear in the air, farmers can be quickly informed and implement protective procedures. Such information may appear, for example, in everyday online information for farmers, or they may be informed by e-mail or SMS.

In all three of the aspects mentioned above, monitoring and statistical modeling, and in particular the creation of forecasting models, are among the most important challenges for statisticians today.

The important issue of aerobiological studies is to find correlations among the characteristics of the pollen or spore season, weather variables, bioclimatic conditions, ecological properties, and selected environmental factors. It is important to analyze all factors so that the resulting models can accurately describe the complex dependencies that occur in nature. Until now, only a few forecasting models for selected genera of spores have been developed. Most of them are characterized by relatively low verifiability, and they do not give the exact values of the weather factors that are responsible for causing the threshold concentrations or specify which of them is the most significant.

2. Modeling

In statistical modeling of fungal spores, the most common input variables are the concentrations of spores in the air and environmental factors such as meteorological, ecological, and biogeographical parameters. Output variables are numbers that determine the extent to which environmental variables affect spore concentrations. Modeling of the concentration of airborne particles is a relatively difficult issue. Due to the complexity of the study object (a large number of analyzed variables, very irregular changes in the concentrations of airborne pollen or fungal spores of a large variety of species, nonlinear correlations between parameters), multidimensional techniques and other advanced statistical methods of exploring data are preferred. Linear models are a commonly used mathematical technique for describing various objects and processes. However, there is no basis for using linear approximation for a given problem and linear models do not work, leading to conclusions being drawn too quickly about the “inability” of the mathematical description of the system.

2.1. Artificial neural networks

2.1.1. What are neural networks?

Artificial neural network (ANN) models are a very sophisticated modeling technique, capable of mapping extremely complex functions. They are nonlinear, which significantly enriches their applicability, and they are among the many combined techniques currently being developed. Utilizing models created using neural networks may be the fastest and most convenient solution to a problem. ANNs also allow control over the complex problem of multidimensionality, which with other approaches significantly impedes attempts to model nonlinear functions with a large number of variables. In practice, neural networks construct the models needed by the user themselves, because they automatically learn from the examples given users. The level of theoretical knowledge necessary to successfully build a model is much lower when using neural networks than in the case of the use of traditional statistical methods. ANNs arose as a result of research conducted in the field of artificial intelligence (building models of basic structures found in the brain). Research carried out in the field of symbolic artificial intelligence in 1960–1980 led to the creation of so-called expert systems. Characteristics of biological nervous systems that were particularly technically useful were the resistance of biological systems to the damage of even a significant part of their elements and their extraordinary ability to learn. The most important advantages of this method are the ability to work with incomplete information, no requirement to know the solution algorithm (automatic learning), and processing of information in a highly parallel manner. Additionally, they can generalize to unknown cases, they are resistant to partial damage, and they can implement associative memory, similar to how human memory works, as opposed to addressed memory, characteristic of classic computers (Osowski, 1996). In addition to aerobiology, these methods are used in many other fields of research, e.g., forecasting pollution (Feng et al., 2015), ground water levels (Daliakopoulus et al., 2005), wind speed (Li and Shi, 2010), and tourism time series (Palmer et al., 2006).

2.1.2. Network structure and operation

This method can potentially be used wherever problems occur with data processing and analysis, prediction, classification, or control. The construction of an ANN is based on the structure and operation of neurons in the human brain. Thus, the question is: how do ANNs learn? In Pavlov’s classic experiment on conditional reflexes, in which a bell is sounded before serving a dog’s dinner, the dog learns to combine the bell’s sound with expectations for food very quickly. This is done due to specific synaptic connections being strengthened as a result of the learning process. First, a number of input signals (values) reach the neuron (e.g., N,4 in Figure 1 means that there are 4 input variables entering the network). Each value is introduced to the neuron by a connection of some strength (Figure 1; weight: W_1, W_2, W_0); these weights correspond to the synapse efficiency in a biological neuron. Each neuron also has a single threshold value, determining how strong the stimulation must be for excitation to occur. The neuron calculates the weighted sum of the inputs (i.e. the sum of the values of the input signals multiplied by the appropriate weighting factors), and then the threshold value is subtracted (Osowski, 1996; Tadeusiewicz, 2001) (Figure 1).

The value obtained in this manner determines the stimulation of the neuron. This is a very advanced approximation of real biological phenomena. The signal representing the total stimulation of the neuron is in turn transformed by the established function of neuron activation (neuron transition function) (Figure 1; f_1, f_2).

The value calculated by the activation function is ultimately the output value (output signal) of the neuron. A neural network must have inputs to have value in use (used to derive the values of variables observed on the outside) as well as output (which means the result of the calculation, e.g., N,3 in Figure 1 means that there are 3 output variables coming out of the network).

Inputs and outputs in the brain correspond to selected nerves: sensory for input and motoric for output. The hidden neurons are important in building the network because they perform internal functions in the network, mediate in the analysis of information provided by sensory nerves, and take part in the processing of sensory signals into decisions activating specific executive elements. An external observer cannot access them. Input, hidden, and output neurons must remain interconnected, posing a problem for the network creator in

choosing its structure. The key issue in choosing a network structure is the presence or absence of feedback in this structure (Osowski, 1996; Tadeusiewicz, 2001).

Simple networks have a one-way structure (feedforward): the signal flows in them only in one direction, from the inputs through subsequent hidden neurons, eventually reaching the output neurons. Such a structure is always characterized by stable behavior, which is its advantage. The network may also have feedback that contains return connections from later to earlier neurons and can do more complicated calculations. This is not “painless”; due to circulation signals in feedback networks, from input to output and via feedback back to the input,

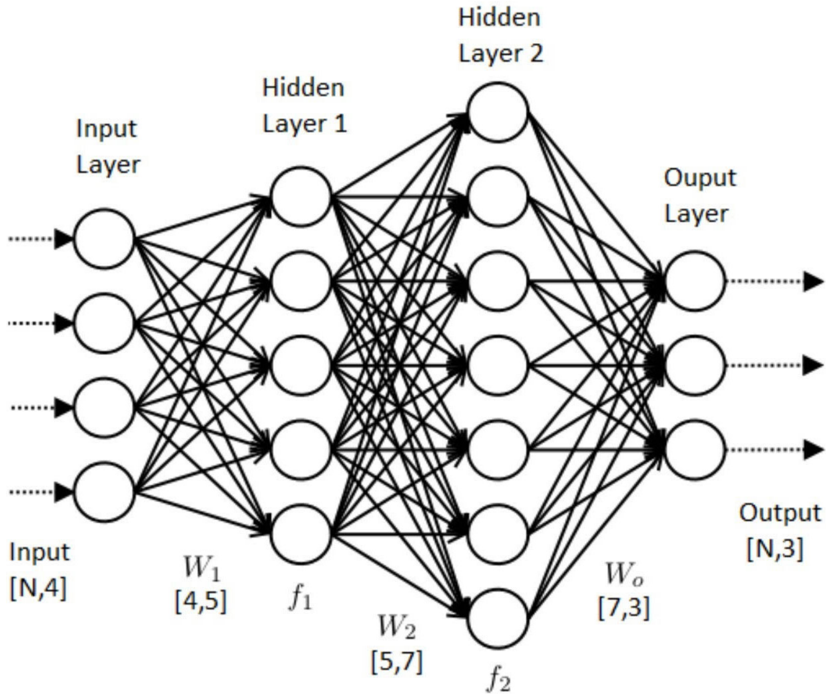

Figure 1. Scheme of building an artificial neural network.

it may behave unstably and have very complex dynamics in which the most complex forms of behavior can be expected, e.g., deterministic chaos.

Quite a lot of practical utility is obtained in networks with relatively many feedback loops, specifically in networks in which all connections are feedback. Such networks are known as Hopfield networks (Osowski, 1996; Tadeusiewicz, 2001).

In addition to the results illustrated in the figures in this article, based on the data obtained in the process of forecasting, they can also be included in tables. In addition to the significance of the impact of a given parameter on an input variable, a ranking of the most important to least significant variables may also be created.

2.1.3. How does a neural network work?

In a typical case, from a mathematical point of view, a single neuron performs the operations of the product of the scalar vector of input signals and the weight vector. As a result, the neuron response depends on the geometrical relationships between signal vectors and weight vectors. The correct geometry of the position of the weight vectors, which guarantees correct operation, is obtained as a result of the learning process (a method of automatic search for such a set of weight coefficients occurring in all neurons of the entire network that guarantees the lowest value of the total error made by the network). As a result of using the appropriate learning algorithm (the best known here are backpropagation algorithm errors), the network can systematically reduce the error during the learning process; as a result, a gradual improvement in its performance during learning is observed. The network reacts as it is commanded by its current knowledge, i.e. some conditions are approved while others are not. The "teacher" (i.e. the computer conducting the training), having a map of the desired behavior of the network, gives it a reference signal.

After completing a set number of steps, the learning process is interrupted and the network is tested. During this test, grades for all possible points must be provided. In each of these applications, the neural network plays the role of a universal approximator of multivariable functions, implementing a nonlinear function of the form y = f (x), where x is the input vector and y is the vector function of many variables. The network learns basic features such as the geometric mapping of the pixel pattern, the distribution of the main components of the pattern, Fourier transformation components, and other properties.

The learning emphasizes the differences occurring in different patterns, which are the basis for making the decision to assign them to the appropriate class. Prediction is the task of the network to determine future system responses based on a string of values from the past.

Having information about the values of the variable x at the times preceding the predictions x (k-1), x (k-2), ..., x (kN), the network decides what the estimated value of x (k) of the tested string at current moment k will be (Osowski, 1996; Tadeusiewicz, 2001).

Adaptation of network weights uses the current prediction error and the value of this error in the preceding moments. The network is a nonlinear model of control and dynamic processes, allowing the development of an appropriate control signal. It also acts as a tracking and following system, adapting to environmental conditions.

At the end, the ANN allows for two criteria in choosing a neural model. The SD ratio is the ratio of the standard error deviation to the standard deviation of the experimental data (good model: <0.7), while correlation refers to Pearson's linear correlation coefficient between data calculated by the network and experimental data (good model: >0.7) (Osowski, 1996; Tadeusiewicz, 2001).

In summary, the most important feature of ANNs, which determines their huge advantages and wide application possibilities, is the parallel processing of information by all neurons. With the mass scale of neural connections, this results in a significant acceleration of information processing. In many cases real-time signal processing is possible. The second equally important feature is the ability to learn and generalize the acquired knowledge.

The network has the so-called property of artificial intelligence. Trained on a limited group of learning data, it can associate acquired knowledge and demonstrate expected operations on data not participating in the learning process (Osowski, 1996; Tadeusiewicz, 2001) (Figures 2 and 3).

Figure 2 shows how experimental data—in this case, daily *Ganoderma* concentrations and selected meteorological parameters of temperature (mean, min, max, and dew point), relative humidity, wind speed, and precipitation—coincide with the results processed by the model. In other words, it reveals the quality of the model. Figure 3 shows the influence of the most important parameter, in this case the dew point of temperature, on the concentration of *Ganoderma* spores. The obtained results reveal that the presence or absence of *Ganoderma* spores depends mostly on dew point temperature, with a threshold value of about 9 °C. Similar results were gained for *Alternaria* and *Cladosporium*. Models created by ANNs for the most important parameters affecting the concentrations of both types indicated 4 temperature parameters and then relative humidity, wind speed, and precipitation. This is understandable in this case, because the concentrations of both types of spores occur in the air in high amounts in similar periods of time.

Models using ANNs with spore concentrations and environmental parameters have been described in several aerobiological articles (Grinn-Gofroń and Strzelczak, 2008a, 2008b, 2009, 2011, 2013; Jedryczka et al., 2015).

3. Conclusion

As a general conclusion it may be stated that aerobiology is at the beginning of a long road to achieving a method of analyzing aerobiological data that allows the creation of comprehensive models with high quality, accuracy, and verifiability.

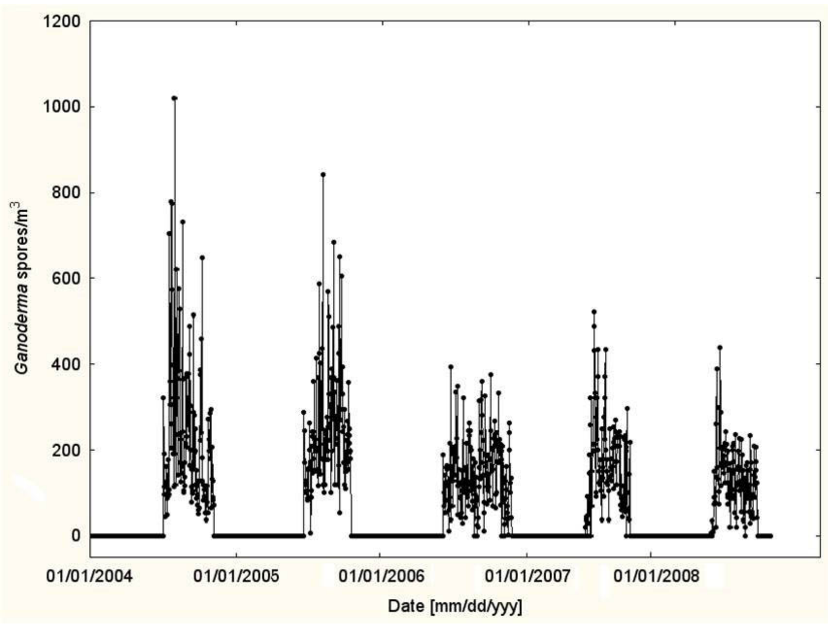

Figure 2. *Ganoderma* sp. spore concentration in the years 2004–2008 (Szczecin, Poland); 8 comparisons of experimental and model's calculated data (Grinn-Gofroń and Strzelczak, 2011).

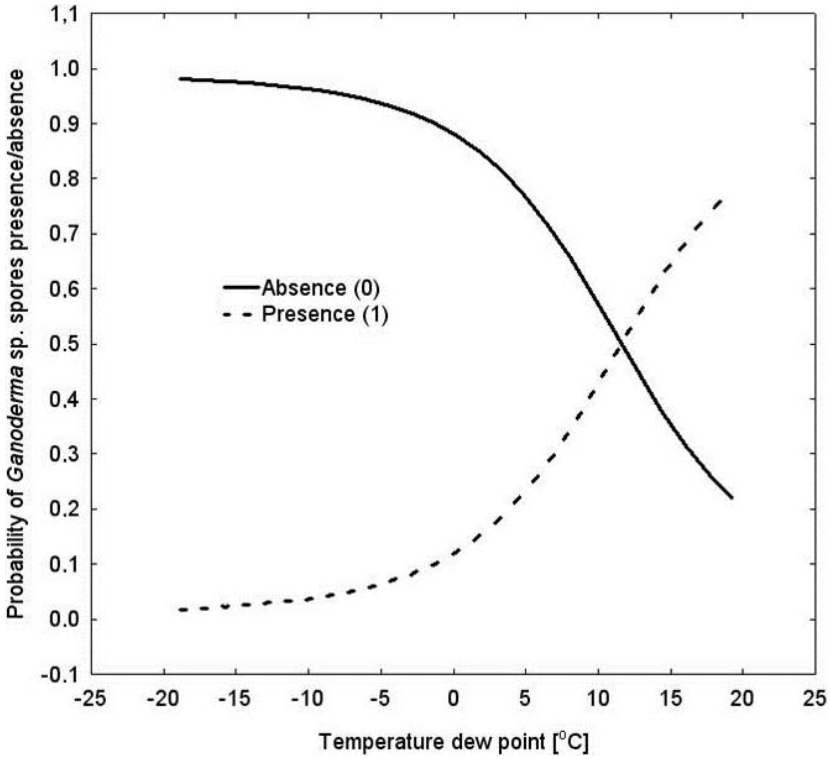

Figure 3. Response plot for *Ganoderma* sp. spore concentration dependent on dew point temperature from multilayer perceptron 8:8-11-1:1 classification neural model (Grinn-Gofroń and Strzelczak, 2011).

References

Agrios GN (2005). Plant Pathology. 5th ed. London, UK: Elsevier Academic Press.

Brown JK, Hovmøller MS (2002). Aerial dispersal of pathogens on the global and continental scales and its impact on plant disease. Science 297 (5581): 537-541.

Burge HA (1989). Airborne allergenic fungi classification, nomenclature, and distribution. Immunology and Allergy Clinics of North America 9: 307-319.

Daliakopoulus IN, Coulibaly P, Tsanis IK (2005). Groundwater level forecasting using artificial neural networks. Journal of Hydrology 309: 229-240.

D'Amato G, Spieksma FT (1995). Aerobiologic and clinical aspects of mould allergy in Europe. Allergy 50: 870-877.

Feng X, Li Q, Zhu Y, Hou J, Jin L et al. (2015). Artificial neural networks forecasting of $PM_{2.5}$ pollution using air mass trajectory based geographic model and wavelet transformation. Atmospheric Environment 107: 118-128.

Fischer G, Schwalbe R, Moller M, Ostrowski R, Dott W (1999). Species-specific production of microbial volatile organic compounds (MVOC) by airborne fungi from a compost facility. Chemosphere 39: 795-810.

Grinn-Gofroń A, Strzelczak A (2008a). Artificial neural network models of relationships between *Alternaria* spores and meteorological factors in Szczecin (Poland). International Journal of Biometeorology 52: 859-868.

Grinn-Gofroń A, Strzelczak A (2008b). Artificial neural network models of relationships between *Cladosporium* spores and meteorological factors in Szczecin (Poland). Grana 47: 304-314.

Grinn-Gofroń A, Strzelczak A (2009). Hourly predictive artificial neural network and multivariate regression tree models of *Alternaria* and *Cladosporium* spore concentrations in Szczecin (Poland). International Journal of Biometeorology 53: 555-562.

Grinn-Gofroń A, Strzelczak A (2011). The effects of meteorological factors on the occurrence of *Ganoderma* sp. spores in the air. International Journal of Biometeorology 55: 235-241.

Grinn-Gofroń A, Strzelczak A (2013). Changes in concentration of *Alternaria* and *Cladosporium* spores during summer storms. International Journal of Biometeorology 57 (5): 759-776.

Grinn-Gofroń A, Strzelczak A, Wolski T (2011). The relationships between air pollutants, meteorological parameters and concentration of airborne fungal spores. Environmental Pollution 159: 602-608.

Hardham AR (2001). The cell biology behind *Phytophthora* pathogenicity. Australasian Plant Pathology 30: 91-98.

Ingold GT (1953). Dispersal in Fungi. New York, NY, USA: Clarendon Press.

James TY, Letcher PM, Longcore JE, Mozley-Standridge SE, Porter D et al. (2006). A molecular phylogeny of the flagellated fungi (Chytridiomycota) and description of a new phylum (Blastocladiomycota). Mycologia 98 (6): 860-871.

Jedryczka M, Strzelczak A, Grinn-Gofron A, Nowak M, Wolski T et al. (2015). Advanced statistical models commonly applied in aerobiology cannot accurately predict the exposure of people to *Ganoderma* spore-related allergies. Agricultural and Forest Meteorology 201: 209-217.

Kauffman HF, Tomee JF, van de Riet MA, Timmerman AJ, Borger P (2000). Protease-dependent activation of epithelial cells by fungal allergens leads to morphologic changes and cytokine production. Journal of Allergy and Clinical Immunology 105: 1185-1193.

Kurup VP (2003). Fungal allergens. Current Allergy and Asthma Reports 3: 416.

Lacey L (1981). The aerobiology of conidial fungi. In: Cole GT, Kendrick B (editors). Biology of Conidial Fungi. New York, NY, USA: Academic Press, pp. 123-128.

Levetin E, Horner WE (2002). Fungal aerobiology: exposure and measurement. In: Breitenbach M, Crameri R, Lehrer SB (editors). Fungal Allergy and Pathogenicity. Chemical Immunology, Vol. 81. Basel, Switzerland: Karger, pp. 10-27.

Li G, Shi J (2010). On comparing three artificial neural networks for wind speed forecasting. Applied Energy 87 (7): 2313-2320.

Osowski S (1996). Algorithmic Approach to Artificial Neural Networks. Warsaw, Poland: WNT (in Polish).

Palmer A, Monatonó JJ, Sesé A (2006). Designing an artificial neural network for forecasting tourism time series. Tourism Management 27 (5): 781-790.

Ščevková J, Hrabovsky M, Kováč J, Rosa S (2019). Intradiurnal variation of predominant airborne fungal spore biopollutants in the Central European urban environment. Environmental Science and Pollution Research (in press). doi: 10.1007/s11356-019-06616-7

Simon-Nobbe B, Denk U, Pöll V, Rid R, Breitenbach M (2008). The spectrum of fungal allergy. International Archives of Allergy and Immunology 145: 58-86.

Tadeusiewicz R (2001). Introduction to Neural Networks. Krakow, Poland: Statsoft Polska (in Polish).

Trail F (2007). Fungal cannons: explosive spore discharge in the Ascomycota. FEMS Microbiology Letters 276: 12-18.