Science of the Total Environment 476–477 (2014) 542–552

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Predicting daily ragweed pollen concentrations using Computational Intelligence techniques over two heavily polluted areas in Europe

Zoltán Csépe [a], László Makra [a,*], Dimitris Voukantsis [b], István Matyasovszky [c], Gábor Tusnády [d], Kostas Karatzas [b], Michel Thibaudon [e]

[a] Department of Climatology and Landscape Ecology, University of Szeged, HU-6701 Szeged, P.O.B. 653, Hungary
[b] Department of Mechanical Engineering, Informatics Systems & Applications Group, Aristotle University, P.O. Box 483, GR-54124 Thessaloniki, Greece
[c] Department of Meteorology, Eötvös Loránd University, HU-1117 Budapest, Pázmány Péter st. 1/A, Hungary
[d] Mathematical Institute of the Hungarian Academy of Sciences, HU-1364 Budapest, P.O.B. 127, Hungary
[e] RNSA (Aerobiology Network of France), La Parličre, F-69610 Saint Genis l'Argentière, France

HIGHLIGHTS

- For Szeged, MLP and tree-based models, while for Lyon only MLP performs well for predicting pollen concentration.
- When predicting alarm levels, the performance of MLP is the best for both cities.
- When forecasting high pollen episodes, the more complex CI methods prove better for both cities.
- The selection of the optimal method depends on climate, as a function of geographical location and relief.

ARTICLE INFO

Article history:
Received 5 December 2013
Received in revised form 14 January 2014
Accepted 15 January 2014
Available online 1 February 2014

Keywords:
Ragweed pollen
Allergy
Forecasting
Neural networks
Multi-Layer Perceptron
Tree based methods

ABSTRACT

Forecasting ragweed pollen concentration is a useful tool for sensitive people in order to prepare in time for high pollen episodes. The aim of the study is to use methods of Computational Intelligence (CI) (Multi-Layer Perceptron, M5P, REPTree, DecisionStump and MLPRegressor) for predicting daily values of *Ambrosia* pollen concentrations and alarm levels for 1–7 days ahead for Szeged (Hungary) and Lyon (France), respectively. Ten-year daily mean ragweed pollen data (within 1997–2006) are considered for both cities. 10 input variables are used in the models including pollen level or alarm level on the given day, furthermore the serial number of the given day of the year within the pollen season and altogether 8 meteorological variables. The study has novelties as (1) daily alarm thresholds are firstly predicted in the aerobiological literature; (2) data-driven modelling methods including neural networks have never been used in forecasting daily *Ambrosia* pollen concentration; (3) algorithm J48 has never been used in palynological forecasts; (4) we apply a rarely used technique, namely factor analysis with special transformation, to detect the importance of the influencing variables in defining the pollen levels for 1–7 days ahead. When predicting pollen concentrations, for Szeged Multi-Layer Perceptron models deliver similar results with tree-based models 1 and 2 days ahead; while for Lyon only Multi-Layer Perceptron provides acceptable result. When predicting alarm levels, the performance of Multi-Layer Perceptron is the best for both cities. It is presented that the selection of the optimal method depends on climate, as a function of geographical location and relief. The results show that the more complex CI methods perform well, and their performance is case-specific for ≥2 days forecasting horizon. A determination coefficient of 0.98 (*Ambrosia*, Szeged, one day and two days ahead) using Multi-Layer Perceptron ranks this model the best one in the literature.

1. Introduction

Warming of the climate system is obvious, as it is now evident from observations of increases in global average air and ocean temperatures, widespread melting of snow and ice, and rising global average sea level (IPCC, 2013). Recent climate warming is associated with the modification of the distribution areas of plants producing allergenic pollen (Laaidi et al., 2011; Ziska et al., 2011), furthermore, with an earlier

* Corresponding author at: Department of Climatology and Landscape Ecology, University of Szeged, Hungary, PO Box 653, HU-6701 Szeged, Hungary. Tel.: +36 62 544856; fax: +36 62 544624.
E-mail addresses: csepzol@geo.u-szeged.hu (Z. Csépe), makra@geo.u-szeged.hu (L. Makra), voukas@isag.meng.auth.gr (D. Voukantsis), matya@ludens.elte.hu (I. Matyasovszky), tusnady.gabor@renyi.mta.hu (G. Tusnády), kkara@eng.auth.gr (K. Karatzas), michel.thibaudon@wanadoo.fr (M. Thibaudon).

0048-9697/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.scitotenv.2014.01.056

onset (Frei, 2008; Rodríguez-Rajo et al., 2011), an earlier end dates (Stach et al., 2007; Recio et al., 2010), a longer pollen season (Stach et al., 2007; Ariano et al., 2010), an increase in the total annual pollen load (Cristofori et al., 2010; Ariano et al., 2010; Laaidi et al., 2011), as well as an increase of patient number sensitised to pollen throughout the year (Ariano et al., 2010).

The genus of ragweed (*Ambrosia* spp.) comprises 42 species. They are the best known weeds for the most severe and widespread allergies caused by its pollen (Béres et al., 2005). However, in Europe, common ragweed (*Ambrosia artemisiifolia*) is predominant of all *Ambrosia* species (Makra et al., 2005; Bullock et al., 2010; Vinogradova et al., 2010). The most important habitat areas of common ragweed in Europe are the Rhône valley in France (Chauvel et al., 2006; Gladieux et al., 2011), north-western Milan and south Varese (Lombardy, Po River valley) in Italy (Bonini et al., 2012), the Pannonian Plain including Hungary and some parts of Serbia, Croatia, Slovenia, Slovakia and Romania (Kiss and Béres, 2006; Makra et al., 2005), furthermore Ukraine (Rodinkova et al., 2012) and the south-western part of the European Russia (Reznik, 2009).

Several evidences of the association between ragweed pollen counts and ragweed induced pollen allergy have been confirmed in the literature. Based on clinical investigations, pollen of ragweed (*Ambrosia* spp.) is the most important cause of allergy-associated respiratory diseases (Kadocsa and Juhász, 2002). Harf and Déchamp (2001) found a steep rise in anti-allergic drug sales (eye drops, nasal spray, oral antihistamines) in July, August and September over an area of high infestation in France. In the Pannonian Plain, about 30% of the Hungarian population has some type of allergy, 65% of them have pollen-sensitivity, and at least 60% of this pollen-sensitivity is caused by *Ambrosia* (Járai-Komlódi, 1998; Makra et al., 2004). Furthermore, in Szeged, 83.7% of the patients were sensitive to *Ambrosia* in 1998–1999 (Kadocsa and Juhász, 2000). In addition, due to the recent climate change (D'Amato and Cecchi, 2008; Ziska and Beggs, 2012), pollen counts of *Ambrosia* show a slight increase according to linear trends as moderate warming is favourable for warm-tolerant *Ambrosia* (Makra et al., 2011a).

Common ragweed and its pollen cause serious losses in the economy and several fields of everyday life. The current costs of *A. artemisiifolia* in terms of human health and agriculture were estimated by Bullock et al. (2010) for 40 European countries. All the costs are given in Euros at 2011 prices. The human health impacts were estimated to affect around 4 million people with total estimated medical costs of €2136 million per year. Furthermore, total estimated workforce productivity losses due to *A. artemisiifolia* as high estimates were €529 million. The estimated total costs are valued at €2.665 billion per year (Bullock et al., 2010).

The above-mentioned facts make unavoidable producing ragweed pollen concentration forecasts in order to help sensitised people prepare for days of severe airborne pollen load. Different techniques have been applied for modelling daily *Ambrosia* pollen concentrations. Makra et al. (2011b) developed time-varying nonparametric regression methods that combine regression analysis with the method of summing temperatures (Laaidi et al., 2003). Furthermore Makra and Matyasovszky (2011) introduced time-varying parametric linear and time-varying nonparametric regression models, as well as a time-varying nonparametric median regression model to predict the daily pollen concentration for Szeged in Hungary using previous-day meteorological parameters and the daily pollen concentration. The models were applied to rainy days and non-rainy days, respectively. Matyasovszky and Makra (2011) used a time-varying first order autoregressive [AR(1)] model to describe daily ragweed pollen levels based on previous-day pollen concentration values and previous-day meteorological variables. Laaidi et al. (2003) used two forecasting models, namely (1) summing the temperatures and (2) a multiple regression to forecast pollen season characteristics. Some further papers using multiple regression analysis for modelling daily pollen concentration of different taxa include Angosto et al. (2005), Ribeiro et al. (2008), Stach et al. (2008), Rodríguez-Rajo et al. (2009) and Myszkowska (2013). Furthermore, selection of a suitable statistical clustering method may help in improving, among other things, the accuracy of the ratio of the transported pollen by long-range air currents in the measured pollen concentration over a target area (Kassomenos et al., 2010).

More advanced techniques such as neural networks, Multi-Layer Perceptron and the support vector regression learning methods have also been used for forecasting air quality parameters (Kassomenos et al., 2006; Juhos et al., 2009; Paschalidou et al., 2011; Vlachogianni et al., 2011; Voukantsis et al., 2011; Kassomenos et al., 2013). However, methods of Computational Intelligence (CI) have only been scarcely applied in airborne pollen related studies. They were used for forecasting (a) daily pollen concentrations (Delaunay et al., 2004, cedar pollen; Aznarte et al., 2007, olive pollen; Rodríguez-Rajo et al., 2010, Poaceae pollen; Voukantsis et al., 2010, Oleaceae, Poaceae and Urticaceae pollen; Puc, 2012; *Betula* pollen), (b) pollen-induced symptoms (Voukantsis et al., 2013), (c) risk level of *Betula* pollen in the air (Castellano-Méndez et al., 2005) and (d) the severity of the Poaceae pollen season (Sánchez Mesa et al., 2005). Furthermore, Aznarte et al. (2007) used neuro-fuzzy models for forecasting olive pollen concentrations. The above-mentioned applications of neural networks and neuro-fuzzy models produced better results than traditional statistical methods (Sánchez Mesa et al., 2005).

These methods of Computational Intelligence 1) can deal with the complexity of the mechanisms concerning the release and dispersion of the airborne pollen, 2) can be applied for different tasks (e.g. optimization and forecasting), 3) are computationally efficient and can be easily integrated into operational use of the models (Voukantsis et al., 2010).

In this paper we use factor analysis with special transformation, a technique for detecting the importance of the influencing variables in defining the pollen levels for 1–7 days ahead. Furthermore, data-oriented models are applied for (1) predicting daily concentration of ragweed pollen that shows the highest allergenicity of all taxa and (2) comparing the efficiency of different prediction techniques over two heavily polluted areas in Europe, i.e. over Lyon (France) and Szeged (Hungary), respectively. The main objectives are: i) development of accurate forecasting models for operational use, ii) evaluation of CI methods that have not been previously applied for *Ambrosia* pollen, such as Multi-Layer Perceptron and regression trees and iii) obtaining a forecast of highest accuracy among CI methods based on input data of former prediction algorithms. Note that (1) data-driven modelling methods including neural networks have never been used in forecasting daily *Ambrosia* pollen concentration, (2) daily alarm thresholds are firstly predicted in the aerobiological literature; furthermore (3) algorithm J48 has never been used in palynological forecasts.

2. Materials and methods

2.1. Study area

Two European cities, namely Lyon (Rhône Valley, France) and Szeged (Pannonian Plain, Hungary) were considered as they represent heavily polluted areas with ragweed pollen in Europe.

These cities differ in their topography and climate as well as in ragweed pollen characteristics. Szeged (46.25N; 20.10E), the largest settlement in South-eastern Hungary, is located at the confluence of the rivers Tisza and Maros (Fig. 1). The area is characterised by an extensive flat landscape of the Great Hungarian Plain with an elevation of 79 m AMSL. The city is the centre of the Szeged region with 203,000 inhabitants. In the Köppen system the climate of Szeged is the Ca type (warm, temperate climate), with relatively mild and short winters and hot summers (Köppen, 1931). Lyon (45.77N; 4.83E) lies in the Rhône-Alpes of France. The city is located in the Rhône valley with an elevation of 175 m AMSL at the confluence of the Rhône and Saône rivers (Fig. 1). Lyon has the second largest metropolitan area in France, with a population of 1.8 million in the urban area,