Available at www.scientevents.com/proscience/

ScientEvents

ProScience 7 (2020) 75-83

digilabs

ProScience

Conference Proceedings

International Meeting of Geohealth Scientists - GHC 2020 - Virtual Meeting

Soft computing techniques highlighting the correlation between air pollution and man health in Benevento (Italy)

Salvatore Rampone[1], Alessio Valente[2*]

[1]*DEMM, University of Sannio, Benevento,Italy*
[2]*DST, University of Sannio, Benevento, Italy*
[*]*valente@unisannio.it*

Abstract

Every day huge quantities of pollutants coming mainly from combustion processes continue to be poured into the atmosphere. Such emission products can interact with each other, favoured by physic-chemical conditions, resulting in new pollutants (ozone). An accumulation of pollutants in the lowest part of atmosphere can be dangerous for human health. Despite the fact that the quality of the air has recently improved in Italy and throughout Europe, to thisday pollution is still recognized as one of the main environmental risk factors. Administrations are obliged to prepare an air quality plan, when the levels of pollutants exceed the assigned limits, and this helps to activate recovery measures in the urban area.

The aim of this research is to correlate trough soft computing techniques (Artificial Neural Networks - ANN and Genetic Programming - GP) the data of the tumours registered by the Local Health Authority (ASL) of the city of Benevento (Italy) with those of the concentrations of pollutants detected in the air quality monitoring stations. Such stations are equipped with instruments able to monitor the following components: NO_2, CO, PM_{10}, $PM_{2.5}$, O_3 and Benzene (C_6H_6). For this research, the data relating to pollutants are from the 2012-2014 period while, assuming possible effects on human health in the medium term, the tumour data, provided by local hospitals, refer to the years 2016-2018. The ANN result, confirmed by GP, shows a high correlation between the cases of lung tumours and the exceedance of atmospheric particulate matter and ozone.

Keywords: Air pollution; Tumour data; Soft computing techniques; Benevento; Italy

ISSN: 2283-5954
Selection and peer-review under responsibility of GeoHealth 2020 Scientific Committee
DOI:10.14644/ghc2020.012

1. Introduction

The economic, industrial and demographic development that took place over the last two centuries has allowed a definite improvement in the quality of human life. However, it has caused deep and rapid changes in the environment and its components (air, water, soil and subsoil, etc.). In particular, huge quantities of pollutants coming mainly from combustion processes (transport, domestic heating, industrial production, etc.) continue to be poured into the atmosphere (Filippelli et al., 2012). Meteorological conditions are not always effective in removing these pollutants, and therefore in the cities these substances tend to remain in the lowest part of the atmosphere in the so-called Planetary Bounded Layer (PBL). In addition, it is known that in air the emission products can interact with each other, favoured by physic-chemical conditions (i.e. sunlight), resulting in new pollutants (i.e. ozone). An accumulation of pollutants is generated in the PBL, when the dilution capacity of pollutants in the atmosphere exceed the emissive capacity (Seigneur, 2019). Thus, dangerous concentrations are reached for human health and balance of ecosystems. Despite the fact that the quality of the air has improved (a strong decrease in emissions in Italy and in Europe has been recorded in the last decades), air pollution is still recognized as one of the main environmental risk factors.

The criticality in the concentration of pollutants is revealed in urban areas, where the anthropization of the territory reaches its maximum values and where the levels of some pollutants cause concern. However, the monitoring of air quality, which in recent years has achieved greater development, has played an important role in this action of attention and mitigation. If sulphur dioxide, carbon monoxide, benzene and lead are not currently a problem, except at local level and in specific circumstances, particulate matter (PM_{10} and $PM_{2.5}$), ozone (O_3) and dioxide nitrogen (NO_2) are among those that are most alarming (Seigneur, 2019). In particular, atmospheric particulate matter, which are carrier of pollutants, and ozone, which represent a secondary pollutant, are recognized as the main culprits of human health effects. In fact, according to WHO (World Health Organization) the incidence of mortality due to exposure to particulate matters and ozone in the air is high and increasing, especially in some urban areas (WHO, 2016). More specifically, it is important to underline that the harmfulness to human health does not depend only on the concentration of the particulate matters, but also on the chemical composition and the size of the particles. For instance, those with a diameter between 5 and 10 μm reach the trachea and bronchi, while those with a diameter of less than 5 μm can penetrate to the pulmonary alveoli. Moreover, for an increase of 5 $\mu g/m^3$ of fine dust ($PM_{2.5}$) there would be a significant increase in the risk of anticipated mortality by 7%. Generally, the dust dispersed in the air obstructs the correct air flow inspiration, or can cause respiratory diseases, pulmonary or bronchial bleeding, and even cancer (Raaschou-Nielsen et al., 2013). However, it is believed that polluted air has harmful effects not only on humans, but also on animals, vegetation, materials and ecosystems as a whole (Seigneur, 2019).

The obligation for the administrations to prepare a plan for the quality of the air, when the levels exceed their assigned limits, can be a help to activate recovery measures in the urban area. However, to implement effective actions that lead to a reduction in the impact of environmental pollution, it is necessary to understand to what extent the contaminants cause negative effects on humans. The purpose of this work is, in fact, to correlate the epidemiological data of the tumours, namely lung tumours registered at the Local Health Authority (ASL) of the city of Benevento (Italy) with the data of the concentration of pollutants detected in the air quality monitoring stations of the same city. The methodology used to calculate this estimate is given by soft computing techniques, more precisely by

artificial neural networks and the genetic programming. These techniques have already been tested both to predict air conditions and to establish possible effects on environmental components and human health (Rampone & Valente, 2017; 2019).

2. Study Area

Our study area is the urban area of Benevento, located at 135 m a.s.l. in a vast basin in the inner area of Campania in southern Italy. Such basin is bordered to the north and west by reliefs, whose nearest peaks reach 1400 m and is surrounded to the east and south by hilly ridges, with altitude which generally remain below 600 m. The city develops at the confluence of the Calore and Sabato rivers. In the basin there are numerous minor watercourses as well, which, in some cases, markedly influence the local morphology.

These geomorphological features strongly affect the climate of Benevento, which can be defined as temperate based on meteorological data recorded for over seventy years. The average temperature recorded relative to Benevento is 14.9 °C. The average of the coldest month (January) is 6.9 °C, and the one for the warmest month (July) is 23.9 °C, with a temperature range of 16.9 °C and four months with temperatures above 20 °C. Overall rainfalls do not reach 800 mm and it is concentrated mainly in autumn and subordinately in winter. This quantity of rain is sharply lower than in the other inland areas of Campania. Humidity in winter is high, frequently exceeding 70%, while in summer it is significantly reduced. The reigning winds are those coming from the south-west, while those that blow with more intensity (dominant winds) are from the north.

Atmospheric stability is a frequent condition in Benevento, deriving from baric configurations, like of an anticyclone type, which often generate the phenomenon of thermal inversion. In this phenomenon, the temperature rather than decreasing with the altitude, is reversed both in the early evening, night and early morning, and less frequently throughout the day. Due to the strong nocturnal radiation, the ground cools, as well as the layer in contact with it, especially in correspondence with the orographic depressions. This phenomenon causes frosts and fog, which can persist even during the day. Such conditions would prevent vertical dispersion of any contaminants present.

In the vast municipal area of over 13,000 m^2, the possible sources of contaminants can be identified in the vehicles that cross the often congested urban centre and the peripheral area with some important roads (Appia, Telesina, etc.), as well as in the residential areas and industrial plants. The latter gather at the edge of the inhabited centre in a sort of radial pattern, which has been supplemented by smaller residential settlements and commercial areas. In the last twenty years the occupation of land has definitely grown up to almost 100 hectares, despite the fact that the population that actually lives in the Municipality of Benevento is in decline (59,031 inhabitants in 2019 with a reduction of 6.4% in 15 years). This situation would point to a heavy increase in vehicular traffic favoured by continuous travel to and from the city and therefore a possible increase in contamination. To avoid this deterioration, advantageous solutions have recently been initiated by the municipal administration (construction of parking lots, increase in green areas, etc.).

The decrease in the population of Benevento is partially due to the negative balance between births and deaths, and also to the migration flow to other municipalities. In particular, we note that the difference between births and deaths is not due only to a decrease in the birth rate, but to a significant increase in deaths. Among these deaths, those due to neoplasms are on the rise (cancer register period 2010-2015), and 24% of these are related to the lung. Obviously, these cases are not all attributable to air pollution problems, but there may be a certain correlation.

3. Data description

In order to develop this research, data from the air quality control system, established by European Decisions and adopted by National Ministerial Laws, were used. This legislative framework involves not only public administrations, but also environmental agencies and public research bodies. This consistent involvement aims a high homogeneity and comparability in the assessment and management of air quality in the national territory. Therefore, the control systems not only provide information to verify compliance with the regulatory limits, but also highlight and inform about the general state of the air quality of a territory. In this case, this research has considered the measurement stations present in the urban area of Benevento belonging to the Monitoring Network of Campania and managed by the Regional Agency of Environment Protection (ARPAC). As the regulations underline, they are mainly influenced by emissions from nearby roads. Their location, in fact, is to be related to areas characterized by considerable concentration of pollutants. The Benevento stations are equipped with instruments able to monitor the following components: NO_2, CO, PM_{10}, $PM_{2.5}$, O_3 and Benzene (C_6H_6).

In order to highlight the consistency of the data of the Benevento stations some of the trends in the period 2011-2019 are reported (Fig. 1). For example, the PM_{10} that would derive from the resuspension of inert dusts from construction sites, uncovered areas and road surfaces, as well as from aggregates of unburnt particles from combustion plants and vehicle engines, were measured in Benevento stations. The measured values revealed a significant decrease in the number of exceedance respect to the official reference value which is 50 µg/m^3. Moreover, in the years 2011-2016 overruns have exceeded 35 times allowed by Italian regulation for one year.

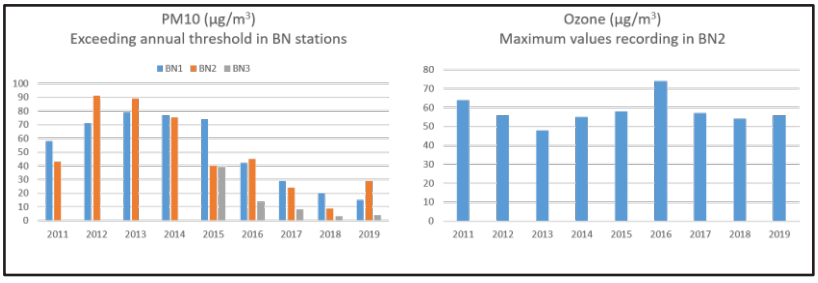

Fig. 1. Trends of PM_{10} and O_3 in Benevento stations from 2011 to 2019

Values above the limit of ozone, a triatomic oxygen molecule, were recorded only in the most peripheral station in Benevento. O_3 is a highly reactive gas, with a pungent odour and a high oxidizing power. It is generated starting from the action of solar radiation mainly on nitrogen dioxide molecules and subordinately on reactive hydrocarbons present in the atmosphere; consequently, ozone is a typical secondary pollutant more frequent in periods of greater insolation. The trend in Fig. 1 shows a cyclicality in the maximum values (74 µg/m^3 in July 2016 at 3 pm and 58 µg/m^3 in the beginning of August 2015 again at 3 pm) and therefore of exceedances (25 in 2017). Such trend is probably related to the variations of solar radiation measured in the spring and summer periods of the examined years. Ozone is particularly irritating to the respiratory tract and eyes. In addition, it causes lesions on the leaves of some plants and a reduction in elasticity on rubber and textile fibres. It should also be noted that the concentrations of very fine particles ($PM_{2.5}$), the measurement of which

has become increasingly refined, increased to reach values higher than 60 μg/m^3 in July 2017.

The data used for this research are those from the period 2012 - 2014 and were acquired by the regional site. The choice of the period is derived from the possibility to correlate the data of pollutants with those of the possible effects on human health. Although knowing, as reported in the literature, that the latency of lung cancer is quite large (Beverland et al., 2012; Raaschou-Nielsen O. et al., 2013), we tried to opt for the consequences on health in the medium term, i.e. in the years 2016-2018. In this period, tumour outbreaks were probably intercepted even earlier than the air quality data considered. However, they are indicative of a negative situation in the city of Benevento which persisted and which had been exposing its citizens for long time. Information regarding cancer was provided by hospitals in Benevento, as the city cancer registry is not yet able to cover this period.

4. Processing techniques

In order to highlight the correlation between the cases of lung tumours and the levels of air pollution two soft computing methods are applied. Such techniques are Artificial Neural Networks (ANNs) (Bishop, 1996; Haykin, 2008) and Genetic Programming (GP) (Koza, 1992). As ANN, we use a feedforward Multi-Layer Perceptron (MLP) Neural Network, a computational structure made by many processing elements (units), the neurons, operating in parallel, aiming to approximate an unknown function of its inputs (Beale and Jackson, 1990). To configure the MLP, we use the training procedure called "back propagation" (Beale & Jackson, 1990). It uses a subset of the data set feature vectors, each one labelled with the correct output, as examples of the correct input/output relationship. ANNs have already been successfully applied in different environmental contexts (Rampone, 2013; Rampone & Valente, 2012).

Table 1. Characteristics (inputs and output) and unit of measure (Air Data from January 2012 to December 2014, Tumour data from January 2016 to December 2018). Legend: M: average of maximum values; Exc.: exceedance of threshold; Avg.: average of values; Max: the maximum value registered; BN1 and BN2: monitoring stations

Input/Output	Characteristic	Unit of measure
1	M PM_{10} BN1	μg/m³
2	M PM_{10} BN2	μg/m³
3	Exc. PM_{10} BN1	Integer in (0:31)
4	Exc. PM_{10} BN2	Integer in (0:31)
5	Avg. $PM_{2,5}$	μg/m^3
6	M NO_2 BN1	μg/m^3
7	M NO_2 BN2	μg/m^3
8	Exc. NO_2 BN1	Integer in (0:31)
9	Exc. NO_2 BN2	Integer in (0:31)
10	M CO BN2	μg/m^3
11	Exc. CO BN2	μg/m^3
12	Max O_3 BN2	μg/m^3
13	Exc. O_3 BN2,	Integer in (0:31)
14	Benzene	μg/m^3
Output	Tumours	Number

The *ANN topology* (number of hidden layers, number of neurons in the hidden layers) is select by a *pruning/growing* methodology (Rampone, 2013), starting from an initial random choice. The resulting topology consists of 14 input, two hidden layers – made up of 2 and 4 neurons, respectively, and one output. The *initial network weights Wt* are randomly chosen in a fixed range. The *learning rate*, a measure of the influence degree, in the formula for

updating weights of the actual error, and the *momentum term*, that determines the influence of the past history of weight changes, are determined by a *trials-and-errors* methodology.

We also employ GP in order to improve the understanding of the neural method performance. GP is an evolutionary computational technique proposed by Koza (1992) in order to extract automatically intelligible relationships in a system without being explicitly programmed. It has been used in many applications such as symbolic regressions (Davidson et al., 2003) and classifications (De Stefano et al., 2002; Zhang & Bhattacharyya, 2004). It works by using genetic algorithms (Goldberg, 1989) to generate and evolve automatically composite functions, traditionally represented as tree structures (Cramer, 1985). GP is then able to show the relation between input data and output data by an explicit formula. This soft computing method has been also successfully applied to many problems of practical order (Makkeasorn et al., 2008; Shiri et al., 2012; Stanislawska et al., 2012; D'Angelo et al. 2019). In the Genetic Programming (GP) experiments, we are looking for a formula *f (...)* that satisfies

$$Tumors = f(M\ PM_{10}\ BN1,\ M\ PM_{10}\ BN2,\ Exc\ PM_{10}\ BN1,\ Exc\ PM_{10}\ BN2,\ Avg\ PM_{2,5},\ M\ NO_2\ BN1,\ M\ NO_2\ BN2,\ Exc\ NO_2\ BN1,\ Exc\ NO_2\ BN2,\ M\ CO\ BN2,\ Max\ O_3\ BN2,\ Exc\ O_3\ BN2,\ Benzene)$$

GP relies on a set of component functions. The set of possible component functions is limited to the *arithmetic operators (+, -, *, /),* some trigonometric functions (*sine, cosine and tangent and hyperbolic versions*) including their inverse, the *exponential* and the *natural logarithm*, the *logistic function*, and the *gauss function*. The function quality (*fitness measure*) is the Absolute Error. The samples are divided in training and validation as in the previous MLP experiments. The experiments are performed by a genetic programming software tool called Eureqa (Schmidt & Lipson, 2009).

4.1 ANNs-MLP: Training and validation

The MLP training is made by the *back propagation* procedure (Beale & Jackson, 1990), and a 10-fold cross validation methodology (Devijver & Kittler, 1982) is applied. So 10 independent experiments are performed for each validation set choice, and we use the resulting average performance as %-misclassification error. The number of epochs (training cycles) is fixed to 500.

The experiments are performed on the basis of the dataset described in Tab.1, by using a neural network Excel-based simulation environment developed by Angshuman Saha (available on-line). As performance indicators, we use both the error percentage and the correlation coefficient resulting from the 10-fold cross-validation application.

Table 2. Results of 10-fold cross-validation application

Trial	%-Average error	Correlation coefficient
1	0.26%	0.95
2	0.74%	0.92
3	0.19%	0.95
4	0.00%	0.91
5	0.27%	0.91
6	0.96%	0.98
7	0.82%	0.96
8	0.18%	0.97
9	0.77%	0.98
10	8.87%	0.97
Mean	**0.31%**	**0.95**

The %-misclassification error is 0,31% and the correlation coefficient is 0,95, both significant values. The predicted and observed tumour values are comparatively represented in two trials, among the worst and best.

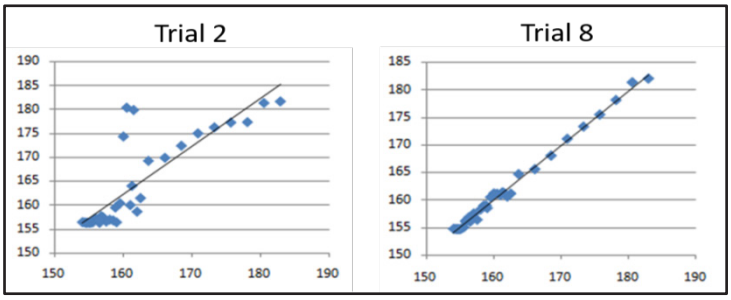

Fig. 2. MLP results in two trials: Observed (Y-axis) and Predicted (X-axis) tumours

4.2 GP: Training and validation

As previously said, we use a genetic programming software tool called Eureqa, for detecting equations and hidden mathematical relationships in a given data set. Eureqa works in order to reduce the error function given by the discrepancy between the data and the generated model (Schmidt & Lipson, 2009). Such software uses evolutionary search to determine mathematical equations that describe sets of data in their simplest form. It also allows to evidence the behaviour of each solution respecting its size.

After about 300000 generations we found the best solution, with a correlation coefficient of 0,99 and %Average Error of 1,65:

$$(Tumours\ cases) = a + b*(M.\ CO\ BN2) + c*(Exc.\ PM_{10}\ BN1) + -d*(Avg.\ PM_{2,5}) * \sin(e - (M.NO_2\ BN2))/(f + g*(M.CO\ BN2) + h*(Exc.\ PM_{10}\ BN1) - (Exc.\ PM_{10}\ BN2)) - i*(Exc.\ PM_{10}\ BN2) - j*(Exc.\ O_3\ BN2) - k*(Exc.\ NO_2\ BN1)$$

A plot of Expected vs Predicted tumours on the whole data set by using the best solution is reported in Fig. 3. We also measure the relevance of each considered characteristic in determining the solution result, evidencing the major role of PM_{10}, NO_2 and O_3.

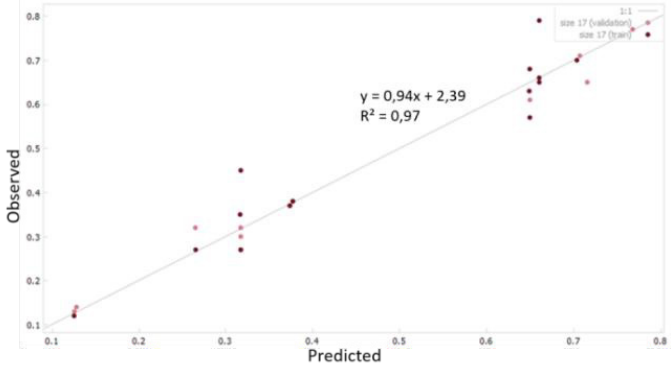

Fig. 3. Plot of Expected vs Predicted tumours on the whole data set

5. Conclusion

As evidenced by the soft computing techniques, it appears there is a direct relationship between air pollution and the occurrence of lung tumours cases. This relationship is highlighted above all for some specific pollutants that lower air quality and can increase the risk of getting sick among the Benevento population. In the absence of strong industrialization in the city area, it plausible that mainly urban traffic and, subordinately, domestic heating are responsible for the levels of pollution recorded. From the traffic point of view, this area represents an important crossroads between the sides of the Tyrrhenian Sea and the Adriatic Sea, as well as a town that is often congested with traffic related to a significant commute. An important effect of traffic had already been highlighted in the past through the presence of metals such as lead, cadmium and zinc in the soils of the city (Zuzolo et al., 2020), and believed to be produced by car exhaust and transported by particulate matter. However, also the meteorological and topographical conditions of Benevento have a significant weight on the formation of secondary pollutants as recorded in the peripheral areas. In conclusion, pollutants such as PM_{10}, NO_2 and O_3 pose a health risk that must be considered to avoid an increase in lung cancers.

References

Beale R., Jackson T. (1990). *Neural Computing: An Introduction.* Taylor & Francis, New York, 256 pp.

Beverland I.J., Robertson C., Yap C., Heal M.R., Cohen G.R., Henderson D.E.J., Hart C.L., Agius R.M. (2012). Comparison of models for estimation of long-term exposure to air pollution incohort studies Atmospheric Environment 62, 530-539.

Bishop C.M. (1996). *Neural Networks for Pattern Recognition.* Oxford University Press., USA, 502 pp.

Cramer N.L. (1985). A representation for the Adaptive Generation of Simple Sequential Programs. In: John J. Grefenstette (ed.) Proceedings of an International Conference on Genetic Algorithms and the Applications, Carnegie Mellon University. Pittsburg, PA, USA, pp. 183-187.

D'Angelo G., Pilla R., Tascini C., Rampone S. (2019). A proposal for distinguishing between bacterial and viral meningitis using genetic programming and decision trees. Soft Computing, 23(22), 11775-11791.

Davidson J.W., Savic D.A., Walters G.A. (2003). Symbolic and numerical regression: Experiments and applications. Information Sciences, 150(1/2), 95–117.

De Stefano C., Della Cioppa A., Marcelli A. (2002). Character preclassification based on genetic programming. Pattern Recognition Letters, 23(12), 1439–1448.

Devijver P.A., Kittler J. (1982). *Pattern Recognition: A Statistical Approach*. Prentice-Hall, London, 448 pp.

Filippelli G.M., Morrison D., Cicchella D. (2012). Urban Geochemistry and human health. Elements 8, 439-434.

Goldberg D.E. (1989). *Genetic Algorithms in Search, Optimization & Machine Learning*. AddisonWesley, 432 pp.

Schmidt M., Lipson H. (2009). Distilling Free-Form Natural Laws from Experimental Data. Science, 324, April.

Haykin S. (2008). *Neural Networks and Learning Machines*. Prentice Hall, 906 pp.

Koza J.R. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, Cambridge, MA, USA, 819 pp.

Raaschou-Nielsen O. et al. (2013). Air pollution and lung cancer incidence in 17 European cohorts: prospective analyses from the European study of cohorts for air pollution effects. The lancet oncology, 14(9), 813–822.

Rampone S. (2013). Three-and-six-month-before forecast of water resources in a karst aquifer in the Terminio massif (Southern Italy), Applied Soft Computing, 13(10), 4077–4086.

Rampone S., Valente A. (2017). Prediction of seasonal temperature using soft computing techniques: application in Benevento (Southern Italy) area. J Ambient Intell. Human Comput., 8 (1), 147–154.

Rampone S., Valente A. (2019). Assessment of desertification vulnerability using soft computing methods. J Ambient Intell Human Comput., 10 (2), 701–707.

Rampone S., Valente A. (2012). Neural Network Aided Evaluation of Landslide Susceptibility in Southern Italy, International Journal of Modern Physics C, Vol. 23, No. 01, 1250002.

Seigneur C. (2019). *Air Pollution*. Cambridge University Press, Cambridge, 370 pp.

Shiri J., Kişi O., Landeras G., Lopez J.J., Nazemi A.H., Stuyt L.C.P.M. (2012). Daily reference evapotranspiration modeling by using genetic programming approach in the Basque Country (Northern Spain). Journal of Hydrology, 414–415, 302–316.

Stanislawska K., Krawiec K., Kundzewicz Z.W. (2012). Modelling global temperature changes with genetic programming. Computers and Mathematics with Applications 64, 3717–3728.

WHO (2016). *Ambient air pollution: a global assessment of exposure and burden of desease*, World Health Organization, Geneva. (https://www.who.int/gho/phe/outdoor_air_pollution/en/, accessed September 2020).

Zhang Y., Bhattacharyya S. (2004). Genetic programming in classifying large-scale data: an ensemble method. Information Science, 163(1/3), 85–101.

Zuzolo D., Cicchella D., Lima A., Guagliardi I., Cerino P., Pizzolante A., Thiombane M., De Vivo B., Albanese S. (2020). Potentially toxic elements in soils of Campania region (Southern Italy): Combining raw and compositional data. Journal of Geochemical Exploration, 213, 106524.