Ecological Indicators 72 (2017) 759–765

Contents lists available at ScienceDirect

Ecological Indicators

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

Research paper

The application of artificial neural networks to the problem of reservoir classification and land use determination on the basis of water sediment composition

Izabela Świetlicka[a,*], Agnieszka Sujak[a], Siemowit Muszyński[a], Michał Świetlicki[b]

[a] Department of Physics, University of Life Sciences, Akademicka 13 St., 20-950 Lublin, Poland
[b] Department of Applied Physics, Lublin University of Technology, Nadbystrzycka 38 St., 20-618 Lublin, Poland

ARTICLE INFO

Article history:
Received 22 March 2016
Received in revised form 23 August 2016
Accepted 8 September 2016

Keywords:
Sediments
Reservoir
Area usage
Artificial neural networks
Lublin Province

ABSTRACT

Presented work reports on the use of artificial neural networks to recognize and classify water reservoir types (lakes, rivers) and the nature of their surroundings (forests, fields, meadows) based on the chemical composition of sediments. The quantitative content of a selection of elements (Ag, As, Ba Ca, Cd, Co, Cr, Cu, Fe, Hg, Mg, Mn, Ni, P, Pb, S, Sr, TOC – Total Organic Cabon, V and Zn) in the sediments of lakes and rivers in the Lublin Province (Poland) were taken and used as working data file. Statistical analysis suggested that both reservoir types and area usage differ in terms of the quantity of studied determinants (elements) and thus might be distinguished on their basis. Artificial neural networks were then examined with respect to their ability to recognize and classify the data. Multilayer perceptron was used as the statistical model. Constructed models were able to give correct answers in 74% of cases when classifying reservoir's area usage and 100% for the type of body of water.

1. Introduction

Sediments are characteristic of the catchment area and the water reservoir they are found in. Sediment collected at the bottom of a river or lake can come from both the erosion of rocks and soils occurring in the catchment area, and from the suspensions of mineral and organic products brought to the surface water with industrial and municipal waste water (Mielnik, 2005). Dead plants, animal organisms, precipitated inorganic and organic materials, e.g.: calcium carbonate, iron and manganese hydroxides and phosphorus compounds, which fall to the bottom are also responsible for the creation of sediment (Calmano et al., 1996; Julien, 2010). Furthermore, the chemical composition of sediment does not depend solely on the geological structure of the catchment area and its surroundings, or on climate conditions determining the processes of weathering, but also on land and catchment usage and by pollution reaching the surface water (Brack et al., 2001; Calmano et al., 1996; Julien, 2010). For example, high concentrations of some trace elements found in sediments of rivers and lakes are mainly a result of sewage spilled from factories, cities and water transports (Brack et al., 2001; Calmano et al., 1996; Lindström, 2001). Pollution also infiltrates water reservoirs as effluents from surficial flows (Mangani et al., 2005; Mecray et al., 2001; Reiss et al., 2004; Rocher et al., 2004). Deposition of some elements from the atmosphere (e.g. lead, arsenic, cadmium, mercury and organochloride compounds) and surface runoff from urbanized (heavy metals) and agricultural (arsenic, mercury, organochloride pesticides) areas contribute significantly to the increase in concentration of potentially harmful elements and persistent organic pollutants in sediments formed today (Birch et al., 2001; Reiss et al., 2004; Rocher et al., 2004; Yamashita et al., 2000). As a result, in the sediments considered the most polluted, elevated levels of metals having wide applications in industry are detected. These metals include silver, arsenic, beryllium, zinc, chromium, copper, mercury, nickel, lead, antimony, selenium, thallium and zinc.

Due to the fact that the majority of potentially harmful metals and organic compounds in sediments reaching the rivers and lakes are retained (Hansen, 1996; Reichardt, 1996; Smal and Salomons, 1995), the water becomes contaminated and poses a potential threat not only to the aquatic environment, but also for humans and animals. In the case of flooding, contamination can enter farmlands and other areas previously uncontaminated (Gabler and Schneider, 2000; Gocht et al., 2001), and become a source of secondary contamination (Ciszewski and Malik, 2003). Thus, the diversity of sediments in terms of their chemical composition can be a valuable source of information about the area that surrounds the collection

* Corresponding author.
E-mail address: izabela.swietlicka@up.lublin.pl (I. Świetlicka).

http://dx.doi.org/10.1016/j.ecolind.2016.09.012

Table 1
Number of samples divided in accordance with reservoir type and nature of its surrounding area.

	Lake	River	Σ
Forest	43	74	117
Field	26	56	82
Meadow	21	102	123
Unknown	37	49	86
Σ	127	281	408

point as well as the reservoir itself. However, since such data are characterized by high nonlinearity and complexity, it is important to develop autonomous statistical models that can determine the reservoir type and area usage on the basis of the chemical composition of sediments with a very high probability.

Artificial neural networks (ANNs) architecture and working principle are inspired by biological neural networks. ANNs usually consist of several computational nodes built with artificial neurons. The role of a neuron is to accept a weighted input, which the neuron sums with a bias and gives a corresponding output. Artificial neural networks are widely used in the fields of recognition and classification (Bishop, 1995) as well as data compression and prediction (Werea et al., 2015). Unlike traditional statistical methods ANNs adjust to data without the necessity of defining any additional function or distribution of input variables. They are also able to determine the probability of object membership in a group, which brings the possibility of using ANNs as a kind of *a posteriori* probability estimators (Burrascano, 1993; Kline and Berardi, 2005). It is known that ANNs are capable of mapping the probability distribution even if only a limited number of learning cases are applied (Hung et al., 1996). In contrast to, for example, agglomerative cluster analysis, or discriminant analysis, artificial neural networks can recognize and classify objects which are seemingly not similar within the group (Bishop, 1995). ANNs are successfully applied, among others, in the field of environmental data processing, where they proved to be a useful and valuable tool for data analysis (Valipour, 2016; Valipour et al., 2015), classification (Crnković et al., 2016; Kzar et al., 2016) and prediction (Valipour, 2016; Valipour et al., 2013, 2015; Vié et al., 2014), especially in situations where the data has been shown to be non-linear, uncorrelated and inconsistent.

Taking into consideration the above facts, the authors decided to apply artificial neural networks to solve the problem of recognition and classification of reservoir types and the nature of the surrounding area on the basis of the concentration of elements in sediment.

2. Materials and methods

The analysed input data consisted of quantitative measurements of the elements content (Ag, As, Ba Ca, Cd, Co, Cr, Cu, Fe, Hg, Mg, Mn, Ni, P, Pb, S, Sr, TOC – Total Organic Carbon, V and Zn) found in samples of sediments collected from lakes and rivers in the Lublin Province, Poland, flowing through farmlands, woodlands, meadows and other (Table 1).

The samples were collected over a period of 22 years (from 1990 to 2012) at different measuring points (Fig. 1) and were made available to the research team by the Chief Inspectorate for Environmental Protection in Warsaw, Poland.

2.1. Initial statistics

The obtained samples were analysed according to their ability to differentiate among considered groups i.e. the way in which the surrounding area is used (fields, meadows or forests) – Experiment I, and the type of water reservoir (lakes or rivers) – Experiment II.

Table 2
Kruskal-Wallis ANOVA values (H) according to calculated determinants. Effects are significant with $p < 0.05$.

Element	Experiment I (Area usage)		Experiment II (Reservoir type)	
	H	p	H	p
Ag	1.13	0.2877	8.25	0.1428
As	63.57	0.0000	38.25	0.0000
Ba	34.69	0.0000	24.98	0.0001
Ca	4.29	0.0384	32.21	0.0000
Cd	22.80	0.0000	36.10	0.0000
Co	0.49	0.4851	37.30	0.0000
Cr	2.69	0.1010	23.70	0.0002
Cu	3.76	0.0000	42.98	0.0000
Fe	13.45	0.0000	13.23	0.0213
Hg	54.65	0.0000	47.27	0.0000
Mg	0.02	0.8932	33.46	0.0000
Mn	3.06	0.0000	16.68	0.0052
Ni	0.08	0.7836	28.89	0.0000
P	130.27	0.0000	47.02	0.0000
Pb	2.05	0.1518	58.03	0.0000
S	231.06	0.0000	59.05	0.0000
Sr	0.43	0.5127	25.64	0.0001
TOC	238.16	0.0000	68.57	0.0000
V	79.31	0.0000	22.12	0.0005
Zn	31.53	0.0000	50.76	0.0000

For the statistical calculations STATISTICA 10.0 application (StatSoft Inc., Tulsa, OK, USA) was used. Differences in the mean values calculated using Welch's test proved that groups differ significantly according to most of the dependent variables. However Levene's test as well as the robust test of means equality (Brown-Forsythe) showed that the homogeneity of variances was not saved which led to the conclusion that data should be analysed using non-parametric methods. Due to that, Kruskal-Wallis ANOVA (Table 2) was applied in order to check if groups differ in respect of the analysed variables.

The values of H statistic and high significance of the test ($p < 0.05$) for the majority of elements indicate that the determined variables made a significant contribution to the groups' differentiation and provided the basis for classifier construction.

2.2. ANN models and architecture

Classifying models were designed with artificial neural networks (ANNs) application. In the first step the analysed variables were subjected to standardization according to the Eq. (1):

$$z_{ij} = \frac{x_{ij} - \bar{x}}{\sigma} \tag{1}$$

where: z_{ij} – standardized variable element, x_{ij} – original variable element, $\bar{x}$ – a mean of original variable, σ – original variable standard deviation. The table built from the obtained data was used in the teaching process. The chart was randomly divided into three groups: teaching, verifying and testing, and then introduced to the network's input (Fig. 2). Classification was conducted with the use of a multilayer perceptron network (MLP). Multilayer perceptron is one of the oldest artificial neural network models but is still successfully applied today. The basic element of MLP network is a neuron which is sigmoidal nature, for which the output signal takes values from 0 to 1 for unipolar functions or from −1 to 1 for bipolar ones. MLP's are strictly organized and neurons are arranged into three main layers: input, hidden and output one. Usually between 1 and 5 hidden layers are used.

The number of hidden layers as well as the number of neurons in each of them was determined in an experimental way, using the growth method while taking into consideration validation errors (Burrascano, 1993; Castellano et al., 1997). The acceptance and