Available online at www.sciencedirect.com
ScienceDirect
Talanta 72 (2007) 1–12
Talanta
www.elsevier.com/locate/talanta

Review

Chemometrics based on fuzzy logic principles in environmental studies

Aleksander Astel *

Environmental Chemistry Research Unit, Biology and Environmental Protection Institute, Pomeranian Pedagogical Academy, 22a Arciszewskiego Str., 76-200 Słupsk, Poland

Received 12 April 2006; received in revised form 14 September 2006; accepted 27 September 2006
Available online 30 October 2006

Abstract

Nowadays, environmetrics based on the principles of fuzzy logic has become an important and sophisticated statistical instrument in modern science, being an adequate tool to investigation the principles of interaction of elements/variables and their integration into a system. This paper is intended mainly for a wide community of ecologist, which are interested in principles of using fuzzy logic in environmental researchers. Properties of environmetrics based on fuzzy logic principles as a useful tool of investigation in environmental studies, are considered and discussed both in the terms of objectives and the examples of using techniques under interest concerning air, water and land ecological systems.

Keywords: Environmetrics; Fuzzy logic; Air; Water; Land; Forecasting; Modeling; Classification

Contents

* Tel.: +48 59 8405 423; fax: +48 59 8405 475.
E-mail address: astel@pap.edu.pl.

0039-9140/$ – see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.talanta.2006.09.026

1. Environmetrics based on fuzzy logic principles—common tendencies

Chemometrics has been evolving as a subdiscipline in chemistry for over 30 years as the need for advanced statistical and mathematical methods has increased with sophistication of chemical instrumentation and processes [1]. As defined by Massart et al. [2], "chemometrics is a chemical discipline that uses mathematics, statistics, and formal logic to design or select optimal experimental procedures, to provide maximum relevant chemical information by analyzing chemical data and to obtain knowledge about chemical systems". Nowadays, the chemometrics is not only limited to formal logic. Named as environmetrics and often based on fuzzy logic principles is concerned with the application and development of statistical methodologies in environmental sciences.

Fuzzy logic was introduced in the 60s [3]. It simplifies the process of taking decisions by simulating the way of reasoning of a human expert in environments characterized by uncertainty and imprecision. The idea behind fuzzy logic is that an element can belong partially to several subsets, unlike classical logic (Boolean) where belonging or not to a set are mutually exclusive. The degree of belonging to a set is a value between 0 and 1, usually determined by to what extent an element belongs to a fuzzy subset or a category of a variable. The concept of a fuzzy set is illustrated schematically by a simple example connected with environmental research in Fig. 1. The horizontal axis represents an environmental quality, such as a pollutant concentration which is always positive, and the vertical dashed line is the environmental standard, or objective. In this co-ordinate system three fuzzy sets are defined representing good, medium and poor environmental quality. The vertical axis shows the membership degree (membership function), which can have various equations of each fuzzy set. Scientific details on the methodology of various chemometrical techniques with implemented fuzzy logic rules can be found elsewhere [4–9] and are not presented in details in this review due to the difficulty of making it representative as the list of available techniques and possible variations is huge. Despite this we decided to present briefly the possible major objectives which can be accomplished successfully using chemometrics based on fuzzy logic principles: forecasting, modeling and classification.

1.1. Forecasting

Time series (TS) are sequences of independent random variables (observations) describing a phenomenon at successive points in time. The main purpose of TS analysis is: (i) to make a statistical analysis of relations between successive observations; (ii) to predict (forecast) future values of the TS; (iii) to detect mechanisms that govern changes in the observed phenomenon in time, in other words, to determine the nature of the phenomenon represented by the sequence of observations. Nowadays, the time series analysis is commonly connected with neural network approach. Artificial neural networks (ANNs) are mathematical models consisting of a network of computation nodes called neurons and the connections between them. Adaptation of a neural network to solving a particular task takes place through a training process using typical stimulations and required responses corresponding to them, and not through defining an algorithm and writing it down in the form of a program, as would be the case in traditional modeling methods [10].

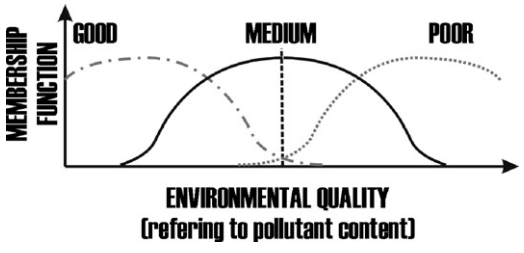

Fig. 1. Schematic explanation of fuzzy sets.

1.2. Classification

A basic problem that arises in a wide variety of fields, including analytical chemistry is the so-called clustering problem. Classification is useful, since it allows meaningful generalizations to be made about large quantities of data by recognizing among them a few basic patterns. It plays a key role in searching for structures in data. Each of these structures is called cluster or class. A class is a group of individuals (e.g., soil samples or pixels of a image) which resemble each other more strongly, in terms of particular properties, than they resemble members of other classes. There are various ways in which classification may be carried out. Generally, two types of algorithm are distinguished, these being hierarchical and non-hierarchical or relocation clustering. Both methods require the calculation of a (dis)similarity matrix. This (dis)similarity which is really a measure of the proximity of the pair of objects (points) in the p-dimensional characteristic space, defined by the p properties measured for each individual, is usually expressed in terms of either the Euclidean or the Mahalanobis distance between the two points.

1.3. Modeling

Fuzzy systems, including fuzzy logic and fuzzy set theory, provide a rich and meaningful addition to standard logic. The mathematics generated by these theories is consistent, and fuzzy logic may be a generalization of classic logic. The applications which may be generated from or adapted to fuzzy logic are wide-ranging, and provide the opportunity for modeling of conditions which are inherently imprecisely defined, despite the concerns of classical logicians. Many systems may be modeled, simulated, and even replicated with the help of fuzzy systems, not the least of which is human reasoning itself. Fuzzy logic is the way the human brain works, and we can mimic this in machines so they will perform somewhat like humans [11].

Uncertainty analysis, common in environmental studies can be dealt with fuzzy methods where uncertainty is due to vagueness or "fuzziness" rather than only to randomness. Therefore, environmental data or classes of ecological objects can be defined as fuzzy sets with not sharply defined boundaries