



“Gheorghe Asachi” Technical University of Iasi, Romania



COMPREHENSIVE COMPARISON OF A FEW VARIANTS OF CLUSTER ANALYSIS AS DATA MINING TOOL IN SUPPORTING ENVIRONMENTAL MANAGEMENT

Aleksander Astel^{1*}, Karolina Astel¹, Stefan Tsakovski², Marek Biziuk³,
Krystian Obolewski⁴, Katarzyna Glińska-Lewczuk⁵, Katarzyna Bigus¹,
Ioan Craciun⁶, Cristina Maria Timofte⁵

¹Pomeranian University, Biology and Environmental Protection Institute, Environmental Chemistry Research Unit, 22b
Arciszewskiego Str., 76-200 Słupsk, Poland

²University of Sofia, Faculty of Chemistry, 1 J. Bourchier Blvd., 1164 Sofia, Bulgaria

³Gdansk University of Technology, Chemical Faculty, Department of Analytical Chemistry, 11/12 Narutowicza Str., 80-952,
Gdansk, Poland

⁴Kazimierz Wielki University in Bydgoszcz, Department of Hydrobiology, 30 Chodkiewicza Str., 85-064 Bydgoszcz, Poland

⁵University of Warmia and Mazury in Olsztyn, Department of Land Reclamation and Environmental Management, 2 Łódzki Sq.,
10-719 Olsztyn-Kortowo, Poland

⁶Gheorghe Asachi Technical University of Iasi, Department of Hydraulic Structures, 63-65 D. Mangeron Blvd., 700050 Iasi,
Romania

Abstract

A few variants of hierarchical cluster analysis (CA) as tool of assessment of multidimensional similarity in environmental dataset are compared. The dataset consisted of analytical results of determination of metals (Na, K, Ca, Sc, Fe, Co, Zn, As, Br, Rb, Mo, Sb, Cs, Ba, La, Ce, Sm, Hf and Th) in ambient air dried and kept alive, by the means of hydroponics, moss baskets collected in 12 locations on the area of Tricity (Poland). The selected urban locations covered sites of various anthropogenic impact (cargo transit, heavy traffic, apartments, shipyard, etc.). Factor analysis in the variant of Principal Component Analysis (PCA) is offered as a validating method for CA. The results of three scaling variants are presented: column scaling, row scaling and column scaling followed by row scaling. Moreover, the results of two linkage methods with squared Euclidean distance as the similarity measure are compared: Ward's linkage and complete linkage. The cluster significance is indicated basing on Sneath's index. Some urban pollution profiles (fertilizer plant emission, fossil fuel burning, municipal waste dump, mineral soil dust) were identified and the possibility of their application by the environmental decision makers or managers is pointed out. The complete environmental data exploration procedure is recommended.

Key words: data scaling, environmetrics, moss, pollution source impact assessment, *Sphagnum palustre*

Received: August, 2015; *Revised final:* March, 2016; *Accepted:* April, 2016

* Author to whom all correspondence should be addressed: e-mail: astel@apsl.edu.pl or AliAst@poczta.fm; Phone: (+4859) 8405423; Fax: (+4859) 8405375