Positive Matrix Factorization and GIS approach to perform data mining on groundwater and surface water quality dataset.

Chiara Zanotti[1], Marco Rotiroti[1], Letizia Fumagalli[1], Gennaro Stefania[1], Francesco Canonaco[2], Giulia Stefenelli[2], Andre Prévôt[2], Barbara Leoni[1], and Tullia Bonomi[1]

[1]University of Milano - Bicocca, Milano, Italy (c.zanotti@campus.unimib.it)

[2]Laboratory of Atmospheric Chemistry, Paul Scherrer Institute, 5232 Villigen-PSI, Switzerland

Positive Matrix Factorization (PMF) is a multivariate analysis aimed at source identification and apportionment, specifically designed to cope with environmental data and manage their uncertainty and distributions. The aim of this work is to test the effectiveness of PMF as a tool to perform data mining and define hydrochemical features of groundwater and surface water and to understand their relationship. Here PMF is applied to a dataset concerning groundwater, springs, rivers and a lake. Factor contributions to the samples are spatially analysed trough GIS approach and their interpretation is supported by considering the land use and hydrogeological features.

The study area is a part of the Oglio River basin (N Italy) and consists in a ~2000 km^2 area around the Oglio River part that outflows from Lake Iseo until the confluence with Mella River. The available dataset is the result of four field surveys conducted in February 2016, June 2016, September 2016 and March 2017, for a total amount of 270 samples collected on 68 monitoring points.

The first factor represents water with a high content of oxygen, with SO4 and a small content of major ions. In this factor neither nitrate nor trace elements are represented. The profile of this factor overlaps with the characteristics of lake Iseo and surface water bodies directly connected with it. Its spatial variability highlights a contribution of this factor, also in several wells even if they are not close to the river itself. This was related to the channels net collecting water from the Oglio River and spreading it to the nearby fields.

The second factor explains the whole chemical variability of As, Fe, P-tot and NH_4. Furthermore, it has a significant contribution of Mn and it can be associated to the advanced stages of reducing conditions due to degradation of natural organic matter.

The third factor is characterized mainly by Mn and SO_4, with a contribution of the major ions. Based on the profile and on the spatial distribution, it is possible to associate this factor to the early stages of the reducing process.

The fourth factor represents only major ions. Major ion concentration in groundwater is mainly determined by water – rock interactions, which increase the concentrations especially with

increasing residence times. This factor shows an increasing trend from north to south, which is the flow direction, confirming its relationship with residence time.

The fifth factor represents the total variability of the variable NO_3 with contributions also in terms of Cl, SO_4, Ca and Mg. Higher NO_3 concentrations in groundwater are mostly related with the use of organic or synthetic fertilizers; recent studies reported that this kind of anthropogenic impact affects also major ions concentrations such as Cl, SO_4, Ca and Mg. Thus, the profile of this fifth factor was associated to the anthropogenic impact related to the agricultural land use, not just in terms of NO_3 but also considering the contribution of different elements.