

Multivariate statistical analyses for the source apportionment of groundwater pollutants in Apulian agricultural sites

Pierina Ielpo, Daniela Cassano, Antonio Lopez, Pasquale Abbruzzese De Napoli, Giuseppe Pappagallo, Vito Felice Uricchio

Water Research Institute – National Research Council, via F. De Blasio, 5 – 70132
Bari, Italy; phone + 39 080 5820512; fax +39 080 5313365

piera.ielpo@ba.irsra.cnr.it

Abstract: Multivariate statistical techniques, such as Principal Component Analysis, Absolute Principal Component Scores, Cluster Analysis and Discriminant Function Analysis were applied to data set (pH, Electrical Conductivity, Total Dissolved Solids (TDS), Dissolved Oxygen (O₂), Chemical Oxygen Demand (COD), the major ions (i.e. Na⁺, Ca²⁺, Mg²⁺, K⁺, Cl⁻, NO₃⁻, SO₄²⁻ and HCO₃⁻), vital organism at 22 °C and 36 °C) of ground waters collected in 473 sites of the Apulia region during the “Expansion of regional agro-meteorological network” project. Multivariate statistical techniques allowed to identify for each province sites with different characteristics as respect to similar characteristics ones. Moreover Absolute Principal Component Scores allowed to identify generally three pollutant sources.

Keywords: ground water, water pollutants, source apportionment, statistical analyses

1. Introduction

During the years 2004-2007 the Agricultural and Food Authority of Apulia Region has implemented the project “Expansion of regional agro-meteorological network” in order to assess, monitor and manage the regional groundwater quality. The wells monitored during this activity amounted to 473 and the water samples analyzed were 998.

This resulted in a huge and complex data matrix comprised of a large number of physical-chemical parameters, which are often difficult to interpret and draw meaningful conclusions. Further, for effective pollution control and water resource management, it is required to identify the pollution sources and their quantitative contributions. The application of different multivariate statistical techniques such as cluster analysis (CA), principal component analysis (PCA), source apportionment by multiple linear regression on absolute principal component scores (APCS) for interpretation of the complex databases offers a better understanding of water quality in the study region. Moreover Discriminant Function Analysis (DA) was used in order to identify the characteristics of the all sites investigated in the Apulia region.

2. Materials and Methods

Groundwater samplings were performed under dynamic conditions, after flushing a large amounts of water for about 30 minutes. Samples were collected in polyethylene tanks with cap and under cap, filled to the brim in order to prevent the transfer of the analytes in the headspace and their loss at the opening of the tanks. After collection, samples were stored in cooled bags and transported to the laboratory as soon as possible.

The samples were analyzed for pH, Electrical Conductivity (Electr. Cond.), Total Dissolved Solids (TDS), Dissolved Oxygen (O₂), Chemical Oxygen Demand (COD), the major ions (ie. Na⁺, Ca²⁺, Mg²⁺, K⁺, Cl⁻, NO₃⁻, SO₄²⁻ and HCO₃⁻), vital organism at 22 °C and 36 °C, according to the official guideline proposed by the Ministero delle Politiche Agricole (the national agriculture authority) in a specific law (Decreto Ministeriale del 23 Marzo 2000 “*Metodi ufficiali di analisi delle acque per uso agricolo e zootecnico*”). Each parameter was analyzed in three replicates. In table 1 the number of monitored wells and collected samples for each Apulian province have shown.

Province	Wells	Samples collected
BARI	96	260
BRINDISI	89	102
FOGGIA	85	219
LECCE	84	165
TARANTO	119	252

Table 1: Groundwater quality monitoring

3. Results

DFA applied to all data set allowed to individuate the variables with bigger discriminatory power. The results are shown in table 2: among variables those with bigger discriminatory power are highlighted in bold.

PCA, CA, APCS methods were firstly applied to the samples collected in each Apulian province separately. Form results obtained by PCA and CA was evident that for each province some sampling sites investigated showed dissimilarities, mostly due to the location of the site (close to the sea, close to not purified water channels), the land use and management techniques (fertilizing and nourishing techniques) and groundwater overuse of the investigated sites. For all these reasons several natural and anthropogenic sources affect the groundwater quality of the investigated sites. As example some results of PCA and APCS for Taranto province are shown in figure 1 and 2.

Considering the score plot (figure 1) in the plane of the first and second Principal Component it is possible to note some scattered samples, highlighted in rectangular lines. Moreover considering the loading plot (not shown here) the samples (sites) enclosed in the red lines (figure 1) show high loading values for vital organism at 22 °C and 36°C; samples enclosed in the green line show high loading values for TDS, Electr. Cond., Cl⁻, Na⁺, Mg²⁺; samples enclosed in the brown line show high loading values for COD, SO₄²⁻, Ca²⁺ and those in magenta line show high values for K⁺.

Variables	Wilks Λ	Partial $\square \Lambda$
pH	0.3488	0.6743
Elect Cond	0.2429	0.9684
TDS	0.2376	0.9900
O ₂	0.2379	0.9886
Na ⁺	0.2355	0.9987
Ca ²⁺	0.2359	0.9971
Mg ²⁺	0.2438	0.9649
K ⁺	0.2529	0.9301
COD	0.2365	0.9945
Cl ⁻	0.2353	0.9995
NO ₃ ⁻	0.2388	0.9851
SO ₄ ²⁻	0.2382	0.9876
HCO ₃ ⁻	0.2416	0.9737
Vit. Org. 22°C	0.2362	0.9956
Vit. Org. 36°C	0.2356	0.9983

Table 2: Variables selected by Wilks lambda

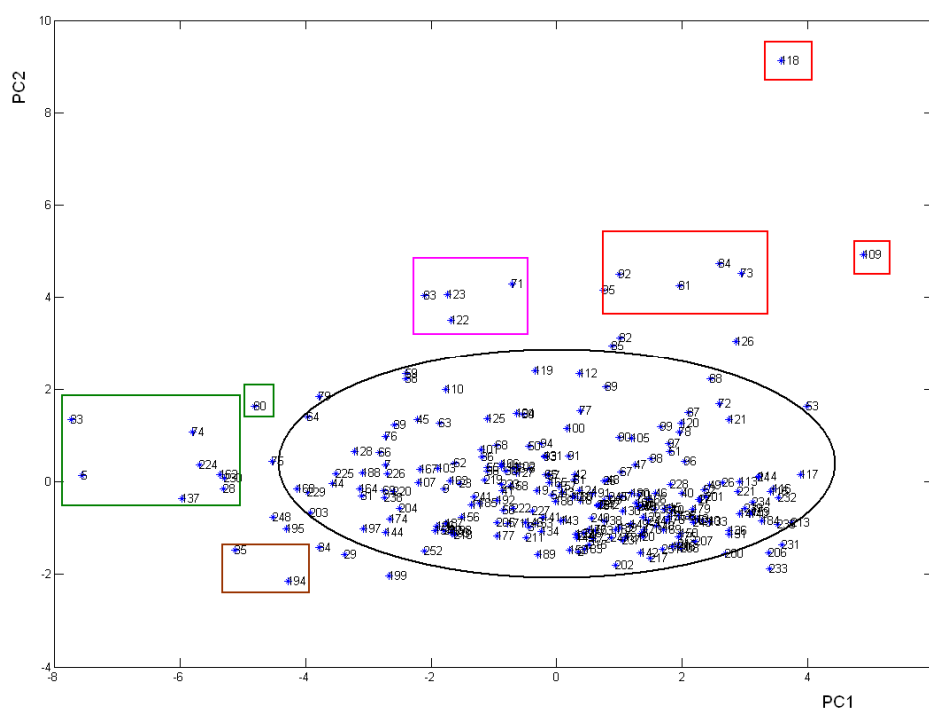


Figure 1: Score plot for Tarano province data matrix

The sites in red line are located close to a channel collecting waters from municipal purifier plant; those in green line are located close to the coast and for those it's possible

suppose an intrusion of marine water. The sites enclosed in brown line are located in an area with high agricultural impact: this means high use of fertilizers and nutrients. The CA results support the PCA ones.

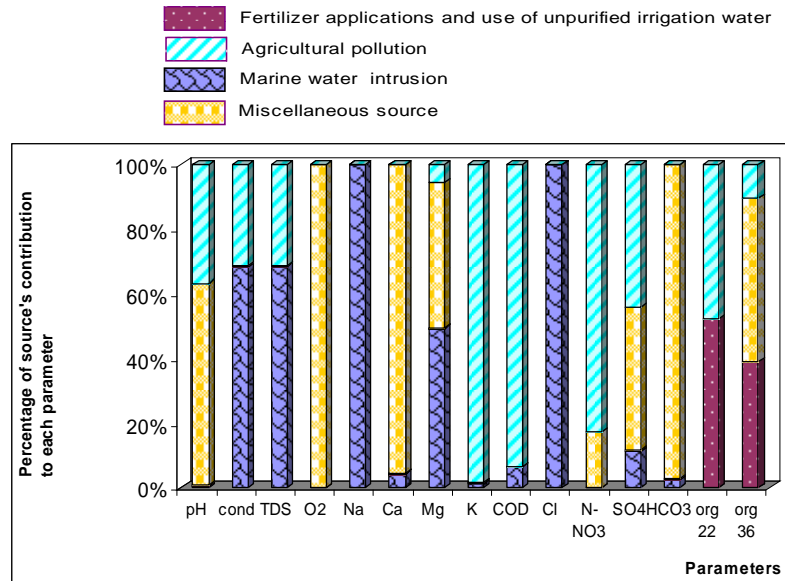


Figure 2: Percentage of source's contribution to each parameters for Taranto province data matrix

In order to individuate the pollutant sources the APCS method was applied to the data matrix of physical-chemical parameters collected. By APCS method it's been possible generally to identify three pollutant sources. About Taranto province data matrix (see figure 2) the pollution sources identified were: a source due to fertilizer applications and use of unpurified irrigation water; a pollution source due to agricultural techniques, marine water intrusion in the site one and a source mostly due to the calcareous characteristics of the soil in that area.

4. Concluding remarks

Multivariate statistical methods represent a valid tool to understand complex nature of groundwater quality issues, determine priorities in the use of ground waters as irrigation water and suggest interactions between land use and irrigation water quality. The results obtained by multivariate statistical methods can be used to suggest to stakeholders, for example, a mitigation in the groundwater overuse of some wells mostly in dry seasons and to require orderly quality tests of the channel waters when they are used for crop irrigation.

References

Bengraïne K., Marhaba T. F. (2003) Using principal component analysis to monitor spatial and temporal changes in water quality, *Journal of Hazardous Materials*, 100, 179-195.