Multivariate statistical analysis of hydrogeochemical data towards understanding groundwater flow systems in granites

A. P. Matos1* & C. Alves2

1 CERENA/DECivil, Instituto Superior Técnico, University of Lisbon, Av. Rovisco Pais 1, 1049-001 Lisbon, Portugal
2 Lab2PT (FCT; FEDER COMPETE POCI) and School of Sciences, University of Minho, 4710-057 Braga, Portugal

*Correspondence: patricia.matos@tecnico.ulisboa.pt

Abstract: The present study focuses on hydrogeochemical data from a Hercynian granitic mass where a new underground hydraulic circuit for a hydropower plant is being developed between two reservoirs and near another existing and operating underground hydraulic circuit. The main goal of this work is to assess the use of multivariate statistics to distinguish groundwater from different flow systems, to aid in evaluating the possible infiltration of water from the existing hydraulic circuit into the new one. Sampling points were selected inside the tunnels (eight points at the new circuit and one point at the existing circuit), near the surface (five points along the new circuit major axis) and on Venda Nova dam reservoir (one point). Based on laboratory determinations of major ions, three distinct groups of waters with different physicochemical characteristics and spatial distribution were established by cluster analyses. The relations of these groups with the characteristics of the occurrences are discussed with the support of principal component analysis. A hydrogeochemical conceptual model for groundwater flow systems related to infiltration and water–rock interaction in this granite rock mass is proposed based on multivariate statistical analyses and geological information.

Supplementary material: Sample point coordinates, data results of physical–chemical parameters and elements, statistical data results of physical–chemical parameters and elements, and isotope data are available at https://doi.org/10.6084/m9.figshare.c.2857303.v1.

Received 21 January 2016; revised 21 January 2016; accepted 01 February 2016

Underground geotechnical works such as tunnels include a range of issues related to the hydrogeological setting. Groundwater inflow can cause obvious difficulties to the construction works, but underground geotechnical works can also affect groundwater characteristics by contamination owing to infiltration from the tunnels (Chae et al. 2008; Mossmark et al. 2015) and by disturbing the drainage system (when there is inflow to the tunnel). Interference by underground works on hydrogeological systems could have adverse effects on the regional groundwater resources, including ecological impacts such as drying of springs and stream flow reduction (Vincenzi et al. 2009; Raposo et al. 2010). The inflow of groundwater in hydraulic tunnels could also have a negative effect on a dam reservoir (Turkm 2003; Ünal et al. 2008).

Therefore, characterization of the groundwater circulation paths is crucial to assess the impact of underground excavations on the hydrogeological conditions and to determine the mitigation measures to be undertaken (namely, impermeabilization measures). This is particularly critical in the case of fractured rocks, where the identification of the geological structures that control groundwater flow paths can be more complex. The present study aims to assess the use of multivariate statistics to distinguish between groundwaters related to different flow systems. For this purpose water samples were collected from an underground hydraulic circuit, a dam reservoir and some superficial water occurrences, so that laboratory and field analysis could be performed.

Materials and methods

Study object

A new repowering boost for an hydropower plant, located in the region of Entre Douro and Minho, in northern Portugal, is currently under way in the left bank of the Rabagão river, between the Venda Nova and Salamonde reservoirs. For this enterprise, a new underground hydraulic circuit c. 4700 m long and of 12 m circular diameter was excavated with explosives (smooth blasting), near another previously operating hydraulic circuit with similar length but 8 m in diameter. Both hydraulic tunnels have, as their main flexible support structures, steel fibre reinforced shotcrete and rock bolting. Some sections, where the main geological faults were intersected, have reinforced concrete lining. Geological and geotechnical monitoring of excavations for the new hydraulic circuit included the detection of potential inflows of water to underground excavation fronts, as well as attempts to establish their origin. It was of the uttermost importance to establish whether these inflows to the new circuit, during its excavation, represented seepage from the existing circuit, as this could imply changes in the hydraulic conditions that are necessary for the proper functioning of the existing and operating hydraulic circuit.

Geological mapping of the tunnels showed that the new hydraulic circuit is constructed in a porphyritic, medium- to coarse-grained, two-mica Hercynian granite designated by Noronha & Ribeiro (1983) as the Pondras and Borralha granite. Within this granite there are masses of rosy granite and episyenite, medium- to fine-grained granodiorite, biotite and mica schist enclaves, and quartz schist levels (Fig. 1). Several faults were intersected by the tunnels; some of these faults are long and continuous, some are of great thickness, and some have mylonitic fills. The presence of several joints was observed and the main and most continuous ones had approximately north–south to NW–SE trends. In the less deep tunnel sections and near the reservoirs (downstream and upstream portals of the hydraulic circuit) there was evidence of water percolation in the joints, such as oxidation stains. As the excavation progressed in depth, there was evidence of water dripping (associated with fault zones or joints) but without any significant flow.
Water sampling and analysis

Sixteen sampling points were selected, but only 15 were monitored, as one of the points (point 2) was obstructed at the beginning of the monitoring period. The depths of the sampling points are variable, from surface level (free surface of reservoir) to a depth of 473.7 m (Fig. 1). The elevations of these points vary between 838 and 207 m. Points 3–11 (increasing number corresponds to increasing depth) are drain pipes constructed for this purpose and located along the new hydraulic circuit. Point 14 is a piezometer with manometer, located in the operating hydraulic circuit. In this manometer calcite stalactites were observed. The other sample points correspond to superficial water: point 1 corresponds to a borehole well with water pump at 50 m depth; point 12 is located in the Venda Nova reservoir; points 13 and 15 are water springs; point 16 is a drain pipe located 20 m deep in an access tunnel for the new circuit. The presence of water at this last sampling point (number 16) was observed only when it was raining.

Measurements of physical and chemical parameters with field kits (pH, conductivity, temperature, redox potential) were carried out approximately once every 2 weeks between December 2012 and July 2013. In addition, a sample collection campaign for laboratory analyses of pH, conductivity, alkalinity, major anions, major cations and stable isotopes ($\delta^{18}O$ and $\delta^{2}H$) was carried out in April 2013 (at the end of the rainy season).

Multivariate analyses

The multivariate analyses used in this work comprised Piper or diamond diagram and statistical tools (cluster and principal component analyses).

A Piper diagram (Piper 1944) allows plotting of the co-joint relative variation of six variables (expressed as equivalents per unit volume): three major anions (bicarbonate, chloride and sulphate) and three variables related to major cations (calcium, magnesium and the sum of the major alkaline cations, sodium and potassium). It is a classical hydrogeochemical tool to study groundwater composition and its variations but it is limited to a proportional perspective of a limited (and fixed) set of variables.

Multivariate statistical techniques in hydrogeochemical studies attempt to analyse relationships between dissolved constituents in groundwater and environmental factors such as the composition of the bedrock or human intervention. These statistical studies do not necessarily establish relations of cause and effect, but present information in a simplified form, allowing the development of plausible hypotheses for data interpretation (Ashley & Lloyd 1978).

The multivariate statistical techniques used in this study were cluster analysis and principal component analysis. Cluster analysis is used for the identification of groups in a set of objects based on the results of diverse characteristics (variables). It is based on the
similarity between objects and ranks them hierarchically into groups
considering joint results from a set of variables for each object
(Davis 1986). To use this technique, the dataset must be
heterogeneous and present similar variance (Usunoff & Guzman-
Guzman 1989). To achieve this premise, the data were standardized.

Principal component analysis (PCA) is a method that uses a series
of mathematical transformations to reduce the number of variables
in a data matrix by linear combination of one or more variables. In
this way it is possible to select the parameters that explain the
variance observed in the matrix (Güller et al. 2002). To avoid the
excessive influence of variables with higher variance, data for this
analysis were also standardized. In PCA, variables can be treated as
active or supplementary (StatSoft 2013). The principal components
will be computed using only active variables. Supplementary
variables can eventually be projected later onto the vector subspace
generated by the factors thus computed and the relations of these
supplementary variables with the proposed principal components
can then be discussed.

Results and discussion

Descriptive statistics

Results of field monitoring at 2 week intervals did not show any trend
in the monitored parameters. Isotopic analysis did not distinguish
samples as all of them showed meteoric water characteristics.

Laboratory results showed that all groundwater samples had a
relatively low mineralization, with electrical conductivity lower
than 274 µS cm\(^{-1}\). Most samples showed pH values between 6.7
and 8 and temperature values between 10 and 21°C.

Bicarbonate levels stood out from those of other ions, reaching
almost 120 mg l\(^{-1}\), followed by silica with a maximum value of
50 mg l\(^{-1}\). Potassium presented the lowest contents, followed by
fluoride and chloride. In the case of nitrate, some high values
differed by one order of magnitude from the third quartile value.
These nitrate values are possibly a consequence of contamination by
explosive materials used in excavation, and hence this anion was not
considered for this study.

Hydrochemical facies

Laboratory results for major ions were plotted in a Piper diagram
using AquaChem software (Fig. 2) for the characterization of the
hydrochemical facies. The trilinear plot showed two distinct groups of
water with hydrochemical affinity: one group of sodium bicarbonate–
chloride type (Na–Ca–Cl–HCO\(_3\)), which included samples 1, 12, 13,
15 and 16; and a second group in which samples 3–11 and 14 were
identified as sodium–calcium bicarbonate type (Na–Ca–HCO\(_3\)).

Upon preliminary analysis, these results suggest that the
superficial water that initially infiltrates the rock mass is of a
sodium bicarbonate–chloride type, which gradually changes to a
bicarbonate type, ranging between calcium–sodium and sodium–
calcium waters as it travels along its path through the rock mass.

Cluster analysis

Agglomerative cluster analysis was applied using Euclidean
distance for degree of similarity and Ward’s method for grouping,
and considering the following variables: sodium, potassium,
calcium, magnesium, chloride, sulphate, silica, fluoride and
bicarbonate. The resulting tree diagram shows the existence of
groups 3 (Fig. 3). Group 1 includes sampling points 1, 12, 13, 15
and 16, related to more superficial waters (spring, well and reservoir
water, and collected rainwater). Group 3 is formed by the deeper
samples (7–10), whereas group 2 is formed by sampling points with
intermediate depth (3–6 and 11) and by sampling point 14 (located
very near the operating hydraulic circuit and at a greater depth than
the other samples in this group).

Principal component analysis

For principal component analysis nine active variables were
considered (sodium, potassium, calcium, magnesium, chloride,
sulphate, silica, fluoride and bicarbonate) and three supplementary
field variables (pH, temperature and electrical conductivity). The
criterion proposed by Cattel (1966), based on a graphic method, was
used to define the number of components (or factors) for the model,
in which the principal components 1–3 accounted for more than 90% of the variance.

In the plot (Fig. 4a) of principal component 1 (PC1) against principal component 2 (PC2), one can distinguish three groups of samples, which coincide with the results of cluster analysis: one group with positive coordinates in PC1 (group 1, superficial points 1, 12, 13, 15 and 16); a second group of samples surrounding the PC1 value of zero (sampling points 3–6, 11 and 14); and a third group with negative coordinates of PC1, which includes samples 7–10.

The plot of PC1 against depth (Fig. 4b) seems to suggest that the more negative values of PC1 correspond to a lesser influence of surface water. Plotting PC1 coordinates against the distance from the existing circuit (Fig. 4c), a good correlation was also observed. These results might indicate that PC1, which represents 71.89% of variance, corresponds to the mineralization of the water and reflects the water–rock interaction. Thus, in general, PC1 values decrease with increasing distance from potential water sources, either from the surface (vertical recharge) or from the existing and operating circuit (lateral recharge). This suggests progressive sodium and bicarbonate enrichment, followed by a similar increase of field parameters (used as supplementary variables) such as pH, temperature and electrical conductivity.

PC2, although representing only 10.87% of variance, clearly separates sample 14 from the remaining samples. Taking into
consideration that this point lies where the existing hydraulic circuit intersects a fault with mylonitic clay fill, these results might reflect the influence of this fault. The consolidation and impermeabilization treatment with cement grout injections, which was carried out on the section where point 14 is located, could also contribute to the distinction of this point from the remaining members of group 2. Considering that the raw materials commonly used for cement production are mainly limestone, silica, clays and small amounts of iron oxides (Taylor 1997), the cement in the surroundings of point 14 could cause the high pH values of the collected water samples (Mossmark et al. 2015) and be an important source of calcium, thus explaining the calcium carbonate stalactites observed on the piezometer.

PC3 represents only 9.57% of variance but it proved to be a relevant component to support the distinction between group 2 (with negative coordinates of PC3) and the other groups (Fig. 4d). It appears that PC3 separates samples with strongly negative silica coordinates (higher silica content in group 2) and with more positive chloride coordinates (higher chloride content in group 3).

Final considerations

Statistical analysis contributed to understanding the groundwater flow systems in a Hercynian granite rock mass intersected by two underground hydraulic circuits. It was possible to distinguish three hydrochemical groups of groundwaters by cluster analysis and explain them by principal component analysis. Group 1 consists of points at the surface (or very near the surface), and is characterized by low conductivities (being the least mineralized of all the 15 points monitored) and an acidic pH. The distinction of the other two groups of samples was shown by principal component analysis to be related to (1) distance from the existing hydraulic circuit, (2) depth from the surface and (3) chloride content. The distance from the existing circuit and the depth from the surface reflect the distance from the recharging source. In both types of recharge, water circulation takes place through the network of fractures and faults that cross the region.

Considering the location of the sampling sites and their geological and structural framework, two distinct hydrogeochemical regions bounded by important and significant regional structural alignments are suggested, referred to here as region A and region B. This proposed conceptual model is presented in the model for the new hydraulic circuit (Fig. 5).

It is possible that in region A there are two recharge fronts of groundwater: one coming from the surface (infiltration of rainfall) and the other resulting from a lateral recharge induced by the existing and operating hydraulic circuit. The recharge process can be facilitated by the existence of subvertical fractures that, given their north–south to NW–SE orientation, intersect the two hydraulic circuits. In these short subvertical and sublateral paths, groundwater become more mineralized as distance from source and depth increase. However, water samples from group 2 and from group 1 (surface) maintain similar chemical and physical characteristics given that the time of contact with the rock for sampling points of group 2 is not very high.

Groundwater sampling points from group 3 are located in region B, where they reach greater depths and distances from the existing hydraulic circuit. These groundwaters show effects of a more evolved water–rock interaction with higher mineralization and higher temperatures (which are commonly found in deep groundwater of this region), conditioned by two major geological faults, F14 and F31, as shown in Figure 5.

With the application of multivariate data analysis techniques and the consideration of geological information from the excavations of the new hydraulic circuit, it was, hence, possible to propose two
distinct hydrogeochemical regions (A and B), defining a conceptual model of groundwater flow path. This model suggests that there could be hydrogeological contributions from the existing hydraulic circuit until a certain depth and distance from the new hydraulic circuit. However, it seems that at greater depths and greater distances between the two hydraulic circuits, and considering also the local geological environment, any seepage coming from the existing and operating hydraulic circuit tends to decrease or even ceases to exist.

Acknowledgements and Funding
The Lab2PT, Landscape, Heritage and Territory Laboratory (UID/AUR/04509/2013) and the Centre of Geological Research, Management and Valorisation of Resources of the School of Sciences (CIG-R; PEst-OE/CTE/UI0697/2011; PEst-OE/CTE/UI0697/2014) received support from the FCT – Fundação para a Ciência e Tecnologia (Portugal) with Portuguese and European Union funds (FEDER, Programa Operacional Factores de Competitividade – COMPETE, project POCI-01-0145-FEDER-007528). The authors thank A. da Silva Lima for scientific advice, and the owner of the Venda Nova Hydropower Plant, EDP, for access to its construction elements.

Scientific editing by Stephen Buss, Daren Gooddy

References
Ashley, R.P. & Lloyd, J.W. 1978. An example of the use of factor analysis and cluster analysis in groundwater chemistry interpretation. Journal of Hydrology, 39, 355–364.
Cattel, R.B. 1966. The scree test for the number of factors. Multivariate Behavioral Research, 1, 245–276.
Chae, G., Yun, S., et al. 2008. Hydrochemistry of urban groundwater, Seoul, Korea: The impact of subway tunnels on groundwater quality. Journal of Contaminant Hydrology, 101, 42–52.
Davis, J.C. 1986. Statistics and Data Analysis in Geology. Wiley, New York.
Güller, C., Thyne, G.D., McCray, J.E. & Turner, A.K. 2002. Evaluation of graphical and multivariate statistical methods for classification of water chemistry data. Hydrogeology Journal, 10, 455–474.
Mossmark, F., Ericsson, L.O., Norin, M. & Dahlström, L. 2015. Hydrochemical changes caused by underground constructions—a case study from the Kattleberg rail tunnel. Engineering Geology, http://dx.doi.org/10.1016/j.engeo.2015.03.004.
Noronha, F. & Ribeiro, M.L. 1983. Carta Geológica de Portugal na escala de 1/500000. Notícia Explicativa da folha 6-A (Montalegre). Serviços Geológicos de Portugal, Lisboa [in Portuguese].
Piper, A.M. 1944. A graphic procedure in geochemical interpretation of water analyses. Transactions, American Geophysical Union, 25, 914–923.
Raposo, J.R., Molinero, J. & Dafonte, J. 2010. Quantitative evaluation of hydrogeological impact produced by tunnel construction using water balance models. Engineering Geology, 116, 323–332.
StatSoft. 2013. Electronic Statistics Textbook. StatSoft, Tulsa, OK.
Taylor, H.F.W. 1997. Cement Chemistry, 2nd edn. Thomas Telford, London.
Turkmens, S. 2003. Treatment of the seepage problem at the Kalecik Dam (Turkey). Engineering Geology, 68, 159–169.
Ünal, B., Çelik, M. & Yildirim, T. 2008. Evaluation of seepage problems of the Armagan Dam, Kirkkaleli, northwestern Turkey. Bulletin of Engineering Geology and the Environment, 67, 335–344.
Usunoff, E.J. & Guzman-Guzman, A. 1989. Multivariate analysis in hydrochemistry: an example of the use of factor and correspondence analyses. Groundwater, 27, 27–34.
Vincenzi, V., Gargni, A. & Goldscheider, N. 2009. Using tracer tests and hydrological observations to evaluate effects of tunnel drainage on groundwater and surface waters in the Northern Apennines (Italy). Hydrogeology Journal, 17, 135–150.

Multivariate statistical analysis of hydrogeochemical data