Development of a powerful approach for classification of surface waters by geochemical signature

FILELLA, Montserrat, POMIAN-SRZEDNICKI, Ian, NIREL, Pascale M.

FILELLA, Montserrat, POMIAN-SRZEDNICKI, Ian, NIREL, Pascale M. Development of a powerful approach for classification of surface waters by geochemical signature. *Water Research*, 2014, vol. 50, p. 221-228

DOI : 10.1016/j.watres.2013.11.046

WATER RESEARCH 50 (2014) 221–228

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/watres

Development of a powerful approach for classification of surface waters by geochemical signature

Montserrat Filella [a,*], *Ian Pomian-Srzednicki* [b], *Pascale M. Nirel* [c]

[a] *Institute F.-A. Forel, University of Geneva, Route de Suisse 10, CH-1290 Versoix, Switzerland*
[b] *Rue du Gothard 2, CH-1225 Chêne-Bourg, Switzerland*
[c] *Service de l'écologie de l'eau, Av. Sainte-Clotilde 23, CH-1211 Geneva 8, Switzerland*

ARTICLE INFO

Article history:
Received 18 June 2013
Received in revised form
25 November 2013
Accepted 28 November 2013
Available online 7 December 2013

Keywords:
Rivers
Chemical signature
Water classification
Geology
Uranium
Barium

ABSTRACT

Easy identification of chemical signatures characteristic of water systems has become a major issue in the field of environmental protection and management. We propose an exploratory method, exclusively based on the statistical analysis of river water composition, capable of characterizing river waters in a given watershed through their chemical composition, as well as of detecting modifications, even when not related to pollution sources. Although the method is based on well-known statistic techniques (Principal Component Analysis and Linear Discriminant Analysis), and therefore is very simple and straightforward to apply, it goes far beyond the common data reduction use of these techniques. Its capabilities are illustrated through its application to rivers in Canton Geneva, Switzerland, a hydrographical network consisting of 310 km of waterways with 250 streams and rivers. The procedure results in a very satisfactory classification of watersheds, in our case by using only two geochemical indicators: U and Ba concentrations. The method also makes it possible to follow the seasonal evolution of river regimes or the effect of wetlands on river water composition.

1. Introduction

It is widely-accepted that knowledge of the 'natural' chemistry of surface waters is fundamental to identifying substances associated with pollution (Menzi et al., 2009). Moreover, current legislation on water bodies (e.g. the European Water Framework Directive (European Union, 2000), Swiss LEau (Swiss Confederation, 1991)) emphasizes the need for a "good chemical status" in order to achieve the desired "good ecological status". Since current legislation is mostly based on River Basin Management Planning, catchment management plans are required to address these questions in each catchment. In practice, this means that the possibility of easily identifying the chemical signature characteristic of a water system has become a major issue with significant impact in the field of protection and management of our environment.

Although water chemical composition results from multiple factors (i.e., atmospheric, geological, biological, etc.), in the absence of anthropogenic pollution, geology is the main supplier of water components. The chemical signature of geological origin mainly originates in the interactions

---

* *Corresponding author.*
E-mail address: montserrat.filella@unige.ch (M. Filella).

0043-1354/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.watres.2013.11.046

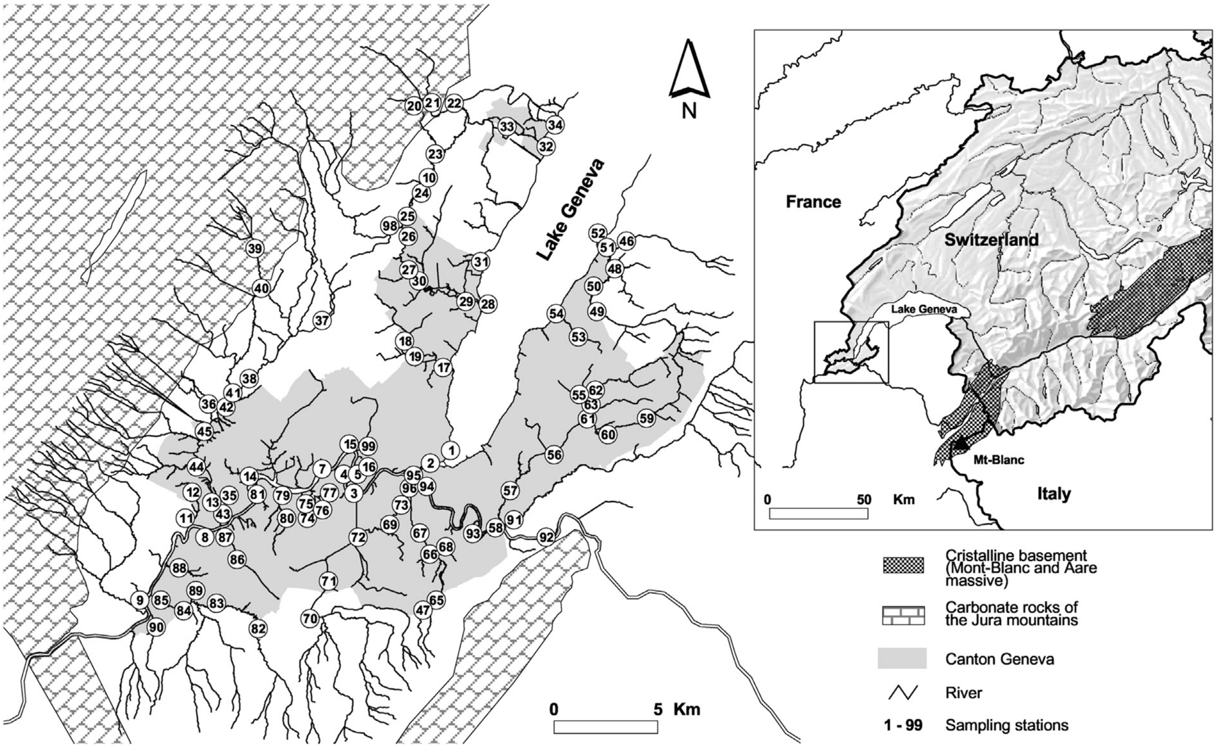

Fig. 1 – Simplified geological map of the Geneva canton with location of sampling sites. A detailed list of sampling sites can be found in Table S3.

between water, soil and subsoil (Drever, 1997) and, in practice, identifying it is far from straightforward because of the many confounding factors. Indeed, the processes controlling the chemical composition of a given water body are dependent on both bioclimatic variables (e.g., temperature, frequency and amplitude of rainfall, type and abundance of microflora) and properties of the physical environment (e.g., porosity, hydraulic conductivity) (Garrels and Mackenzie, 1967; Meybeck, 1987).

Here we propose a method exclusively based on the statistical analysis of river water composition which is able to classify river waters in a given watershed through their chemical composition. The method is explained by applying it to rivers in Canton Geneva, Switzerland. These rivers are an ideal test system because they are in contact with a wide range of terrains with contrasting geochemical, hydrological and structural properties. Although statistical tools, such as Principal Component Analysis (PCA), have been extensively used in the environmental sciences, they have been mostly used for data reduction purposes only. Here we propose a new methodological approach that, on the basis of these techniques, allows the easy and robust classification and follow-up of water bodies without the need of performing endless and expensive analysis. This approach is a powerful tool for the characterization of watersheds according to their hydrological and ecological region mandatory under the above-mentioned legislation (Omernik, 2004; Menzi et al., 2009).

Table 1 – Description of sites studied.

Type	Number of streams	Number of sampling sites	Number of samples
Alps	2	11	452 (22%)
Jura	16	30	804 (40%)
Plain	19	31	598 (29%)
Wetland	7	17	175 (9%)
Total	44	89	2029

2. Materials and methods

2.1. *Sites studied*

The hydrographical network of the Geneva area consists of 310 km of waterways, comprising 250 streams and rivers. The major watercourse, and final collector for all waters, is the Rhone River. The sources of all the streams, with the single exception of the Seymaz, are outside the canton, for the most part in neighbouring France. The vast majority of the streams (196) are less than 1 km long. The surface area of the Canton of Geneva (282 km^2) is primarily devoted to agriculture (40%), buildings (33%) and woodland (12%), with the remaining area being occupied by the lake and surface waters (14%) and miscellaneous uses (1%) (République et Canton de Genève, 2012).

The geological features of the Canton of Geneva also vary considerably (Wildi and Pugin, 1998; www.toposhop.admin.ch/fr/shop/products/maps/geology/gk500/vector_1#). Located at the south-western tip of Lake Geneva, it is formed by a drainage basin surrounded by the Jura and Alps mountain ranges (Fig. 1). Waters flowing in Canton Geneva are

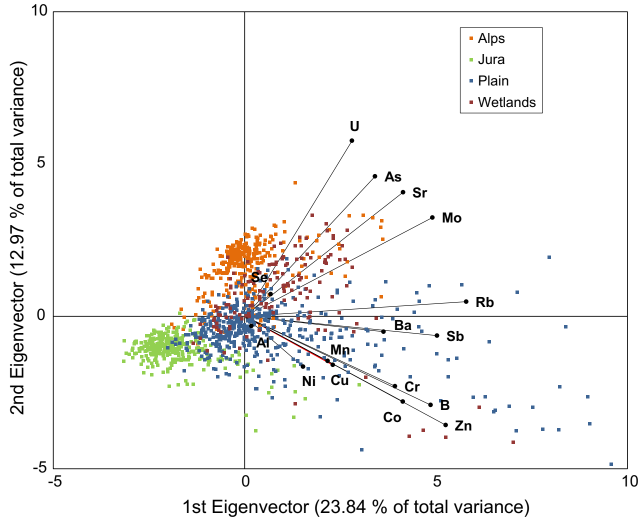

Fig. 2 – PCA analysis. Scores and loads on PC2 versus PC1.

Table 2 – Median and 80% Percentile values of Ba and U concentrations.

Type	Ba/μg L^{-1}		U/μg L^{-1}		*n*
	Median	80th percentile	Median	80th percentile	
Alps	27.9	57.0	1.0	1.6	452
Jura	12.2	20.6	0.4	0.5	804
Plain	53.0	67.9	0.8	1.1	598
Wetland	57.0	66.9	1.3	2.0	175

Table 3 – Results from method testing obtained with linear discriminant analysis.

	Types	Predicted group membership				Total
		Alps	Jura	Plain	Wetland	
Original count	Alps	449	0	2	1	452
	Jura	3	777	24	0	804
	Plain	3	20	511	64	598
	Wetland	1	1	48	125	175
Percentage	Alps	99.3	0.0	0.4	0.2	100
	Jura	0.4	96.6	3.0	0.0	100
	Plain	0.5	3.3	85.5	10.7	100
	Wetland	0.6	0.6	27.4	71.4	100

therefore also in contact with a wide range of terrain types which have markedly different hydrological and lithological properties. Four main waters can be identified: (i) waters under the influence of the Alpine belt, which is quite heterogeneous and includes a large variety of igneous, metamorphic, evaporite and carbonate rocks, but is characterized mainly by large bodies of igneous rock where water flows through fractures under conditions of elevated pressure and temperature (Arve and Rhone rivers and Lake Geneva); (ii) rivers of the right bank, which are fed by springs emerging from karst carbonate rocks of the Jura mountains where waters are mineralized by dissolution of relatively soluble limestone and dolomite but show relatively low levels of dissolved elements due to their short residence time in the karst (Allondon and Versoix watersheds); (iii) rivers flowing through zones with heterogeneous detrital sediments, mainly tertiary (molasses) and quaternary (glacial and alluvial deposits),

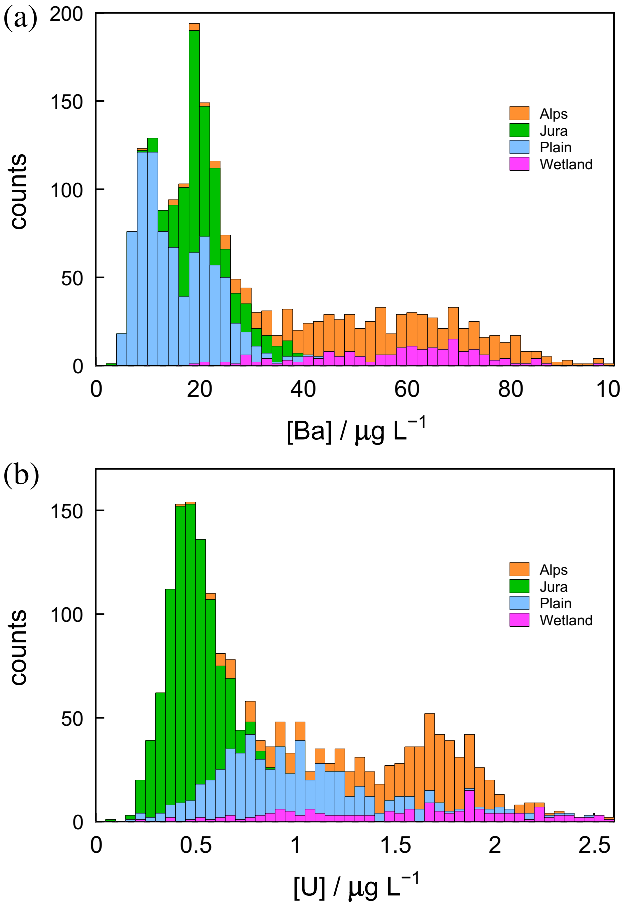

Fig. 3 – Distribution of uranium (a) and barium (b) concentration values in rivers of Canton Geneva, Switzerland, from 1994 to 2002, classified according to four different types of watershed.

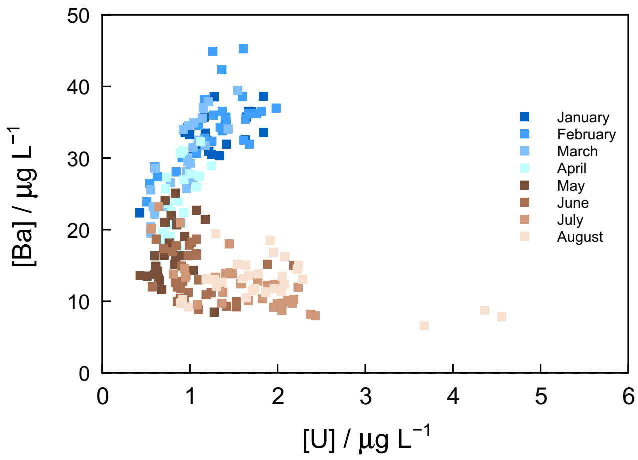

Fig. 4 – Seasonal evolution of the chemical composition of River Arve waters from 1994 to 2006. The change from a pluvial to a nival system along the year can be observed.

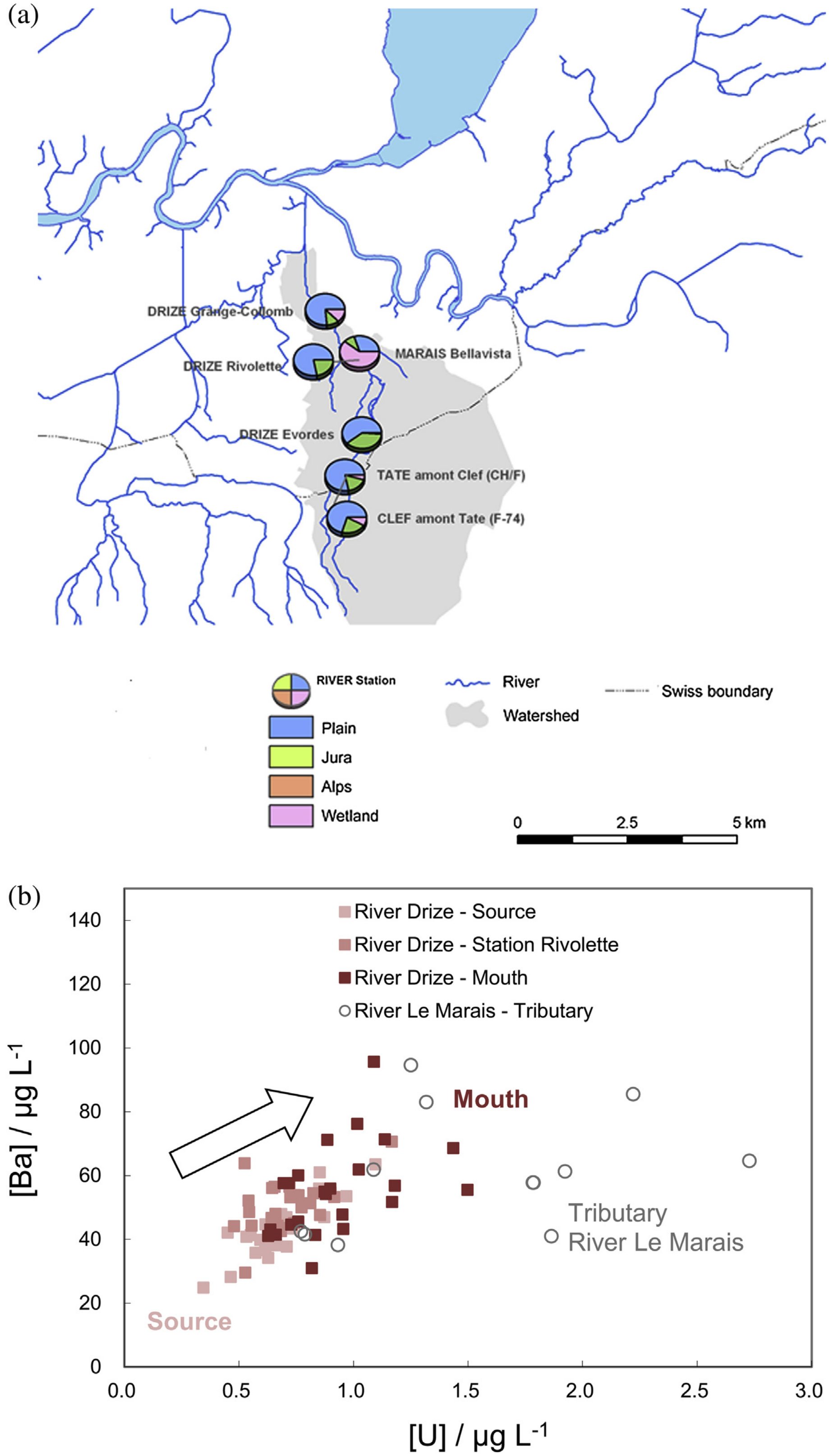

Fig. 5 – (a) Automatic classification of water samples from the Drize river obtained with LDA. Pie-charts show the results in terms of probability of belonging to one of the four water types. (b) U–Ba diagrams where the effect of the presence of wetlands in some areas of the watershed is shown.

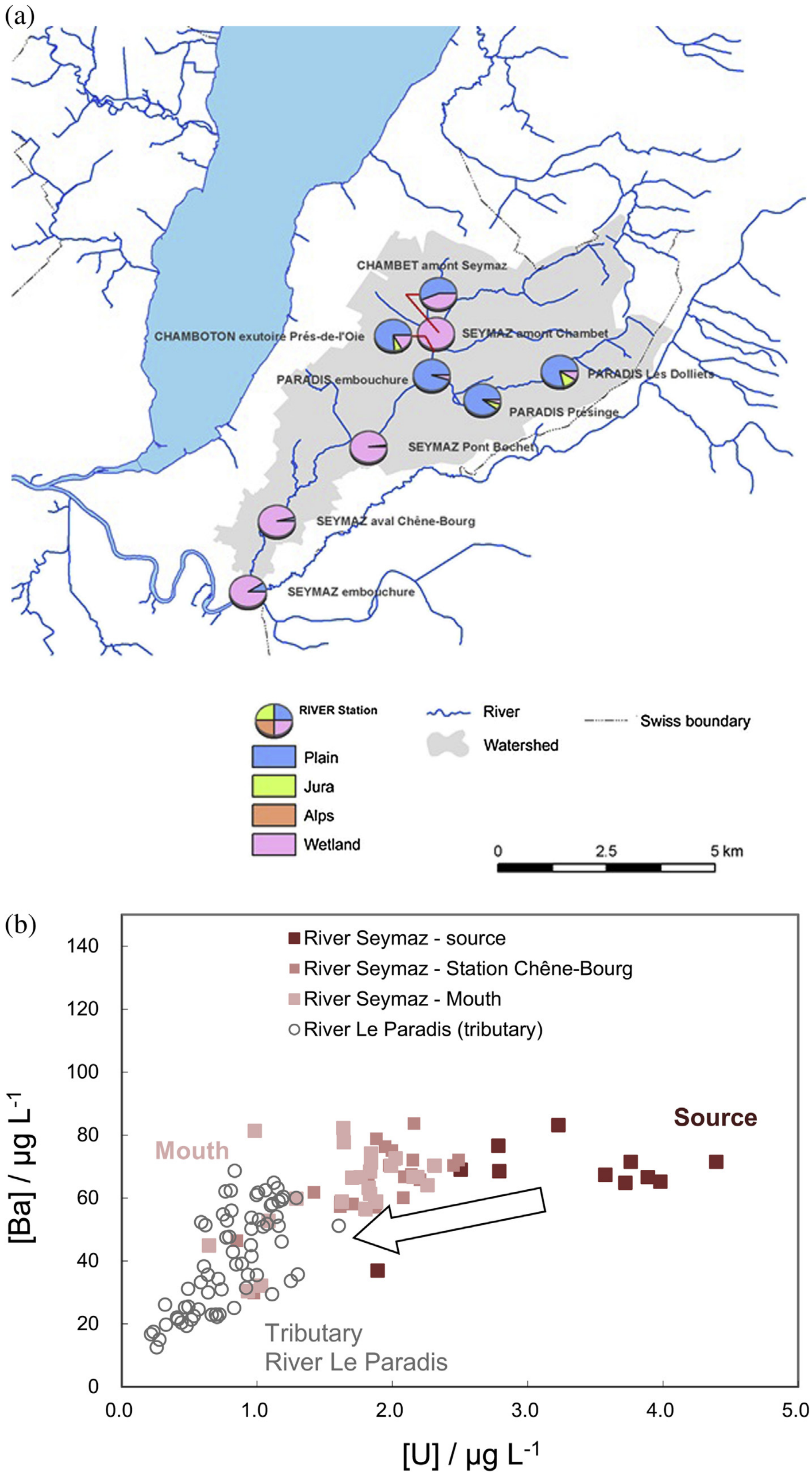

Fig. 6 – **(a) Automatic classification of water samples from the Seymaz river obtained with LDA. Pie charts have the same meaning as in Fig. 5a. (b) U–Ba diagrams where the effect of the presence of wetlands in some areas of the watershed is shown.**

originating in the Alps and the Jura mountains and enclosed in a clay matrix; waters and groundwater circulate slowly in the plain, giving them time to acquire a considerable load of dissolved elements (Laire, Aire, Drize rivers and Rhone tributaries); (iv) waters flowing on organic-rich sediments accumulated in wetland environments (Hermance, Seymaz rivers). The location of the rivers and sampling points is shown in Figure S1 (Supporting Information).

2.2. *Sampling sites*

The element concentration data used in this study were collected as part of a water-quality monitoring program (1994–2002) set up by the *Service de l'Ecologie de l'Eau* (SECOE) of the Canton of Geneva, Switzerland. The survey network comprises ninety sampling sites, shown in Fig. 1. The origin and number of samples are presented in Table 1. These 89 sites, covering 44 streams accounting for approximately two thirds of the network, show a wide variation in both physico-chemical and biological qualities (AFNOR, 2004; Liechti et al., 2004; Hurlimann and Niederhauser, 2007).

2.3. *Sample collection and treatment*

Individual water samples were hand-collected monthly in polyethylene vials that had been previously soaked in 10% v/v nitric acid (Suprapur, Merck) and thoroughly rinsed with Milli Ω-Milli Q water (Millipore). Samples were filtered within few hours after sampling through 0.45 μm filters (Millex Durapore, Millipore) and acidified prior to analysis with 2% v/v nitric acid (Suprapur, Merck).

2.4. *Analytical method*

Concentrations of 15 trace elements (Al, As, B, Ba, Co, Cr, Cu, Mn, Mo, Ni, Rb, Sb, Sr, U, Zn) were measured by Inductively Coupled Plasma Mass Spectrometry (ICP-MS) on a VG-PQ2 spectrometer in its standard configuration, with a Meinhard nebulizer. Concentrations were determined in semi-quantitative mode (with internal standard calibration). The analytical accuracy of 2029 measurements was estimated from replicated analysis ($n = 136$) of the certified standard SLRS-4 (NRC). The certified and measured values are given in Table S1 in the Supporting Information and are satisfactory.

2.5. *Data processing*

Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) calculations were performed with the SPSS 12 software package. These two multivariate techniques were applied to a data matrix consisting of 2029 observations (samples) containing the values of 15 dissolved trace elements. We performed all calculations on the raw dataset without applying any prior normalization, exclusion, or weighting pre-processing. As a result, the procedure presented here is very simple and straightforward. Our aim of keeping the approach as simple as possible was not achieved at the cost of rigor: PCA does not require data pre-processing when using the correlation matrix since data are intrinsically normalised, while LDA requires normal distributions but is known to be robust to modest violations of normality (Davis, 1986; Huberty, 1994; Tabachnick and Fidell, 1996; Brown, 1998).

3. Results and discussion

The procedure we suggest follows three steps:

1) Application of PCA as a tool for data reduction (i.e., removal of redundant variables) and detection of underlying structures in the dataset;
2) Choice of relevant tracer elements and definition of a model linking the chemical signature of surface waters with the underlying geology;
3) Model testing using LDA.

3.1. *Data reduction*

PCA was applied to the whole dataset. Fig. 2 shows the projection of variables on the factorial plane defined by the first two PCA eigenvectors where several groups of well-correlated elements appear. These groups correspond to families of elements with similar behaviour in aqueous media: alkali and alkaline earth metals (Rb and Ba), divalent metals (Co, Ni, Cu and Zn) and anion-forming elements (As, Mo and U). Aluminium contributes very little to the definition of the two eigenvectors. Some elements are often combined in waters in contact with specific geological formations, as is the case for Rb, Sr, and Ba, commonly associated with sedimentary rocks, and for Mo, U and W, often associated with igneous rocks. Projecting observations on the factorial plane (Fig. 2) makes it possible to distinguish several clusters of water samples with a similar chemical composition. These initial results suggest that a classification of waters according to their chemistry is possible; the aim of the procedure thus becomes to relate it to the geology.

3.2. *Choice of natural tracers and model definition*

A natural tracer is a substance present in water that comes from a defined natural source. Ideally, a natural tracer must satisfy the following conditions: (i) be typical of certain geological formations, (ii) be soluble in water, (iii) behave conservatively (under little or no influence from changes in physico-chemical conditions and no interaction with biological cycles), (iv) be rarely used in human activities (no pollution sources); (v) have reliable methods of analysis.

Members of the alkali and alkaline earth metal groups and elements with valences V and VI are those that best meet these requirements. PCA shows that U, a valence VI element, is the main contributor to the 2nd component and Ba, an alkaline earth metal, contributes heavily to the 1st component (see Fig. 2). U and Ba are therefore good candidates as natural tracers. Both are geogenic elements subject to little influence from human activities and show a behaviour essentially conservative in freshwaters (Windom et al., 2000; Seyler and Boaventura, 2003; Schmidt, 2005; Féraud et al., 2009; Roeske et al., 2012).

Fig. 3a and b show U and Ba concentrations in rivers of Canton Geneva, classified according to different types of watershed. Their distribution clearly results from the contribution of several groups/populations. Mean U and Ba concentrations for the different groups are shown in Table 2. Distinct patterns, corresponding to clusters observed with PCA, can be recognized, thus confirming the suitability of the chosen tracers. The origin of the clustering has therefore been explained by a model able to relate the structures observed in the data to the geology of the area being studied. When locating the samples on the geological map (Fig. 1), it can be observed that: (i) *Alpine* river waters are characterized by high U and low Ba concentrations, (ii) *Jura* river waters (e.g., Allondon and Versoix) have low U and Ba concentrations, (iii) *Plain* river waters have medium U and high Ba concentrations and (iv) streams draining former wetlands, such as Seymaz or Nant des Marais, show concentrations up to five times higher than those usually measured in *Plain* rivers. This may be due to the fact that U accumulated in rich peat organic matter is being released by peat mineralization after wetland draining. A further measure of the discriminatory ability of each variable has been obtained by applying LDA. Results are shown in Table S2 and again confirm that the tracers chosen are appropriate.

3.3. *Procedure testing*

As a final testing step, an automatic classification of each observation (2029 samples) was performed by LDA and compared to the classification previously established. Results are shown in Table 3. The resulting score of 92% correctly classified waters is very satisfactory and proves that both Ba and U are reliable geochemical indicators for the watersheds studied and are preserved from anthropogenic and bioclimatic interference. Although the general applicability of these particular tracers is expected, it needs to be tested system by system. Nonetheless, it should be noted that the method is independent of the particular tracers chosen.

3.4. *Examples of application*

The method proposed not only allows the classification of water bodies but also the easy follow-up of seasonal evolutions of a river regime or the evaluation of the effect of the presence of wetlands on river water composition.

It is well established that the regime of a river can change along its course (e.g., from glacial to pluvial). However, the change of regime for a given river over a year is less well-characterized. This change is not without consequences when evaluating the river's good chemical status because the chemical characteristics of the waters can be markedly different. Our method allows such changes to be followed in a simple way. Fig. 4 shows the case of the Arve River, whose regime clearly shifts from pluvial in winter to nival in summer. In winter the river is fed by rain (or groundwaters) draining detritic zones rich in Ba while in summer it is mainly fed by high-altitude waters draining granitic zones rich in U in the Montblanc Massif. In this particular case, since the Arve river waters are used to load an aquifer which provides drinking water for the city of Geneva, the availability of a simple method that makes it possible to check the status of the river is clearly very important.

As shown in Table 3, wetland-type waters are not as well characterized as the others. A closer look at this type of river suggests that the underlying reason is the limited extent of the wetland areas in the watersheds studied, which, in practice, implies that no river's watercourse lies entirely within a wetland zone. Interestingly, the method we propose is very sensitive and makes it possible to follow the changing type of the zone simply by monitoring the relationship between U and Ba concentrations. Two examples are shown in Figs. 5 and 6. Fig. 5a shows the case of the Drize River with most of the measuring stations showing a predominant Plain character, except when it receives the waters of the Nant de Marais, a small affluent draining a wetland area (Fig. 5b). The opposite phenomenon can be observed in the Seymaz River (Fig. 6a and b), where the upper part of the catchment shows a strong wetland character that is lost after its junction with the Paradis affluent.

4. Conclusions

When large amounts of data are available, as is often the case for long-term or large-scale water survey programs set up by governmental institutions such as the U.S. Geological Survey in the United States or cantonal and federal bodies in Switzerland, application of the methodology developed and tested in this study will allow surface waters to be classified as a function of their location and chemical status despite the interference caused by the mixing of different tributaries and dilutions that occur during floods. Any impact affecting the geochemical composition of a hydrological and ecological region can be evidenced by this approach (e.g., significant inputs of water from an unknown origin, changes in the aquifer recharge regime, etc.). Water composition has a major influence on organisms and ecosystems and is a major factor in defining hydrological and ecological regions. Three important advantages of our method are that (i) it can be applied without knowing discharge values – which are often missing or measured with low precision –, (ii) its simplicity and robustness make it suitable for automatic identification of incidents in big data sets, and (iii) it can be used to monitor data even in highly populated zones. In this sense, it should be pointed out that the method has performed very well in Canton Geneva, a particularly 'difficult' case, with watercourses located in a small geographic but densely populated area. On the other hand, note that our method is not suitable for detecting point toxic element pollution sources.

Acknowledgements

We thank an anonymous reviewer for his/her valuable comments.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.watres.2013.11.046

REFERENCES

AFNOR, 2004. Détermination de l'indice biologique global normalisé (IBGN). NF T 90–350.

Brown, C.E., 1998. Applied Multivariate Statistics in Geohydrology and Related Sciences. Springer, Berlin.

Davis, J.C., 1986. Statistics and Data Analysis in Geology, second ed. John Wiley & Sons.

Drever, J.I., 1997. The Geochemistry of Natural Waters. Surface and Groundwater Environments, third ed. Prentice Hall.

European Union, 2000. Directive 2000/60/EC of the European Parliament and of the Council of 23 October 2000 Establishing a Framework for Community Action in the Field of Water Policy. Official Journal (OJ L 327).

Féraud, G., Potot, C., Fabretti, J.-F., Guglielmi, Y., Fiquet, M., Barci, V., Maria, P.-C., 2009. Trace elements as geochemical markers for surface waters and groundwaters of the Var River catchment (Alpes Maritimes, France). C. R. Chimie 12, 922–932.

Garrels, R.M., Mackenzie, F.T., 1967. Origin of the chemical compositions of some springs and lakes. In: Stumm, W. (Ed.), Equilibrium Concepts in Natural Water Systems, Advances in Chemistry Series, vol. 67. American Chemical Society, Washington DC, pp. 222–242.

Huberty, C.J., 1994. Applied Discriminant Analysis. John Wiley & Sons, New York.

Hurlimann, J., Niederhauser, P., 2007. Méthodes d'analyse et d'appréciation des cours d'eau. In: Diatomées Niveau R (région). État de l'environnement UV-0740-F. OFEV, Bern, Switzerland. www.environnement-suisse.ch/uv-0740-f. in German: UV-0740-D.

Liechti, P., Frutiger, A., Zobrist, J., 2004. Méthodes d'analyse et d'appréciation des cours d'eau en Suisse: Module chimie. Analyses physico-chimiques niveau R & C. OFEFP, Bern, Switzerland.

Meybeck, M., 1987. Global chemical weathering of surficial rocks estimated from river dissolved loads. Am. J. Sci. 287, 401–428.

Menzi, C.A., Ziccardi, L.M., Lowney, Y.W., Fairbrother, A., Shock, S.S., Tsuji, J.S., Hamai, D., Proctor, D., Henry, E., Su, S.H., Kierski, M.W., McArdle, M.E., Yost, L.J., 2009. Importance of considering the framework principles in risk assessment for metals. Environ. Sci. Technol. 43, 8478–8482.

Omernik, J.M., 2004. Perspectives on the nature and definition of ecological regions. Environ. Manage. 34 (Suppl.1), S27–S38.

République et Canton de Genève, 2012. Annuaire Statistique du Canton de Genève 2012. Office cantonal de la statistique (OCSTAT). www.ge.ch/statistique.

Roeske, T., Bauch, D., Rutgers, M., Loeff, V.D., Rabe, B., 2012. Utility of dissolved barium in distinguishing North American from Eurasian runoff in the Arctic Ocean. Mar. Chem. 132–133, 1–14.

Schmidt, S., 2005. Investigation of dissolved uranium content in the watershed of Seine River (France). J. Environ. Radioact. 78, 1–10.

Seyler, P.T., Boaventura, G.R., 2003. Distribution and partition of traces metals in the Amazon basin. Hydrol. Proc. 17, 1345–1361.

Swiss Confederation, Loi fédérale du 24 janvier 1991 sur la protection des eaux (LEaux), 814.20, http://www.admin.ch/ch/f/rs/c814_20.html.

Tabachnick, B.G., Fidell, L.S., 1996. Using Multivariate Statistics. Harper and Collins.

Wildi, W., Pugin, A., 1998. Histoire géologique du relief du Bassin Lémanique. Arch. Sci. 51, 5–12.

Windom, H., Smith, R., Niencheski, F., Alexander, C., 2000. Uranium in rivers and estuaries of globally diverse, smaller watersheds. Mar. Chem. 68, 307–321.