

Title page

- i. Identification of potential end-members and their apportionment from downstream high-frequency chemical data
- ii. Identification of hydrochemical end-members
- iii. José Manuel Tunqui Neira^{1,2}, Jean-Marie Mouchel², Gaëlle Tallec^{1,*} Vazken Andréassian¹.
- iv. ⁽¹⁾Université Paris-Saclay, INRAE, HYCAR Research Unit, 92761 Antony, France.
⁽²⁾Sorbonne Université, CNRS, EPHE, UMR Metis 7619, Paris, France.
^(*)Corresponding author : gaelle.tallec@inrae.fr
- v. *Acknowledgments.* The first author acknowledges the Peruvian Scholarship Cienciactiva of CONCYTEC (*grant no. 099-2016-FONDECYT-DE*) for supporting his PhD study at INRAE and Sorbonne University. The authors acknowledge the EQUIPEX CRITEX program (*grant no. ANR-11-EQPX-0011*) for the data availability, and thank Dr Charles Perrin and Dr Guillaume Thirel for their review.

Abstract

High-frequency chemical measurements in streamwater give detailed information on the different hydrochemical processes occurring in the catchment. They can provide clues about the end-members (water sources) whose mixing produces the ionic stream concentrations observed. To identify and quantify the end-members and their respective contribution to streamwater chemistry, we developed a new methodology, termed “IQEA” (Identification and Quantification of End-members and their Apportionment). It consists in the resolution by a minimization process of an objective function based on downstream high-frequency chemical data and the size of the shape of a pre-defined fixed number of three end-members. We applied the IQEA method on the high-frequency concentration dataset of nitrates and calcium from the ORACLE-Orgeval Observatory in France. The results were compared with field data obtained from various potential end-members. The IQEA methodology yielded robust and very satisfactory results, apart from an exceptional flood event that occurred during the 2 years of measurements used in this study. While the IQEA model treats concentration and discharge datasets independently, its results show a strong relationship between the dominant end-member computed and the hydrological features.

Keywords

High-frequency chemical data, ORACLE-Orgeval Observatory, cluster analysis, end-member quantification, nitrates, calcium.

1 Introduction

Where do the ionic concentrations in streamwater come from? Since it is possible to measure them, this question has both intrigued and challenged researchers (Lenz & Sawyer, 1944). Knowledge of water sources (named “end-members”) and their contribution in the stream is especially important for detailed analysis of the hydrochemical behavior of a catchment and the related processes. Hydrological pathways, biochemical processes, non-point source pollution, or ecological functioning, among others, can be studied in an efficient manner with end-members (Genereux et al., 1993; Liu et al., 2008; Miller et al., 2017). Rivers display different chemical signatures resulting from several end-members, collected from different pools, spatially and temporarily distributed across the catchment (Chanat et al., 2002; Evans & Davies, 1998; Hrachowitz et al., 2016; Miller et al., 2017; Probst, 1985). These pools, including near-surface and deep groundwater areas, such as soils and fractured bedrock aquifers, respectively, are connected to each other (Dwivedi et al., 2019). Because of these

connections, the chemical solutes, collected from the end-members, interact and mix with each other (Genereux et al., 1993).

These interactions and mixing processes are subject to various factors:

- i. the precipitation that provides exogenous chemical solutes to the runoff and soil pools (Neal & Kirchner, 2000; Pearce et al., 2015);
- ii. bedrock geology that provides the endogenous chemical signature of the groundwater and soil pools (Frisbee et al., 2013; Gaillardet et al., 1999);
- iii. the hydrological regimes (wet, dry season) that determine whether chemical contributions to the stream water are produced by interaction and mixing of all the pools (Chanat et al., 2002; Rose et al., 2018) or of a selected number of them (Muñoz-Villers & McDonnell, 2012; Zhang et al., 2013);
- iv. the catchment size, which determines the residence time of the solutes in the pools (Ameli et al., 2017; Frisbee et al., 2011);
- v. the human activities, which modify and/or disturb the chemical contributions of different end-members (Botter et al., 2019; Garnier et al., 2016).

High-frequency measurements provide a detailed record of the interactions between the different pools and both natural and external factors (Duncan et al., 2017; Flourey et al., 2017; Neal et al., 2012). However, the mixing processes and the mixing equation of end-members that follows, used to this day, assume a strong hypothesis and its resolution still raises several questions.

1.1 Formulation of the problem and main resolution techniques

The mixing equation states that the composition of any so-called mixed sample in the hydrological system studied results from the mixing of an unknown proportion of end-members. It also makes the strong hypothesis that: (i) the number of end-members is finite, (ii) their composition is constant, and (iii) chemicals are conservative within the hydrological system after emerging from the end-member waters. Thus, the mixing equation can be written as a mass balance equation for all chemicals in all streamwater samples:

$$C_{ij} = \sum (D_{ik} \cdot \delta_{kj}) \quad \text{Eq. (1)}$$

where the matrix C stands for the composition of the mixed samples (i is an index for the chemical components and j for the mixed samples), the matrix D stands for the composition in i chemical component of end-member waters (k is an index for the potential end-members), and the matrix δ stands for the mixing ratios of end-member waters.

An additional equation is the conservation of water. This equation states that the sum of mixing ratios for any sample must be 1 (i.e., in Eq. (1), $\sum_k \delta_{ik} = 1$). Moreover, all contributions of any end-member to any mixed sample, namely, each of the terms of δ , must be inside the $[0,1]$ interval (i.e., in Eq. (1), $0 \leq \delta_{ik} \leq 1$). This leads to simple inequalities (inside $[0, \infty]$) if the sum of the mixing ratio is set to 1. It is also expected that all terms in matrix C and D are positive.

Note that the set of unknowns in Eq. (1) depends on the question posed. Likewise, partial information may be available for some of the unknowns and errors terms can be added to any available value. In this approach, we can resume three principal problems/questions:

1. What is the number (n_k) of end-members to consider? This first problem is related to the way the system is or should be represented. This representation, wrong in most situations, depends first on the objective of the study, the available data, and the data themselves. It largely influences the optimal or chosen n_k .
2. How to derive the values of δ for each end-member? Once n_k is defined, the second problem is to derive the values of δ for each of the mixed samples, knowing that they can be treated independently. This linear problem can be over-defined when n_k is strictly larger than the number of chemical species (n_i) plus 1, or under-defined in the opposite situation. An additional complexity is added due to the need to respect inequalities.
3. How to deal with the end-member lack of information? The third more complex and nonlinear problem can be derived when the end-member composition itself is unknown or only partly known. Since end-member waters and their composition are generally poorly defined, this situation would be very common in practical applications.

Several methods exist to identify end-members and estimate their contributions. There are analytical methods such as the enrichment factors, chemical mass balance, and linear regression and also statistical multivariate methods. The statistical multivariate methods include, among others, eigenvector analysis (also termed "principal component analysis (PCA) and factor analysis (FA)), cluster analysis (CA), discriminant analysis (DA), multiple linear regression, or neural networks (Le et al., 2017; Li et al., 2019; Singh et al., 2004; Zanotti et al., 2019). Eigenvector and CA analysis have been widely used to identify the end-members and typology of pollution (Clow & Mast, 2010; Christophersen & Hooper, 1992; Pfister et al., 2018; Simeonov et al., 2003). Several authors used mixed analysis such as PCA with CA, or multiple regressions on principal components (i.e., end member mixing analysis (EMMA)) (Christophersen et al., 1990; Liu et al., 2008; Simeonov et al., 2003). However, although these methods make it possible to identify the end-members, none allows them to be quantified. Understanding the composition of streamwater requires the identification not

only of geochemically distinct water stores but also of time-dependent variation in their relative contributions to streamflow. Carrera et al. (2004) and Vázquez-Suñé et al. (2010) have developed other statistical approaches such as the MIX (from MIXing ratios) method to quantify these end-members and deal with their spatial and temporal variability. Specifically developed for the chemical streamwater topic, here we present explicitly only the EMMA and the MIX method. These methods are intended to solve the three problems mentioned above. EMMA aims at identifying the minimum number of end-members required to explain the variability of measured concentrations in time or space (Tubau et al., 2014). The MIX method solves the third problem, incorporated positiveness constraints and the unicity of the solution (Carrera et al., 2004).

EMMA is a multiple regression on principal components, developed from work carried out by Christophersen et al. (1990), Christophersen and Hooper (1992), Hooper (2003), and modified by Liu et al. (2004) & (2008) and Barthold et al. (2011). It reduces the dimensionality of the analysis, with projections on a space of much smaller dimensions (Pelizardi et al., 2017). The explanation of the variance and the contribution of each species to the mixture are obtained from the analysis of information provided by the calculation of the eigenvalues (Tubau et al., 2014). The main problem with EMMA is that the outcomes (end-members compositions) can be positive or negative. Another disadvantage is that the concentrations of end-members must be different and precisely known (Vázquez-Suñé et al., 2010). Finally, EMMA does not explicitly compute mixing contributions of end-members and complementary tools such as the MIX method are needed to calculate them.

The MIX method was developed for hydrology, with the main objective of evaluating groundwater inflow to surface water bodies and reversely. The methodology proposed by Carrera et al. (2004) assumes the existence of a known number of end-member waters, and looks for the mixing ratios that will better explain the composition of a set of mixed water samples. The MIX method adds a starting postulate with information regarding the composition of the end-members. The main drawback of this calculation tool is the uncertainty regarding the end-member definition, caused by insufficiently documented spatial and temporal variability. The limited experience in selecting appropriate tracers and the co-linearity can also make the solution unidentifiable or unstable.

1.2 Scope of this paper

Because none of the existing methods could solve completely the three problems mentioned in Section 1.1 and since new high-frequency long-term chemical datasets are now available, we aimed to develop a methodology from a purely chemical point of view, to identify and quantify end-members and their contribution to water chemistry.

A series of reasonable hypotheses led to an objective function, applied to the high-frequency chemical dataset of the ORACLE-Orgeval Observatory (France). The new method developed here, called the “IQEA method” (Identification and Quantification of End-Members and their Apportionment), without any preliminary assumption on the composition of the potential end-members, allows us to analyze the temporal variability of the end-members and their relationship to the different flow regimes.

2 Material and method

2.1 Study site and datasets

The Avenelles catchment, located 70 km east of Paris (France), is a sub-catchment of the ORACLE-Orgeval Observatory (<https://gisoracle.inrae.fr/>), covering an area of 46 km² (Figure 1). It is under temperate and oceanic climate conditions, with an annual average temperature of 11 ± 1 °C and a mean annual rainfall of 674 ± 31 mm (Tallec et al., 2013). The average streamflow measured at the Avenelles outlet is approximately 0.2 m³s⁻¹ (1962–2017), with minimum flows in summer (~ 0.1 m³s⁻¹) and floods up to 10 m³s⁻¹ in spring and winter. The Avenelles catchment presents a rather homogeneous topography (Figure 1a), with a deep silt layer (Tallec et al., 2013). With respect to geology, the catchment is underlain entirely by limestone rocks, with two aquifers: the shallower Brie limestone aquifer and the deeper Champigny limestone aquifer (Mouhri et al., 2013) (Figure 1b). Land use (Figure 1c) is mostly agricultural with few villages, and with intensive farming practices, mainly based on mineral nitrogen fertilization (Garnier et al., 2016; Garnier et al., 2014). Nearly 60% of the surface of the catchment is drained with tile drains.

[Insert Figure 1]

Streamflow is measured continuously by a gauging station at the outlet of the catchment and precipitation by rain gauges (see Tallec et al., 2015). The chemical concentrations are measured in situ, continuously, by the *River Lab* (Floury et al., 2017). To test the methodology developed, we chose two ions among all those measured by the *River Lab*: calcium and nitrate (Table 1). These ions are the most chemically representative of our catchment (Floury et al., 2017) with a marked seasonal variation, which is an advantage for testing the model efficiency. The main half-hourly (30 min) datasets (flow rates and chemical concentrations) cover the period between June 2015 and July 2017, i.e., 17,300 measurements over 26 months.

We also used concentrations measured from January 2016 to April 2018 in the framework of the PIREN Seine program in bank piezometers, drains, deep piezometers (plateau), and springs from the ORACLE-Orgeval Observatory (Table 1) (see Mouchel et al., 2016; Mouhri et al., 2013).

[Insert Table 1]

Because it is computationally complicated to work with all data simultaneously (i.e., 17,300), we first subdivided our dataset into monthly samples. To reduce computing time, we then made a random draw: For each month, we took 10 random samples, each containing 100 pairs of calcium and nitrate concentrations (see example in Figure 2). This approach allows us to evaluate the progression of potential source composition on a monthly scale and to assess the robustness of the method by replicate computations. In order to avoid possible bias caused by the different orders of magnitude between the two ions studied, both concentrations are standardized (i.e., concentrations are divided by their mean, see Table 1).

[Insert Figure 2]

2.2 Methodology

Optimization function

Before developing our function and computing its parameters, we have to set a series of starting points and hypotheses.

First, we have to choose the n -dimension of the function. To identify the number of dimensions, and to avoid misinterpretations and possible negative results, we worked without the previous PCA. Regarding the Avenelles catchment, the one-dimensional situation (i.e., two end-members) has already been discussed in previous work, introducing the question of the hydrograph separation methods (see Tunqui Neira, 2019; Tunqui Neira et al., 2020). To explain the streamwater chemical concentrations, geochemists usually identify three main water sources: groundwater, soil water, and runoff (e.g. Bowes et al., 2009; Chanat et al., 2002; Evans & Davies, 1998; Rose et al., 2018). The two-dimensional choice in small catchments has been used repeatedly (e.g. Genereux et al., 1993; Miller et al., 2017; Probst, 1985). With more than three end-members, the level of uncertainty would increase giving an infinite number of possible apportionments, equally correct for each of the samples. For all these reasons, and with the goal of parsimony, we limited the size of the end-members to three ($k = 3$ end-members in Eq. (1)).

Second, because Eq. (1) does not have a single solution, we have to impose adjustment constraints to the end-members compared with the observation data (see Figure 3). To this end, we minimize two terms in an objective function (F):

1. the residual variance of the chemical concentrations of mixed samples after apportionment to the end-members:

$$\left(\frac{1}{n}\right) \sum_{i,j} \left(C_{ij} - \sum_k (D_{ik} \cdot \delta_{kj})\right)^2$$

with C_{ij} the measured concentration and n the number of measured chemical concentrations.

2. the variance of the set of end-member composition:

$$\left(\frac{1}{k}\right) \cdot \sum_{k,j} (D_{kj} - G_j)^2$$

where G_j is the center of gravity of the end-member composition.

[Insert Figure 3]

Because these two terms have different weights in our objective function, we have assigned them two respective coefficients, A' and B' . The objective function F is written as:

$$F = A' \cdot \left(\frac{1}{n}\right) \cdot \sum_{i,j} \left(C_{ij} - \sum_k (D_{ik} \cdot \delta_{kj})\right)^2 + B' \cdot \left(\frac{1}{k}\right) \cdot \sum_{k,j} (D_{kj} - G_j)^2 \quad \text{Eq. (2)}$$

The ratio A'/B' has to be chosen to fully define the objective function. The optimization problem also includes the three constraints of the conservation of water (see Section 1.1):

$$\sum_k \delta_{kj} = 1; 0 \leq \delta_{kj} \leq 1; D_{ik} > 0$$

For easier writing, we define A and B as follow:

$$A = \frac{A'}{n}$$

$$B = \frac{B'}{k}$$

Solution method and evaluation criteria

We do not have a priori information on the best values for constants A and B . To evaluate the quality of the estimation of the optimization of F (Eq. (2)), we used the ratio of the residual variance to initial variance of the dataset (VR):

$$VR = \frac{\sum_{i,j} \left(C_{ij} - \sum_k (D_{ik} \cdot \delta_{kj}) \right)^2}{\sum_{i,j} (C_{ij} - G'_j)^2} \quad \text{Eq. (3)}$$

where G'_j is the center of gravity of the C_{ij} chemical dataset.

VR is equivalent to the term $1 - R^2$ in a regression, where R^2 is the coefficient of determination. In this paper, we have chosen an acceptable threshold for the VR ratio of 0.05 as a rough estimate of the chemical measurement errors.

VR is dependent on the A/B ratio, whose a priori values are unknown. For this reason, the method was applied to 100 different A/B ratios ranging from 0 to 10 to each of the monthly random samples. Then, with the 100 results of VR obtained for each month from the optimization of Eq. (3), we established a power-law relationship between the ratio A/B and VR . For the 10 different samples of each month (22 months), the relationships between VR and A/B ratio show a well-fitted power-law relationship (e.g., in Figure 4 for the month of April 2016). Note that the different samples of the month (and also observed in the other months of the study) have different values of A and B (i.e., A/B ratio) (Figure 4).

[Insert Figure 4]

From this relationship, we obtain the value of A and B for a VR of 0.05. With these optimal values, we resolve the optimization function (Eq. (2)) again so as to calculate the set of end-members (D_{kj}) and their respective apportionment (δ_{kj}) for each month. To solve Eq. (2), we used the "solnp" function of the RSOLNP package (version 1.16) on R (<https://cran.r-project.org/web/packages/Rsolnp/Rsolnp.pdf>). The "solnp" algorithm is based on the nonlinear optimization of the parameters using the augmented Lagrangian method (Ye, 1988). The main benefit of this algorithm for our work is that it allows us to optimize simultaneously two variables with their respective equality and inequality constraints (D_{kj} and δ_{kj}).

Sensitivity analysis of end-members

The optimization of each of the 10 samples with $VR=0.05$ produces a set of three sources, whose good positioning must be verified before the reproducibility of the source characterization. A cluster analysis was performed for this purpose. Then, an analysis of variance compares the variance in each cluster of estimated source with the total variance. The following determination coefficient (R^2) was used:

$$R^2 = 1 - \frac{\sum_k \sum_i d^2(D_{ik}, G_k)}{\sum_{g=1} d^2(D_{ig}, G_g)} \quad \text{Eq. (4)}$$

where the term $\sum_k \sum_i d^2(D_{ik}, G_k)$ represents the sum of variances of each end-member in the cluster, with respect to the center of gravity of the cluster (G_k), and the term $\sum_{g=1} d^2(D_{ig}, G_g)$ represents the variance between the whole set of end-members, with respect to the center of gravity of the whole set of end-members (G_g). Note that the optimal value of Eq. (4) is 1.

To further enhance the robustness of the cluster analysis, we also calculated the value of R^2 of the set of end-members calculated for a set of a variable \sqrt{R} ($0.02 < \sqrt{R} < 0.10$).

Finally, we compared (graphically) the resulting end-members with $\sqrt{R} = 0.05$ of each month with the field data (mean \pm SD) shown in Table 1.

3 Results and discussion

3.1 Cluster analysis and kinetics of the potential end-members

For each month, we calculated the average value of each set of the three potential end-members (i.e., blue, red, and green, see Figure 5). Figure 5 illustrates the cluster analysis performed for the different set of $k=3$ end-members, computed using the IQEA model for $\sqrt{R} = 0.05$, from the 10 different samples of each month. In total, for each month, we have 30 end-members. With a cluster analysis, we were able to group them (blue, red and green) according to their nearby position. The R^2 obtained from the cluster analysis for each month is quite satisfactory, as almost all of them are close to 1 (see Figure 5). To check the robustness of the cluster analysis, we also calculate the R^2 of the different end-members ($k=3$) for a variable \sqrt{R} ($0.02 < \sqrt{R} < 0.10$, see Figure 8 in Appendix 5.1). Even with a variable \sqrt{R} , the R^2 values remain high. Thus, we can support that the imposition of three potential sources for the two ions tested is well defined.

We also calculated the average monthly apportionment given by each monthly potential end-member, presented in Table 2. Note that the few sample points available (i.e., less than 25 % of measurement points recorded for the month) during the months of July and December 2016, and March 2017, did not allow us to calculate the monthly average potential end-members in a representative way. In addition, from October 2016 to February 2017, there were no measurements over several months from the *River Lab*.

Figure 6 shows the contribution of each of the potential end-members at each time point, as well as the dominant end-member.

The apportionment and the time series of the potential end-members give us a rich description of the hydrological and chemical coupling within the catchment. Whereas the IQEA model treats concentration and discharge datasets independently, its results show a strong relationship between the dominant end-member computed and the hydrological features.

[Insert Figure 5]

[Insert Table 2]

[Insert Figure 6]

The time-series identify successive well-defined periods, with a single highly dominant (or unique) end-member (Figure 6), even if in specific situations the end-members may overlap (e.g., August 2016, Figure 6). We note, however, that the more diluted end-member, for calcium and nitrate, is dominant in all samples for most conditions of peak discharge (see red end-member, Figure 5 and Figure 6).

From June to October 2015, a first period could be identified; with similar end-member compositions almost aligned, in particular during June and July 2015 (see Figure 5). Consequently, it is difficult to define precisely the contribution of each one for a given sample. This first period is dominated by a blue end-member and corresponds to a hydrological period, characterized by a low base flow ($0.1 \text{ m}^3\text{s}^{-1}$ or less) and rare higher discharge peaks (less than $0.5 \text{ m}^3\text{s}^{-1}$) (Figure 6 and Table 2).

November 2015 is a transition period where the predominant potential end-member changes into the green one (see Figure 5 and Figure 6). According to Tallec et al. (2013), it should be noted that this corresponds to the beginning of the wet season and of the water recharge of the Brie aquifer. During this transition period, unlike the first period, the red potential end-member, which always dominates peak discharges, is diluted in terms of calcium and enriched in nitrate (see November 2015, Figure 5). This indicates a complex concentration pattern during floods. The following months (December 2015 to February 2016) return to a situation where all discharge peaks appear more diluted (see Figure 5 and Figure 6).

From March to June 2016, we observed a very different period, characterizing an exceptional event: the late-May flood ($\sim 12 \text{ m}^3\text{s}^{-1}$), with a 20-year return period, very unusual in late spring in Paris (and in the Orgeval catchment) (see Figure 6). During this period, the peak flow corresponds to a mixed predominance of two potential end-members, characterized by low calcium concentrations, with more strongly fluctuating nitrate concentrations (see Figure 5). This explains why two potential end-members are necessary to visualize this particular event (Figure 6).

The last period (since July 2016) has similar end-members to the previous low-flow period (i.e., June to October 2015) (see Figure 5 and Figure 6). Knowing that the main hydrological contributor of our catchment is groundwater (Floury et al., 2018; Mouchel et al., 2016; Tallec et al., 2013), we can reasonably state that the predominant potential end-members (green and blue) are representative of groundwater. Note that during this year, there is no change in predominant potential end-member, because of the previous exceptional event produced in the catchment (see Figure 6). The enormous volume produced by this phenomenon was easily stored in the aquifer (paradox of the small catchment, see Kirchner, 2003), with the green end-member chemical characteristics (main contributor at that time). The end-member predominance does not change until June 2017 (the blue end-member returns as the predominant one, see Table 2).

3.2 Potential end-members versus pre-identified possible end-members

Figure 7 compares possible end-members (mean \pm SD of field data, see Table 1) with the potential end-members calculated using the IQEA model. Most of the possible end-members are rather well characterized and define four highly distinct poles (Table 1 and Figure 7):

- i) a pole with high calcium and nitrate concentrations representing plateau and spring end-members,
- ii) a pole with high calcium and lower nitrate concentrations representing bank (i.e., riparian) end-members,
- iii) a pole with low calcium and high nitrate concentrations representing tile-drain end-members,
- iv) a pole with very low calcium and nitrate concentrations representing rainwater end-members.

It must be noted that some poles can be highly variable (with significant SD, see Table 1 and Figure 7). Indeed, rainwater chemistry is subjected to oceanic climate winds (Floury et al., 2018). Because of a gypsum dissolution process occurring on the Orgeval catchment, spring and plateau pole concentrations may be very different from one period to another, mainly during high-water periods (Floury et al., 2018; Mouchel et al., 2016; Tallec et al., 2013). Likewise, tile-drain pole concentrations depend on the nitrate fertilization phase during wet periods (Garnier et al., 2016; Garnier et al., 2014), and bank pole concentrations on the biogeochemical processes (Mouchel et al., 2016). We should also note that these poles can only be the pale reflection of the water pools of the catchment. For example, whereas the rainwater was sampled from rain gages, a theoretical composition of the rainwater reservoir should take into account evaporation and transpiration (about 2/3 of precipitation on average in this region, see Floury et al., 2018; Mouchel et al., 2016). Thus, the possible rainwater end-member presented here should be restricted to non-evaporated or slightly evaporated water.

Despite this variability or imprecision, the potential end-members are only barely outside of the polygon formed by the four distinct poles (Figure 7). This suggests that, during most of the monthly periods studied, a mixture of the pre-identified end-members can be explained by the downstream concentration (Figure 7).

However, we note specific situations with a notable contribution over a given period of one or another pole. In April and May 2016, we observe very high nitrate and limited calcium concentrations (see months of April and May 2016, Figure 7). This suggests an important contribution of the tile-drain pole characterized by an above-average nitrate concentration. During the month of June, the high calcium concentrations observed are only explainable by the high variability in concentrations of the plateau pole (see June 2016 and 2017, Figure 7). We also note that the potential end-members identified, inside the polygon, exclude the bank pole (i.e., riparian water) (see Figure 7). Except in

specific situations, especially during very dry periods (around $0.04 \text{ m}^3\text{s}^{-1}$, see Figure 6), i.e., from May to July 2017 (see Figure 7), riparian water would generally represent a minor contribution.

[Insert Figure 7]

4 Conclusion

In this article we developed a methodology (the IQEA model) that only uses downstream chemical data and a pre-defined fixed number of end-members ($k=3$). We can identify and quantify these end-members (and their apportionment) allowing the formation of the chemical concentrations of stream. We have been able to verify, through random sampling in the dataset and a further analysis of variance of the end-member compositions obtained, that the procedure developed was robust regarding the composition of the sources. For unusual situations (e.g., exceptional event in March 2016), some possible overlap of potential end-members could not be excluded, which also demonstrates complex hydrological features.

These calculated end-members are slightly different from the possible end-members deduced from field data. However, they can be considered as secondary potential end-members, describing the way primary sources are combined over a given period. It is an important point of the method proposed, because primary sources are never visible and neither can they be decrypted from downstream data. The other major point of this method is that it allows us to determine, at a monthly scale, the main potential sources that feed the chemical concentrations of the stream.

However, although the IQEA model is an interesting and promising tool for understanding the ionic concentrations of streams, more work is required in this direction. There must be a relationship between the number of chemical components (n_C) and the number of potential end-members (n_D): $n_D \leq n_C + 1$. The additional 1 comes from an additional constraint that states the sum of discharges from all end-members must be the total discharge. Solutions may be obtained with $n_D < n_C + 1$; our experience was that these solutions are generally unstable, i.e., different random samples, leading to different and overlapping solutions. In the example presented here, we tried as a way of simplification that n_D was exactly equal to $n_C + 1$ (2 chemical components and 3 sources).

The IQEA model has to overcome the same problem faced by other factorization methods. Indeed, the problem of the non-negative matrix factorization (NMF) is that it has no unique solutions and an additional constraint is needed. We chose to reduce the variance of the source composition. Carrera et al. (2004) proposed adding a hypothesis regarding the position of the end-members, which is another possibility requiring additional information. More work is needed to improve the

optimization algorithm, as we were practically limited to approximately 100 points of data. Several NMF algorithms have been proposed to solve NMF problems; the faster ones are based on alternate direction minimization but their convergence has been debated and depends on the additional constraint. More work is required in this area.

5 Appendix

5.1 Cluster analysis for evolution of end-members using a variable VR ($0.02 < VR < 0.10$)

[Insert Figure 8]

6 References

- Ameli, A. A., Beven, K., Erlandsson, M., Creed, I. F., McDonnell, J. J., & Bishop, K. (2017). Primary weathering rates, water transit times, and concentration-discharge relations: A theoretical analysis for the critical zone. *Water Resources Research*, 53(1), 942-960. doi: 10.1002/2016wr019448
- Barthold, F. K., Tyralla, C., Schneider, K., Vaché, K. B., Frede, H.-G., & Breuer, L. (2011). How many tracers do we need for end member mixing analysis (EMMA)? A sensitivity analysis. *Water Resources Research*, 47(8). doi: 10.1029/2011wr010604
- Botter, M., Burlando, P., & Fatichi, S. (2019). Anthropogenic and catchment characteristic signatures in the water quality of Swiss rivers: a quantitative assessment. *Hydrol. Earth Syst. Sci.*, 23(4), 1885-1904. doi: 10.5194/hess-23-1885-2019
- Bowes, M. J., Smith, J. T., & Neal, C. (2009). The value of high-resolution nutrient monitoring: A case study of the River Frome, Dorset, UK. *Journal of Hydrology*, 378(1-2), 82-96. doi: <http://dx.doi.org/10.1016/j.jhydrol.2009.09.015>
- Carrera, J., Vázquez-Suñé, E., Castillo, O., & Sánchez-Vila, X. (2004). A methodology to compute mixing ratios with uncertain end-members. *Water Resources Research*, 40(12). doi: 10.1029/2003wr002263
- Clow, D. W., & Mast, M. A. (2010). Mechanisms for chemostatic behavior in catchments: Implications for CO₂ consumption by mineral weathering. *Chemical Geology*, 269(1), 40-51. doi: <https://doi.org/10.1016/j.chemgeo.2009.09.014>
- Chanat, J. G., Rice, K. C., & Hornberger, G. M. (2002). Consistency of patterns in concentration-discharge plots. *Water Resources Research*, 38(8), 22-21-22-10. doi: 10.1029/2001wr000971
- Christophersen, N., & Hooper, R. P. (1992). Multivariate analysis of stream water chemical data: The use of principal components analysis for the end-member mixing problem. *Water Resources Research*, 28(1), 99-107. doi: 10.1029/91WR02518
- Christophersen, N., Neal, C., Hooper, R. P., Vogt, R. D., & Andersen, S. (1990). Modelling streamwater chemistry as a mixture of soilwater end-members—a step towards second-generation acidification models. *Journal of Hydrology*, 116(1-4), 307-320.
- Duncan, J. M., Band, L. E., & Groffman, P. M. (2017). Variable nitrate concentration-discharge relationships in a forested watershed. *Hydrological processes*, 31(9), 1817-1824. doi: 10.1002/hyp.11136
- Dwivedi, R., Meixner, T., McIntosh, J. C., Ferré, P. A. T., Eastoe, C. J., Niu, G.-Y., . . . Chorover, J. (2019). Hydrologic functioning of the deep critical zone and contributions to streamflow in a high-elevation catchment: Testing of multiple conceptual models. *Hydrological processes*, 33(4), 476-494. doi: 10.1002/hyp.13363
- Evans, C., & Davies, T. D. (1998). Causes of concentration/discharge hysteresis and its potential as a tool for analysis of episode hydrochemistry. *Water Resources Research*, 34(1), 129-137.
- Floury, P., Gaillardet, J., Gayer, E., Bouchez, J., Tallec, G., Ansart, P., . . . Roubaty, J. L. (2017). The potamochemical symphony: new progress in the high-frequency acquisition of stream chemical data. *Hydrol. Earth Syst. Sci.*, 21(12), 6153-6165.
- Floury, P., Gaillardet, J., Tallec, G., Ansart, P., Bouchez, J., Louvat, P., & Gorge, C. (2018). Chemical weathering and CO₂ consumption rate in a multilayered-aquifer dominated watershed under intensive farming: The Orgeval Critical Zone Observatory, France. *Hydrological processes*, 1-19.
- Frisbee, M. D., Phillips, F. M., Campbell, A. R., Liu, F., & Sanchez, S. A. (2011). Streamflow generation in a large, alpine watershed in the southern Rocky Mountains of Colorado: Is streamflow generation simply the aggregation of hillslope runoff responses? *Water Resources Research*, 47(6).

- Frisbee, M. D., Phillips, F. M., White, A. F., Campbell, A. R., & Liu, F. (2013). Effect of source integration on the geochemical fluxes from springs. *Applied Geochemistry*, 28, 32-54. doi: <https://doi.org/10.1016/j.apgeochem.2012.08.028>
- Gaillardet, J., Dupre, B., Louvat, P., & Allegre, C. J. (1999). Global silicate weathering and CO₂ consumption rates deduced from the chemistry of large rivers. *Chemical Geology*, 159(1-4), 3-30. doi: 10.1016/S0009-2541(99)00031-5
- Garnier, J., Anglade, J., Benoit, M., Billen, G., Puech, T., Ramarson, A., . . . Tallec, G. (2016). Reconnecting crop and cattle farming to reduce nitrogen losses to river water of an intensive agricultural catchment (Seine basin, France): past, present and future. *Environmental Science & Policy*, 63, 76-90.
- Garnier, J., Billen, G., Vilain, G., Benoit, M., Passy, P., Tallec, G., . . . Kao, C. (2014). Curative vs. preventive management of nitrogen transfers in rural areas: Lessons from the case of the Orgeval watershed (Seine River basin, France). *Journal of Environmental Management*, 144, 125-134.
- Genereux, D. P., Hemond, H. F., & Mulholland, P. J. (1993). Use of radon-222 and calcium as tracers in a three-end-member mixing model for streamflow generation on the West Fork of Walker Branch Watershed. *Journal of Hydrology*, 142(1), 167-211. doi: [https://doi.org/10.1016/0022-1694\(93\)90010-7](https://doi.org/10.1016/0022-1694(93)90010-7)
- Hooper, R. P. (2003). Diagnostic tools for mixing models of stream water chemistry. *Water Resources Research*, 39(3).
- Hrachowitz, M., Benettin, P., van Breukelen, B. M., Fovet, O., Howden, N. J. K., Ruiz, L., . . . Wade, A. J. (2016). Transit times—the link between hydrology and water quality at the catchment scale. *Wiley Interdisciplinary Reviews: Water*, 3(5), 629-657. doi: 10.1002/wat2.1155
- Kirchner, J. W. (2003). A double paradox in catchment hydrology and geochemistry. *Hydrological processes*, 17(4), 871-874.
- Le, T. T. H., Zeunert, S., Lorenz, M., & Meon, G. (2017). Multivariate statistical assessment of a polluted river under nitrification inhibition in the tropics. *Environmental Science and Pollution Research*, 24(15), 13845-13862. doi: 10.1007/s11356-017-8989-2
- Lenz, A., & Sawyer, C. N. (1944). Estimation of stream-flow from alkalinity-determinations. *Eos, Transactions American Geophysical Union*, 25(6), 1005-1011. doi: 10.1029/TR025i006p01005
- Li, T., Sun, G., Yang, C., Liang, K., Ma, S., Huang, L., & Luo, W. (2019). Source apportionment and source-to-sink transport of major and trace elements in coastal sediments: Combining positive matrix factorization and sediment trend analysis. *Science of the Total Environment*, 651, 344-356. doi: <https://doi.org/10.1016/j.scitotenv.2018.09.198>
- Liu, F., Bales, R. C., Conklin, M. H., & Conrad, M. E. (2008). Streamflow generation from snowmelt in semi-arid, seasonally snow-covered, forested catchments, Valles Caldera, New Mexico. *Water Resources Research*, 44(12).
- Liu, F., Williams, M. W., & Caine, N. (2004). Source waters and flow paths in an alpine catchment, Colorado Front Range, United States. *Water Resources Research*, 40(9). doi: 10.1029/2004wr003076
- Miller, M. P., Tesoriero, A. J., Hood, K., Terziotti, S., & Wolock, D. M. (2017). Estimating Discharge and Nonpoint Source Nitrate Loading to Streams From Three End-Member Pathways Using High-Frequency Water Quality Data. *Water Resources Research*.
- Mouchel, J.-M., Rocha, S., Rivière, A., & Tallec, G. (2016). *Caractérisation de la géochimie des interfaces nappe-rivière du bassin des Avenelles*. France: PIREN Seine.
- Mouhri, A., Flipo, N., Réjiba, F., De Fouquet, C., Bodet, L., Kurtulus, B., . . . Ansart, P. (2013). Designing a multi-scale sampling system of stream-aquifer interfaces in a sedimentary basin. *Journal of Hydrology*, 504, 194-206.
- Muñoz-Villers, L. E., & McDonnell, J. J. (2012). Runoff generation in a steep, tropical montane cloud forest catchment on permeable volcanic substrate. *Water Resources Research*, 48(9).

- Neal, C., & Kirchner, J. W. (2000). Sodium and chloride levels in rainfall, mist, streamwater and groundwater at the Plynlimon catchments, mid-Wales: inferences on hydrological and chemical controls. *Hydrology and Earth System Sciences*, 4(2), 295-310.
- Neal, C., Reynolds, B., Rowland, P., Norris, D., Kirchner, J. W., Neal, M., . . . Armstrong, L. (2012). High-frequency water quality time series in precipitation and streamflow: From fragmentary signals to scientific challenge. *Science of the Total Environment*, 434, 3-12.
- Pearce, C. R., Parkinson, I. J., Gaillardet, J., Chetelat, B., & Burton, K. W. (2015). Characterising the stable ($\delta^{88}\text{Sr}/\delta^{86}\text{Sr}$) and radiogenic ($^{87}\text{Sr}/^{86}\text{Sr}$) isotopic composition of strontium in rainwater. *Chemical Geology*, 409, 54-60. doi: <https://doi.org/10.1016/j.chemgeo.2015.05.010>
- Pelizardi, F., Bea, S. A., Carrera, J., & Vives, L. (2017). Identifying geochemical processes using End Member Mixing Analysis to decouple chemical components for mixing ratio calculations. *Journal of Hydrology*, 550, 144-156. doi: <https://doi.org/10.1016/j.jhydrol.2017.04.010>
- Pfister, L., Thielen, F., Deloule, E., Valle, N., Lentzen, E., Grave, C., . . . McDonnell, J. J. (2018). Fresh-water pearl mussels as a stream water stable isotope recorder. *Ecohydrology*, 11(7), e2007. doi: 10.1002/eco.2007
- Probst, J. L. (1985). Nitrogen and Phosphorus exportation in the Garonne basin (France). *Journal of Hydrology*, 76(3-4), 281-305. doi: 10.1016/0022-1694(85)90138-6
- Rose, L. A., Karwan, D. L., & Godsey, S. E. (2018). Concentration-discharge relationships describe solute and sediment mobilization, reaction, and transport at event and longer timescales. *Hydrological processes*, 32(18), 2829-2844. doi: 10.1002/hyp.13235
- Simeonov, V., Stratis, J. A., Samara, C., Zachariadis, G., Voutsas, D., Anthemidis, A., . . . Kouimtzis, T. (2003). Assessment of the surface water quality in Northern Greece. *Water Research*, 37(17), 4119-4124. doi: [https://doi.org/10.1016/S0043-1354\(03\)00398-1](https://doi.org/10.1016/S0043-1354(03)00398-1)
- Singh, K. P., Malik, A., Mohan, D., & Sinha, S. (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study. *Water Research*, 38(18), 3980-3992. doi: <https://doi.org/10.1016/j.watres.2004.06.011>
- Talleg, G., Ansard, P., Guérin, A., Delaigue, O., & Blanchouin, A. (2015). Observatoire Oracle [Data set]. doi: <https://dx.doi.org/10.17180/obs.oracle>
- Talleg, G., Ansart, P., Guérin, A., Derlet, N., Pourette, N., Guenne, A., . . . Loumagne, C. (2013). Introduction. In C. Loumagne & G. Talleg (Eds.), *L'observation long terme en environnement, exemple du bassin versant de l'Orgeval* (pp. 11-33). Versailles: QUAE.
- Tubau, I., Vázquez-Suñé, E., Jurado, A., & Carrera, J. (2014). Using EMMA and MIX analysis to assess mixing ratios and to identify hydrochemical reactions in groundwater. *Science of the Total Environment*, 470-471, 1120-1131. doi: <https://doi.org/10.1016/j.scitotenv.2013.10.121>
- Tunqui Neira, J. M. (2019). *Revisiting the concentration-discharge (C-Q) relationships with high-frequency measurements*. Sorbonne University, Paris, France. Retrieved from https://webgr.inrae.fr/wp-content/uploads/2020/01/thesis_TunquiJose_EN.pdf
- Tunqui Neira, J. M., Talleg, G., Andréassian, V., & Mouchel, J.-M. (2020). A combined mixing model for high-frequency concentration-discharge relationships. *Journal of Hydrology*, 591, 125559. doi: <https://doi.org/10.1016/j.jhydrol.2020.125559>
- Vázquez-Suñé, E., Carrera, J., Tubau, I., Sánchez-Vila, X., & Soler, A. (2010). An approach to identify urban groundwater recharge. *Hydrol. Earth Syst. Sci.*, 14(10), 2085-2097. doi: 10.5194/hess-14-2085-2010
- Ye, Y. (1988). *Interior algorithms for linear, quadratic, and linearly constrained convex programming*. Stanford University.
- Zanotti, C., Rotiroti, M., Fumagalli, L., Stefania, G. A., Canonaco, F., Stefenelli, G., . . . Bonomi, T. (2019). Groundwater and surface water quality characterization through positive matrix factorization combined with GIS approach. *Water Research*, 159, 122-134. doi: <https://doi.org/10.1016/j.watres.2019.04.058>
- Zhang, R., Li, Q., Chow, T. L., Li, S., & Danielescu, S. (2013). Baseflow separation in a small watershed in New Brunswick, Canada, using a recursive digital filter calibrated with the conductivity mass balance method. *Hydrological processes*, 27(18), 2659-2665.

Table 1: Chemical concentrations (mean and standard deviation) measured in hydrological compartments in the framework of the PIREN Seine program (January 2016 to April 2018) and at the Avenelles outlet by the River Lab (June 2015 to July 2017), from the ORACLE-Orgeval Observatory. Rainfall concentrations were taken from (Floury et al., 2018).

Solute	Unit	Stream (River Lab)	Bank	Drains	Plateau	Springs	Rainfall (Floury et al., 2018)
		mean ± SD					
Nitrate	mgNL ⁻¹	12 ± 1	5 ± 5	15 ± 4	13 ± 4	15 ± 3	0.75
Calcium	mg L ⁻¹	124 ± 13	141 ± 31	70 ± 8	142 ± 60	134 ± 29	1.56

Table 2 : Average monthly concentrations of the potential end-members and their respective apportionment calculated using the IQEA model

Month	Average \pm SD end-member concentration for each month						Average end-member apportionment for each month (%)			Comment/predominant end-member
	Calcium (mgL ⁻¹)			Nitrate (mgNL ⁻¹)			D_{blu} e	D_{re} d	D_{gree} n	
	D_{blue}	D_{red}	D_{green}	D_{blue}	D_{red}	D_{green}				
Jun-15	122 \pm 1	117 \pm 1	125 \pm 0	12.8 \pm 0.1	11.6 \pm 0.1	12.5 \pm 0.2	42.8	18.1	39.1	Blue
Jul-15	127 \pm 2	111 \pm 2	126 \pm 2	12.8 \pm 0.1	11.1 \pm 0.4	11.7 \pm 0.4	71.0	11.0	18.0	Blue
Aug-15	128 \pm 0	103 \pm 2	117 \pm 2	12.6 \pm 0.4	9.8 \pm 0.1	11.1 \pm 0.2	79.5	15.8	4.7	Blue
Sep-15	131 \pm 1	96 \pm 4	115 \pm 5	13.4 \pm 0	10.1 \pm 0.3	11.6 \pm 0.4	52.7	16.4	30.8	Blue
Oct-15	125 \pm 0	99 \pm 1	118 \pm 4	12.8 \pm 0.1	9.4 \pm 0.2	11 \pm 0.2	59.8	17.1	23.0	Blue
Nov-15	123 \pm 4	90 \pm 3	131 \pm 0	12.7 \pm 0.4	13.5 \pm 0.3	10.3 \pm 0.1	21.0	29.7	49.3	Green
Dec-15	103 \pm 2	90 \pm 3	119 \pm 1	12.8 \pm 0.1	10.5 \pm 0.2	11.8 \pm 0.1	37.6	11.2	51.2	Green
Jan-16	105 \pm 2	99 \pm 9	123 \pm 1	13.1 \pm 0.1	10.7 \pm 0.7	13.1 \pm 0.1	41.6	18.8	39.6	Blue
Feb-16	103 \pm 1	79 \pm 3	117 \pm 1	12.0 \pm 0.1	8.4 \pm 0.3	10.9 \pm 0.1	47.0	15.8	37.2	Blue
Mar-16	80 \pm 3	95 \pm 12	128 \pm 1	12.2 \pm 1.5	9 \pm 0.3	10.9 \pm 0.1	7.7	21.0	71.3	Green
Apr-16	95 \pm 3	104 \pm 2	128 \pm 1	15.9 \pm 0.4	10.3 \pm 0.1	10.5 \pm 0	10.8	35.5	53.7	Green
May-16	101 \pm 20	66 \pm 2	118 \pm 11	14 \pm 1.9	10.5 \pm 0.6	10.9 \pm 0.7	6.2	9.0	84.8	Green
Jun-16	105 \pm 2	91 \pm 3	158 \pm 2	13.3 \pm 0.2	9.1 \pm 0.3	10.9 \pm 0.1	19.9	24.5	55.6	Green
Jul-16	114 \pm 1	111 \pm 1	116 \pm 1	11.7 \pm 0.1	11.5 \pm 0.1	12.0 \pm 0.1	21.8	31.0	47.2	Insufficient data
Aug-16	124 \pm 1	124 \pm 0	129 \pm 0	13.3 \pm 0.1	12.6 \pm 0	13.4 \pm 0.1	32.4	27.2	40.5	Green
Sep-16	133 \pm 0	120 \pm 2	130 \pm 0	12.8 \pm 0	12.3 \pm 0.2	13.8 \pm 0.1	36.0	13.6	50.4	Green
Oct-16	---	---	---	---	---	---	---	---	---	No data
Nov-16	---	---	---	---	---	---	---	---	---	No data
Dec-16	118 \pm 1	111 \pm 2	128 \pm 1	12.5 \pm 0.1	11.3 \pm 0.2	13.1 \pm 0.1	31.4	13.1	55.5	Insufficient data
Jan-17	---	---	---	---	---	---	---	---	---	No data
Feb-17	---	---	---	---	---	---	---	---	---	No data
Mar-17	107 \pm 1	110 \pm 1	113 \pm 1	9.4 \pm 0.1	8.8 \pm 0.2	9.1 \pm 0.1	45.5	22.5	32.0	Insufficient data
Apr-17	111 \pm 1	114 \pm 0	119 \pm 1	12 \pm 0.2	9.8 \pm 0.1	12 \pm 0.1	40.1	13.2	46.7	Green
May-17	113 \pm 4	110 \pm 1	131 \pm 1	15.3 \pm 0.7	11.1 \pm 0.1	11.3 \pm 0.1	7.8	31.5	60.7	Green
Jun-17	150 \pm 0	117 \pm 4	130 \pm 8	12 \pm 0.1	11.1 \pm 0.3	9.8 \pm 0.2	48.2	33.9	17.9	Blue
Jul-17	147 \pm 1	129 \pm 1	127 \pm 1	13.4 \pm 0.1	13.4 \pm 0.1	9.9 \pm 0.2	42.5	15.4	42.1	Blue

Figure 1: Catchment location and measurement sites (a), geology (b), and land use (c) of the Avenelles catchment (ORACLE-Orgeval Observatory).

Figure 2: Example of the random draw from monthly sample: on the left, the cloud of the standardized concentration measured in April 2016 at the *River Lab* station (Avenelles catchment, ORACLE-Orgeval Observatory). On the right, 10 random samples, each containing 100 pairs of concentrations of the calcium and nitrate ions.

Figure 3: Schematized optimization function of the IQEA method projected on the observed data (gray point cloud) of nitrate and calcium ($j=2$): example with three end-members ($k=3$), where C is measured concentration, D the chemical concentrations of end-members, δ the mixing ratios of end-members, and G represents the center of gravity of the point cloud.

Figure 4: Relationship between the VR index and A/B ratio for the 10 different samples of the month of April 2016. The numbers inside the red squares are the values obtained from the ratio A/B for a VR = 0.05.

Figure 5: Cluster analysis (ellipses shown in the graph) and R^2 of the different end-members computed with the IQEA model for VR = 0.05.

Figure 6 : Predominant end-member in the water sample (i.e., δ^*Q) calculated with the IQEA method from high-frequency measurements, flow, and precipitation from the Avenelles station and subcatchment (ORACLE-Orgeval Observatory).

Figure 7 : Comparison between the end-members calculated using the IQEA model ($VR = 0.05$) and the field measurements from plateau, drains, banks, springs, and rainwater from the Orgeval catchment.

Figure 8 : Cluster analysis (ellipses shown in the graph) and R^2 of the evolution of different end-members computed with the IQEA model for a variable VR ($0.02 < VR < 0.10$).