Technical note: Uncertainty in multi-source partitioning using large tracer data sets

Alicia Correa[1,2], Diego Ochoa-Tocachi[3], Christian Birkel[1,4]

[1] Department of Geography and Water and Global Change Observatory, University of Costa Rica, San José, 2060, Costa Rica

[2] Institute for Applied Sustainability Research (iSUR), Quito, 170503, Ecuador

[3] Department of Mathematics, Universidad San Francisco de Quito, Quito, 170901, Ecuador

[4] Northern Rivers Institute, University of Aberdeen, Aberdeen, AB24 3UF, UK

Correspondence to: Alicia Correa (alicia.correa@ucr.ac.cr)

Abstract

The availability of large tracer data sets opened up the opportunity to investigate multiple source contributions to a mixture. However, the source contributions may be uncertain and apart from Bayesian approaches, to date there are only solid methods to estimate such uncertainties for two and three sources. We introduce an alternative uncertainty estimation method for four sources based on multiple tracers as input data. Taylor series approximation is used to solve the set of linear mass balance equations. We illustrate the method to compute individual uncertainties in the calculation of source contributions to a mixture, with an example from hydrology, using a 14-tracer set from water sources and streamflow from a tropical, high-elevation catchment. Moreover, this method has the potential to be generalized to any number of tracers across a range of disciplines.

1. Introduction

Tracer applications have dramatically increased over recent years across a wide range of disciplines (West et al., 2010). Applications in hydrology (Hooper, 2003; James and Roulet, 2006; Kirchner and Neal, 2013), ecology (Phillips and Gregg, 2003; Semmens et al., 2009b), anthropology (Ehleringer et al., 2008), conservation biology (Bicknell et al., 2014), nutrition (Magaña-Gallegos et al., 2018), environmental and ecosystem science (Bartov et al., 2013; Granek et al., 2009), and erosion and sediment transportation (Davies et al., 2018) have been the most prominent. Such a widespread use of tracers was mainly facilitated by the availability of analytical techniques that provide high sensitive, rapid multi-element analysis at lower cost (Falkner et al., 1995). For example, the use of inductively coupled plasma mass spectrometry (ICP-MS) as one of the leading analytical techniques for elemental analysis (Helaluddin et al., 2016), led to the availability and use of large tracers sets (elements) in hydrological studies (Barthold et al., 2017; Belli et al., 2017; Correa et al., 2017; Kirchner and Neal, 2013; Mimba et al., 2017). Trace elements together with water stable isotopes (Cavity Ringdown Laser Absorption Spectroscopy paved the way: Berman et al., 2009; Lis et al., 2008) as well as physical-chemical water parameters (e.g. electrical conductivity and pH) are now often used to improve understanding of hydro-geochemical cycles, flow pathways and runoff generation in hydrology. Furthermore, mixing models based on tracer mass balance equations are widely-applied to identify the dominant sources of a mixture and their contribution dynamics.

In hydrological mixing models the composition of the stream is assumed to be an integrated mixture of signatures of different sources (Christophersen et al., 1990). The proportional contributions of n+1 sources to the stream can be uniquely determined using n different tracers (Christophersen & Hooper, 1992). Bayesian methods have been developed to identify multiple (> 3) sources and compute their contributions to a mixture in a two-dimensional mixing space (Parnell et al., 2010; Stock et al., 2018). In this case a unique solution is not feasible and a higher uncertainty is attributed to the model (Phillips and Gregg, 2001, 2003). On the other hand, End Member Mixing Analysis (EMMA) (Hooper, 2003) was developed to use multiple tracers as input, and therefore, allows for a multi-dimensional space that potentially increases the number of identifiable sources (Barthold et al., 2011; Inamdar et al., 2013; Liu et al., 2004). Additionally, the use of multiple tracers can avoid bias and subjectivity in the input information. Therefore, EMMA provides a robust and complete conceptualization of catchment functioning and source interactions during runoff generation (Iwasaki et al., 2015). However, despite its benefits, the EMMA approach lacks a formal methodology to assess the uncertainty of multiple end-members (Delsman et al., 2013) and their individual uncertainties in the calculation of source contributions to a stream.

To our knowledge, the uncertainty estimation of source contributions to streams is based on Gaussian error propagation (Genereux, 1998) and was so far only calculated using one or two tracers simultaneously (MixSIAR: Parnell et al., 2010; Phillips & Gregg, 2001; Semmens, Moore, et al., 2009). Alternatively, when the number of sources is higher, the uncertainty is usually based on the sum of analytical errors, elevation effects and the spatial variability of end-member concentrations (Uhlenbrook and Hoeg, 2003). Hence, we propose an alternative methodology based on the first-order Taylor series approximation to estimate the uncertainty of individual end-member or sources (e.g., precipitation, soil water, groundwater) to a mixture (e.g., streamflow).

We illustrate this application using a multi-tracer data set from Correa et al. (2019b), in a three-dimensional space defined by a Principal Component Analysis (PCA). In Correa et al. (2019b), the authors computed the uncertainties but without disclosing any details in the calculation and methodology used. The main objective of this Technical Note is therefore to explicitly describe the mathematical development that allows the calculation of partial derivatives, degrees of freedom and confidence interval limits for each source fraction contribution. Moreover, to provide the code and several examples for their calculation and reproducibility.

2. Uncertainty estimation method development

In this section, the uncertainty estimation method presented in Phillips and Gregg, (2001) is expanded for four source contributions to the mixture.

Let $\mathcal{C}$ represent the set of sources: A, B, C and D, and mixture M, $\mathcal{C} = \{A, B, C, D, M\}$. In the following equations, $x \in \mathcal{C}$, $y \in \{\overline{\delta}, \overline{\lambda}, \overline{\phi}\}$ and $z \in \{A, M, C\}$. *x*, *y* and *z* are variables that belong to the sets: *x* to the set of A, B, C, D and mixture M, *y* to the set of medians of every projected source and mixture in each principal component $\overline{\delta}, \overline{\lambda}, \overline{\phi}$ respectively of the used sub index and *z* to the set of A, M and C. Furthermore, f_A, f_B, f_C and f_D represent the contribution fraction of sources A, B, C and D respectively to the mixture M.

The data required for this analysis are the median and standard deviations (σ) of each of the sources (A, B, C and D) and the mixture M, projected and expressed in the coordinates of the three-dimensional PCA space. In addition, the sample size (n) of each source is required. Details of the application are presented in section 3.2.

If the system is composed of Eq. (1)

$$\begin{cases} \overline{\delta}_A f_A & + & \overline{\delta}_B f_B & + & \overline{\delta}_C f_C & + & \overline{\delta}_D f_D & = & \overline{\delta}_M \\ \overline{\lambda}_A f_A & + & \overline{\lambda}_B f_B & + & \overline{\lambda}_C f_C & + & \overline{\lambda}_D f_D & = & \overline{\lambda}_M \\ \overline{\phi}_A f_A & + & \overline{\phi}_B f_B & + & \overline{\phi}_C f_C & + & \overline{\phi}_D f_D & = & \overline{\phi}_M \\ f_A & + & f_B & + & f_C & + & f_D & = & 1 \end{cases} \qquad \text{Eq.(1)}$$

and has solution[1] for $f_A, f_B, f_C, f_D > 0$, the contribution fractions take the following form:

$$\begin{aligned} f_A &= \frac{(\Phi_M - \Delta_M)(\Lambda_C - \Delta_C) - (\Lambda_M - \Delta_M)(\Phi_C - \Delta_C)}{(\Phi_A - \Delta_A)(\Lambda_C - \Delta_C) - (\Lambda_A - \Delta_A)(\Phi_C - \Delta_C)} = \frac{Num}{Den} \\ f_C &= \frac{(\Delta_M - \Lambda_M) - (\Delta_A - \Lambda_A) f_A}{(\Delta_C - \Lambda_C)} \\ f_B &= \Delta_M - (\Delta_C f_C + \Delta_A f_A) \\ f_D &= 1 - (f_C + f_B + f_A) \end{aligned} \qquad \text{Eq.(2)}$$

where

$$\Delta_x = \frac{\overline{\delta}_x - \overline{\delta}_D}{\overline{\delta}_B - \overline{\delta}_D}, \; \Lambda_x = \frac{\overline{\lambda}_x - \overline{\lambda}_D}{\overline{\lambda}_B - \overline{\lambda}_D}, \; \Phi_x = \frac{\overline{\phi}_x - \overline{\phi}_D}{\overline{\phi}_B - \overline{\phi}_D}. \qquad \text{Eq.(3)}$$

The partial derivatives of Eq. (2) are given by:

---

[1] The system has a solution if the vector of mixture M is on the polyhedron generated by the vectors of sources A, B, C, D such that $\sum_x f_x = 1$.

$$\begin{aligned}
\frac{\partial f_A}{\partial y_x} &= \frac{1}{Den^2}\Bigg[\Bigg[(\Lambda_C-\Delta_C)\left(\frac{\partial \Phi_M}{\partial y_x}-\frac{\partial \Delta_M}{\partial y_x}\right)+(\Phi_M-\Delta_M)\left(\frac{\partial \Lambda_C}{\partial y_x}-\frac{\partial \Delta_C}{\partial y_x}\right) \\
&\quad -(\Phi_C-\Delta_C)\left(\frac{\partial \Lambda_M}{\partial y_x}-\frac{\partial \Delta_M}{\partial y_x}\right)-(\Lambda_M-\Delta_M)\left(\frac{\partial \Phi_C}{\partial y_x}-\frac{\partial \Delta_C}{\partial y_x}\right)\Bigg] Den \\
&\quad -\Bigg[(\Lambda_C-\Delta_C)\left(\frac{\partial \Phi_A}{\partial y_x}-\frac{\partial \Delta_A}{\partial y_x}\right)+(\Phi_A-\Delta_A)\left(\frac{\partial \Lambda_C}{\partial y_x}-\frac{\partial \Delta_C}{\partial y_x}\right) \\
&\quad -(\Phi_C-\Delta_C)\left(\frac{\partial \Lambda_A}{\partial y_x}-\frac{\partial \Delta_A}{\partial y_x}\right)-(\Lambda_A-\Delta_A)\left(\frac{\partial \Phi_C}{\partial y_x}-\frac{\partial \Delta_C}{\partial y_x}\right)\Bigg] Num\Bigg] \\
\frac{\partial f_C}{\partial y_x} &= \frac{1}{(\Delta_C-\Lambda_C)^2}\Bigg[\Bigg[\left(\frac{\partial \Delta_M}{\partial y_x}-\frac{\partial \Lambda_M}{\partial y_x}\right)-\left(\frac{\partial \Delta_A}{\partial y_x}-\frac{\partial \Lambda_A}{\partial y_x}\right)f_A-(\Delta_A-\Lambda_A)\frac{\partial f_A}{\partial y_x}\Bigg](\Delta_C-\Lambda_C) \\
&\quad -\left(\frac{\partial \Delta_C}{\partial y_x}-\frac{\partial \Lambda_C}{\partial y_x}\right)[(\Delta_M-\Lambda_M)-(\Delta_A-\Lambda_A)f_A]\Bigg], \\
\frac{\partial f_B}{\partial y_x} &= \frac{\partial \Delta_M}{\partial y_x}-\frac{\partial \Delta_C}{\partial y_x}f_C-\Delta_C\frac{\partial f_C}{\partial y_x}-\frac{\partial \Delta_A}{\partial y_x}f_A-\Delta_A\frac{\partial f_A}{\partial y_x}, \\
\frac{\partial f_D}{\partial y_x} &= -\frac{\partial f_C}{\partial y_x}-\frac{\partial f_B}{\partial y_x}-\frac{\partial f_A}{\partial y_x}
\end{aligned} \qquad \text{Eq.(4)}$$

It is trivial that

$$\frac{\partial \Delta_z}{\partial w_x}=0,\ w\in\{\overline{\lambda},\overline{\phi}\};\ \frac{\partial \Lambda_z}{\partial w_x}=0,\ w\in\{\overline{\delta},\overline{\phi}\};\ \frac{\partial \Phi_z}{\partial w_x}=0,\ w\in\{\overline{\delta},\overline{\lambda}\}. \qquad \text{Eq.(5)}$$

where

$$\frac{\partial \Delta_z}{\partial \overline{\delta}_x} = \left(\overline{\delta}_B-\overline{\delta}_D\right)^{-1}\begin{cases} 1 & z\in\{A,C,M\} \text{ and } x=z \\ -\Delta_z & z\neq B \text{ and } x=B \\ \Delta_z-1 & z\neq D \text{ and } x=D \\ 0 & otherwise \end{cases}, \qquad \text{Eq.(6)}$$

$$\frac{\partial \Lambda_z}{\partial \overline{\lambda}_x} = \left(\overline{\lambda}_B-\overline{\lambda}_D\right)^{-1}\begin{cases} 1 & z\in\{A,C,M\} \text{ and } x=z \\ -\Lambda_z & z\neq B \text{ and } x=B \\ \Lambda_z-1 & z\neq D \text{ and } x=D \\ 0 & otherwise \end{cases} \quad \text{and} \qquad \text{Eq.(7)}$$

$$\frac{\partial \Phi_z}{\partial \overline{\phi}_x} = \left(\overline{\phi}_B-\overline{\phi}_D\right)^{-1}\begin{cases} 1 & z\in\{A,C,M\} \text{ and } x=z \\ -\Phi_z & z\neq B \text{ and } x=B \\ \Phi_z-1 & z\neq D \text{ and } x=D \\ 0 & otherwise \end{cases}. \qquad \text{Eq.(8)}$$

For example, for f_A we have

$$\frac{\partial f_A}{\partial \bar{\delta}_x} = \frac{1}{Den^2}\left[\left[\frac{\partial \Delta_M}{\partial \bar{\delta}_x}(\Phi_C - \Lambda_C) - \frac{\partial \Delta_C}{\partial \bar{\delta}_x}(\Phi_M - \Lambda_M)\right] Den - \left[\frac{\partial \Delta_A}{\partial \bar{\delta}_x}(\Phi_C - \Lambda_C) - \frac{\partial \Delta_C}{\partial \bar{\delta}_x}(\Phi_A - \Lambda_A)\right] Num\right].$$

$$\frac{\partial f_A}{\partial \bar{\lambda}_x} = \frac{1}{Den^2}\left[\frac{\partial \Lambda_C}{\partial \bar{\lambda}_x}(\Phi_M - \Delta_M) - \frac{\partial \Lambda_M}{\partial \bar{\lambda}_x}(\Phi_C - \Delta_C)\right] Den - \left[\frac{\partial \Lambda_C}{\partial \bar{\lambda}_x}(\Phi_A - \Delta_A) - \frac{\partial \Lambda_A}{\partial \bar{\lambda}_x}(\Phi_C - \Delta_C)\right] Num\Bigg].$$

$$\frac{\partial f_A}{\partial \bar{\phi}_x} = \frac{1}{Den^2}\left[\left[\frac{\partial \Phi_M}{\partial \bar{\phi}_x}(\Lambda_C - \Delta_C) - \frac{\partial \Phi_C}{\partial \bar{\phi}_x}(\Lambda_M - \Delta_M)\right] Den - \left[\frac{\partial \Phi_A}{\partial \bar{\phi}_x}(\Lambda_C - \Delta_C) - \frac{\partial \Phi_C}{\partial \bar{\phi}_x}(\Lambda_A - \Delta_A)\right] Num\right]. \qquad \text{Eq.(9)}$$

Using Eq. (9), the first-order Taylor series approximation (Taylor, 1982) for the variance (σ^2) of f_A evaluated at the mean can be calculated by:

$$\sigma_{f_A}^2 = \sum_x \left(\frac{\partial f_A}{\partial \bar{\delta}_x}\right)^2 \sigma_{\bar{\delta}_x}^2 + \sum_x \left(\frac{\partial f_A}{\partial \bar{\lambda}_x}\right)^2 \sigma_{\bar{\lambda}_x}^2 + \sum_x \left(\frac{\partial f_A}{\partial \bar{\phi}_x}\right)^2 \sigma_{\bar{\phi}_x}^2 = \sum_y \sum_x \left(\frac{\partial f_A}{\partial y_x}\right)^2 \sigma_{y_x}^2. \qquad \text{Eq.(10)}$$

To calculate γ_A (the Satterthwaite (1946) approximation for the degrees of freedom), we define $f_{Ay_x} = c_A \left(\frac{\partial f_A}{\partial y_x}\right)^2$, where c_A is a scale constant that relates f_{Ay_x} with the respective derivative. It means that f_A with respect to y_x can be a scalar multiple of the derivative value.

In this case, we get:

$$\gamma_A = \frac{\left(\sum_y \sum_x f_{Ay_x} \sigma_{y_x}^2\right)^2}{\sum_y \sum_x \frac{\left(f_{Ay_x} \sigma_{y_x}^2\right)^2}{n_{y_x} - 1}}. \qquad \text{Eq.(11)}$$

Note that whatever the value of c_A is, Eq. (11) leads to:

$$\gamma_A = \frac{\left(\sum_y \sum_x \left(\frac{\partial f_A}{\partial y_x}\right)^2 \sigma_{y_x}^2\right)^2}{\sum_y \sum_x \frac{\left(\left(\frac{\partial f_A}{\partial y_x}\right)^2 \sigma_{y_x}^2\right)^2}{n_{y_x} - 1}}$$

and if we set $f_{Ay_x}^* = \left(\frac{\partial f_A}{\partial y_x}\right)^2$ then the numerator of the last equation can be replaced by $\left(\sigma_{f_A}^2\right)^2$. In other words, we can use Eq. (10) and the derivatives Eq. (9) to estimate the value of γ_A resulting in $f_{Ay_x} = c_A f_{Ay_x}^*$. Of course, it is required that c_A is constant w.r.t. y_x. Then,

$$\gamma_A = \frac{\left(\sigma_{f_A}^2\right)^2}{\sum_y \sum_x \frac{\left(\left(\frac{\partial f_A}{\partial y_x}\right)^2 \sigma_{y_x}^2\right)^2}{n_{y_x} - 1}} \qquad \text{Eq.(12)}$$

Let $w \in \mathcal{C} \setminus \{A\}$. The first-order Taylor series approximation for the variance of f_w, can be calculated by (as above Eq.10):

$$\sigma_{f_w}^2 = \sum_x \left(\frac{\partial f_w}{\partial \overline{\delta}_x}\right)^2 \sigma_{\overline{\delta}_x}^2 + \sum_x \left(\frac{\partial f_w}{\partial \overline{\lambda}_x}\right)^2 \sigma_{\overline{\lambda}_x}^2 + \sum_x \left(\frac{\partial f_w}{\partial \overline{\phi}_x}\right)^2 \sigma_{\overline{\phi}_x}^2 = \sum_y \sum_x \left(\frac{\partial f_w}{\partial y_x}\right)^2 \sigma_{y_x}^2. \qquad \text{Eq.(13)}$$

If we construct γ_w as γ_A, we get:

$$\gamma_w = \frac{\left(\sum_y \sum_x f_{w y_x}^* \sigma_{y_x}^2\right)^2}{\sum_y \sum_x \frac{\left(f_{w y_x}^* \sigma_{y_x}^2\right)^2}{n_{y_x} - 1}}$$

where $f_{w_{y_x}} = c_w f_{w_{y_x}}^*$ and $f_{w_{y_x}}^* = \left(\frac{\partial f_w}{\partial y_x}\right)^2$ with c_w constant w.r.t. y_x, then we finally get:

$$\gamma_w = \frac{\left(\sigma_{f_w}^2\right)^2}{\sum_y \sum_x \frac{\left(\left(\frac{\partial f_w}{\partial y_x}\right)^2 \sigma_{y_x}^2\right)^2}{n_{y_x} - 1}} \qquad \text{Eq.(14)}$$

The upper and lower confidence interval limits for each end-member fraction can be calculated using partial derivatives and the 95% confidence intervals constructed as:

$$f_w \pm t_{0.05\gamma_w} \sigma_{fw} \qquad \text{Eq.(15)}$$

Where $t_{0.05,\gamma}$ is the Student's t for α=0.05 (two-tailed) (Walpole et al., 2017) and γ degrees of freedom related with σ_{fw}.

3. Application

3.1. Study site and data

This methodology was tested using data from a high elevation (3,500 - 3,900 m a.s.l.) tropical catchment (7.53 km^2) located in southern Ecuador (3°4′38″S, 79°15′30″O). The mean annual precipitation for this study site is 1,300 mm (Padrón et al., 2015), the mean annual discharge is 860 mm yr^{-1}. The catchment is of a volcanic origin and dominated by volcanic Histosol (24%) and Andosol (72%) soils (Quichimbo et al., 2012), both with high percentage of organic matter content (450 and 310 g kg, respectively) (Quichimbo et al., 2012) and large water-holding capacities (Buytaert et al., 2006). Histosols are primarily located at the valleys and covered by cushion plants, while Andosol soils are predominated on the hillslopes under a cover of tussock grass. Nearly-saturated conditions of the riparian zone are observed year-round, and a spring is located in the north-western part of the catchment. Streamwater samples from 5 nested streams were collected weekly from March 2013 to April 2014 (n=257) and at a higher frequency during experimental campaigns. Additionally, bi-weekly water samples from 4 potential end-members: rainfall (RF), soil water from Andosols (AN) and Histosols (HS) and spring water (SW) (n ~ 30, for each end-member) were collected. A multi-tracer (14 tracers) data set of conservative tracers was obtained from each water sample (Na, Mg, Al, Si, K, Ca, Rb, Sr, Ba, Ce, V, Y, Nd) at the Institute for Landscape Ecology and Resource Management of the Justus Liebig University using an ICP-MS (Agilent 7500ce, Agilent Technologies) and the electrical conductivity (EC) was measured in situ. More detailed information on the study site and data set can be found in Correa et al., (2017, 2019b).

3.2. Uncertainty estimation of water source contributions

Using the classic EMMA approach (Christophersen and Hooper, 1992), data from end-members SW, HS, AN, RF and stream M, were projected into a three-dimensional space (Correa et al., 2019b) and presented in Figure 1. The resulting median and standard deviation of end-members and stream coordinates are shown in Table 1. Furthermore, Figure 2 shows the distribution of projected samples from individual end-members in the PCA coordinates.

The uncertainty range of each of the four end-member contributions to the stream was determined using the above developed Eq. 15 based on the first-order Taylor series approximation (Eq. 14) (MatLab code in Correa et al., 2019a). The f_w gives the proportion of w in M and $\sigma_{f_w}^2$, the variances of w. The upper uncertainty limit was computed as $f_w + t_{0.05,\gamma}\,\sigma_{fw}$ and the lower limit as $f_w - t_{0.05,\gamma}\,\sigma_{fw}$. This procedure was applied to all end-members. The resulting uncertainty estimates for each source end-member are shown in Table 5.

Note that the set of sources: A, B, C and D used for the development of the equations are represented here by SW, HS, AN and RF in this specific order. U1, U2 and U3 represent the principal components PC1, PC2 and PC3, respectively.

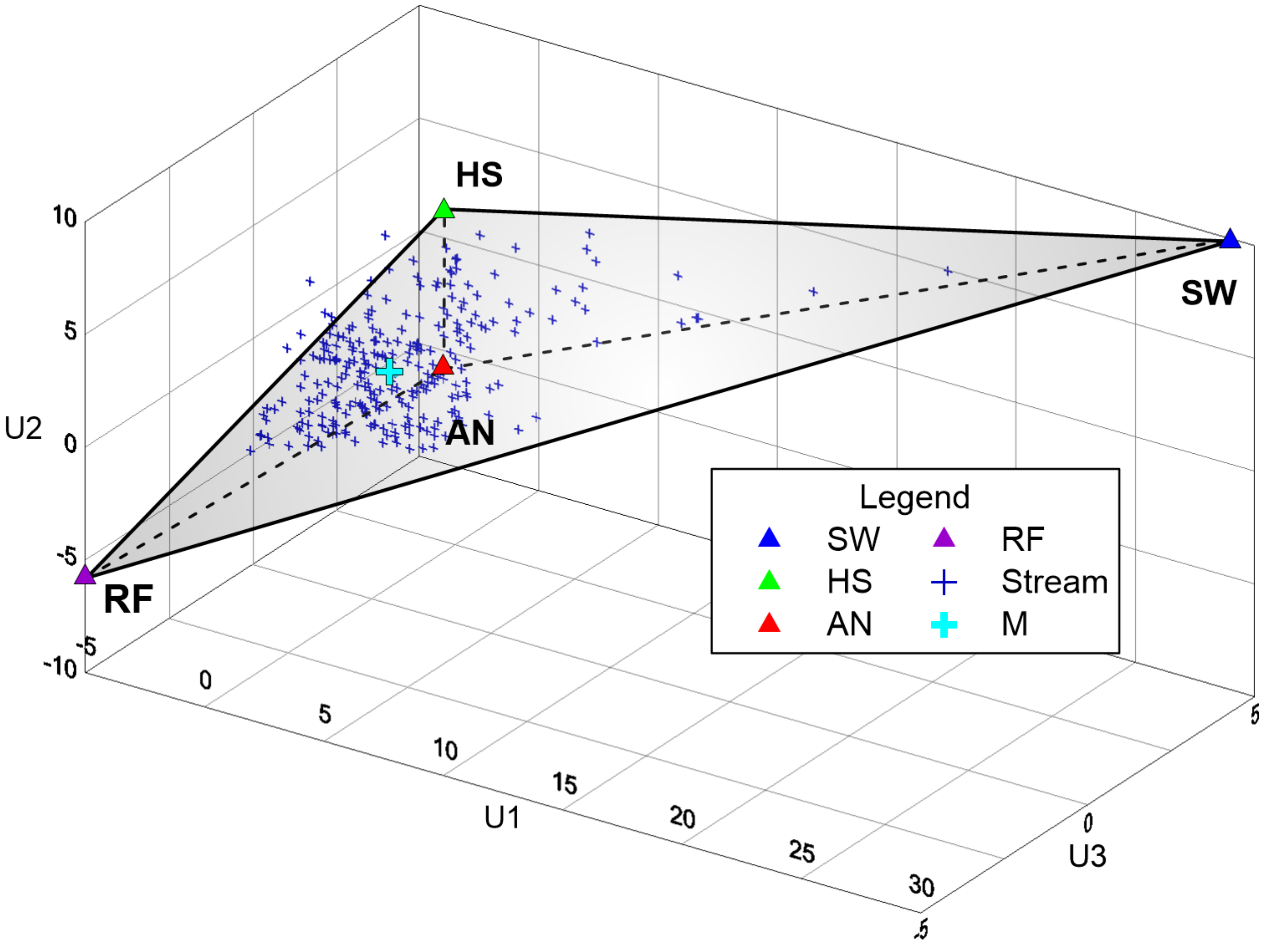

Figure 1. Three-dimensional mixing space generated using stream data, where the median of end-members are projected. U1 represents 59.6% of the variance, U2 19.7%, and U3 7.4% (From PCA); RF, rainfall; AN, Andosols; HS, Histosols; SW, spring water; M, median of stream data (mixture)

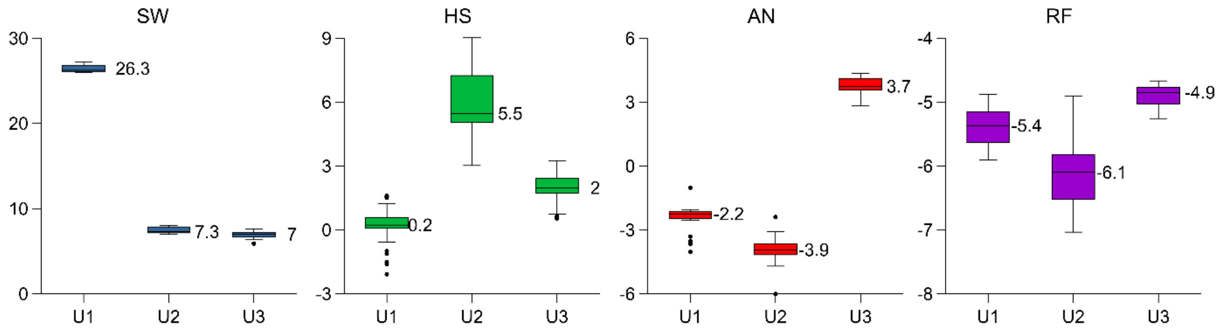

Figure 2. Boxplots of end-members projected in the three-dimensional mixing space for the study period 2013–2014, the Y-axis represents the coordinates of the mixing space and the X-axis the principal components U1, U2 and U3 (the central bar in the box represents the median; notches represent the 95% confidence intervals; whiskers 1.5 times the interquartile range and circles represent outliers). SW, spring water; HS, Histosol; AN, Andosol; RF, rainfall.

3.3. Sensitivity of water sources uncertainty to input data

From the above-mentioned data set, we have generated 6 examples to assess the sensitivity of the uncertainty calculation to the source sample size, the artificial inclusion of outliers (upper and lower extremes) and the increased standard deviations of the source datasets.

- The first example considers 50% of the samples (collected in the first half of the monitoring period) from each source. The median, standard deviation and sample size are input data (Table 2) to calculate the uncertainty bands (Table 6).
- The second example considers the remaining 50% of samples and was similarly executed (Table 2).
- In the third example, outliers were artificially included at the upper positive end of data sets for each source at each coordinate, respectively. The outliers consisted of twice the maximum positive value of the observed data (Table 3).
- Using the same criteria, the negative extremes were included in the fourth example (Table 3).
- Sources affected by dispersed data clouds were taken into account by an increase in the standard deviation. We considered two cases, the first, in the example five, increasing three times the value of the standard deviation of the initial data set (Table 4) and finally, increasing the standard deviation five times for the sixth example (Table 4).

The results of this analysis are presented in Tables 6-8. In examples 1 and 2 the sample size reduction from 24 to 12 and 13 samples respectively (Table 6), had a minimal effect (less than 3%) on the calculation of the uncertainty ranges compared to the original complete set (Table 1). The fractions of source contributions did not experience changes. The inclusion of outliers affected the values of the medians at levels of the second decimal (Table 3) in relation to the median of the initial data (Table 2). However, the standard deviations increased in a range of 1.2 to 2.5 times the original value for AN and HS, and more for RF (2.5 to 10.5) and drastically for SW (4 to 20 times wider). These variations were reflected in the widening (1% to 12%) of uncertainty bands for all existing cases (Table 7) in comparison with those calculated from the original data set (Table 5). Furthermore, the widening of the standard deviations to three and five times their initial values resulted in an increase in the range of uncertainty between 2% and 22% for the first case and between 5% and 37% for the second case. For the latter, the minimum limit of the uncertainty range was reached in all the reported cases. The large number of samples used in these exercises were reflected in high degrees of freedom.

4. Summary and remarks

Our methodology was developed to calculate the contribution of sources to the mixture and its associated uncertainty (based on multiple tracer sets) showing to be effective in real application cases. The application of the method reflected that the calculations of the uncertainty ranges of multiple source contributions to a mixture do not experience significant changes with sample size reduction or inclusion of outliers. Rather, it shows marginally different results by incorporating standard deviations from widely dispersed data.

The methodology, based on Phillips and Gregg, (2001) combined with EMMA applications (Hooper, 2003) presents high potential for use as an alternative method to the simple sum of analytical errors (Uhlenbrook and Hoeg, 2003) or the Bayesian approach (Parnell et al., 2010; Stock et al., 2018). We provide a tool to close the gap in studies of mixing processes, when a larger number of source contributions (>3) and related uncertainty estimates are needed for a more complete conceptualization(Iwasaki et al., 2015).

The MatLab code provided and the illustrative examples facilitate the understanding of the methodology and promote future scientific applications. We are confident that the use of this methodology will help the scientific community that is increasingly using large tracer sets in its research to obtain results supported by a sound uncertainty analysis..

5. Code and data availability

A MatLab code to calculate the fractions of end-members contribution to the mixture and their associated uncertainties is freely available in *https://zenodo.org/record/2649201*. As well as input data (used in this study) as an example for the code run and an instruction note.

6. Author contribution

AC and CB conceptualized the methodology. AC was responsible for the data collection and analysis. DO AC programmed and evaluated the MatLab code with collected data. AC wrote the manuscript with contributions from all co-authors.

7. Competing interests

The authors declare that they have no conflict of interest

Acknowledgements

AC and CB would like to acknowledge support by a UCR postdoctoral fellowship awarded to AC, financial support by UCREA awarded to CB and the Water and Global Change Observatory at the Department of Geography, UCR. The authors thank the Central Research Office (DIUC) of the Universidad de Cuenca for making available part of the tracer data sets. We are especially grateful for the constructive comments that were provided by the Referees, which greatly improved the quality of the Technical Note.

References:

Barthold, F. K., Tyralla, C., Schneider, K., Vaché, K. B., Frede, H.-G. and Breuer, L.: How many tracers do we need for end member mixing analysis (EMMA)? A sensitivity analysis, Water Resour. Res., 47(8), W08519, doi:10.1029/2011WR010604, 2011.

Barthold, F. K., Turner, B. L., Elsenbeer, H. and Zimmermann, A.: A hydrochemical approach to quantify the role of return flow in a surface flow-dominated catchment, Hydrol. Process., 31(5), 1018–1033, doi:10.1002/hyp.11083, 2017.

Bartov, G., Deonarine, A., Johnson, T. M., Ruhl, L., Vengosh, A. and Hsu-Kim, H.: Environmental Impacts of the Tennessee Valley Authority Kingston Coal Ash Spill. 1. Source Apportionment Using Mercury Stable Isotopes, Environ. Sci. Technol., 47(4), 2092–2099, doi:10.1021/es303111p, 2013.

Belli, R., Borsato, A., Frisia, S., Drysdale, R., Maas, R. and Greig, A.: Investigating the hydrological significance of stalagmite geochemistry (Mg, Sr) using Sr isotope and particulate element records across the Late Glacial-to-Holocene transition, Geochimica et Cosmochimica Acta, 199, 247–263, doi:10.1016/j.gca.2016.10.024, 2017.

Berman, E. S. F., Gupta, M., Gabrielli, C., Garland, T. and McDonnell, J. J.: High-frequency field-deployable isotope analyzer for hydrological applications: RAPID COMMUNICATION, Water Resources Research, 45(10), doi:10.1029/2009WR008265, 2009.

Bicknell, A. W. J., Knight, M. E., Bilton, D. T., Campbell, M., Reid, J. B., Newton, J. and Votier, S. C.: Intercolony movement of pre-breeding seabirds over oceanic scales: implications of cryptic age-classes for conservation and metapopulation dynamics, Diversity and Distributions, 20(2), 160–168, doi:10.1111/ddi.12137, 2014.

Buytaert, W., Iñiguez, V., Celleri, R., Bièvre, B. D., Wyseure, G. and Deckers, J.: Analysis of the water balance of small páramo catchments in south Ecuador, in Environmental Role of Wetlands in Headwaters, pp. 271–281, Springer, Dordrecht., 2006.

Christophersen, N. and Hooper, R. P.: Multivariate analysis of stream water chemical data: The use of principal components analysis for the end-member mixing problem, Water Resour. Res., 28(1), 99–107, doi:10.1029/91WR02518, 1992.

Christophersen, N., Neal, C., Hooper, R. P., Vogt, R. D. and Andersen, S.: Modelling streamwater chemistry as a mixture of soilwater end-members - A step towards second-generation acidification models, J. Hydrol., 116(1–4), 307–320, 1990.

Correa, A., Windhorst, D., Tetzlaff, D., Crespo, P., Célleri, R., Feyen, J. and Breuer, L.: Temporal dynamics in dominant runoff sources and flow paths in the Andean Páramo, Water Resour. Res., doi:10.1002/2016WR020187, 2017.

Correa, A., Ochoa-Tocachi, D. and Christian, B.: MatLab code to calculate fractions of contribution to the mixture and associated uncertainties, Zenodo., 2019a.

Correa, A., Breuer, L., Crespo, P., Célleri, R., Feyen, J., Birkel, C., Silva, C. and Windhorst, D.: Spatially distributed hydro-chemical data with temporally high-resolution is needed to adequately assess the hydrological functioning of headwater catchments, Science of The Total Environment, 651, 1613–1626, doi:10.1016/j.scitotenv.2018.09.189, 2019b.

Davies, J., Olley, J., Hawker, D. and McBroom, J.: Application of the Bayesian approach to sediment fingerprinting and source attribution, Hydrological Processes, 32(26), 3978–3995, doi:10.1002/hyp.13306, 2018.

Delsman, J. R., Essink, G. H. P. O., Beven, K. J. and Stuyfzand, P. J.: Uncertainty estimation of end-member mixing using generalized likelihood uncertainty estimation (GLUE), applied in a lowland catchment: Uncertainty Estimation of End-Member Mixing, Water Resources Research, 49(8), 4792–4806, doi:10.1002/wrcr.20341, 2013.

Ehleringer, J. R., Bowen, G. J., Chesson, L. A., West, A. G., Podlesak, D. W. and Cerling, T. E.: Hydrogen and oxygen isotope ratios in human hair are related to geography, PNAS, 105(8), 2788–2793, doi:10.1073/pnas.0712228105, 2008.

Genereux, D.: Quantifying uncertainty in tracer-based hydrograph separations, Water Resour. Res., 34(4), 915–919, doi:10.1029/98WR00010, 1998.

Granek, E. F., Compton, J. E. and Phillips, D. L.: Mangrove-Exported Nutrient Incorporation by Sessile Coral Reef Invertebrates, Ecosystems, 12(3), 462–472, doi:10.1007/s10021-009-9235-7, 2009.

Helaluddin, A., Khalid, R. S., Alaama, M. and Abbas, S. A.: Main Analytical Techniques Used for Elemental Analysis in Various Matrices, Tropical Journal of Pharmaceutical Research, 15(2), 427, doi:10.4314/tjpr.v15i2.29, 2016.

Hooper, R. P.: Diagnostic tools for mixing models of stream water chemistry, Water Resour. Res., 39(3), 1055, doi:10.1029/2002WR001528, 2003.

Inamdar, S., Dhillon, G., Singh, S., Dutta, S., Levia, D., Scott, D., Mitchell, M., Van Stan, J. and McHale, P.: Temporal variation in end-member chemistry and its influence on runoff mixing patterns in a forested, Piedmont catchment, Water Resour. Res., 49(4), 1828–1844, doi:10.1002/wrcr.20158, 2013.

Iwasaki, K., Katsuyama, M. and Tani, M.: Contributions of bedrock groundwater to the upscaling of storm-runoff generation processes in weathered granitic headwater catchments, Hydrol. Process., 29(6), 1535–1548, doi:10.1002/hyp.10279, 2015.

James, A. L. and Roulet, N. T.: Investigating the applicability of end-member mixing analysis (EMMA) across scale: A study of eight small, nested catchments in a temperate forested watershed, Water Resources Research, 42(8), doi:10.1029/2005WR004419, 2006.

Kirchner, J. W. and Neal, C.: Universal fractal scaling in stream chemistry and its implications for solute transport and water quality trend detection, Proceedings of the National Academy of Sciences, 110(30), 12213–12218, doi:10.1073/pnas.1304328110, 2013.

Lis, G., Wassenaar, L. I. and Hendry, M. J.: High-Precision Laser Spectroscopy D/H and 18O/16O Measurements of Microliter Natural Water Samples, Anal. Chem., 80(1), 287–293, doi:10.1021/ac701716q, 2008.

Liu, F., Williams, M. W. and Caine, N.: Source waters and flow paths in an alpine catchment, Colorado Front Range, United States, Water Resour. Res., 40(9), W09401, doi:10.1029/2004WR003076, 2004.

Magaña-Gallegos, E., González-Zúñiga, R., Cuzon, G., Arevalo, M., Pacheco, E., Valenzuela, M. A. J., Gaxiola, G., Chan-Vivas, E., López-Aguiar, K. and Noreña-Barroso, E.: Nutritional Contribution of Biofloc within the Diet of Growout and Broodstock of Litopenaeus vannamei, Determined by Stable Isotopes and Fatty Acids, Journal of the World Aquaculture Society, 919–932, doi:10.1111/jwas.12513@10.1111/(ISSN)1749-7345.focus_on_aquaculture_research_in_latin_america, 2018.

Mimba, M. E., Ohba, T., Nguemhe Fils, S. C., Wirmvem, M. J., Numanami, N. and Aka, F. T.: Seasonal Hydrological Inputs of Major Ions and Trace Metal Composition in Streams Draining the Mineralized Lom Basin, East Cameroon: Basis for Environmental Studies, Earth Syst Environ, 1(2), 22, doi:10.1007/s41748-017-0026-6, 2017.

Padrón, R. S., Wilcox, B. P., Crespo, P. and Célleri, R.: Rainfall in the Andean Páramo: New Insights from High-Resolution Monitoring in Southern Ecuador, J. Hydrometeor, 16(3), 985–996, doi:10.1175/JHM-D-14-0135.1, 2015.

Parnell, A. C., Inger, R., Bearhop, S. and Jackson, A. L.: Source Partitioning Using Stable Isotopes: Coping with Too Much Variation, PLOS ONE, 5(3), e9672, doi:10.1371/journal.pone.0009672, 2010.

Phillips, D. L. and Gregg, J. W.: Uncertainty in source partitioning using stable isotopes, Oecologia, 127(2), 171–179, doi:10.1007/s004420000578, 2001.

Phillips, D. L. and Gregg, J. W.: Source partitioning using stable isotopes: coping with too many sources, Oecologia, 136(2), 261–269, doi:10.1007/s00442-003-1218-3, 2003.

Quichimbo, P., Tenorio, G., Borja, P., Cárdenas, I., Crespo, P. and Célleri, R.: Efectos sobre las propiedades físicas y químicas de los suelos por el cambio de la cobertura vegetal y uso del suelo: Páramo de Quimsacocha al Sur del Ecuador, Sociedad Colombiana de la Ciencia del Suelo, 2(42), 138–153, 2012.

Satterthwaite, F. E.: An Approximate Distribution of Estimates of Variance Components, Biometrics Bulletin, 2(6), 110–114, doi:10.2307/3002019, 1946.

Semmens, B. X., Moore, J. W. and Ward, E. J.: Improving Bayesian isotope mixing models: a response to Jackson et al. (2009), Ecology Letters, 12(3), E6–E8, doi:10.1111/j.1461-0248.2009.01283.x, 2009a.

Semmens, B. X., Ward, E. J., Moore, J. W. and Darimont, C. T.: Quantifying Inter- and Intra-Population Niche Variability Using Hierarchical Bayesian Stable Isotope Mixing Models, PLOS ONE, 4(7), e6187, doi:10.1371/journal.pone.0006187, 2009b.

Stock, B., Semmens, B., Ward, E., Parnell, A., Jackson, A. and Phillips, D.: MixSIAR: Bayesian Mixing Models in R. [online] Available from: https://CRAN.R-project.org/package=MixSIAR (Accessed 20 March 2019a), 2018.

Stock, B. C., Jackson, A. L., Ward, E. J., Parnell, A. C., Phillips, D. L. and Semmens, B. X.: Analyzing mixing systems using a new generation of Bayesian tracer mixing models, PeerJ, 6, e5096, doi:10.7717/peerj.5096, 2018b.

Taylor, J. R.: An introduction to error analysis. The study of uncertainties in physical measurements. [online] Available from: http://adsabs.harvard.edu/abs/1982aite.book.....T (Accessed 17 October 2018), 1982.

Uhlenbrook, S. and Hoeg, S.: Quantifying uncertainties in tracer-based hydrograph separations: a case study for two-, three- and five-component hydrograph separations in a mountainous catchment, Hydrol. Process., 17(2), 431–453, doi:10.1002/hyp.1134, 2003.

Walpole, R., Myers, R., Myers, S. and Ye, K.: Probability & Statistics for Engineers & Scientists, MyLab Statistics Update, 9th Edition, 9th Edition., Pearson. [online] Available from: /content/one-dot-com/one-dot-com/us/en/higher-education/product.html (Accessed 3 September 2019), 2017.

West, J. B., Bowen, G. J., Dawson, T. E. and Tu, K. P., Eds.: Isoscapes: Understanding movement, pattern, and process on Earth through isotope mapping, Springer Netherlands. [online] Available from: https://www.springer.com/la/book/9789048133536 (Accessed 20 March 2019), 2010.

Table 1. Median and standard deviation (std.dev.) of end-members and stream projected in three-dimensional space for the study period 2013–2014.

End-member		Coordinates*			Naming
		U1	U2	U3	in equations
SW (n = 25)	median	26,25	7,29	7,00	A
	std.dev.	0,46	0,36	0,39	
HS (n = 33)	median	0,23	5,48	1,97	B
	std.dev.	0,85	1,29	0,69	
AN (n = 37)	median	-2,24	-3,93	3,71	C
	std.dev.	0,55	0,58	0,45	
RF (n = 36)	median	-5,38	-6,10	-4,84	D
	std.dev.	0,27	0,56	0,15	
Stream (n = 257)	median	-0,61	-1,04	0,94	M
	std.dev.	2,06	1,10	0,66	

* Coordinates of end-members and stream (mixture) medians in three-dimensional space (U1, U2 and U3). n represents the sample size.

Table 2. Median and standard deviation (std.dev.) of end-members and stream projected in three-dimensional considering 50% of the data sets

Naming in equations		1) End member	Coordinates* U1	U2	U3	2) End member	Coordinates* U1	U2	U3
A	median	SW	26.18	7.29	6.66	SW	26.28	7.29	7.1
	std.dev.	(n = 12)	0.34	0.39	0.48	(n = 13)	0.51	0.36	0.21
B	median	HS	0.23	5.41	1.87	HS	0.28	5.9	2.26
	std.dev.	(n = 17)	0.74	1.19	0.52	(n = 17)	0.96	1.33	0.74
C	median	AN	-2.37	-3.93	3.69	AN	-2.2	-3.94	3.89
	std.dev.	(n = 19)	0.59	0.4	0.49	(n = 19)	0.46	0.73	0.41
D	median	RF	-5.37	-6.26	-4.78	RF	-5.35	-5.99	-5.01
	std.dev.	(n = 18)	0.26	0.58	0.07	(n = 18)	0.28	0.53	0.15
M	median	Stream	-0,61	-1,04	0,94	Stream	-0,61	-1,04	0,94
	std.dev.	(n = 257)	2,06	1,10	0,66	(n = 257)	2,06	1,10	0,66

The example 1) considers the initial 50% and 2) the remaining 50% of the sample sets.* Coordinates of end-members and stream (mixture) medians in three-dimensional space (U1, U2 and U3). n represents the sample size.

Table 3. Median and standard deviation (std.dev.) of end-members and stream projected in three-dimensional including artificial outliers

Naming in equations		3) End member	Coordinates* U1	U2	U3	4) End member	Coordinates* U1	U2	U3
A	median	SW	26.25	7.3	7.02	SW	26.21	7.29	6.95
	std.dev.	(n = 26)	5.51	1.73	1.68	(n = 26)	10.28	2.87	2.54
B	median	HS	0.27	5.47	1.98	HS	0.23	5.45	1.97
	std.dev.	(n = 34)	0.99	2.45	1.03	(n = 34)	1.12	1.99	0.8
C	median	AN	-2.24	-3.92	3.79	AN	-2.26	-3.95	3.74
	std.dev.	(n = 38)	0.78	1.17	0.92	(n = 38)	1.07	1.43	1.15
D	median	RF	-5.36	-6.08	-4.84	RF	-5.37	-6.11	-4.86
	std.dev.	(n = 37)	1.7	1.89	1.58	(n = 37)	1.09	1.42	0.94
M	median	Stream	-0,61	-1,04	0,94	Stream	-0,61	-1,04	0,94
	std.dev.	(n = 257)	2,06	1,10	0,66	(n = 257)	2,06	1,10	0,66

The example 3) considers outliers included at the positive extreme of the dataset of each source and 4) outliers included at the negative extreme.* Coordinates of end-members and stream (mixture) medians in three-dimensional space (U1, U2 and U3). n represents the sample size.

Table 4. Median and enlarged standard deviation (std.dev.) of end-members and stream projected in three-dimensional

Naming in equations		5) End member	Coordinates* U1	U2	U3	6) End member	Coordinates* U1	U2	U3
A	median	SW	26,25	7,29	7,00	SW	26,25	7,29	7,00
	std.dev.	(n = 25)	1.39	1.07	1.19	(n = 25)	2.32	1.78	1.99
B	median	HS	0,23	5,48	1,97	HS	0,23	5,48	1,97
	std.dev.	(n = 33)	2.56	3.87	2.06	(n = 33)	4.27	6.45	3.43
C	median	AN	-2,24	-3,93	3,71	AN	-2,24	-3,93	3,71
	std.dev.	(n = 37)	1.65	1.73	1.34	(n = 37)	2.75	2.88	2.24
D	median	RF	-5,38	-6,10	-4,84	RF	-5,38	-6,10	-4,84
	std.dev.	(n = 36)	0.8	1.69	0.46	(n = 36)	1.34	2.81	0.77
M	median	Stream	-0,61	-1,04	0,94	Stream	-0,61	-1,04	0,94
	std.dev.	(n = 257)	2,06	1,10	0,66	(n = 257)	2,06	1,10	0,66

The example 5) considers 3-times the standard deviation of the original data set and 6) 5-times the standard deviation of the original data set.* Coordinates of end-members and stream (mixture) medians in three-dimensional space (U1, U2 and U3). n represents the sample size.

Table 5. Uncertainty of individual end-member contributions to the stream and Satterthwaite (1946) approximation for the degrees of freedom calculated for the study period 2013–2014

Naming in equations End-member	A SW	B HS	C AN	D RF
Fraction of end-members contribution	0.06	0.3	0.35	0.29
Upper 95% confidence limit	0.21	0.57	0.58	0.46
Lower 95% confidence limit	0.00	0.03	0.12	0.12
Degrees of freedom	291	536	749	628

Table 6. Uncertainty of individual end-member contributions to the stream and Satterthwaite (1946) approximation for the degrees of freedom computed considering 50% of the data sets

Naming in equations End-member	1)	A SW	B HS	C AN	D RF	2)	A SW	B HS	C AN	D RF
Fraction of end-members contribution		0.06	0.3	0.35	0.28		0.06	0.28	0.35	0.3
Upper 95% confidence limit		0.21	0.57	0.58	0.45		0.21	0.55	0.58	0.46
Lower 95% confidence limit		0.00	0.03	0.12	0.11		0.00	0.02	0.12	0.14
Degrees of freedom		289	493	676	589		288	491	679	537

The example 1) was computed considering the initial 50% and 2) the remaining 50% of the sample sets.

Table 7. Uncertainty of individual end-member contributions to the stream and Satterthwaite (1946) approximation for the degrees of freedom computed after including artificial outliers

Naming in equations End-member	3)	A SW	B HS	C AN	D RF	4)	A SW	B HS	C AN	D RF
Fraction of end-members contribution		0.06	0.3	0.35	0.29		0.06	0.3	0.35	0.29
Upper 95% confidence limit		0.22	0.62	0.64	0.5		0.22	0.61	0.63	0.49
Lower 95% confidence limit		0.00	0.00	0.06	0.08		0.00	0.00	0.07	0.08
Degrees of freedom		350	448	640	529		353	554	757	621

The example 3) was computed after including outliers at the positive extreme of the dataset and 4) including outliers at the negative extreme.

Table 8. Uncertainty of individual end-member contributions to the stream and Satterthwaite (1946) approximation for the degrees of freedom computed with enlarged standard deviations

Naming in equations End-member	5)	A SW	B HS	C AN	D RF	6)	A SW	B HS	C AN	D RF
Fraction of end-members contribution		0.06	0.3	0.35	0.29		0.06	0.3	0.35	0.29
Upper 95% confidence limit		0.23	0.68	0.69	0.52		0.26	0.83	0.83	0.61
Lower 95% confidence limit		0.00	0.00	0.01	0.05		0.00	0.00	0.00	0.00
Degrees of freedom		372	225	362	312		335	122	211	172

The example 5) was computed considering 3-times the standard deviation of the original data set and 6) 5-times the standard deviation of the original data set.