



Original Paper

# Assessing Indices Tracking Changes in River Geochemistry and Implications for Monitoring

Caterina Gozzi <sup>1,2</sup> and Antonella Buccianti<sup>1</sup>

Received 24 August 2021; accepted 1 January 2022

Published online: 5 February 2022

In geochemical data analysis, assessing the potential of new techniques to identify compositional time–space changes is of great interest for monitoring purposes. This work aims to evaluate, in the light of the compositional data analysis perspective, the performance of different statistical indices in tracing the evolution of a geochemical composition and the relationships among its parts. To reach this goal, source-to-sink chemical changes in water and stream sediment composition of the Tiber river (central Italy) are analyzed using three indices: (i) the cumulative sum of unclosed perturbation factors of each composition (row sum) with respect to a reference composition; (ii) the robust Mahalanobis distance, describing the compositional differences from the same reference and, (iii) the geometric mean of each composition as a measure able to capture the interactions among the parts. The results highlight the major compositional changes downriver, allowing to explore geochemical footprints' propagation and their natural or anthropogenic origin. The tested indices are consistent in most cases, particularly if high-variability species are treated separately and low values are rare. Under this latter condition, the geometric mean of the composition shows a close connection with the cumulative sum of unclosed perturbation factors. This indicates that both indices inherit the complex history of the changes, well capturing the interactions among the parts under the influence of environmental drivers. With this awareness, the application of these methods in monitoring and applied geochemical studies could offer new insights into the inner workings of river systems and their resilience to environmental pressures.

**KEY WORDS:** Unclosed perturbation factors, Geometric mean, Mahalanobis distance, Compositional data, Geochemical changes.

## INTRODUCTION

One of the main challenges in geochemical data analysis is to describe and model time–space changes in the composition of environmental media (e.g., water, soil and stream sediment) and to link them to

the dynamics of natural or anthropic drivers. Compositional changes are often driven by thermodynamic and kinetical laws and can modify a given natural matrix in a transient or permanent way (Bethke, 2008). In this framework, most natural systems are open and classified as dissipative. This means they contain subsystems that permanently fluctuate, in an apparently stationary way, until fluctuations become strong enough to move the system to a new state, often unexpectedly (Kondepudi & Prigogine, 1998; Shvartsev, 2009; Roduner & Radhakrishnan, 2016). Earth history has been

Caterina Gozzi and Antonella Buccianti contributed equally to this study.

<sup>1</sup>Department of Earth Sciences, University of Florence, Via G. La Pira 4, Florence 50121, Italy.

<sup>2</sup>To whom correspondence should be addressed; e-mail: caterina.gozzi@unifi.it

marked by occasional and unexpected shifts in some characteristics, for example, temperature and climatic conditions, able to generate alternative states characterized by dynamic regimes at different spatial and temporal scales (Kleidon, 2010; Heinze et al., 2021; Li et al., 2021). Gradual changes in some underlying conditions (drivers) can bring a system closer to a catastrophic bifurcation point (tipping point) causing a loss of resilience. Even small perturbations may therefore trigger an abrupt shift to an alternative system state. In this context, fractal structures in time and space offer an optimal non-linear means to dissipate energy gradients developing spontaneous emergent properties, as typical of complex dissipative systems (Seely & Macklem, 2012). Thus, the presence of scale-invariant structures are examples of how the increasing energy dissipation and entropy production create order, structure, and complexity, as demonstrated in spatial patterns of geochemical data (Zuo & Wang, 2016). The identification of early warning signals (or leading indicators) represents a fundamental challenge to limit restoration costs or irreparable damages to ecosystems and to understand if a critical threshold is being approached in a gradual or a critical (sudden) way (Weinan et al., 2021). Several investigations have been undertaken on the mathematical properties of such indicators and on the identification of tools able to highlight the governing (additive or multiplicative) dynamics of a system (van Rooij et al., 2013). The presence of modularity and heterogeneity appears to favor adaptive capacity and gradual changes, while connectivity and homogeneity increase the resistance to change, thus opening the path to sudden critical transitions (Scheffer et al., 2012). It is evident that the nature of the interconnections among the parts of a system is fundamental to understanding its behavior. In our case, the parts that can interact, simultaneously or at a different hierarchical scale, are the single variables of the composition and the whole is the composition itself, a condition replicated for all geological samples. However, when geochemical data are analyzed by using several mathematical and statistical tools, it is important to recognize that concentrations are relative data (% , ppm, mg/L and similar measurement units) and that the investigation of the variance-covariance structure of the whole cannot ignore the presence of induced relationships (Chayes, 1960). In fact, it is not sufficient to apply some multivariate procedure to capture the joint behavior of the whole, since compositional data do

not pertain to a Euclidean space. Implications of using an incorrect sample space are evident especially when comparing confidence regions in a constrained space with the real ones (Buccianti & Magli, 2011). Hence, it is necessary to consider the principles of the compositional data analysis (CoDA, Aitchison 1982) and work in the *simplex* sample space by using its algebraic structure (the so-called “stay in the simplex approach”) or transforming data to move variables from the simplex to the Real space, (“log-ratio approach”). In this research, the “stay in the simplex approach” was adopted by focusing on the use of the perturbation operator to investigate compositional changes from source to mouth in surficial waters and stream sediments of the Tiber river (central Italy). Measuring the “nearness” or “farness” between two compositions requires the use of a distance/difference measure, and the results may depend on how this distance is measured. For this reason, testing the performance of different statistical methods to trace the evolution of a geochemical composition is of critical importance. Starting from a  $n \times p$  data matrix ( $n$  number of cases,  $p$  number of variables or parts) unclosed perturbation factors were estimated considering the difference with respect to a selected reference composition, thus obtaining a new  $n \times p$  matrix. The cumulative sum of these factors for each composition ( $n$  row sums) was then compared with  $n$  values of the robust Mahalanobis distance (RMD), calculated in a compositional context, and with  $n$  geometric means. The latter, representing the compositional barycenter of the composition (row barycenter), assumes in CoDA a central role in the log-ratio approach, tracing the interactions among the parts inside each composition (Pawlowsky-Glahn et al., 2015). The goal of this research is to assess the potential of these techniques for their applications to environmental monitoring, with special attention to the role of the geometric mean as a potential quick monitoring index.

## MATERIALS AND METHODS

### Overview on the Tiber River and Its Catchment

The Tiber river was chosen as a study area due to the extensive knowledge acquired from previous studies (Gozzi et al., 2019) on the different sources of riverine variability (i.e., natural, seasonal, anthropogenic). A deep understanding of the

catchment processes is important in order to evaluate the performance of the tested methods. The Tiber river is the second-largest Italian river in terms of catchment size (ca. 17,375 km<sup>2</sup>), second only to the Po river. It contributes, with a mean annual discharge of about 240 m<sup>3</sup>/s, to almost 20% of the riverine inputs into the Tyrrhenian Sea. The basin has a mean altitude of 520 m and a climate regime from sub-coastal to marine with a mean annual precipitation of about 1200 mm (Bagnini et al., 2005). During its 409 km course from the Mt. Fumaiolo to the sea, the Tiber river is fed by several major tributaries: Paglia-Chiani and Treia on the right bank, Chiascio-Topino, Nera-Velino and Aniene on the left (Fig. 1).

A torrential and turbulent regime characterizes the upper stretch of the river, whereas a slow and meandering flow dominates after the confluence with the Nera. The natural variability of the catchment is linked to its heterogeneous morphological, geological and hydrogeological setting which defines sub-basins with important differences in terms of lithology, discharge rate, summer base flow and winter run-off (Boni et al., 1986). The upper basin is dominated by impermeable flysch sediments and surface run-off processes prevail over infiltration, leading to an irregular outflow regime (Fig. 1). Conversely, the SE sector is characterized by Apennine limestones which ensure high infiltration and a more stable outflow with reduced seasonal variations. Potassic and ultrapotassic volcanic rocks prevail to the SW, with intermediate hydrogeological features compared to the previous ones.

Anthropic pressures are significantly high due to the presence of important urban centers (e.g., Rome), industrial activities and diffuse pollution from agricultural areas, representing 53% of the total land use. The discharge distribution of the Tiber river has been significantly impacted by the construction of multiple hydropower dams and weirs (Fig. 1). A decrease in the amount of suspended load is being observed over the last decades (Iandanza & Napolitano, 2006) together with a reduction of sediment supply to the coast. This is strongly contributing to the erosion of the beach at Ostia, the coastal environs of Rome (Ruggiero, 2020).

### Geochemical Dataset

The geochemical dataset consists of: (i) 38 water samples collected from 19 sampling sites along

the river course in different hydrological conditions; and, (ii) 17 stream sediments taken from the river bank at the same locations, except for two points at which it was not possible to collect the samples due to safety reasons (Fig. 1). Water samples refer, with a good approximation, to high and low flow rate conditions, even if distributed over two years (2017–2018), as highlighted by Gozzi et al. (2021). All major cations and ions were considered for a total of 10 chemical variables with concentrations expressed in mg/L. For stream sediments, 8 major and 8 trace elements were selected for their high variability and environmental interest. Concentrations were reported in % and ppm for major and trace elements, respectively. In CoDA, it is a good practice to consider all parts of the composition in the same units if a joint analysis is planned. In our case, major and trace elements were analyzed separately; therefore, the use of common measurement units is not compulsory. The only two values below the lower detection limit in the water dataset were replaced by 2/3rds of the detection limit, while none are found for the stream sediment dataset. In the presence of a significant number of values below the detection limit, the use of imputation methods such as those described by Templ et al. (2016) is suggested. More detailed information about sampling and analytical methods is reported in Gozzi (2020).

### Computational Methods

#### *Perturbation Operator*

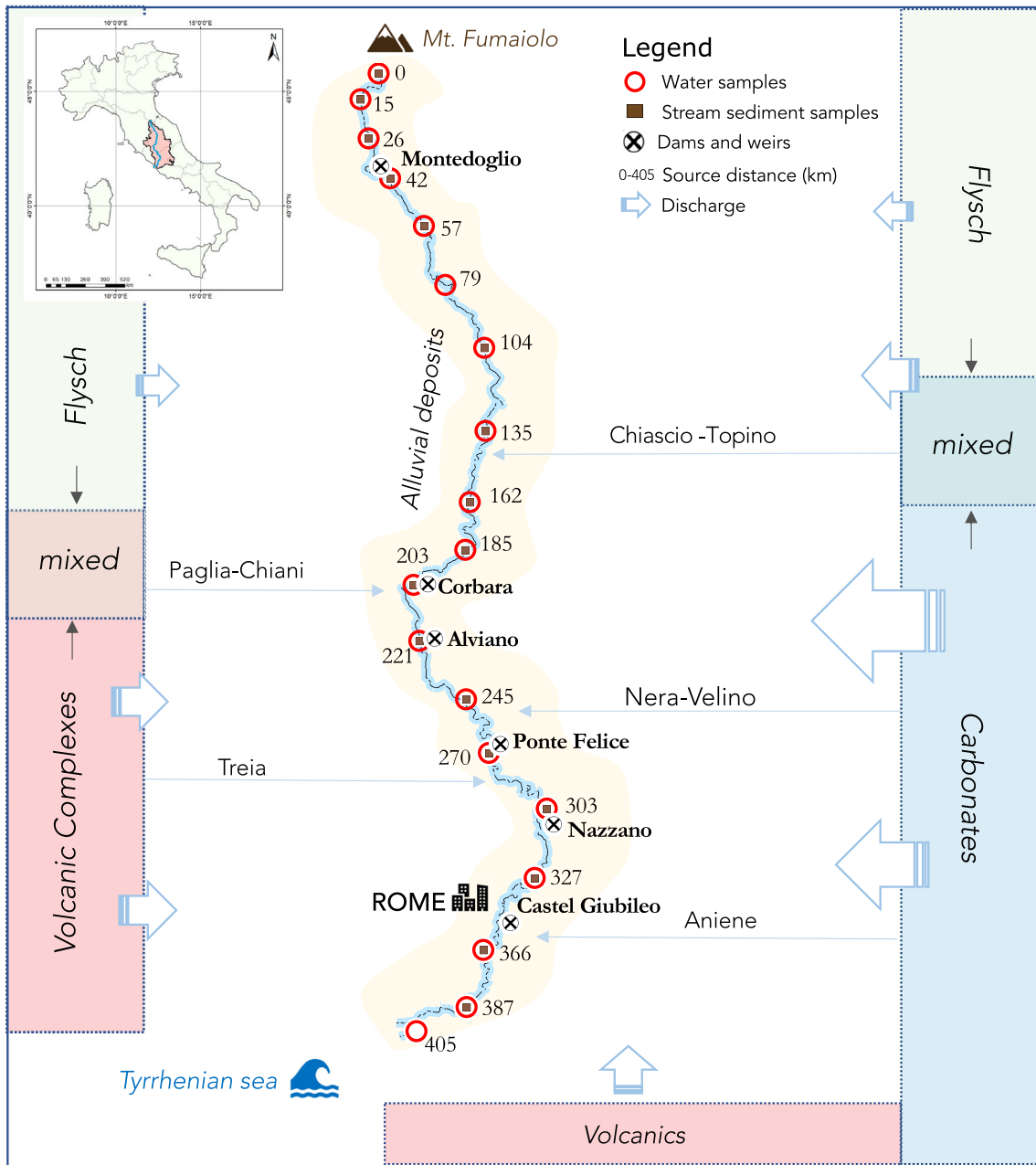
Since the sample space associated with a  $D$ -part composition is the unit simplex:

$$S^D = \{[x_1, \dots, x_D] : x_i > 0 \ (i = 1, \dots, D), \ x_1 + \dots + x_D = 1\} \quad (1)$$

in order to trace compositional changes, the perturbation is the group operation able to counterpart translation in real space (classical Euclidean statistics) and rotation on the sphere (circular statistics).

Given two  $D$ -part compositions  $x, y \in S^D$ ,  $x \oplus y$  is defined by:

$$x \oplus y = \frac{[x_1 y_1, \dots, x_D y_D]}{(x_1 y_1 + \dots + x_D y_D)} = C[x_1 y_1, \dots, x_D y_D] \quad (2)$$



**Figure 1.** Location of the sampling points along the Tiber river together with a simplified scheme depicting the geology and water circulation in the catchment.

where the “closure” operator  $C$  standardizes the perturbation vector by dividing by the sum of its components so that the sum is equal to unity. Following Aitchison and Ng (2005), the operation  $\oplus$  defines an Abelian group on the simplex with identity  $e = (1/D)[1, \dots, 1]$  and inverse

$x^{-1} = C[x_1, \dots, x_D]$ . This mathematical property is fundamental to characterize the operation that changes a  $D$ -part composition  $x$  into a  $D$ -part composition  $y$ , thus answering the geochemical question: what is the perturbation  $p$  such that a composition  $x$  is transformed, by the action of several environ-

mental and anthropic factors, in a composition  $y$ ? The answer is given by the inverse operation, also called perturbation difference by analogy with standard operations in real space (Pawlowsky-Glahn et al., 2015):

$$p = y \ominus x = C \left[ \frac{y_1}{x_1}, \dots, \frac{y_D}{x_D} \right] \quad (3)$$

The perturbation, together with the powering operation, analogous to scalar multiplication in real space, both defined in  $S^D$ , satisfies the requirements for operations of a vector space for which an inner product, a norm and a distance can also be defined (Billheimer et al., 2001; Pawlowsky-Glahn & Egozcue, 2001). Consequently, any measure of difference between compositions  $x$  and  $y$  must be expressible in terms of perturbations. In geochemistry, the perturbation approach finds easy application where an initial specimen of composition  $x_0$  is subjected to a sequence of perturbations  $p_1, \dots, p_n$  to reaching its current state  $x_n$ . For example, compositional changes can be traced for river water and stream sediment chemistry starting from the source to the sink following the landscape morphology. The results can be interpreted in the light of the dynamics of geochemical processes, thus discovering areas experiencing a major rate of change or more resilient zones. The outcomes could be useful for both monitoring environmental changes, in the light of climate changes modifying erosion rates and hydrological cycle, and applied environmental geochemical studies. Perturbations may be computed sequentially after ranking the data according to a spatiotemporal sequence or a physical-chemical criterion (Gozzi et al., 2020; Sauro Graziano et al., 2020). Alternatively, they can be determined with respect to a reference composition with cases previously ordered following some criteria (e.g., spring distance, conductivity, elevation, or other landscape parameters). Common examples of reference compositions are those of spring water for rivers or the average Earth crust for stream sediments.

Perturbation can be represented as: (i) a composition expressed in percentage (%), as in Eq. 3; (ii) multiplicative non-closed factors by considering ratios of Eq. 3 without the closure application; and (iii) percentage of increase/decrease. A detailed description of the use of these modalities is presented in Egozcue and Pawlowsky-Glahn (2011).

In this work, perturbations are represented as multiplicative non-closed factors (Pfs) by consider-

ing the ratios  $\left[ \frac{y_1}{x_1}, \dots, \frac{y_D}{x_D} \right]$  of Eq. 3 without applying the closure operator  $C$ . For every factor,  $x_i$  represents each term of the reference composition, the choice of which is discussed in Section “The Reference Composition: a Central Issue”, while  $y_i$  the terms of the sample composition from the ordered sequence along the river course. Since Pfs take into account the ratios  $y_i/x_i$  for  $i = 1, \dots, D$ , the closer the latter are to 1, for each  $D$ -part, the more alike the two compositions are. On the contrary, if a ratio is  $\gg 1$  or  $\ll 1$ , this indicates, respectively, an excess or a diminution of the involved variable compared to the benchmark. The resulting Pfs, representing the contribution of every single term of the composition with respect to the reference, are reported with the letter “p” followed by the considered chemical component (e.g.,  $pHCO_3^-$ ,  $pCa^{2+}$ ). These compositional deviations monitored by the Pfs might be the sentinels of major environmental modifications or stressors. For this reason, Pfs were compared with the variations of potential drivers within the corresponding catchment (e.g., lithotypes and land-use). The latter were derived from Copernicus Land Monitoring Service (2019) and ISPRA Ambiente (2017), respectively, and estimated according to the procedure described in Gozzi et al. (2021). Lithotypes and land use changes are here expressed in terms of drainage area differences between each point and the previous one along the source-to-mouth sequence.

Ultimately, all Pfs of a  $D$ -part composition were summed for each row  $j$  of the data matrix as in Eq. 4 and reported in the text as  $\sum$ Pfs.

$$\sum_{i=1}^D Pf_j = p_{j1} + p_{j2} + \dots + p_{jD} \quad \text{for } j = 1, \dots, n \quad (4)$$

This operation was done in order to obtain a comprehensive measure to trace the downstream evolution of the chemical composition of water and stream sediment samples. Pfs are here reported as multiplicative non-closed factors and their denominators do not contain common parts. Therefore, they are positive real numbers and the sum by row of these factors could be considered a proper operation in the positive Real space. This quantity, representing an interesting cumulative information about the compositional change, will be compared

with the other tested measures illustrated in the following Section.

### *Comparative Methods: Mahalanobis Distance and Geometric Mean*

The Mahalanobis distance is a distance measure which accounts for the variance–covariance matrix of the data (Mahalanobis, 1936). Its robust version calculated through the minimum covariance determinant (MCD) estimator (Rousseeuw & Van Driessen, 1999; Filzmoser & Hron, 2008; Filzmoser et al., 2012) is defined as:

$$RMD_i = \sqrt{(x_i - \mu_{MCD}) \Sigma_{MCD}^{-1} (x_i - \mu_{MCD})^T} \quad \text{for } i = 1, \dots, n \quad (5)$$

where  $x_1, \dots, x_n$  are the observations in the  $p$ -dimensional space having center  $\mu_{MCD}$  and covariance  $\Sigma_{MCD}$ , these last calculated from the sample data  $\bar{x}$  and  $s$ , respectively. To reliably estimate the location and scatter from which distances are computed, a certain number of compositions is required, and a single reference sample (or case) is not sufficient. This number depends on how many variables and cases are present in the dataset. Distances were computed in R with the package *robustbase* (Maechler et al., 2018) after having transformed the data in pivot coordinates (Filzmoser et al., 2018), a special case of the isometric log-ratio coordinates, calculated by means of the function *pivotCoord* within the R package *robCompositions* (Templ et al., 2011). Therefore,  $\mu_{MCD}$ , distances to  $\mu_{MCD}$  and  $\Sigma_{MCD}$  are here estimated from the pivot coordinates.

The geometric mean may represent an alternative measure of location to the mean when data of a given geochemical variable are skewed and near-lognormal, but it is equally susceptible to outliers (Rock 1988). As pointed out by Reimann and Filzmoser (2000), the geometric mean can be used, but it may be problematic in some cases. Rock (1988) highlights that its use requires caution since it vanishes or becomes imaginary in the event of zero or negative values in the data. This occurs in geochemistry with all standardized data, values lower than a detection limit or with real negative data (e.g., stable isotopes values). In CoDA, the geometric mean is well known since it enters the calculation of the commonly applied centered and isometric log-ratios transformations (Aitchison,

1986; Egozcue et al., 2003). In this framework, it is usually calculated not for the values of a given variable but for a composition or its subsets; in other words, the calculus is performed on the rows of a data matrix. The geometric mean is also used to describe, with bar plots, differences between groups in compositional datasets (Martín-Fernández et al., 2015). To summarize, when computed for data belonging to the same variable, it could measure, under some assumptions, the univariate barycenter of the values. However, the geometric mean  $g(x)$  can also be computed in a multivariate framework for a  $D$ -part chemical composition  $x = x_1, x_2, \dots, x_D$ , as follows:

$$g(x) = \sqrt[p]{x_1 \cdot x_2 \cdots x_D}. \quad (6)$$

In this case,  $g(x)$  is calculated for the entire composition of each sample from data pertaining to different chemical variables. What does this measure represent from a geochemical perspective if calculated in a multivariate context? Similarly, it is expected that it will measure the compositional barycenter of the entire composition. The main difference is that, in this case, the barycenter is not univariate but multivariate and the effects of multiple variables are taken into consideration altogether. To better answer this question, the geometric mean of water and sediment compositions was compared with both  $\sum$ Pfs and RMD. This was done with the aim of gaining a greater understanding of its geochemical meaning, its relation with single variable fluctuations and with other multivariate measures.

The consistency of the three different approaches,  $g(x)$ ,  $\sum$ Pfs and RMD, was cross-checked, after having removed outliers, by means of H-scatterplots (at lag 0) created using the R package: *Astsa* (Stoffer, 2021). All the considered measures represent potential tools to monitor the spatial or temporal evolution of a composition, taking into consideration the interaction among its constituent parts. By comparing these measures, we expect to verify if all these quantities equally capture the fundamental compositional changes downriver and, if differences are found, the potential influences will be examined.

### **The Reference Composition: A Central Issue**

The choice of the reference composition is critical since it influences the nature and magnitude

of the monitored changes. To capture all variations, the objective is to identify a suitably homogeneous composition not too far from the environmental conditions of the target samples. As concerns rivers, the spring water composition is often considered as a benchmark to detect downstream hydro-geochemical variations linked to water–rock interaction processes as well as various anthropogenic pressures. However, its composition might significantly differ from that of streams since it flows directly from an aquifer, not being affected by the multiple interactions with the surficial environment. This is the case of the Tiber source which, with a total dissolved solids (TDS) of about 250 mg/L, is compositionally inhomogeneous. This is mainly due to the high and low relative concentrations of  $NO_3^-$  and  $K^+$ , respectively. For this reason, the mean composition of 14 samples collected over the two-year period of study from the upper reaches of the catchment and having the lowest TDS (310–510 mg/L) was chosen as a reference from which Pfs are calculated. These include 3 samples collected from the main course and 11 from tributaries, characterized by a low variability and absence of outlying observations. In the presence of outliers and therefore a significant difference between the mean and median value, the use of the latter as estimator for the reference is recommended. The resulting composition represents

a local pristine surficial water of calcium-bicarbonate type with a low content of the other ions (Table 1a).

The same waters were used to estimate the robust compositional center and covariance matrix from which RMDs were calculated (Gozzi et al., 2019). The results measure how each observation differs from the benchmark condition, taking into account the multivariate information. In this way, Pfs and RMDs were computed starting from common reference compositions, thus helping to compare the potentials of the two methods in disclosing compositional changes.

A similar approach was used for stream sediments by considering 15 homogeneous and uncontaminated samples, 3 from the main course and 12 from tributaries. The final reference compositions for major and trace elements are reported in Tables 2a and 3a.

## RESULTS

### Compositional Perturbations in River Water

The medians of the Pfs exceed 1 for most of the variables, indicating an increased ion content with respect to the reference (Table 1b–c). The sole excep-

**Table 1.** (a) Reference water composition from which unclosed perturbation factors are calculated; summary statistics of unclosed perturbation factors for each term of the composition in flood (b) and drought (c) periods. Median values <1, >1, >1.5 and >2 are highlighted in green, yellow, orange and red colors, respectively

a)	$HCO_3^-$	$F^-$	$Cl^-$	$NO_3^-$	$SO_4^{2-}$	$Ca^{2+}$	$Mg^{2+}$	$Na^+$	$K^+$	$NH_4^+$
Ref. (mg/L)	64.81	0.02	2.67	0.28	6.59	18.29	3.75	3.19	0.39	0.02

b)	$pHCO_3^-$	$pF^-$	$pCl^-$	$pNO_3^-$	$pSO_4^{2-}$	$pCa^{2+}$	$pMg^{2+}$	$pNa^+$	$pK^+$	$pNH_4^+$
Median	0.94	1.39	1.17	3.46	1.37	0.98	0.88	1.22	1.52	1.12
Min.	0.77	0.70	0.87	0.64	0.52	0.89	0.29	0.60	0.01	0.47
Max.	1.02	3.56	3.4	6.63	1.95	1.24	1.13	2.29	2.63	3.32

c)	$pHCO_3^-$	$pF^-$	$pCl^-$	$pNO_3^-$	$pSO_4^{2-}$	$pCa^{2+}$	$pMg^{2+}$	$pNa^+$	$pK^+$	$pNH_4^+$
Median	0.84	2.39	1.95	2.99	1.92	0.87	0.83	1.52	1.72	1.44
Min.	0.22	0.46	1.06	0.66	0.47	0.31	0.24	0.61	0.01	0.09
Max.	1.03	10.1	14.0	9.46	2.24	1.24	1.08	7.48	4.89	17.5

**Table 2.** (a) Reference stream sediment composition for major elements from which unclosed perturbation factors are calculated; (b) summary statistics of the unclosed perturbation factors for each element of the composition

a)	Si	Fe	Mn	Mg	Ca	Na	K	P
% wt	22.58	2.00	0.07	1.03	12.79	0.35	1.27	0.07

b)	pSi	pFe	pMn	pMg	pCa	pNa	pK	pP
Median	1.06	1.29	1.01	1.11	0.81	1.26	1.14	0.88
Min.	0.66	0.33	0.23	0.21	0.06	0.06	0.98	0.20
Max.	1.58	1.79	1.61	2.34	1.46	2.53	1.76	1.64

Median values <1 and >1 are highlighted in green and yellow, respectively

**Table 3.** (a) Reference stream sediment composition for trace elements from which unclosed perturbation factors are calculated; (b) summary statistics of unclosed perturbation factors for each element of the composition

a)	Cu	Sr	Rb	Pb	Zn	Ni	Ba	Cr
ppm	46.4	287.4	71.2	63.5	83.0	42.2	349	67.8

b)	pCu	pSr	pRb	pPb	pZn	pNi	pBa	pCr
Median	0.77	0.86	1.13	0.75	0.82	0.98	1.08	1.15
Min.	0.31	0.29	0.68	0.31	0.37	0.37	0.54	0.70
Max.	1.55	1.48	1.99	2.72	1.28	1.84	1.60	3.29

Median values <1 and >1 are highlighted in green and yellow, respectively

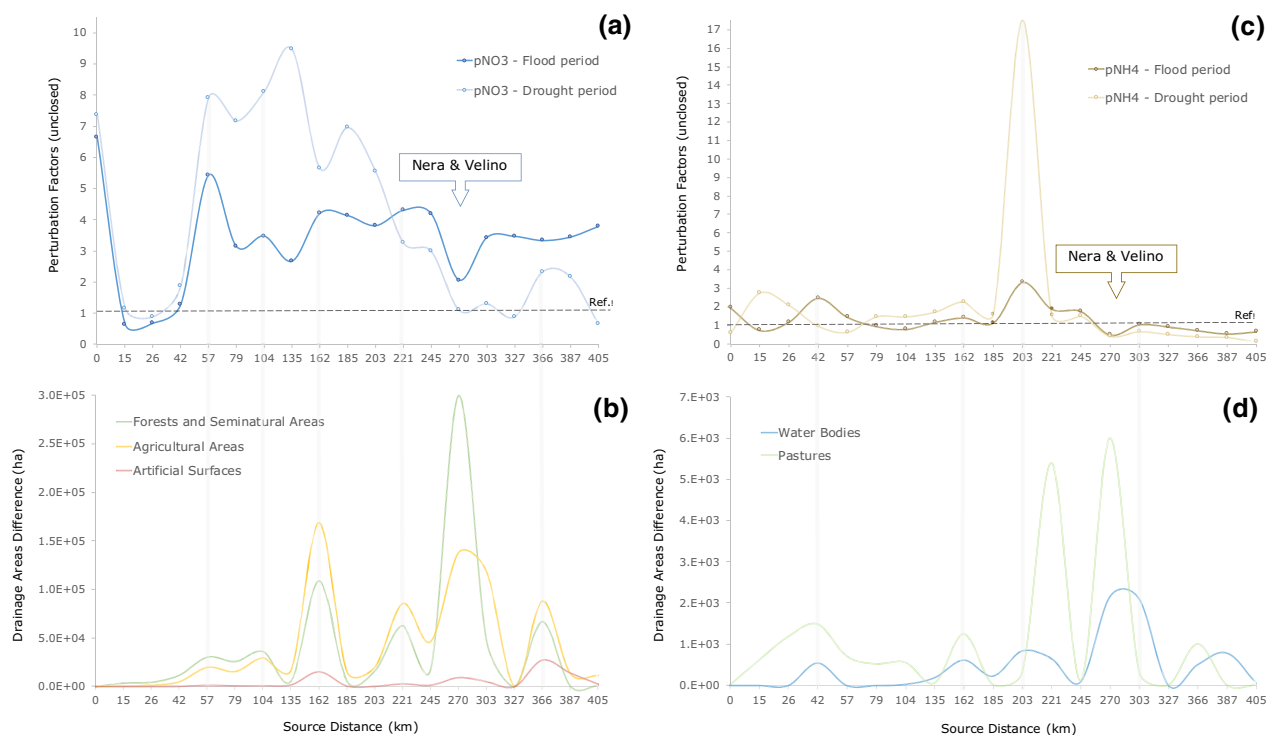
tions are  $pCa^{2+}$ ,  $pHCO_3^-$  and  $pMg^{2+}$  showing medians ranging from 0.83 to 0.94. The higher values are found for  $pNO_3^-$  in flood periods (3.46) and  $pF^-$ ,  $pCl^-$ ,  $pSO_4^{2-}$  in dry conditions (2.39, 1.95 and 1.92, respectively). Pfs' variability significantly increases in the season of minimum flow as well as the skewness of the distributions, with  $pNH_4^+$  showing the highest maximum value. The curves represented in Figure 2a, c highlight the evolution of  $pNO_3^-$  and  $pNH_4^+$  from the source to the mouth. These species are handled separately from the rest of the composition owing to their large range and link to anthropogenic processes.

After a peak at the spring,  $pNO_3^-$  starts to significantly increase at a distance of 57 km. High perturbations are monitored down to 57 km with values up to 9.5 reached in the dry period. At 270 km, when the Nera-Velino river system enters the main course, Pfs decrease, reaching a value similar to the reference in summer. Comparing the

increments of drainage area referring to different land uses (Fig. 2b) with  $pNO_3^-$  fluctuations, several matches are noticed especially with forests and agricultural areas. On the contrary, Pfs of  $NH_4^+$  are overall more restrained (Fig. 2c). High values occur mainly in the upper river stretch and at 203 km, downstream from the Corbara reservoir. At this place, a great spike shows up achieving a value of 17. A decrease is also found in this case at 270 km, after which Pfs remain below the benchmark until the river mouth. Potential associations of these changes with the increasing area covered by pastures and water bodies are inferred from Figure 2d.

The curves depicting the downstream evolution of  $pCa^{2+}$ ,  $pHCO_3^-$  and  $pMg^{2+}$  confirmed their stable pattern with low changes with respect to the reference in both the sampling conditions (Fig. 3-a, b). Conversely, the Pfs of the other species show a higher variability downstream from 221 km. In the





**Figure 2.** Unclosed perturbation factors (Pfs), ordered from source to mouth, for (a)  $NO_3^-$  and (c)  $NH_4^+$  in the Tiber river waters; the horizontal black dashed line marks the reference value. (b) Differences in drainage areas characterized by forests and seminalural areas, agricultural areas, artificial surfaces and (d) water bodies and pastures (Copernicus Land Monitoring Service 2019); the single area is calculated as difference between that of each point and the previous one along the sequence; the vertical gray lines indicate potential associations between the considered Pfs and land use variations.

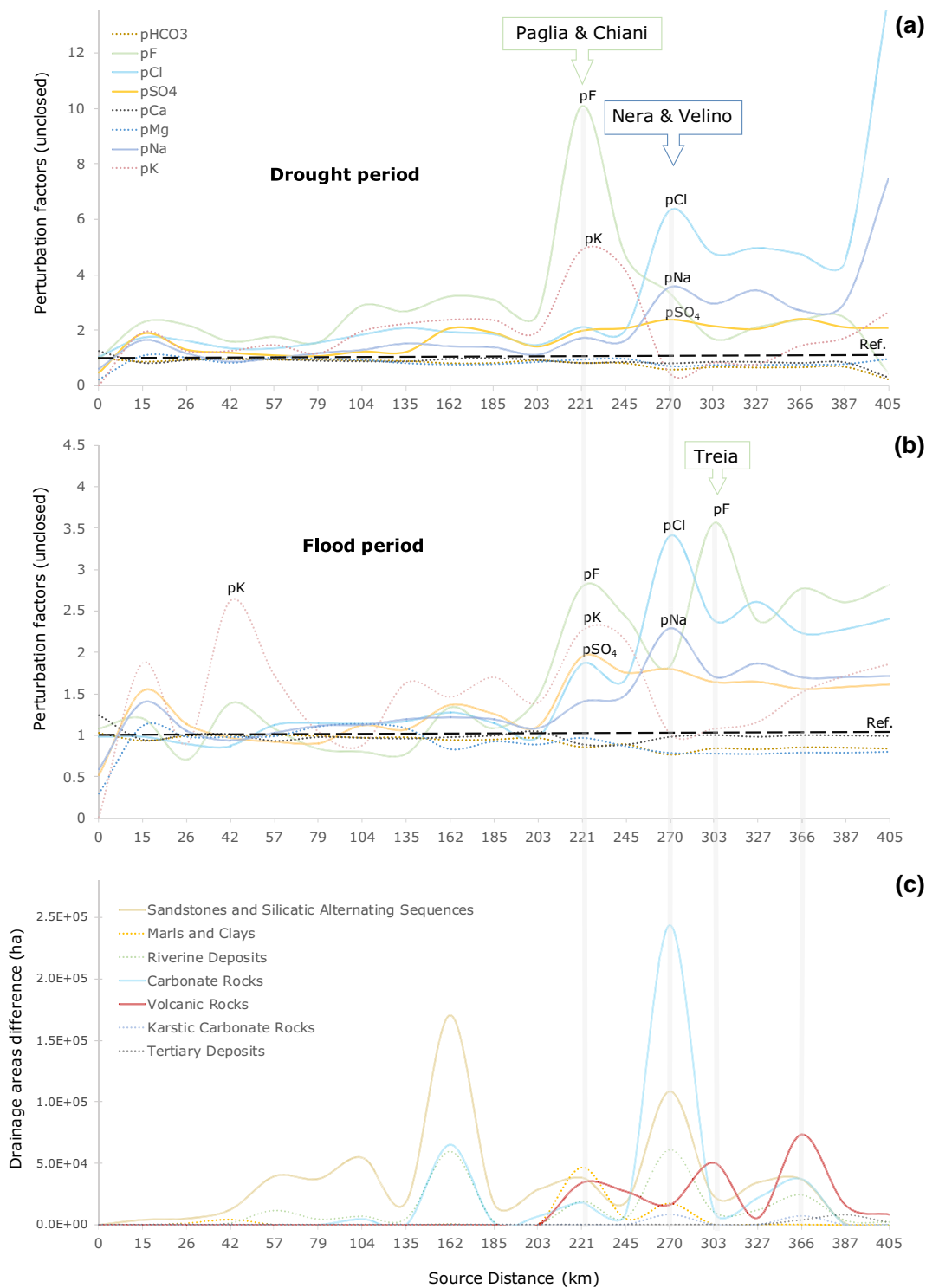
upper-middle course, these variables do not contribute as much to the compositional change. Minor perturbation increments in this area are mainly due to  $K^+$  and  $F^-$ . Values much lower than 1 are detected at the source. This fact, together with the high  $pNO_3^-$  (Fig. 2a), confirmed the spring to be compositionally far from the main river. Corresponding to the input from the Paglia-Chiani river system at 221 km, important deviations take place primarily in terms of  $K^+$ ,  $F^-$  and  $SO_4^{2-}$  with increased perturbations in the dry period. Pfs of  $Cl^-$  and  $Na^+$  are affected as well, but higher values are recorded downriver at 270 km, followed by a further fluctuation at 327 and a peak to the mouth solely in summer (Fig. 3a, b). The shifting points in the river water match fairly well the different drained lithotypes (Fig. 3c). In fact, the major compositional changes appear associated with a greater lithological complexity marked by the overlapping fluctuations of Fig. 3c (e.g., 221 and 270 km). Particularly noteworthy is the perfect correspondence of  $pF^-$  with the increasing areas underlain by the volcanic rocks.

## Compositional Perturbations in Stream Sediments

### Major Elements

The medians of the Pfs for major elements slightly exceed 1 for most of the variables (Table 2b). Values lower than 1 are only found for  $pCa$  and  $pP$  (0.81 and 0.88, respectively). Overall, the medians are smaller compared to those of river waters, with no values over 1.3. Similarly, data variability is also reduced, with the highest range belonging to  $pNa$  and  $pMg$  (Table 2b). Unlike the Pfs of the waters, those of stream sediments are characterized by multiple fluctuations along almost the whole river course (Fig. 4a).

Notable increments of  $Mg$ ,  $Ca$  and  $Fe$  with respect to the reference affect the primary sediments supplied from headwater streams. Downstream, the most relevant perturbation mainly involves  $Na$ , which almost matches the increment of drainage area featured by flysch sediments. As the lithological heterogeneity of the watershed increases, the Pfs of multiple elements show peaks at selected locations.



**Figure 3.** Unclosed perturbation factors (Pfs), ordered from source to mouth, for major elements in the Tiber river waters in **(a)** drought and **(b)** flood conditions; the horizontal black dashed line marks the reference value. **(c)** Differences in drainage areas characterized by different classes of lithotypes (ISPRA Ambiente 2017); the single area is calculated as difference between that of each point and the previous one along the sequence; the vertical gray lines indicate potential associations between Pfs and lithological variations.

The main increases are monitored for: (i) pMn at 203 km; (ii) pMg, pFe and pK at 221 km; (iii) pNa at 270 km; (iv) pFe, pK and pMn at 303 km, and (v) pP from 366 km to the river mouth. The increments of pFe and pK appear well related to the presence of volcanic rocks (Fig. 4c). Moreover, it is worth noting that most of the perturbations are not maintained downstream but instead appear sourced locally, not showing any increasing or decreasing pattern along the river course. Only pNa and pFe seem to partially preserve their perturbation rate in the early-medium and lower river stretch, respectively.

### Trace Elements

Pfs of trace elements are generally lower compared to the major ones, showing medians below 1 for all elements except for Ba, Rb and Cr (Table 3b). The highest range is detected for pCr, which exhibits a significant increase from 57 to 104 km, where it reaches a maximum value of 3.29 (Fig. 4b). The pPb shows the second widest range, with four great spikes at 15, 135, 221 and 366 km. Further notable deviations from the reference are detected for: (i) pSr between 14–26 km, (ii) pCu at 42 km; and (iii) pNi at 162 km. In the lower reaches, Rb and Cr show higher perturbations in correspondence of volcanic lithotypes, thus following the same behavior as pFe and pK. Overall, no trends are found during the source to sink propagation but rather trace element perturbations seem limited to short-range extents. Additionally, unlike major elements, trace element Pfs seem to flatten from 245 km to the river mouth.

### Comparison among Monitoring Indices

#### River Waters

The comparison between the  $\sum$ Pfs with the other methods (i.e., the row geometric mean and the RMD) reveals, for the water chemistry, a good consistency between the different indices (Fig. 5).

The main points of compositional changes along the river course are well exhibited in all cases, with only a few differences. During the high-flow period, these crucial points are located in the upper river course at 15 and 42 km, and in the lower stretch at 221 and 270 km, with an additional variation at 327, only identified by the RMD. The green sections of the curves (79–185 km) are generally characterized

by the absence of significant compositional variations. In the dry period, the magnitude of these changes increases with a more relevant variation detected between 327 and 366 km and spike at the river mouth. It is worth noting that the  $\sum$ Pfs seems to emphasize the peak at 221 km, with respect to  $g(x)$  and RMD, which instead give more prominence to that located at 270 km.

The h-scatterplots represented in Figure 6a–f confirm the consistency of the methods showing moderately high cross-correlations values, ranging from a minimum of 0.72 to a maximum of 0.94.

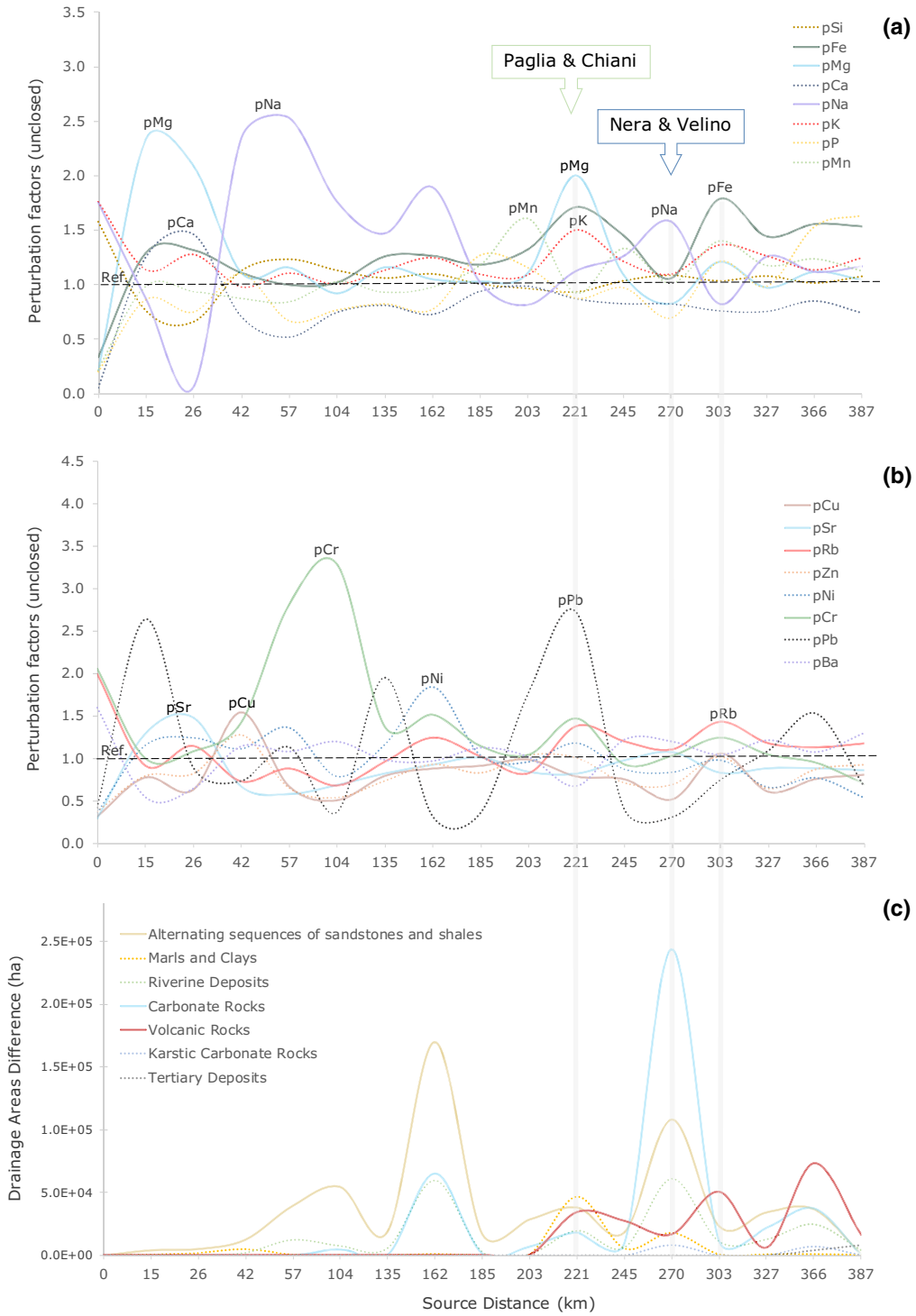
#### Stream Sediments

The statistical indices computed for major elements show common changing points (i.e., those at 15, 42, 221 and 366 km) with river waters (Fig. 7).

The major difference is the presence of an additional spike at 303 km and the absence of that at 270 km. The performance of the  $\sum$ Pfs and  $g(x)$  looks similar, as confirmed by a cross-correlations value of 0.88 (Fig. 8a).

On the contrary, the RMD shows a different pattern, especially in the upper-middle course with displacement of some peaks and a lower correlation with the other indices. This inconsistency is even more pronounced if the trace element compositions are considered, with cross-correlations values close to zero (Fig. 8e–f). One of the main divergences of the RMD is the great spike at 104 km (Fig. 7f) not depicted by the other parameters. The  $\sum$ Pfs and  $g(x)$  of trace elements seem to have a similar behavior, even though the correlation index indicates a weaker association (0.45) between them. Their pattern also agrees with that of major elements, with the sole exception of a greater magnitude of the fluctuations and a more marked spike located around 135–162 km.

It is worth noting that pronounced changes in stream sediment composition and especially, those highlighted by  $\sum$ Pfs and  $g(x)$ , occur most commonly downstream from dams (Fig. 1). This is verified for the sample at 42 km located downflow to the Montedoglio dam and similarly for those located after the Corbara (203 km) and Alviano (221 km) dams, respectively. These represent the largest artificial reservoirs along the river course, having a crucial role in curtailing sediment transport to the sea (Tacconi et al., 2020). The same occurs for stream sediment samples at 303 and 366 km, situ-



**Figure 4.** Unclosed perturbation factors (Pfs), ordered from source to sink, for major (a) and trace elements (b) in the Tiber river sediments; the horizontal black dashed line marks the reference value. (c) Differences in drainage areas characterized by different classes of lithotypes; the single area is calculated as difference between that of each point and the previous one along the sequence; the vertical gray lines indicate potential associations between Pfs and lithological variations.

ated downstream of the Nazzano and Castel Giubileo weirs, respectively.

## DISCUSSION

### Downstream Propagation of Geochemical Signals in the Tiber River

Pfs seem to track quite well the compositional fluctuations in the water chemistry from the source to the river mouth, with a good agreement with the results obtained in previous studies by means of different methods (Panichi et al., 2005; Gozzi et al., 2021; 2019). This is particularly useful when the focus is on the monitoring of a specific contaminant or toxic element. An example for the Tiber river is represented by the analysis of  $pNO_3^-$  (Fig. 2a, b). The pressure of forests and agricultural areas appear to be responsible for its high values, by supplying  $NO_3^-$  from anthropic and natural sources (i.e., synthetic fertilizers or  $NO_2$  from biological decay, respectively). The most prominent instances occur especially during the dry period, as also highlighted by Gozzi et al., (2021). Conversely, the volume of water from the Nera-Velino river system (270 km), draining Apennine limestones, causes a strong dilution of  $NO_3^-$  contamination, as demonstrated by the lowering of values at both sampling times. The same effect also occurs for  $pNH_4^+$ , whose spikes in the upper course highlight a potential link with the presence of pastures, representing a possible source of  $NH_4^+$  through animal waste decomposition. The huge perturbation located just below the Corbara dam and some matches of the increases with the presence of water bodies (Fig. 2b, c) suggest the potential role of reservoirs and dams in rising  $NH_4^+$  concentration downstream. The latter, heavily modifying the hydraulic and sedimentation conditions, reshapes the distributions of microorganisms, affecting the nitrogen retention and transformation in rivers (Xia et al., 2018; Maavara et al., 2020). The effects of water impoundment, such as lack of turbulence and mixing, thermal stratification and anoxic conditions, may contribute, especially with bottom-release dams, in increasing ammonia in downstream waters (Symons et al., 1965; Beutel 2006).

Hence, the perturbation operator highlights the changes to which the investigated variables are subject, making it possible to link them with driving

agents of variation. This facilitates the identification of those species which are more resilient to environmental changes, with relevant implications for water system monitoring plans. The stable patterns shown by  $pCa^{2+}$ ,  $pHCO_3^-$  and  $pMg^{2+}$  most likely depict this condition, showing an efficient response in the face of the sharp increase of carbonatic lithotypes. As highlighted by Gozzi (2020) for European rivers, this state is favored by the rapid dissolution reaction of carbonates compared to silicates which maintain river waters permanently in near-saturation conditions.

By contrast, the other species are significantly perturbed due to elements inputs from surface runoff and groundwater circulation stemming from physicochemical weathering of the existing lithotypes. The three spikes of  $pF^-$  well demonstrate the leaching of the volcanic districts of Mt. Amiata (Paglia river), Mts. Sabatini and Cimino (Treia river) and Albani hills, respectively. Differently,  $K^+$  is perturbed only in correspondence of the first volcanic district, where also  $pSO_4^{2-}$  increase as a result of local mixing process with  $SO_4^{2-}$ -rich ground waters (e.g., Minissale 2004, and references therein). The perturbation of  $K^+$  at 42 km in the flood period could be linked to the use of fertilizers or to the absence of uptakes by vegetation in winter when plants are dormant (Berner & Berner, 1996). Groundwater circulation dissolving gypsum, anhydrite and halite from Triassic evaporites of the Narni-Amelia aquifer system (Nera sub-basin) (Froncini et al., 2012) explain the considerable perturbation values of  $Cl^-$  and  $Na^+$ .

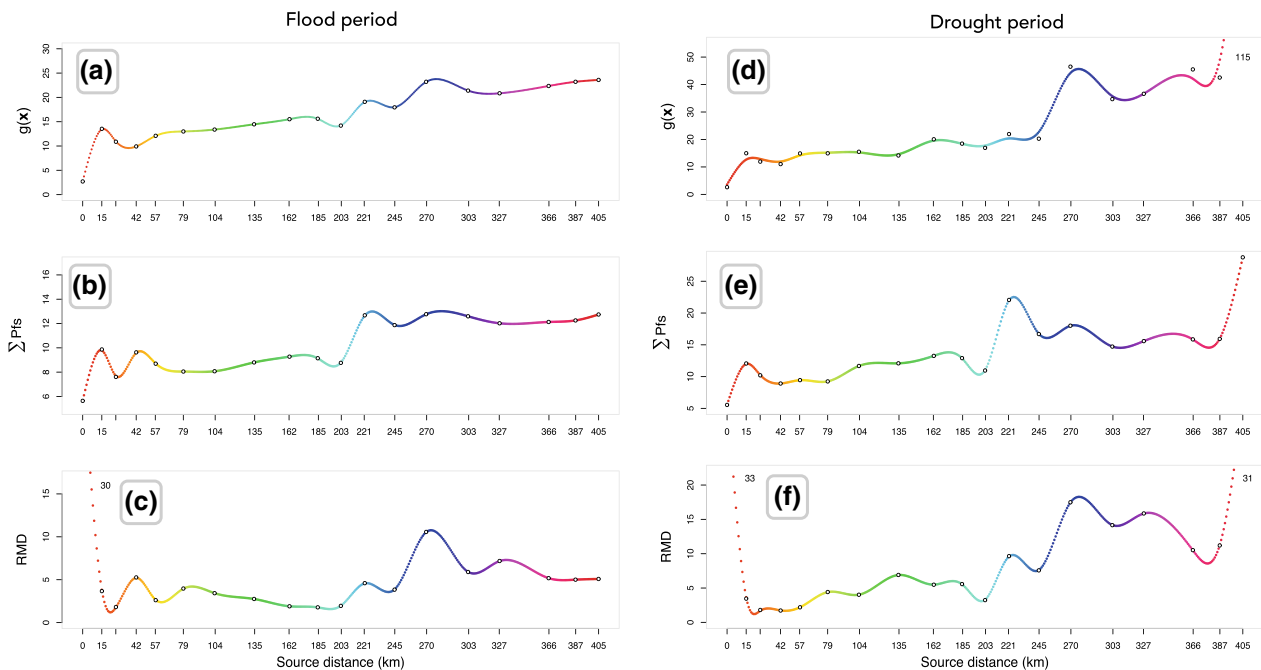
As a result, these compounds appear less resilient as confirmed by the persistence of their perturbations downstream. Possibly, this is due to their conservative nature in the aqueous phase which prevents precipitation from the water column or participation in biochemical reactions and to the lack of strong dilution processes.

The downstream propagation of the water geochemical footprints is sensitive, at least for conservative species, to the effects of upstream signals, thus increasing the magnitude of overall perturbations in the lower river course (Fig. 5b, e). This condition is not exhibited in stream sediments for which upstream perturbations seem to have scarce effects downstream with no monitored source-to-sink trend. This might indicate that the sediment routing system is disrupted or intermittent, thus affecting the transfer of sediments particles and

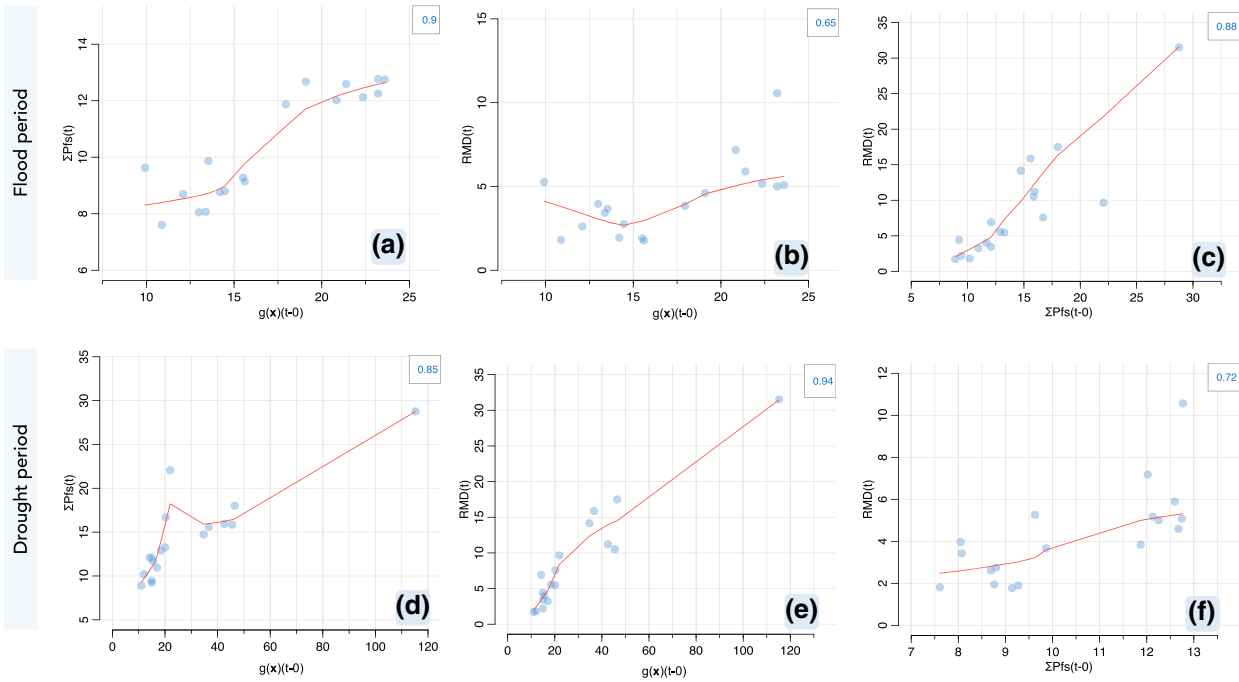
dissolved solids to the sea (Brandt, 2000; Armitage et al., 2013; Allen, 2017). A low propagation of the geochemical footprints from source regions could imply a higher stream sediment vulnerability to geochemical threats at local scale due to the lack or short-range compositional homogenization. Dams most likely play a significant role in the observed behavior by trapping sediments and reducing the natural blending of stream sediments from lithologically different watersheds during their seaward movement. This could explain why some of the major compositional variations are measured just downstream from dams.

The most relevant perturbation in the stream sediment composition involves Na, which perfectly monitors the increments of catchment area drained by flysch as a consequence of silicate weathering processes. Cr shows a high perturbation due to the denudation of the occasional ophiolitic formations outcropping in the high Tiberine valley (Bortolotti, 1961, 1962), while K, Rb and Fe are perturbed in correspondence to the presence of volcanic rocks. The high values of pCa, pSr and pMg detected between 15 and 26 km are likely related to the presence of calcareous sandstones and marls in the

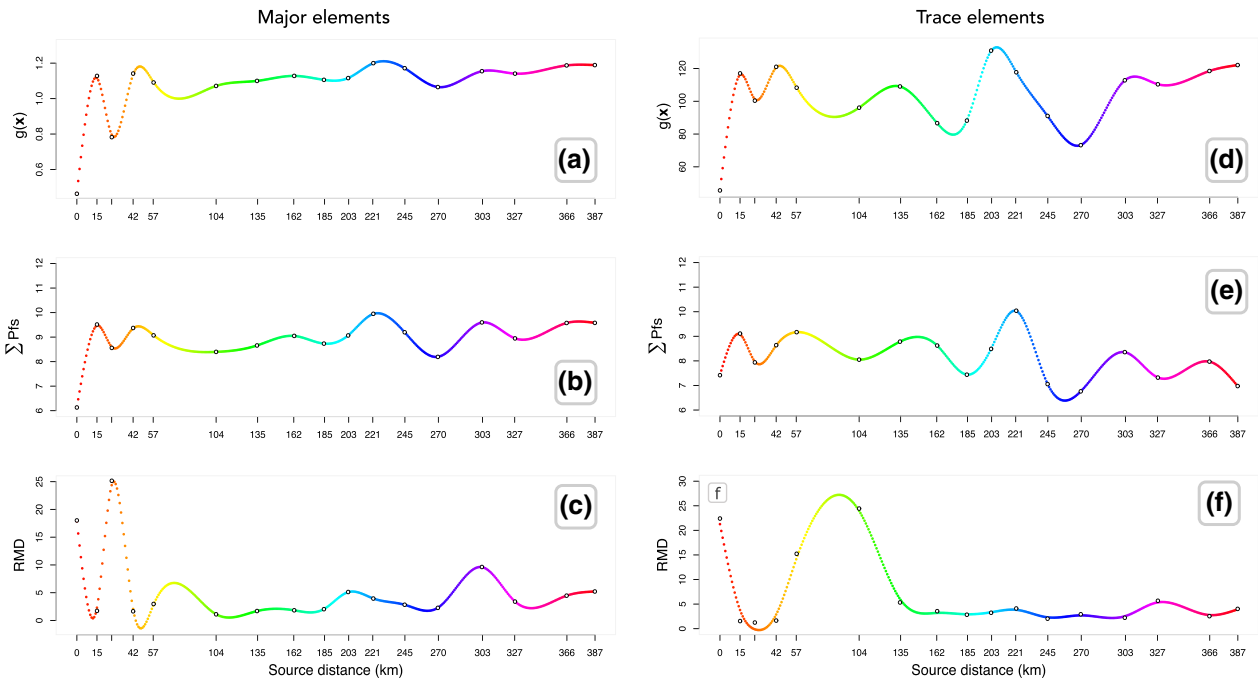
upstream area (Conti et al., 2016). Unexpectedly, no compositional changes in terms of these elements are observed after Apennine limestones were drained. This might be due to the lack of sediment transfer to the Tiber from the Nera whose course is interrupted and impacted by several weirs and dams constructed for hydropower purpose (Panichi et al., 2005; Iadanza & Napolitano, 2006). The high variability of pPb, characterized by 4 great spikes, may be predominantly due to anthropogenic lead contamination since no significant natural mineralizations are known in the areas concerned. Possible sources are represented by industrial emissions, paint production, coal combustion, fertilizers and pesticides (e.g., Komárek et al., 2008; Bur et al., 2009, and references therein). The perturbation at 366 km is most likely relative to Rome's urban and industrial development; nevertheless, volcanic rocks with elevated Pb-background might provide additional Pb inputs to the lower course (Reimann et al., 2012). Delile et al. (2017), by studying the sedimentary profile in the Ostia harbor, show that lead water pipes were the primary source of lead pollution in the city's runoff, reflecting the main stages of ancient Rome's urbanization.



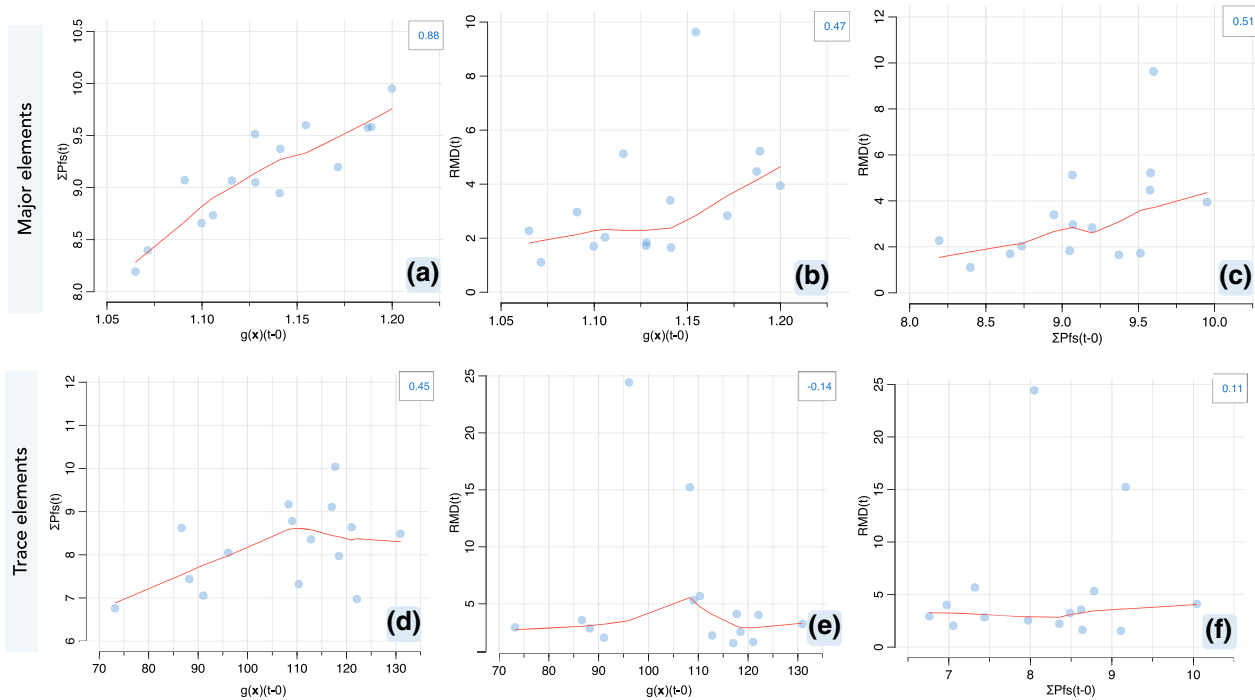
**Figure 5.** Comparison between the geometric mean of the composition ( $g(x)$ ), the sum of unclosed perturbation factors ( $\sum Pfs$ ) and the robust Mahalanobis distance (RMD) of the river water composition in flood (a–c) and drought periods (d–f). The profiles are fitted using the smoothing cubic spline function, and predicted values are depicted with rainbow colors for an easier comparison among the same sections of the curves. The black dots represent the actual values.



**Figure 6.** H-scatterplot at lag 0 relating each other the indices ( $g(x)$ , RMD and  $\sum$  Pfs) of the water composition in flood (a–c) and drought (d–f) periods. The cross-correlation value is reported in the top-right legend, and the red line represents the lowest fit of each scatterplot.



**Figure 7.** Comparison between the geometric mean of the composition ( $g(x)$ ), the sum of unclosed perturbation factors ( $\sum$  Pfs) and the robust Mahalanobis distance (RMD) of the stream sediment composition for the major (a–c) and trace elements (d–f). The profiles are fitted using the smoothing cubic spline function, and predicted values are depicted with rainbow colors for an easier comparison among the sections of the curves. The black dots represent the actual values.



**Figure 8.** H-scatterplot at lag 0 relating each other the different indices ( $g(x)$ , RMD and  $\sum$  Pfs) of the stream sediment composition for major (a–c) and trace elements (d–f). The cross-correlation value is reported in the top-right legend and the red line represents the lowest fit of each scatterplot.

### Potential of Monitoring Indices

The good consistency between the  $\sum$  Pfs with  $g(x)$  and RMD indicates that all indices operate well in tracking compositional changes. This appears generally true if the variables having a high variability are removed from the analysis and treated separately, as in the example proposed for river waters (Fig. 2a, c). Otherwise, the risk is that their high variability will mask other variations occurring on a smaller scale, though they are equally important (Gozzi et al., 2019). This drawback primarily affects the calculus of RMD which accounts for the variance–covariance matrix of the data. This fact is supported by the example of stream sediments (Fig. 7) where all variables were considered, including those having high variance. In this case, the RMD differs considerably from the other two indices and more extensively for the trace element composition. The reason may be found in the high variability of Na, Mg, for major elements, and Pb and Cr, for traces, distorting the variance–covariance structure of the data and consequently the distance measure. This seems to occur in spite of the minimization of the influence of outlying observa-

tions due to the application of a robust method. A higher number of cases compared to the number of variables would increase the stability of the covariance matrix, ensuring a greater consistency of RMD with the other indices. Therefore, in the presence of variables having a mixed natural/anthropogenic origins, not treatable separately, and a low number of cases, the  $\sum$  Pfs and  $g(x)$  could be a more reliable alternative to monitor the overall geochemical fluctuations of a composition. The few discrepancies between these two methods, mainly consisting in differences of peak magnitudes, likely depend on: (1) the choice of the reference composition from which unclosed perturbations factors are calculated; and (2) the major influence that low values have on the geometric mean, by reducing it. The high cross-correlation values of about 0.9 among the two indices, for the major element composition of both media, suggest that the reference composition was correctly selected. Diversely, the presence of low values could have played a role in determining the poorer agreement of the two approaches for trace element composition.

Therefore, the geometric mean could represent a simple and effective monitoring index to quickly



assess comprehensive compositional changes in major element composition where low values are rare. From the obtained results,  $g(x)$  reflects the  $\sum$ Pfs of the single variables calculated by working in the *simplex*. This means that it is able to account for the perturbations experienced by the composition, representing a sort of multivariate barycenter. Its power, with respect to other more robust estimates such as the median, is that it keeps track of the interactions among the constituents and these last play a pivotal role in the understanding of the system resilience. A first explorative framework for a given geochemical system can be thereby realized. The source-to-sink pattern of  $g(x)$  (e.g., being it stable or cumulative) enables a deeper understanding of the inner workings of the river system and how the geochemical footprints are transmitted or buffered along topographic gradients. This awareness might have multiple implications for environmental monitoring, enabling a quick assessment of where or when a given medium is perturbed as a result of natural and/or anthropogenic drivers.

## CONCLUSIONS

The application of the perturbation operator by staying in the *simplex* has led to capture the relevant source-to-mouth compositional changes in the surficial waters and stream sediments of the Tiber river. They mainly originate from water–rock interaction and denudation processes within the corresponding sub-catchments and from pollution sources linked to anthropic activities. The results revealed that the geochemical footprints of stream sediments, unlike those of waters, are poorly transmitted downstream. This suggests a low resilience of this medium to geochemical threats at a local scale due to the disruption or intermittency of the sediment routing system, also enhanced by the anthropic activities (i.e., dams and weirs).

The comparison of the sum of unclosed perturbation factors with the geometric mean of the compositions reveals both indexes providing consistent results, especially for major elements, where low values are rarer. In this case, the geometric mean reflects, with a good approximation, the sum of the unclosed perturbation factors of each composition, thus constituting a simple and effective monitoring index of the complex relationships among the involved parts. Conversely, the robust Mahalanobis distance occasionally diverges from the

other two measures due to the instability of the covariance matrix; therefore, its application is only recommended for larger datasets.

In conclusion, with the awareness of the importance of the correct choice of the reference composition for the calculus of perturbations and the influence of low values on the geometric mean, the use of these methods for environmental monitoring and assessment appears promising. They offer new insights into the inner workings of the river system, its short or long compositional memory, its connectivity and thus its resilience to environmental pressures at catchment scale. Particularly, the use of the geometric mean of a composition as a quick monitoring index of comprehensive compositional variations could find broad applications in environmental studies to uncover time–space changes.

## ACKNOWLEDGMENTS

The International Association for Mathematical Geosciences and Elsevier are thanked for supporting this research with the Computers & Geosciences Research Scholarship 2020 (Grant N. CG-2020-8). The University of Florence is acknowledged for the financial assistance through University funds (A.B). Prof. Gerd Rantitsch, Prof. Orlando Vaselli and Dr. Barbara Nisi are thanked for their scientific and technical aid.

## AUTHOR CONTRIBUTIONS

Caterina Gozzi and Antonella Buccianti were involved in the conceptualization; Caterina Gozzi and Antonella Buccianti contributed to the methodology; Caterina Gozzi was involved in the formal analysis and investigation; Caterina Gozzi contributed to the writing—original draft preparation; Antonella Buccianti was involved in writing—review and editing; Antonella Buccianti acquired the funding; Antonella Buccianti was involved in the supervision. All authors read and approved the final manuscript.

## FUNDING

This research is supported by Computers & Geosciences Research Scholarship 2020 (Grant N.

CG-2020-8) co-sponsored by the International Association for Mathematical Geosciences and Elsevier.

## DECLARATIONS

**CONFLICT OF INTEREST** The authors have no conflicts of interest to declare that are relevant to the content of this article.

## OPEN ACCESS

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## REFERENCES

- Aitchison, J. (1982). The statistical analysis of compositional data (with discussion). *Journal of the Royal Statistical Society Series B*, 44(2), 139–177.
- Aitchison, J. (1986). *The statistical analysis of compositional data*. Monographs on statistics and applied probability. Chapman & Hall Ltd.
- Aitchison, J., & Ng, K. (2005). The role of perturbation in compositional data analysis. *Statistical Modelling*, 5, 173–185.
- Allen, P. (2017). *Sediment routing systems: The fate of sediment from source to sink* (p. 442 pp). Cambridge University Press.
- Armitage, J. J., Dunkley Jones, T., Duller, R. A., Whittaker, A. C., & Allen, P. A. (2013). Temporal buffering of climate-driven sediment flux cycles by transient catchment response. *Earth and Planetary Science Letters*, 369–370, 200–210.
- Bagnini, S., Bruschi, S., Castellano, F., Colatosti, G., Gatta, L., Malvati, P., Moretti, D., Ruisi, M., Terranova, I., Traversa, P., Vitale, V., Villani Conti, C., & Verga, V. (2005). *Tevere Pilot River Basin Article 5 Report: Pursuant to the water framework directive* (p. 175pp). Gangemi Editore spa.
- Berner, E. K., & Berner, R. A. (1996). *Global environment: Water, air and geochemical cycles* (p. 376pp). Prentice Hall.
- Bethke, C. (2008). *Geochemical and biogeochemical reaction modeling* (II, p. 543pp). Cambridge University Press.
- Beutel, M. W. (2006). Inhibition of ammonia release from anoxic profundal sediments in lakes using hypolimnetic oxygenation. *Ecological Engineering*, 28(3), 271–279.
- Billheimer, D., Guttorp, P., & Fagan, W. (2001). Statistical interpretation of species composition. *Journal of the American Statistical Association*, 96(456), 1205–1214.
- Boni, C., Bono, P., & Capelli, G. (1986). Hydrogeological scheme of central Italy. *Memories of the Italian Geological Society*, 35, 991–1012. [in Italian].
- Bortolotti, V. (1961). On the ophiolite-alberese relations between Pieve S. Stefano and Borgo Sansepolcro (Arezzo). *Bollettino della Società Geologica Italiana*, 80(3), 257–306 (in Italian).
- Bortolotti, V. (1962). On the ophiolite series disposition in the Mt. Rognosi (Arezzo). *Bollettino della Società Geologica Italiana*, 81(3), 313–322 (in Italian).
- Brandt, S. (2000). Classification of geomorphological effects downstream of dams. *CATENA*, 40(4), 375–401.
- Buccianti, A., & Magli, R. (2011). Metric concepts and implications in describing compositional changes for world river's water chemistry. *Computers and Geosciences*, 37(5), 670–676.
- Bur, T., Probst, J., N'guessan, M., & Probst, A. (2009). Distribution and origin of lead in stream sediments from small agricultural catchments draining Miocene molassic deposits (SW France). *Applied Geochemistry*, 24(7), 1324–1338.
- Chayes, F. (1960). On correlation between variables of constant sum. *Journal of Geophysical Research*, 65(12), 4185–4193.
- Conti, S., Fioroni, C., Fontana, D., & Grillenzoni, C. (2016). Depositional history of the Epiligurian wedge-top basin in the Val Marecchia area (northern Apennines, Italy): A revision of the Burdigalian-Tortonian succession. *Italian Journal of Geosciences*, 135, 1–37.
- Copernicus Land Monitoring Service (2019). Corine land cover. <https://land.copernicus.eu/pan-european/corine-land-cover/clc-2018>.
- Delile, H., Keenan-Jones, D., Blichert-Toft, J., Goiran, J.-P., Arnaud-Godet, F., & Albarède, F. (2017). Rome's urban history inferred from Pb-contaminated waters trapped in its ancient harbor basins. *Proceedings of the National Academy of Sciences*, 114(38), 10059–10064.
- Egozcue, J. J., & Pawlowsky-Glahn, V. (2011). *Basic concepts and procedures* (chap. 2, pp. 12–28). John Wiley & Sons, Ltd.
- Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., & Barceló-Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, 35(3), 270–300.
- Filzmoser, P., & Hron, K. (2008). Outlier detection for compositional data using robust methods. *Mathematical Geosciences*, 40(3), 233–248.
- Filzmoser, P., Hron, K., & Reimann, C. (2012). Interpretation of multivariate outliers for compositional data. *Computer and Geoscience*, 39, 77–85.
- Filzmoser, P., Hron, K., & Templ, M. (2018). *Applied compositional data analysis: With worked examples in R* (p. 280pp). Springer.
- Fron dini, F., Cardellini, C., Caliro, S., Chiodini, G. G., & Morgantini, N. (2012). Regional groundwater flow and interactions with deep fluids in western Apennine: the case of Narni-Amelia chain (Central Italy). *Geofluids*, 12, 182–196.
- Gozzi, C. (2020). Weathering and transport processes investigated through the statistical properties of the geochemical landscapes: The case study of the Tiber river basin (Central Italy). *PLINIUS*, 46, 48–55.
- Gozzi, C., Dakos, V., Buccianti, A., & Vaselli, O. (2021). Are geochemical regime shifts identifiable in river waters? Exploring the compositional dynamics of the Tiber River (Italy). *Science of The Total Environment*, 785, 147268.
- Gozzi, C., Filzmoser, P., Buccianti, A., Vaselli, O., & Nisi, B. (2019). Statistical methods for the geochemical characterisa-

- tion of surface waters: The case study of the Tiber River basin (Central Italy). *Computers and Geosciences*, 131, 80–88.
- Gozzi, C., Sauro Graziano, R., & Buccianti, A. (2020). Part-whole relations: New insights about the dynamics of complex geochemical riverine systems. *Minerals*, 10(6), 501.
- Heinze, C., Blenckner, T., Martins, H., Rusiecka, D., Doscher, R., Gehlen, M., Gruber, N., Holland, E., Hov, O., Joos, F., Matthews, J., Rodven, R., & Wilson, S. (2021). The quiet crossing of ocean tipping points. *Proceedings of the National Academy of Science of the United States of America*, 118(9).
- Iadanza, C., & Napolitano, F. (2006). Sediment transport time series in the Tiber River. *Physics and Chemistry of the Earth*, 31, 1212–1227.
- ISPRA Ambiente (2017). Geoportale Ispra Ambiente. <http://geoportale.isprambiente.it/sfoglia-il-catalogo/?lang=en>.
- Kleidon, A. (2010). Life, hierarchy, and the thermodynamic machinery of planet Earth. *Physics of Life Reviews*, 7, 424–460.
- Komárek, M., Ettler, V., Chrástný, V., & Mihaljevič, M. (2008). Lead isotopes in environmental sciences: A review. *Environment International*, 34(4), 562–577.
- Kondepudi, D., & Prigogine, I. (1998). *Modern thermodynamics: From heat engines to dissipative structures* (II, p. 552pp). John Wiley & Sons Ltd.
- Li, L., Kang, X., Biederman, J., Hao, Y., & Wang, Y. (2021). Nonlinear carbon cycling responses to precipitation variability in a semiarid grassland. *Science of the Total Environment*, 781, 147062.
- Maavara, T., Chen, Q., Van Meter, K., Brown, L. E., Zhang, J., Ni, J., & Zarfl, C. (2020). River dam impacts on biogeochemical cycling. *Nature Reviews Earth & Environment*, 1(2), 103–116.
- Maechler, M., Rousseeuw, P., Croux, C., Todorov, V., Ruckstuhl, A., Salibián-Barrera, M., Verbeke, T., Koller, M., Conceicao, E. L. T., & Anna di Palma, M. (2018). *robustbase: Basic Robust Statistics*. R package version 0.93-1.1. <http://robustbase.r-forge.r-project.org/>.
- Mahalanobis, P. C. (1936). On the generalized distance in statistics. *Proceedings of the National Institute of Sciences (Calcutta)*, 2, 49–55.
- Martín-Fernández, J. A., Daunis-I-Estadella, J., & Mateu-Figueras, G. (2015). On the interpretation of differences between groups for compositional data. *SORT*, 39(2), 231–252.
- Minissale, A. (2004). Origin, transport and discharge of CO<sub>2</sub> in Central Italy. *Earth-Science Reviews*, 66(1), 89–141.
- Panichi, C., Giuliano, G., Preziosi, E., Gherardi, F., & Droghieri, E. (2005). *Hydrochemical and isotopic characterisation of the base flow in the Tiber basin. Relations between surface waters and groundwaters* (Vol. 124, p. 113pp). Romana Editrice.
- Pawlowsky-Glahn, V., & Egozcue, J. (2001). Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment*, 15(5), 384–398.
- Pawlowsky-Glahn, V., Egozcue, J., & Tolosana-Delgado, R. (2015). *Modeling and analysis of compositional data. Statistics in practice* (p. 272 pp). John Wiley & Sons Ltd.
- Reimann, C., & Filzmoser, P. (2000). Normal and lognormal data distribution in geochemistry: death of a myth. Consequences for the statistical treatment of geochemical and environmental data. *Environmental Geology*, 39(9), 1001–1014.
- Reimann, C., Flem, B., Fabian, K., Birke, M., Ladenberger, A., Négrel, P., et al. (2012). Lead and lead isotopes in agricultural soils of Europe—The continental perspective. *Applied Geochemistry*, 27(3), 532–542.
- Rock, N. M. S. (1988). Summary statistics in geochemistry: A study of the performance of robust estimates. *Mathematical Geology*, 20(3), 243–275.
- Roduner, E., & Radhakrishnan, S. (2016). In command of non-equilibrium. *Chemical Society Reviews*, 45, 2768–2784.
- Rousseeuw, P., & Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. *Technometrics*, 41(3), 212–223.
- Ruggiero, D. (2020). “From river to sea” - Thinking over the Roman coast according to nature. Chap. The “Lungomuro” of Ostia among erosion, mismanagement and coastal emergency ((Vol. 18, pp. 13–23). Culture Territori Linguaggi: University of Perugia.
- Sauro Graziano, R., Gozzi, C., & Buccianti, A. (2020). Is compositional data analysis (CoDA) a theory able to discover complex dynamics in aqueous geochemical systems? *Journal of Geochemical Exploration*, 211, 106465.
- Scheffer, M., Carpenter, S., Lenton, T., Bascompte, J., Brock, W., Dakos, V., van de Koppel, J., van de Leemput, I., Levin, S., van Nes, E., Pascual, M., & Vandermeer, J. (2012). Anticipating critical transitions. *Science*, 338, 344–348.
- Seely, A., & Macklem, P. (2012). Fractal variability: An emergent property of complex dissipative systems. *Chaos*, 22, 13108-1-013108-7.
- Shvartsev, S. (2009). Self-organizing abiogenic dissipative structures in the geologic history of the earth. *Earth Science Frontiers*, 16(6), 257–275.
- Stoffer, D. (2021). *astsa: Applied statistical time series analysis*. R package version 1.13. <https://CRAN.R-project.org/package=astsa>.
- Symons, J., Weibel, S., & Robeck, G. (1965). Impoundment influences on water quality. *Journal (American Water Works Association)*, 57(1), 51–75.
- Tacconi, P., De Rosa, P., Fredduzzi, A., & Cencetti, C. (2020). “From river to sea”—Thinking over the Roman coast according to nature. Chap. The River Tiber: sediment supply, solid transport, hydroelectric power stations and sedimentary deficit at mouth. The “node” of Corbara-Alviano (Vol. 18). Culture Territori Linguaggi, University of Perugia.
- Templ, M., Hron, K., & Filzmoser, P. (2011). *robCompositions: An R-package for robust statistical analysis of compositional data* (p. 378 pp). John Wiley & Sons Ltd.
- Templ, M., Hron, K., Filzmoser, P., & Gardlo, A. (2016). Imputation of rounded zeros for high-dimensional compositional data. *Chemometrics and Intelligent Laboratory Systems*, 155, 183–190.
- van Rooij, M., Nash, B., Rajaraman, S., & Holden, J. (2013). A fractal approach to dynamic inference and distribution analysis. *Frontiers in Physiology*, 4(1), 1–11.
- Weinan, E., Quax, R., vanNes, E., & Leemput, I. (2021). Evaluating the performance of multivariate indicators of resilience loss. *Scientific Reports*, 11(1), 9148.
- Xia, X., Zhang, S., Li, S., Zhang, L., Wang, G., Zhang, L., et al. (2018). The cycle of nitrogen in river systems: Sources, transformation, and flux. *Environmental science: Processes & impacts*, 20(6), 863–891.
- Zuo, R., & Wang, J. (2016). Fractal/multifractal modeling of geochemical data: A review. *Journal of Geochemical Exploration*, 164, 33–41.