Water Resources Management 34: 3879–3894 (2020)

FingerPro: An R package for tracking the provenance of sediment

Ivan Lizaga [a]*, Borja Latorre [a], Leticia Gaspar [a], Ana Navas [a]

a Experimental station of Aula Dei (EEAD-CSIC), Avenida Montañana, 1005, 50059 Zaragoza (Spain).

* ilizaga@eead.csic.es/lizaga.ivan10@gmail.com; +34 670306713/+34 976716143

Author's ORCID: 0000-0003-4372-5901; 0000-0002-6720-3326; 0000-0002-3473-7110; 0000-0002-4724-7532

ABSTRACT

Soil loss by erosion processes is one of the largest challenges for food production and reservoir siltation around the world. Information on sediment, nutrients and pollutants is required for designing effective control strategies. The estimation of sediment sources is difficult to get using conventional techniques, but sediment fingerprinting is a potentially valuable tool. This procedure intends to develop methods that enable to identify the apportionment of sediment sources from sediment mixtures.

We developed a new tool to quantify the provenance of sediments in an agroforest catchment. For the first time, the procedure for the selection of the best combination of tracers was included in the tool package. An unmixing model algorithm is applied to the sediment samples to estimate the contribution of each possible source. The operations are compiled in an R package named **FingerPro**, which unmixes sediment samples after selecting the optimum set of tracers. An example from a well-studied Mediterranean catchment is included in the package to test the model. The sediment source apportionments are compared with previous results of soil redistributions where ^{137}Cs derived rates validate the unmixing results, highlighting the potential of sediment fingerprinting for quantifying the main sediment provenance. Fingerprinting techniques

will allow us to better comprehend sediment transport to water ecosystems and reservoirs and its detrimental effect on the quality of the water and aquatic habitats.

The **FingerPro** package provides further understanding of the unmixing procedure through the use of graphical and statistical tools, offering a broader and easier application of the technique.

KEYWORDS: FingerPro, unmixing model, sediment source fingerprinting, source variability, R package

1. INTRODUCTION

Reliable information on sediment loads transported by a river or stream is crucial to evaluate the severity of reservoir siltation and river pollution. However, determining sediment provenance or sediment budgets in catchments using conventional monitoring techniques is often challenging. However, in most situations, it can be provided by applying tracing techniques. Fingerprinting techniques can be used to recognise sediment sources and to determine their relative contribution, thereby allowing the identification of areas or land uses prone to erosion processes (Schuller et al. 2013). Soil erosion and subsequent sediment transport are related to the loss of nutrients and their distribution in the catchment (Lizaga et al. 2019). To assess this issue, several software and indices have been developed to quantify the effects of different erosion mechanisms, such as connectivity (Shore et al. 2013), runoff (dos Santos et al. 2017), sheet and rill (Molnár and Julien, 1998), wind erosion (Schmidt et al.. 2017) and the subsequent effects on water quality (Panagopoulos et al. 2011; Foucher et al. 2020). However, sediment source fingerprinting has been developed in recent decades for catchment sediment and pollutant investigation as the most powerful tool to assess this problem. The procedure identifies sediment provenance and estimates the relative contribution of each potential source, using the selected tracer properties.

The first fingerprinting approach dates back to the seventies, based on mineralogical and grain size characterisation (Klages and Hsieh, 1975). The earliest fingerprinting studies were fundamentally qualitative in their result, but the introduction of quantitative mixing models was a methodological advance that enabled researchers to obtain quantitative results of the relative contribution from different sediment sources (Collins et al. 1997). Since these early works, sediment source fingerprinting applications have been greatly expanding with the development of new techniques (Owens et al. 2016).

The traditional approach for applying source-tracing methods is to define the relevant tracer properties that provide a particular signature between all source samples and unequivocally discriminate the different sources (Collins and Walling, 2002). Due to the inherent complexity of catchment characteristics, with large variations in climate, geology, land use, vegetation, soil, and management practices, commonly, no unique tracer can discriminate between multiple sediment sources. Consequently, different tracer properties need to be analysed, such as radionuclides (Wallbrink et al. 1998; Evrard et al., 2020; Navas et al., 2020), geochemistry (Martínez-Carreras et al. 2010; Smith and Blake, 2014; Gaspar et al., 2019a), ultraviolet-visible spectra derived tracers (Ramon et al., 2020), CSSI (Reiffarth et al. 2019) and eDNA (Evrard et al. 2019).

The fundamental theory that supports this technique is that the tracer properties of the sediment mixtures are directly comparable to the sediment of the sources. A common procedure, the so-called “range test”, checks if sediment tracers are conservative excluding the tracers of the mixture/s outside the minimum and maximum values in the potential sediment sources. This procedure prevents the inclusion in the optimum tracers of the fingerprint properties exhibiting non-conservative behaviour. However, the exclusion of a great number of fingerprint properties likely suggests that not all sources have been correctly identified or characterised. Thus, the methodology for tracer selection is an open question that is being discussed at present by several authors since different tracer selection methods could lead models to different results (Pulley et al. 2015; Lizaga et al. 2020). Following this assumption, the two-stage statistical procedure previously proposed by Collins and Walling, (2002), is commonly used to assess this conservativeness. Thus, the Kruskal Wallis H test (KW) and discriminant function analysis (DFA) test the ability of individual tracers to discern between sediment sources and select the best combination of tracers. This procedure was used to select the smallest combination of tracers that provided the maximum discrimination of the identified source categories and it is implemented by several authors as a common procedure when using frequentist (Palazón et al. 2015; Lin et al. 2015; Gholami et al., 2020) and Bayesian (Koiter et al. 2013; Barthod et al. 2015) unmixing models. Subsequently, the relative contribution of each identified source is estimated using a linear multivariate unmixing model.

Due to the growing use of fingerprinting methods, other unmixing models, such as SIFT (Pulley and Collins, 2018), MixSIR (Moore and Semmens, 2008) and IsoSource (Phillips and Gregg, 2003), appeared in the last years for pollution and ecological purposes. However, due to operational complexity and the need to use different statistical software not included in the packages, the use of unmixing models is generally restricted to academics with a good knowledge of the procedure. Refinement of the sediment source fingerprinting techniques requires open-source models that help the user in tracer selection decisions and optimise this time-consuming process for non-expert and academics with low programming and statistical skills by including the essential statistical functions and plots.

To fill this gap, our objective was to develop for the first time an R package comprising an unmixing model with the additional tools needed to comprehend the effect of the selected properties on the model outcome. To this purpose, we create a new R package that combines in a novel assemblage the tools needed to unmix sediment samples and the previous statistical tests to select optimum tracers. Thus, we aim to provide an easy and straightforward way to apply the sediment fingerprinting technique aimed at beginners or non R users.

This paper presents the **FingerPro** package, a user-friendly application and freely available software for users with limited or nor expertise in statistics. Thus, any user could implement the fingerprinting procedure with limited previous experience in the technique and with no need for additional software for statistical analyses. Furthermore, unlike previous models, this new tool to identify sediment provenance has been successfully tested with artificial samples (Gaspar et al. 2019b).

The analyses explained in this research are based on 1) a reproducible data set example of small catchment included in the package and 2) another example of ongoing research in a medium-size catchment to further describe the capability of the package. The medium size catchment is selected as representative of mountain headwaters (South Pyrenean region) that supply water to reservoirs as siltation and pollution is one of the main environmental issues worldwide (Valero-Garcés et al. 1999). Through these two examples, the utility of the **FingerPro** package for applying tools aimed at pre-processing input data or combining sources without significant differences before or after running the unmixing model is shown. Through the provided examples, it is evident that fingerprinting methods are necessary to identify sediment sources to establish management strategies for ensuring water supply to the lowlands while preserving water quality.

2. METHODS

Sediment fingerprinting requires a preliminary analysis to select a subset of conservative tracers that discriminate the potential sources. Then, the relative contribution of each source is estimated using a linear multivariate unmixing model. This procedure is iterated considering the variability of the sediment sources to obtain the statistical distribution of the source contribution.

2.1 Statistical analysis for the selection of tracer properties

Several statistical tests can be used to confirm source discrimination and select the optimal subset of conservative tracer properties, such as the procedure suggested by Collins and Walling (2002). However, the use of many tests could remove a considerable number of tracers and therefore restrict the discrimination between sediment sources. Consequently, none of the functions included in the **FingerPro** package are mandatory and the tracer exclusions can be based on 'expert judgement' after visualising boxplots and results from the statistical tests included. The tracer selection methods implemented in the package are:

i) Range test: the minimum and maximum values of the tracer properties in the sediment sources are compared to those of the mixtures. The tracers falling out of the range of the selected sources are removed from subsequent analyses. These properties may not be conservative or their exclusion supports the existence of an additional hidden source.

ii) Kruskal-Wallis H test: this is a rank-based nonparametric test used to determine if there are significant differences between the medians of selected groups or sources. This procedure removes tracers that do not show significant differences between at least two of the sediment sources.

iii) Discriminant Function Analysis identifies the optimum set of tracers that maximises the discrimination between the sediment sources whilst minimising the number of tracers. This function executes a stepwise forward variable selection for classification using the Wilk's Lambda criterion. The function selects the tracers based on how much they decrease Wilks' lambda. At each step, the function includes the variable that minimises the overall Wilks' lambda.

2.2 Mixing model

The relative contribution of each potential sediment source is determined using a standard linear multivariate mixing model:

$$\sum_{j=1}^{m} a_{i,j} \cdot \omega_j = b_i$$

which satisfies:

$$\sum_{j=1}^{m} \omega_j = 1$$

$$0 \leq \omega_j \leq 1$$

where b_i is the tracer property *i* (*i* =1 to *n*) of the sediment mixture, $a_{i,j}$ represents the tracer property *i* in the source type *j* (*j* =1 to *m*), ω_j is the unknown relative contribution of the source type *j*, *m* represents the number of potential sediment sources and *n* is the number of tracer properties selected.

This system of equations is mathematically determined if the number of tracers is greater than or equal to the number of potential sources minus one ($n \geq m - 1$). The procedure tries to find the source proportions that conserve the mass balance for all tracers from the complete exploration of the parameter space (Palazón et al. 2015). All possible combinations of each source contribution (0-100%) are examined in small increments, using Latin hypercube sampling (LHS) (McKay et al. 1979). The quality of each candidate is measured using the following function or goodness of fit (GOF), based on the sum of squares of the relative error:

$$GOF = 1 - \frac{1}{n} \times \left(\sum_{i=1}^{n} \frac{\left| b_i - \sum_{j=1}^{m} \omega_j a_{i,j} \right|}{\Delta_i} \right)$$

where Δ_i is the range of the tracer property i, used as a normalisation factor. The combinations that reproduce the observed sediment mixture with the maximum GOF are selected as the solution.

2.3 Variability analysis of the sources

In small to large size catchments, the heterogeneity of sediment tracers, defined by different land uses, geomorphic processes, soil types or human activity, is always present. For this reason, fingerprinting studies should correctly characterise source variability by collecting several samples from each source. Thus, evaluation of the variability in tracer data used to characterise sediment sources is important to correctly interpret the source apportionment results.

Variability analysis is assessed following classical frequentist inference utilising a Monte-Carlo method (Helton, 1994). A succession of deterministic calculations is executed, each with different input values sampled from their respective distributions, to obtain probability distributions of the targeted outcomes.

The heterogeneity of each source is considered as a t-distribution for each property. The fingerprinting analysis of each sediment mixture is repeated by randomly sampling the source probability distributions. For the first iteration, the central value of the source distributions is used as a reference result. The corresponding output values are gathered to infer the probability distribution of the potential source contributions. Several samples must be collected for characterising each source to compute the mean and the SD of the analysed tracer properties.

3. THE FINGERPRO PACKAGE

Application of the functions in the package allows the user to i) characterise the different tracer properties and select the relevant variables; ii) unmix the sediment samples and quantify the different source apportionment; iii) assess the effect of the source variability; and iv) visualise and export the results. Thus, **FingerPro** package proposes a step by step procedure divided into three main sections to help users in their decisions.

3.1 The example dataset

The package includes a soil dataset from a small Mediterranean catchment (4 km^2) that contains high-quality radionuclide and geochemistry data to test the operation of the functions and help the user to understand the model (Fig.1). This study area was selected due to its heterogeneous land uses/land covers which are likely to exhibit large differences in sediment tracer contents. Furthermore, the study area is located in a well-studied catchment where several studies of soil redistribution ^{137}Cs derived rates were pursued (Quijano et al. 2016; Lizaga et al. 2018). Thus, soil redistribution rates were used to evaluate **FingerPro** model as a suitable tool in the northern-central part of the Ebro basin. The results obtained by Lizaga et al. (2018) found that net soil loss values were 4 times higher in agricultural lands than in pine forest highlighting the importance of the vegetation cover and land management to prevent erosion processes and subsequent land degradation.

The study area dataset is composed of 21 source sediment samples from 4 different sources and 2 mixture samples. The sources are divided into agricultural (AG), old pine forest (PI), recent pine forest (PI1) and degraded soil named subsoil (SS) which occupy 9%, 32%, 58% and 1% of the catchment area, respectively. The agricultural land use is mainly composed of winter cereal crops and the pine afforestation forest is predominantly

Pinus halepensis Mill. The average temperature ranges from 5 °C to 18 °C, and the mean annual rainfall is about 520 mm (AEMET).

3.2 Input data

The input variables need to be stored as an R table object. The dataset must satisfy the following requirements: i) the first column represents the sample *id*; ii) the second column is the source classification, containing target samples in the last place.

3.3 Characterising the sediment samples

One of the advantages of the **FingerPro** package is that it allows the user to analyse and visually compare different tracer properties, using the state of the art of R packages:

The *boxPlot()* function displays a boxplot of each tracer property to help the user in the decision by visualising the different concentrations of each tracer versus the mixture sample. A parameter (*columns*) with the number of tracer properties in the boxplots is provided. The number of columns (*ncol*) refers to the number of plots per row in the display (Fig.2). The boxPlot() function could be used for tracer selection by helping the user to visualise and select the tracers based on the boxplots and its expert knowledge. Thus, the user visualises in the example dataset that most of the $^{210}Pb_{ex}$ in the mixture sample likely comes from PI and PI1 sources and that ^{40}K is almost out of range (Fig.2). Furthermore, by repeating this function after implementing each test for tracer selection, users can envisage how representative the remaining tracers are.

The *correlationPlot()* function displays a correlation matrix of each tracer, divided by the different sources to help the user by testing the conservatism of tracers by visualising the relationships between the different tracers and sediment mixtures following the methodology proposed by Pulley et al. (2015). A parameter (*columns*) with the number of tracer properties in the correlation matrix is provided, along with the possibility to include the sediment mixture (mixtures = *T*) in the matrix or to exclude it (by default). In addition, in the correlation plot, once the users have selected the optimum set of tracers, it is possible to visualise if the mixture samples fit inside the source distributions. If a mixture sample is outside the sources distribution, then no solution exists or the mixing model assumptions are not met.

The *PCAPlot()* function performs a principal components analysis on the given data matrix and displays a biplot of the results, divided by the different sources, to help the user in the decision. A parameter (components) with the number of principal components to display is included.

The *LDAPlot()* function performs a linear discriminant analysis and visualises the data in the relevant dimensions. A parameter (P3D) allows the user to display a 3D LDA graph (Fig.3). This set of functions allows the user to visualise the principal components plot and the linear discriminant plot after the statistical selection procedure. Thus, the plots help the user to visually identify whether the excluded variables increase or decrease the discrimination capacity between sources. Furthermore, the LDAPlot() function was used in the catchment example to visualise the number of sources that show good discrimination with this set of tracers (Fig.3). The function shows a large overlap between PI and PI1 that would suggest merging both sources. Thus, after grouping PI and PI1 the discriminant plot shows better discrimination between the selected sources (Fig.3).

3.4 Statistical test for selecting the optimal set of tracers

The selection of the optimal tracers is usually based on the two-step procedure proposed by Collins and Walling (2002), which includes some previous statistical procedures such as the “range test”, the KW and the DFA test. Thus, **FingerPro** has included these functions to support user decisions. However, this procedure might remove too many tracers or include some inadequate properties and could, therefore, restrict the discrimination between sediment sources. Hence, the procedure is included as an individual and informative function to only use the steps needed and to prevent a reduction in the source discriminations. For this reason, the tracer selection procedure cannot be only based on statistical tests but also on the expert knowledge of the geomorphological and hydrological processes of the catchment (Blake et al. 2018). Thus, the boxplot chart, LDA plots and correlation plot included in the **FingerPro** package were implemented to help the users in the decision.

The *rangeTest()* function excludes the tracer properties of the mixture/s outside the lowest and highest values in the sediment sources.

The *KWTest()* function excludes tracers from the original dataset which do not show a significant difference between sources. This function performs a Kruskal-Wallis rank-sum test using the kruskal.test() function from the R package **stats**. A parameter to select the p-value (*pvalue*) is provided.

The *DFATest()* function executes a stepwise forward variable selection, using the Wilk's Lambda criterion, which maximises the discrimination between the sources whilst minimising the number of tracers. This function performs a stepwise forward variable selection using the greedy.wilks() function from the R package **klaR**. A parameter to select the niveau (*niveau*) for an approximate F-test decision is provided with a default value of 0.1. This value could be reduced to be more restrictive in the tracer selection procedure. However, by reducing the value below 0.05 the statistical test could remove the majority of the tracers with the subsequent decrease in the discrimination of the different sources.

These three tracer selection methods were applied in the example dataset. In Fig.2 the tracers removed by each method can be seen and, based on the boxplot graph, to decide if it is suitable to use all of them or if the selected tracers represent a good approximation of the dataset. After the implementation of the range test function, we can see in the boxplot graph that effectively Pb, Zn and Cr have been removed. However, there are other tracers such as ^{40}K, Sr, Fe, and V that remain in the dataset though they should not be considered as tracers inside the source range. Furthermore, by using the LDA and PCA plots we can decide if the use of other tracer selection methods decreases the discrimination or if by using them we could remove a tracer with specific information. As shown in Fig.4, by removing ^{226}Ra and Mn from the dataset by using the DFA after the KW test, the LDA plot shows similar results. In addition, the arrows of the removed tracers in the PCA plot were parallel to those that remain in the dataset. Thus, in this example, the plot information suggests that including or removing ^{226}Ra and Mn should not produce important variations in the discrimination of sources or the model results as is evident in Fig.4.

3.5 Sediment unmixing

The *unmix()* function assesses the relative contribution of the selected sediment sources for each mixture in the dataset. A parameter (*samples*) with the number of samples of the LHS is provided. The number of iterations (*iter*) in the source variability analysis is also configurable. However, if the number of iterations is set as 1, results are produced in a single analysis considering the sources mean value.

The *plotResults()* function displays a plot with the density distribution of the model solutions and a table with the mean value and the standard deviation of the model solutions (Fig.4). Besides, users can display the results in violin plots instead of density plots by adding the word *True* to the violin option.

After the tracer selection procedure, **FingerPro** results reveal that 18% of the mean sediment supply comes from agricultural land use and 34% and 47% from bare soil and pine forest, respectively. The small standard deviation of the three sources together with the high GOF value shows a good fit of the model to efficiently discriminate the selected sources (Fig.4). However, users should be cautious about using GOF as an assessment of model reliability. Recent research has shown models with a high GOF can still deliver inaccurate results (Palazón et al. 2015; Gaspar et al. 2019b), but also has shown that all models with low GOF always deliver wrong results.

The results of the example dataset are supported by soil erosion rates estimated with ^{137}Cs by Lizaga et al. (2018) in a Mediterranean catchment comprising the one studied here. Thus, 18% of the sediment contribution is supplied from 9% of the area under agricultural management and 47% of the contribution comes from pine forest that occupies 90% of the study catchment. Relatively speaking, the subsoil was the main source with 34% of the contribution for only 1% of the area taken by the bare soil in the study catchment. Our results highlight the hazards that subsoils have on supplying important amounts of sediments to the water systems.

3.6 Application in a Medium-Size catchment

In this section, as an example, the results of applying the **FingerPro** package in a medium-size catchment (Lizaga et al. 2019) are described. Its larger surface area and a higher number of sources result in a more complex unmixing. For this reason, all the tools added in the **FingerPro** package to help the users and characterise the unmixing dataset are essential to reach robust results. Here, we highlight the most important decisions made during the fingerprinting procedure and how the different tools included in the package help the authors to unmix their data. To avoid repetition in this manuscript, only one mixture sample collected at the outlet of the catchment is used to describe the FingerPro utilities.

Following the application of the range test and Kruskal Wallis test, the final selection was made based on expert judgment using the boxplots and correlation plots to finally identify the conservative tracers. Fig.5 illustrates how some tracers pass the selection tests, such as RT, KW and DFA, but show non-conservative behaviour, i.e. LF, Fe, Ti and Ca. In addition, if we analyse the correlation plot of the tracers that shows non-conservative behaviour, the sample mixture is located almost out of the point cloud. On the other hand, the sample mixture is located inside the point cloud of the conservative tracers. Thus, based on this information it was decided to select the tracers after passing the KW test using expert knowledge, thus obtaining more defined results and higher GOF (Fig.6). Hence, all the tools added in the **FingerPro** package to remove the tracers that violate the principles of conservativeness are needed in fingerprinting studies. This methodology suggests that including tracers with non-conservative or discordant information into fingerprinting models does not add valuable information and could lead the model to unpredictable results as it was found by Lizaga et al. (2020).

4. CONCLUSIONS

The application of mixing models it is necessary to understand source-tracer relationships what is generally performed by applying different software's to select the best combination of sediment tracers. With **FingerPro**, diverse test and mechanisms have been incorporated for tracer selection in a single software. Furthermore, the inclusion of several plot functions such as *boxPlot*, *correlationPlot*, *LDAPlot* and *PCAPlot* allows the user to check if the selected tracers are suitable for the unmixing process. This package for sediment source fingerprinting in hydrological systems offers a wider and easier application in catchments affected by natural and human-induced changes.

Due to the increasing attention in tracing sediment methods and the need to select the best tracer combination, an open-source tool that includes all the steps for sediment unmixing is a key tool for the unmixing process. The example dataset included in **FingerPro** provides evidence of the large sediment supply and severe soil loss caused by land degradation and bare soil. In addition, the agreement between the unmixing results obtained from the example dataset with the ^{137}Cs derived rates supports the capability of the model for the sediment fingerprinting task. These results reflect the high importance of creating a low time-consuming and open-source unmixing model that combines the necessary tools to solve environmental issues such as reservoir siltation or soil loss and trace the sediment provenance.

FingerPro provides the users with tools to i) characterise the different sediment sources, establish correlations between the tracers and assist the selection of the optimal tracers, ii) graph the results using the state of the art of R packages iii) unmix sediment samples to estimate the apportionment of the sediment sources and iv) test the model using data from a Mediterranean study catchment included in the package.

In addition, the example dataset and the explained results of a medium-size catchment introduce the users on to the functioning and potential of the tools included in the **FingerPro** package while also showing the advantages of the fingerprinting technique to improve the understanding of sediment supply processes. Future research will concentrate on keeping the **FingerPro** package updated with new and upcoming fingerprinting techniques, such as the recently published Consensus Method (Lizaga et al., 2020) for improving the selection of tracers, avoiding non-conservative and dissenting tracers.

ACKNOWLEDGMENTS

This work has been supported by project PID2019-104857RB-I00 funded by the Spanish Ministry of Science and Innovation and the aid of a predoctoral contract (BES-2015-071780) of the project CGL2014-52986-R.

DECLARATIONS

Funding: This work has been supported by project CGL2014-52986-R funded by the Spanish Ministry of Science, Innovation and Universities and the aid of a predoctoral contract (BES-2015-071780).

Conflict of Interest: The authors declare that they have no conflict of interest.

Availability of data, material and Code availability:

The R package source files are available via:

GitHub platform: https://github.com/eead-csic-eesa/fingerPro

Name of code: fingerPro

Name of the manual: fingerPro_manual 1.2.pdf

Developer and contact address:

Ivan Lizaga Villuendas (lizaga.ivan10@gmail.com) / Borja Latorre (borja.latorre@csic.es)

Year first available: 2018

Software required: R   Program language: R

Authors' contributions: All authors contributed to the study conception and design. Ivan Lizaga: Conceptualization, Writing - Original Draft, Software, Data Curation, Methodology, Formal analysis, Investigation, Resources; Borja Latorre: Conceptualization, Writing - review & editing, Software, Supervision, Methodology, Formal analysis, Investigation, Resources; Leticia Gaspar: Conceptualization, Validation, Data Curation, Investigation, Resources, Writing - review & editing; Ana Navas: Conceptualization, Methodology, Investigation, Writing - review & editing, Supervision, Project administration, Funding acquisition.

5. REFERENCES

1 Barthod LRM, Liu K, Lobb DA, et al (2015) Selecting Color-based Tracers and Classifying Sediment Sources in the Assessment of Sediment Dynamics Using Sediment Source Fingerprinting. Journal of Environmental Quality 44:1605–1616. https://doi.org/10.2134/jeq2015.01.0043

2 Blake, W. H., Boeckx, P., Stock, B. C., Smith, H. G., Bodé, S., Upadhayay, H. R., Gaspar, L., Goddard, R., Lennard, A. T., Lizaga, I., Lobb, D. A., Owens, P. N., Petticrew, E. L., Kuzyk, Z. Z. A., Gari, b. d., Munishi, L., Mtei, k., Nebiyu, A., Mabit, L., Navas, A., Semmens, B.X. (2018). A deconvolutional Bayesian mixing model approach for river basin sediment source apportionment. Scientific Reports, 8: 13073. https://doi.org/10.1038/s41598-018-30905-9.

3 Collins AL, Walling DE, Leeks GJL (1997) Fingerprinting the Origin of Fluvial Suspended Sediment in Larger River Basins: Combining Assessment of Spatial Provenance and Source Type. Geografiska Annaler: Series A, Physical Geography 79:239–254. https://doi.org/10.1111/j.0435-3676.1997.00020.x

4 Collins AL, Walling DE (2002) Selecting fingerprint properties for discriminating potential suspended sediment sources in river basins. Journal of Hydrology 261:218–244. https://doi.org/10.1016/S0022-1694(02)00011-2

5 dos Santos, J.C.N., de Andrade, E.M., Medeiros, P.H.A., Guerreiro, M.J.S., de Queiroz Palácio, H.A., 2017. Effect of Rainfall Characteristics on Runoff and Water Erosion for Different Land Uses in a Tropical Semiarid Region. Water Resour Manage 31, 173–185. https://doi.org/10.1007/s11269-016-1517-1

6 Evrard, O., Chaboche, P.-A., Ramon, R., Foucher, A., Laceby, J.P., 2020. A global review of sediment source fingerprinting research incorporating fallout radiocesium (137Cs). Geomorphology 107103. https://doi.org/10.1016/j.geomorph.2020.107103

7 Evrard, O., Laceby, J.P., Ficetola, G.F., Gielly, L., Huon, S., Lefèvre, I., Onda, Y., Poulenard, J., 2019. Environmental DNA provides information on sediment sources: A study in catchments affected by Fukushima radioactive fallout. Science of The Total Environment 665, 873–881. https://doi.org/10.1016/j.scitotenv.2019.02.191

8 Foucher, A., Evrard, O., Huon, S., Curie, F., Lefèvre, I., Vaury, V., Cerdan, O., Vandromme, R., Salvador-Blanes, S., 2020. Regional trends in eutrophication across the Loire river basin during the 20th century based on multi-proxy paleolimnological reconstructions. Agriculture, Ecosystems & Environment 301, 107065. https://doi.org/10.1016/j.agee.2020.107065

9 Gaspar, L., Lizaga, I., Blake, W.H., Latorre, B., Quijano, L., Navas, A., 2019a. Fingerprinting changes in source contribution for evaluating soil response during an exceptional rainfall in Spanish pre-pyrenees. Journal of Environmental Management 240, 136–148. https://doi.org/10.1016/j.jenvman.2019.03.109

10 Gaspar, L., Blake, W.H., Smith, H.G., Lizaga, I., Navas, A., 2019b. Testing the sensitivity of a multivariate mixing model using geochemical fingerprints with artificial mixtures. Geoderma 337, 498–510. https://doi.org/10.1016/j.geoderma.2018.10.005

11 Gholami, H., Rahimi, S., Fathabadi, A., Habibi, S., Collins, A.L., 2020. Mapping the spatial sources of atmospheric dust using GLUE and Monte Carlo simulation. Science of The Total Environment 723, 138090. https://doi.org/10.1016/j.scitotenv.2020.138090

12 Helton JC (1994) Treatment of Uncertainty in Performance Assessments for Complex Systems. Risk Analysis 14:483–511. https://doi.org/10.1111/j.1539-6924.1994.tb00266.x

13 Klages MG, Hsieh YP (1975) Suspended Solids Carried by the Gallatin River of Southwestern Montana: II. Using Mineralogy for Inferring Sources. Journal of Environmental Quality 4:68–73. https://doi.org/10.2134/jeq1975.00472425000400010016x

14 Koiter AJ, Lobb DA, Owens PN, et al (2013) Investigating the role of connectivity and scale in assessing the sources of sediment in an agricultural watershed in the Canadian prairies using sediment source fingerprinting. J Soils Sediments 13:1676–1691. https://doi.org/10.1007/s11368-013-0762-7

15 Lin, J., Huang, Y., Wang, M., Jiang, F., Zhang, X., Ge, H., (2015). Assessing the sources of sediment transported in gully systems using a fingerprinting approach: An example from South-east China. CATENA 129, 9–17. https://doi.org/10.1016/j.catena.2015.02.012

16 Lizaga I, Quijano L, Gaspar L, Navas A (2018) Estimating soil redistribution patterns with ^{137}Cs measurements in a Mediterranean mountain catchment affected by land abandonment. Land Degrad Dev 29:105–117. https://doi.org/10.1002/ldr.2843

17 Lizaga, I., Gaspar, L., Blake, W., Latorre, B., Navas, A (2019). Fingerprinting changes of source apportionments from mixed land uses in stream sediments before and after an exceptional rainstorm event. Geomorphology. https://doi.org/10.1016/j.geomorph.2019.05.015

18 Lizaga, I., Latorre, B., Gaspar, L., Navas, A., 2020. Consensus ranking as a method to identify non-conservative and dissenting tracers in fingerprinting studies. Science of The Total Environment 720, 137537. https://doi.org/10.1016/j.scitotenv.2020.137537.

19 Martínez-Carreras, N., Krein, A., Udelhoven, T., Gallart, F., Iffly, J.F., Hoffmann, L., Pfister, L., Walling, D.E., (2010). A rapid spectral-reflectance-based fingerprinting approach for documenting suspended sediment sources during storm runoff events. J Soils Sediments 10, 400–413. https://doi.org/10.1007/s11368-009-0162-1

20 McKay MD, Beckman RJ, Conover WJ (1979) A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. Technometrics 21:239–245. https://doi.org/10.2307/1268522

21 Molnár, D.K., Julien, P.Y., (1998). Estimation of upland erosion using GIS. Computers & Geosciences 24, 183–192. https://doi.org/10.1016/S0098-3004(97)00100-3

22 Moore JW, Semmens BX (2008) Incorporating uncertainty and prior information into stable isotope mixing models. Ecology Letters 11:470–480. https://doi.org/10.1111/j.1461-0248.2008.01163.x

23 Navas, A. Lizaga, I., Gaspar, L., Latorre, B., Dercon, G. (2020). Unveiling the provenance of sediments in the moraine complex of Aldegonda Glacier (Svalbard) after glacial retreat using radionuclides and elemental fingerprints. Geomorphology. 367,107304**.** DOI: 10.1016/j.geomorph.020.107304

24 Owens, P.N., Blake, W.H., Gaspar, L., Gateuille, D., Koiter, A.J., Lobb, D.A., Petticrew, E.L., Reiffarth, D.G., Smith, H.G., Woodward, J.C., 2016. Fingerprinting and tracing the sources of soils and sediments: Earth and ocean science, geoarchaeological, forensic, and human health applications. Earth-Science Reviews 162, 1–23. https://doi.org/10.1016/j.earscirev.2016.08.012

25 Palazón L, Latorre B, Gaspar L, et al (2015) Comparing catchment sediment fingerprinting procedures using an auto-evaluation approach with virtual sample

mixtures. Science of The Total Environment 532:456–466. https://doi.org/10.1016/j.scitotenv.2015.05.003

26 Panagopoulos, Y., Makropoulos, C., Mimikou, M., 2011. Diffuse Surface Water Pollution: Driving Factors for Different Geoclimatic Regions. Water Resour Manage 25, 3635. https://doi.org/10.1007/s11269-011-9874-2

27 Phillips DL, Gregg JW (2003) Source partitioning using stable isotopes: coping with too many sources. Oecologia 136:261–269. https://doi.org/10.1007/s00442-003-1218-3

28 Pulley, S., Foster, I., Antunes, P., 2015. The uncertainties associated with sediment fingerprinting suspended and recently deposited fluvial sediment in the Nene river basin. Geomorphology 228, 303–319. https://doi.org/10.1016/j.geomorph.2014.09.016

29 Pulley, S., Collins, A.L., 2018. Tracing catchment fine sediment sources using the new SIFT (SedIment Fingerprinting Tool) open source software. Science of The Total Environment 635, 838–858. https://doi.org/10.1016/j.scitotenv.2018.04.126

30 Quijano L, Gaspar L, Navas A (2016) Spatial patterns of SOC, SON, ^{137}Cs and soil properties as affected by redistribution processes in a Mediterranean cultivated field (Central Ebro Basin). Soil and Tillage Research 155:318–328. https://doi.org/10.1016/j.still.2015.09.007

31 Ramon, R., Evrard, O., Laceby, J.P., Caner, L., Inda, A.V., Barros, C.A.P. de, Minella, J.P.G., Tiecher, T., 2020. Combining spectroscopy and magnetism with geochemical tracers to improve the discrimination of sediment sources in a homogeneous subtropical catchment. CATENA 195, 104800. https://doi.org/10.1016/j.catena.2020.104800

32 Reiffarth, D.G., Petticrew, E.L., Owens, P.N., Lobb, D.A., 2019. Spatial differentiation of cultivated soils using compound-specific stable isotopes (CSSIs) in a temperate agricultural watershed in Manitoba, Canada. J Soils Sediments 19, 3411–3426. https://doi.org/10.1007/s11368-019-02406-3

33 Schmidt S, Meusburger K, de Figueiredo T, Alewell C (2017) Modelling Hot Spots of Soil Loss by Wind Erosion (SoLoWind) in Western Saxony, Germany. Land Degrad Develop 28:1100–1112. https://doi.org/10.1002/ldr.2652

34 Schuller P, Walling DE, Iroumé A, et al (2013) Using 137Cs and 210Pbex and other sediment source fingerprints to document suspended sediment sources in small forested catchments in south-central Chile. Journal of Environmental Radioactivity 124:147–159. https://doi.org/10.1016/j.jenvrad.2013.05.002

35 Shore M, Murphy PNC, Jordan P, et al (2013) Evaluation of a surface hydrological connectivity index in agricultural catchments. Environmental Modelling & Software 47:7–15. https://doi.org/10.1016/j.envsoft.2013.04.003

36 Smith HG, Blake WH (2014) Sediment fingerprinting in agricultural catchments: A critical re-examination of source discrimination and data corrections. Geomorphology 204:177–191. https://doi.org/10.1016/j.geomorph.2013.08.003

37 Valero-Garcés BL, Navas A, Machín J, Walling D (1999) Sediment sources and siltation in mountain reservoirs: a case study from the Central Spanish Pyrenees. Geomorphology 28:23–41. https://doi.org/10.1016/S0169-555X(98)00096-8

38 Wallbrink PJ, Murray AS, Olley JM, Olive LJ (1998) Determining sources and transit times of suspended sediment in the Murrumbidgee River, New South Wales, Australia, using fallout 137Cs and 210Pb. Water Resour Res 34:879–887. https://doi.org/10.1029/97WR03471

FIGURE CAPTIONS

Fig.1 Location of the study area. 3D picture of the catchment created with a DEM and land cover map

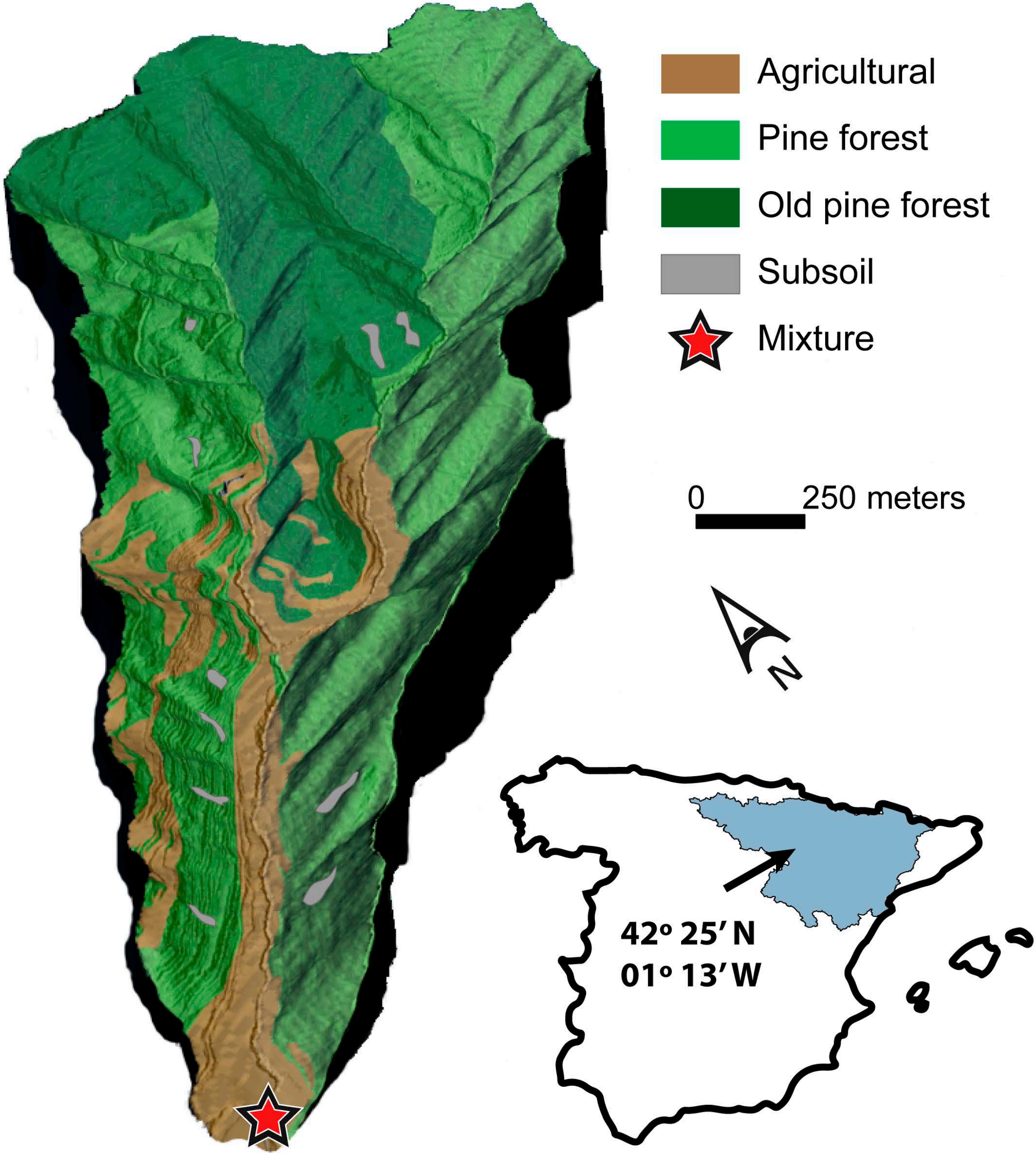
Fig.2 Boxplot of the tracer properties included in the data example of a small catchment. In different colours, the tracers removed by each statistical test

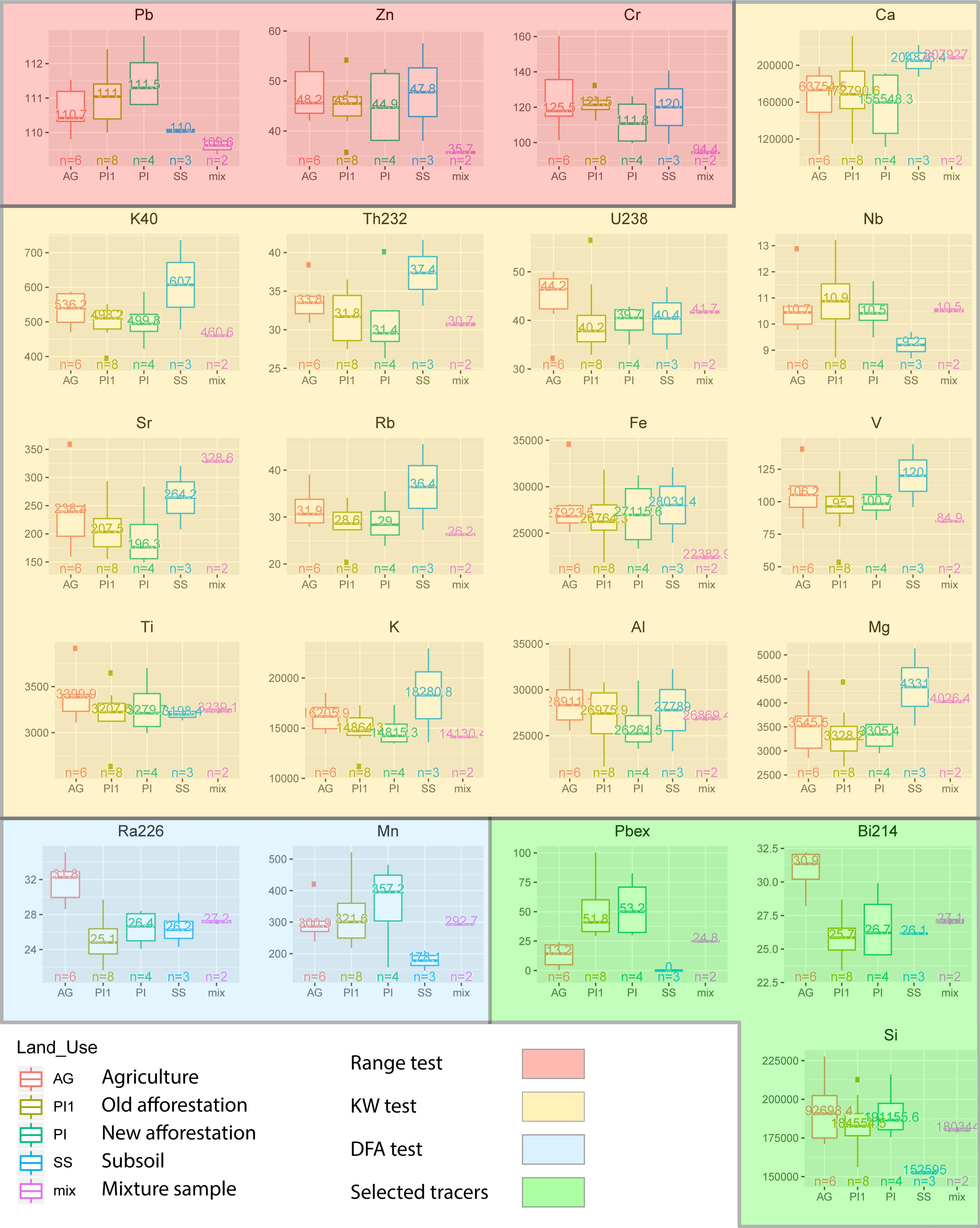
Fig.3 LDA plot of the data example of a small catchment for the different land covers: agricultural (AG), old pine forest (PI); recent pine forest (PI) and subsoil (SS). a) Before running the statistical test, the dataset shows collinearity. b & c) 2D and 3D LDA display of the dataset after running the statistical selection. d) LDA display after merging both pines sources PI and PI1

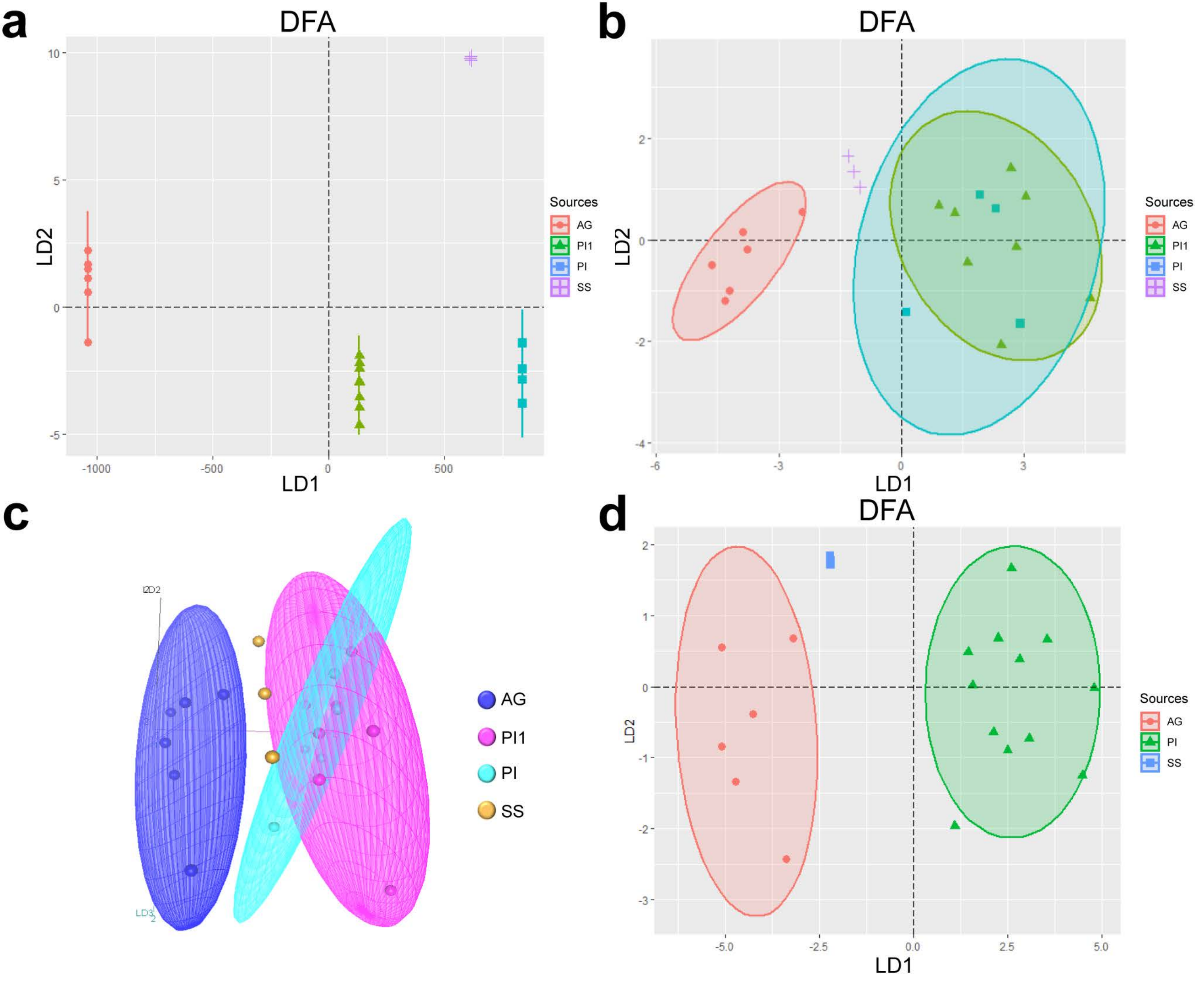
Fig.4 LDA, PCA and density plots of the unmixing process before and after the use of the DFA test for the different land covers: agricultural (AG), pine afforestation (PI) and subsoil (SS)

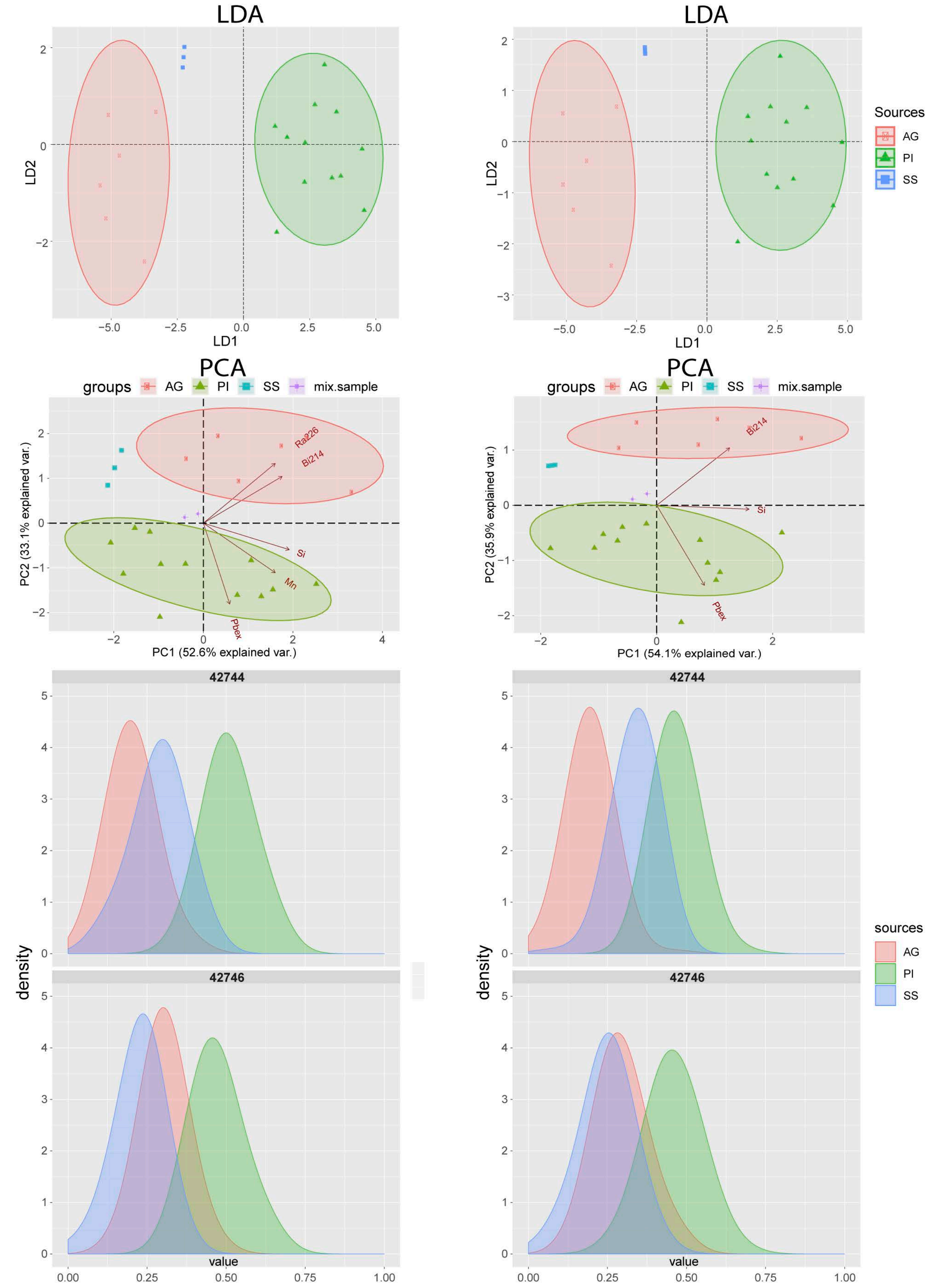
Fig.5 Correlation plots of seven of the tracer properties of the medium size catchment example. Agricultural (AG), Forest (FO), subsoil (SS) and channel bank (CB)

Fig.6 Scaled density plots and results of the unmixing process after the two different tracer selection approaches for the different land covers: Agricultural (AG), Forest (FO), subsoil (SS) and channel bank (CB)


Agricultural
Pine forest
Old pine forest
Subsoil
Mixture
0
250 meters
N
42º 25' N
01º 13' W

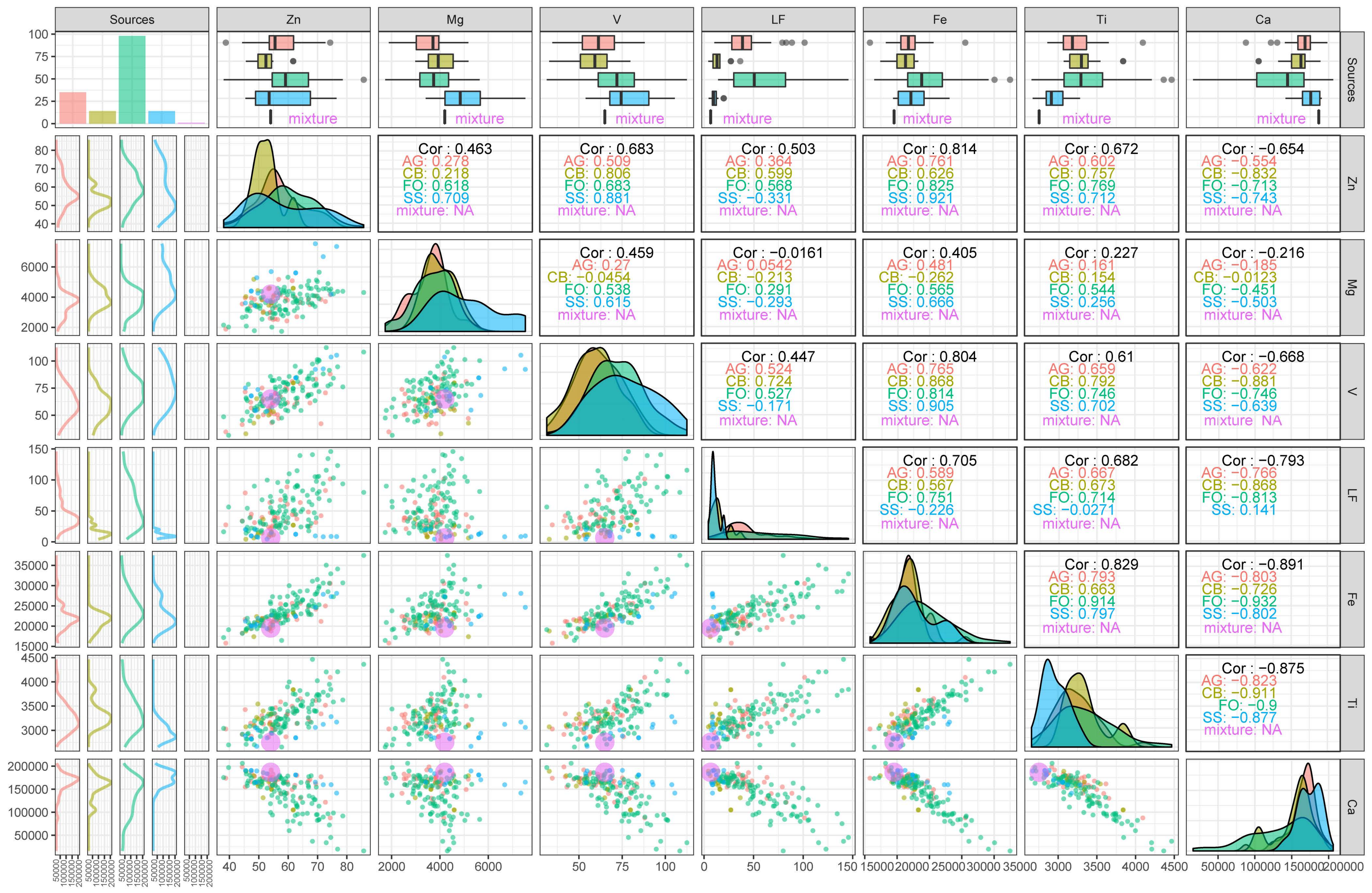

Pb
Zn
Cr
Ca
K40
Th232
U238
Nb
Sr
Rb
Fe
V
Ti
K
Al
Mg
Ra226
Mn
Pbex
Bi214
Si
AG
PI1
PI
SS
mix
n=6
n=8
n=4
n=3
n=2
Land_Use
AG Agriculture
PI1 Old afforestation
PI New afforestation
SS Subsoil
mix Mixture sample
Range test
KW test
DFA test
Selected tracers


a
DFA
LD2
LD1
Sources
AG
PI1
PI
SS
b
DFA
LD2
LD1
Sources
AG
PI1
PI
SS
c
LD2
LD3
AG
PI1
PI
SS
d
DFA
LD2
LD1
Sources
AG
PI
SS

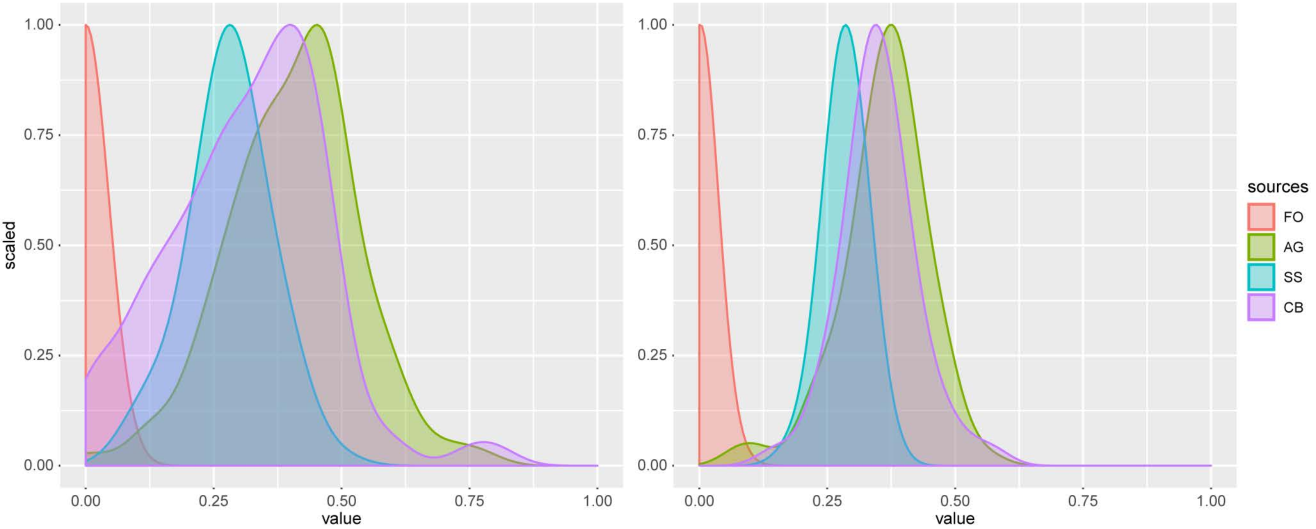


id	GOF.mean	GOF.SD	AG.mean	AG.SD	PI.mean	PI.SD	SS.mean	SS.SD
42744	0.921	0.033	0.202	0.068	0.510	0.073	0.287	0.079
42746	0.926	0.035	0.304	0.061	0.471	0.076	0.225	0.067

id	GOF.mean	GOF.SD	AG.mean	AG.SD	PI.mean	PI.SD	SS.mean	SS.SD
42744	0.952	0.032	0.195	0.064	0.467	0.061	0.338	0.062
42746	0.953	0.037	0.295	0.075	0.458	0.083	0.247	0.077



Sources
Zn
Mg
V
LF
Fe
Ti
Ca
mixture
Cor : 0.463
AG: 0.278
CB: 0.218
FO: 0.618
SS: 0.709
mixture: NA
Cor : 0.683
AG: 0.509
CB: 0.806
FO: 0.683
SS: 0.881
mixture: NA
Cor : 0.503
AG: 0.364
CB: 0.599
FO: 0.568
SS: −0.331
mixture: NA
Cor : 0.814
AG: 0.761
CB: 0.626
FO: 0.825
SS: 0.921
mixture: NA
Cor : 0.672
AG: 0.602
CB: 0.757
FO: 0.769
SS: 0.712
mixture: NA
Cor : −0.654
AG: −0.554
CB: −0.832
FO: −0.713
SS: −0.743
mixture: NA
Cor : 0.459
AG: 0.27
CB: −0.0454
FO: 0.538
SS: 0.615
mixture: NA
Cor : −0.0161
AG: 0.0542
CB: −0.213
FO: 0.291
SS: −0.293
mixture: NA
Cor : 0.405
AG: 0.481
CB: −0.262
FO: 0.565
SS: 0.666
mixture: NA
Cor : 0.227
AG: 0.161
CB: 0.154
FO: 0.544
SS: 0.256
mixture: NA
Cor : −0.216
AG: −0.185
CB: −0.0123
FO: −0.451
SS: −0.503
mixture: NA
Cor : 0.447
AG: 0.524
CB: 0.724
FO: 0.527
SS: −0.171
mixture: NA
Cor : 0.804
AG: 0.765
CB: 0.868
FO: 0.814
SS: 0.905
mixture: NA
Cor : 0.61
AG: 0.659
CB: 0.792
FO: 0.746
SS: 0.702
mixture: NA
Cor : −0.668
AG: −0.622
CB: −0.881
FO: −0.746
SS: −0.639
mixture: NA
Cor : 0.705
AG: 0.589
CB: 0.567
FO: 0.751
SS: −0.226
mixture: NA
Cor : 0.682
AG: 0.667
CB: 0.673
FO: 0.714
SS: −0.0271
mixture: NA
Cor : −0.793
AG: −0.766
CB: −0.868
FO: −0.813
SS: 0.141
Cor : 0.829
AG: 0.793
CB: 0.663
FO: 0.914
SS: 0.797
mixture: NA
Cor : −0.891
AG: −0.803
CB: −0.726
FO: −0.932
SS: −0.802
mixture: NA
Cor : −0.875
AG: −0.823
CB: −0.911
FO: −0.9
SS: −0.877
mixture: NA



GOF.mean	GOF.SD	FO.mean	FO.SD	AG.mean	AG.SD	SS.mean	SS.SD	CB.mean	CB.SD
0.713	0.016	0.002	0.008	0.403	0.128	0.280	0.080	0.315	0.153

GOF.mean	GOF.SD	FO.mean	FO.SD	AG.mean	AG.SD	SS.mean	SS.SD	CB.mean	CB.SD
0.933	0.017	0.003	0.009	0.362	0.084	0.285	0.035	0.351	0.073