

# Joint compositional calibration: an example for U-Pb geochronology

R. Tolosana-Delgado<sup>1</sup>, K. G. van den Boogaart<sup>1</sup>,  
E. Fišerová<sup>2</sup>, K. Hron<sup>2</sup>, and I. Dunkl<sup>3</sup>

<sup>1</sup>Helmholtz Zentrum Dresden-Rossendorf, Helmholtz Institute Freiberg for Resource Technology,  
Freiberg, Germany *r.tolosana@hzdr.de*

<sup>2</sup>Palacky University, Dept. of Mathematical Analysis and Applications of Mathematics,  
Olomouc, Czech Republic

<sup>3</sup>Georg August University, Dept. of Sedimentology and Environmental Geology, Goettingen, Germany

## Abstract

This contribution explores several issues arising in the measurement of a (geo)chemical composition with Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICP-MS), specially in the case that the quantities of interest are linear functions of (log)-ratios. These quantities are scale invariant, but in general cannot be estimated without taking into account possible additive noise effects of the instrumentation, incompatible with a purely compositional approach. The proposed ways to a solution heavily build upon the multi-Poisson distribution, highlighting the counting nature of the readings delivered by these instruments. The model can be fitted using a generalised linear model formalism, and it allows for a joint calibration of all components at once. Relevance of these considerations is shown with some simulation studies and in a real case of multi-isotopic geochronological analyses. Results suggest that the most critical aspect of this analytical technique is the assumption that the amount of ablated mass per second between samples of unknown and of known compositions is similar (*matrix matching*): if this cannot be ensured, absolute estimations of the abundance of each of these isotopes fails, while their (log)ratios are perfectly estimable. This opens the door to using the model for a joint calibration by loosening the condition of matrix matching and using several standards of different composition.

## Abbreviations

	concept	abbreviation (definition)
	indices for variables	$i, j, l$
	index for time slices	$k$
	measuring interval	$T_k$
	background time window	$B_k = [t_A^k, t_b^k]$
	signal time window	$M_k = [t_M^k, t_N^k]$
	a time slice from interval $T_k$	$t_k$
	reading of variable $i$ at time slice $t_n$	$x_i(t_n)$
	background level of variable $i$ at interval $T_k$	$b_{ki}$
	signal of variable $i$ at $t_n$	$\Delta x_i(t_n) = x_i(t_n) - b_{ki}$
	signal level of variable $l$ at interval $T_k$	$y_{kl}$
	average readings of variable $x_i$	$\bar{x}_i$
	standard deviation of readings of variable $x_i$	$s_{xi}$
	average background associated to variable $x_i$	$b_i$
	standard deviation of the background $b_i$	$s_{bi}$
	measuring channel	$i$
	number of channels	$P$
	(random) reading	$X_i$
	true composition	$\mathbf{Z} = [z_1, z_2, \dots, z_D]$
	expected background counts per second at channel $i$	$\lambda_{bi}$
	$i$ -th channel dwell time duration	$\omega_{0i}$
	sensitivity of channel $i$ to isotope $j$	$\lambda_{ij}$
	matrix of sensitivities	$\mathbf{\Lambda}$
	total number of counts produced by isotope $j$	$\lambda_j$
	proportion of counts in each channel $i$ produced by isotope $j$	$\Lambda_{ij}^*$
	number of analytes analysed in this session	$K$
	index for one analyte	$k$
	set of indices for standard analytes	$\mathcal{K}_s$
	set of indices for sample analytes	$\mathcal{K}_m$
	total set of analyte indices	$\mathcal{K} = \mathcal{K}_s \cup \mathcal{K}_m$
	composition-to-counts model	$\mathbf{\Lambda}(\dots; t)$
	vector of expected background counts	$\boldsymbol{\lambda}_b$
	other parameters of the model	$\boldsymbol{\theta}$
	whole measurement period	$T = \bigcup_{k=1}^K T_k$
	true composition of sample $k$	$\mathbf{Z}_k$
	nominal composition, if sample $k$ is a standard	$\mathbf{z}_k$
	union of background windows of the session	$B = \bigcup_k B_k$
	calibration data set	$\mathbf{X}^s$
	prediction data set	$\mathbf{X}^m$

# 1 Introduction

Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICP-MS) is an in-situ analytical technique to quantify the abundance of some isotopes (or their ratios) in a sample of unknown composition. The matter to analyse is ablated with a laser and the resulting aerosol is introduced in a plasma, where its atoms part with their most loosely bound electron to form a (+1)-charged ion. These are then separated by an electro-magnetic field, on the basis of their mass-charge ratio, each colliding with a detector. The result is a vector of counts (number of detected ions per unit of time) for several masses, which must then somehow be related to concentrations or relative abundances of the several isotopes. Usually unknowns (called “samples”) and materials with known compositions (standard reference materials, later shortly as “standards”) are used alternately. This common procedure is called as unknown-standard bracketing and performed in order to derive the proportionality factor between signal and composition. The detectors receive counts even when no ablation happens, forming what is often called a *background* or *blank*. Moreover, the signal received while analysing standards shows systematic drifts at several time scales (along the day, during a measurement, etc), even though the standards have homogeneous compositions. Thus, the desired ratios of abundances must be estimated from data on counts, taking into account the noise in the signal and a proportionality between ratios of signals and ratios of abundances along time, which is in general non-constant. This problem is typically dealt with by establishing a calibration line for each element; these are obtained by subtracting an additive noise to the signal and fitting some parametric function of time to the proportionality values obtained from dividing the readings for the standards by their normative value. Each isotope is calibrated separately. Several functions have been reported to be used, though mostly they are piecewise linear segments or polynomial fits (Cheatham, Sangrey and White, 1993). It is important to mention that conventional calibration procedure of LA-ICP-MS instruments is targetted to obtaining estimates of the amount of each element *individually* with the best precision and accuracy possible, i.e. unbiased and with lowest variability in absolute terms. However, these values are most often used to compute some informative (log)ratios.

None of the existing methods considers in any sense the possible compositional nature of the problem (Pawłowsky-Glahn et al., 2015), namely the fact that the target vector to estimate is a composition, which introduces some modifications on the setting. Two definitions exist of a composition, each with its own implications. The classical definition states that a composition is a vector of positive components and a constant sum to 100%. This definition implies that results of the measurement procedure should deliver vectors of amounts on all elements of the periodic system which are always non-negative and sum to 100%. Because of the lack of a joint calibration/measurement model, one does not know whether this constant sum will be honored, or how can results be corrected to satisfy it. The modern definition states that a (chemical) composition is a vector of positive components reporting the relative abundance of (the set of all) elements in a sample, or alternatively, which total sum is irrelevant (a property called *scale invariance*). This definition begs the question of why should we spend efforts in obtaining *individually* unbiased and lowest variability estimates of each of these quantities in absolute terms, if we are going to interpret them *jointly* in terms of a relative scale.

This contribution presents several models and methods to take the compositional nature of the problem into account, in its two definitions and with all implications outlined before. The paper discusses and shows the potential uses and limitations of each of these models, compared with the classical approach. This work builds upon materials on from Fišerová et al. (2016) and the considerations regarding the background noise in measuring geochemical compositions by van den Boogaart et al. (2013). The keystone of our approach is given by the multi-Poisson distribution (van den Boogaart and Tolosana-Delgado, 2013). This distribution is chosen because it is the most parsimonious model able to describe the number of counts observed in a series of categories when the total number of counts is not known. The multi-Poisson distribution is an extremely simple model, with many shortcomings (e.g.: there is no way to model the dependence between the counts in two categories; or the fact that the variance and the mean must be equal for each component). However, its simplicity makes it a very good model when one does not have enough information to model the actual physics of the phenomenon.

The paper is distributed as follows. Section 2 presents the fundamentals of the LA-ICP-MS analytical technique and its current practice for the non-expert readers. Section 3 puts forward two stochastic models of generation of LA-ICP-MS signals that include both compositional considerations and simplified physics, and gives reasons to the choice of the multi-Poisson model. Section 4 presents a statistical method to work with each of these models, in both the calibration and estimation settings. Section 5 uses a series of simulated scenarios to show the potentials and limitations of these models, specially with regard to their ability to produce unbiased estimates on absolute or relative terms under distortion from the model hypotheses. Section 6 shows the usefulness of one of these models in a real case study. Finally Section 7 discusses the main aspects raised by the simulation and real case studies, as a form of preliminary conclusions. Two appendices are included: one summarising the several geometries involved, and one presenting other compositional calibration models not fitting to the data but included for the sake of completeness.

## 2 Basics of Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA-ICP-MS)

### 2.1 Review of Poisson distribution properties

A random variable  $X$  is said to follow a Poisson distribution with intensity parameter  $\lambda$ , denoted as  $X \sim \mathcal{Po}(\lambda)$  if and only if its probability mass function is

$$f_X(x) = \frac{\lambda^x}{x!} e^{-\lambda}, \quad x = 0, 1, 2, 3, \dots$$

The expected value and population variance of  $X$  are  $E[X] = \text{Var}[X] = \lambda$ , hence this parameter is often called expected number of counts. Note that, in spite of this identification,  $X$  is an integer, while the parameter and these statistics are real positive values  $\lambda \in \mathbb{R}_+$ . Its dispersion coefficient defined as  $D = \text{Var}[X]/E[X] = 1$  always. If  $D > 1$  ( $D < 1$ ), there is evidence of overdispersion (underdispersion) which implies that there is more (less) variability around the model's fitted values than is consistent with a Poisson distribution. Note that this is not the coefficient of variation.

If  $\lambda$  is large, the Poisson distribution can be excellently approximated by a normal distribution with mean and variance both equal to  $\lambda$ .

The sum of two independent Poisson distributed variates  $X_1 \sim \mathcal{P}o(\lambda_1)$  and  $X_2 \sim \mathcal{P}o(\lambda_2)$  follows also a Poisson distribution  $Y = X_1 + X_2 \sim \mathcal{P}o(\lambda_1 + \lambda_2)$ . The difference  $Z = X_1 - X_2$  follows a Skellam distribution (Skellam, 1946), but if  $\lambda_1 \gg \lambda_2$ , then  $Y = X_1 - X_2 \sim \mathcal{P}o(\lambda_1 - \lambda_2)$  approximately.

A vector of  $D$  Poisson variates  $X_i \sim \mathcal{P}o(\lambda_i)$  follows a multi-Poisson distribution (van den Boogaart and Tolosana-Delgado, 2013, p.64) with parameter vector  $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_D]$ . This is a vector of non-negative integer components characterized by the following conditional construction:

1. the sum of these components  $X^T = \sum_i^D X_i \sim \mathcal{P}o(\lambda^T = \sum_i^D \lambda_i)$  gives a total number of counts; note that this holds because the several components are independent by the nature of the Poisson process;
2. conditional on a fixed total number of counts  $x^T$ , the number of counts on each category follows a multinomial distribution with parameters  $\mathbf{p} = \mathcal{C}[\boldsymbol{\lambda}]$  and  $n = x^T$ . Here  $\mathcal{C}[\boldsymbol{\lambda}]$  means the closure of  $\boldsymbol{\lambda}$ .

This construction allows to study the multi-Poisson distributions as the product of a (classical univariate) Poisson distribution times a Multinomial distribution.

## 2.2 Technical procedure

A laser ablation inductively coupled plasma mass spectrometer (LA-ICP-MS) is an *in situ* analytical instrument that is used for determining the elemental and isotopic composition of micrometer-sized areas of solid materials such as minerals, glasses, metal alloys, bones, teeth, calcareous shell and wood, and fluids trapped as inclusions within solids. The analytes are polished and cleaned before introduction into the laser cell. Typically an excimer UV laser (193 nm) is used for ablation, the generated aerosol from the approximately 20 to 100  $\mu\text{m}$  diameter laser spot is transported by a helium-argon gas mixture into the ICP-MS. Here the particles of the aerosol are dried, atomized, and partly ionized by the plasma. The resulting positive ions are accelerated and separated in variable electric and magnetic fields and the ions are distinguished according to their mass-to-charge ratio.

One or more detectors is counting the ions at the end of the trajectories during a certain time period (*dwell time*). Then the detectors move to different positions, corresponding to other masses, and count impacts there. And so on, until the complete set of masses has been visited. Afterwards the detectors comes back to the starting position and counts again, thus starting a new time slice. The dwell time can be variable for the different analytes according to their abundance or importance. Several of such time slices occur during a second, producing a multivariate reading. In spite of the sequential character of the measurements within a time slice, these are considered simultaneous.

This reading procedure is applied in three different situations:

**background** the readings obtained with no material,

**standard reference material** the readings obtained with an analyte of known composition (or short, *standard*),

**sample** (*sensu stricto*) with an analyte of unknown composition.

Given that the counts are registered for each analytes during a fraction of a second, the resulting total counts per time window are linearly upscaled to counts per second. The analytical sequences are usually organized as *bracketing*, i.e. samples and standards are measured alternately.

### 2.3 Conventional data analysis approach

Typically, each measuring interval  $T_k$  contains a period of background readings and a period of sample or standard readings, with more or less sharp transitions between them. The first step is the identification of time windows of background  $[t_A^k, t_B^k] \subset T_k$  and of signal  $[t_M^k, t_N^k] \subset T_k$  (see Fig. 8 later on, for an example), a task that is often manually done by the lab analyst guided by some homogeneity statistics.

The second step is the characterization of the background. This is usually done for each measuring period and for each isotope  $i$  separately. All readings in the background window, denoted  $\{x_i(t_A^k), x_i(t_{A+1}^k), \dots, x_i(t_B^k)\}$ , are considered as independent realizations of a Poisson distribution of unknown parameter  $\lambda_i$ . With the standard assumptions of statistics of Poisson variates, this parameter can be estimated as the mean readings of that variable in the background window, denoted as  $b_{ki}$ . Some labs implement a quality control assessment on the dispersion coefficient within the background window

$$\hat{D}_k = \frac{\text{Var}[x_i(t_A^k), x_i(t_{A+1}^k), \dots, x_i(t_B^k)]}{\text{E}[x_i(t_A^k), x_i(t_{A+1}^k), \dots, x_i(t_B^k)]} = \frac{s_{bi}^2(T_k)}{b_{ki}}$$

with heuristic rules that suggest a too-strong non-Poissonal regime if the hypothesis  $D = 1$  is not acceptable. In this case, typically the analyst reconsiders the choice of background window.

The third step is the definition of the expected readings for each signal window. This is sometimes applied to the absolute readings, sometimes to ratios between two readings. In particular, in U/Pb-geochronological studies the following ratios of interest are commonly used:  $\text{Pb}^{206}/\text{U}^{238}$ ,  $\text{Pb}^{207}/\text{U}^{235}$ ,  $\text{Pb}^{207}/\text{Pb}^{206}$  and eventually  $\text{Pb}^{208}/\text{Th}^{232}$ . A first approach would be to neglect fractionation effects on those ratios and apply the same procedure of the background at the set of readings  $\{\Delta x_i(t_M^k), \Delta x_i(t_{M+1}^k), \dots, \Delta x_i(t_N^k)\}$  vs.  $\{\Delta x_j(t_M^k), \Delta x_j(t_{M+1}^k), \dots, \Delta x_j(t_N^k)\}$ . Note that these values are obtained by subtracting the background levels  $b_{ki}$  and  $b_{kj}$  to the read counts. Some labs work with the arithmetic mean of the ratios  $\Delta x_i(t_m)/\Delta x_j(t_m)$ , while other work with the ratios of the count means  $\overline{\Delta x_i}/\overline{\Delta x_j}$  within the signal window (Jackson et al, 2004). More elaborate approaches consider the fractionation trend as a line or as a curve, and attempt several ways of extracting a representative average ratio. For instance, an option is to fit a linear regression trend to the fractionation drift and extrapolate it to the time when the laser beam hit the sample. Whichever method is used, at the  $k$ -th measurement interval  $T_k$  one has a background value  $b_{kl}$  and a signal value for each quantity of interest  $l$  (ratio or concentration)  $y_{kl}$ .

The fourth step is to study the several measurements  $\{y_{kl}\}$  available for the standards, which, due to their homogeneity, should be “equal”, i.e. ideally realizations of the same random variable. If this can be assumed, then the average of all standard readings  $y_i^{std}$  is compared with the known nominal value  $\mu_i^{std}$ , and all measurements for unknown analytes are upscaled conveniently as

$$y_{kl} = \frac{\mu_i^{std}}{y_i^{std}} y_{kl}.$$

Note that this equation has only sense if one can assume that the sample and the standard behave in the same way during ablation, i.e. that the same amount of mass per second has been ablated and sent to the mass spectrometre. To ensure that, it is common to select standards of the same kind than the sample, something called *matrix matching* (Sylvester).

If the several readings of the standard show too obvious systematic drifts, then it is common to assume some functional model for this drift, fit it to the available measurements of the standard, predict its value  $y_i^{std}(T_k)$  for a measuring interval  $T_k$  and then upscale the value at that interval accordingly

$$y_{kl} = \frac{\mu_i^{std}}{y_i^{std}(T_k)} y_{kl}.$$

Some effort in these data processing steps is devoted to evaluating the “measurement error” (statistical error, uncertainty). Denoting by  $\bar{x}_i$ ,  $s_{xi}^2$  and  $b_i$ ,  $s_{bi}^2$  the means and variances of the signal (no background correction) and of the background (dropping the dependence on  $T_k$  for simplicity) of the isotope  $i$ , the variance of the corrected signal is

$$s_{ci}^2 = s_{bi}^2 + s_{xi}^2,$$

under the hypothesis that the background and the corrected signal are considered as independent Poisson distributions which sums to the uncorrected signal. These concepts and statistics allow as well the definition of *detection limit*, also known as *level of detection (LoD)*. The detection limit is defined as that level of signal which cannot be distinguished from the background. It is customary to take the detection limit of each isotope counts as  $LoD_i = b_i + 3 \cdot s_{bi}$ .

### 3 A compositional calibration model for ICP-MS

To build a compositional calibration model for this kind of measurements it is important to distinguish between the measurements, the several estimates of some abundances of components of interest, and their actual abundances. In what follows, we call a *channel* a particular position of the detector, which is (hopefully) placed at the end of the trajectory taken by ions of one single charge/mass ratio.  $P$  denotes the total number of channels available. The number of collisions counted by the detector on the  $i$ -th channel is called a *reading*, and is denoted as  $X_i$ . The actual vector of abundances of all  $D$  chemical elements in the sample is called its (*actual*) *composition*, and is denoted by the vector  $\mathbf{Z} = [Z_1, Z_2, \dots, Z_D]$ .

Following the ideas of the preceding chapter, we may assume the readings of the background to be

a vector of  $P$  components following a multi-Poisson distribution, i.e. at each time slice while the sample is not being ablated,

$$X_i(t) \sim \mathcal{Po}(\omega_{0i}\lambda_{bi}), \quad t \in [t_A, t_B], \quad (1)$$

where  $\lambda_{bi}$  is the expected counts per second of the background and  $\omega_{0i}$  is the length of the dwell time of channel  $i$ , in seconds. At the moment that ablation starts, though, that channel will produce readings assumed to follow the law

$$X_i(t) \sim \mathcal{Po} \left( \omega_{0i} \left[ \lambda_{bi} + \sum_{j=1}^D \Lambda_{ij}(t) \cdot (\dot{m}(t) \cdot z_j) \right] \right), \quad t \in [t_M, t_N], \quad (2)$$

where  $\dot{m}(t)$  is the mass of sample per second escaping the spot, and  $\Lambda_{ij}(t)$  is the expected number of counts per second produced by one gram of element  $j$  in channel  $i$ . The sample composition is considered constant. The escaped mass per second is often observed to be an exponentially decaying function of time,  $\dot{m}(t) = \dot{m}_0 \cdot \exp(\theta_1 t)$ . This model is called a *full interaction model*, because it allows that each isotope potential influences all channels, in a varying way along time.

The quantities  $\Lambda_{ij}(t)$  deserve a longer discussion. Each can be interpreted as a *sensitivity of channel  $i$  to element  $j$* . It is typically assumed that each channel  $i$  corresponds to one single element  $j$ , i.e. that this matrix is diagonal. In this case,  $P = D$  and  $\Lambda_{ij}(t) := \lambda_j(t)\delta_{ij}$ , thus

$$X_i(t) \sim \mathcal{Po}(\omega_{0i} [\lambda_{bi} + \dot{m}(t) \cdot \lambda_i(t) \cdot z_i]), \quad t \in [t_M, t_N]. \quad (3)$$

Note that Eq. (3) implies that the process is partly non-compositional, as the intensity of the Poisson process is not scale invariant. This effect can be seen later in the application with true data, in Section 6. The expected vector of counts is

$$\mathbb{E}[\mathbf{X}(t)] = \boldsymbol{\lambda}_b + \omega_0 \cdot \dot{m}(t) \cdot [\lambda_1(t) \cdot z_1, \lambda_2(t) \cdot z_2, \dots, \lambda_D(t) \cdot z_D]$$

which is rather an object of  $\mathbb{R}_+^D$ , the multivariate positive real space. The part produced by the ablated mass  $\dot{m}(t)$  is related to the vector  $\boldsymbol{\lambda}(t) \oplus_+ \mathbf{z} = [\lambda_1(t) \cdot z_1, \lambda_2(t) \cdot z_2, \dots, \lambda_D(t) \cdot z_D]$ , where  $\oplus_+$  denotes the component-wise product of two vectors of positive components. This operation is the Abelian group operation of  $\mathbb{R}_+^D$  (Pawlowsky-Glahn and Egozcue, 2001), also known as amount-perturbation (van den Boogaart and Tolosana-Delgado, 2008). Thus, this model will be further referred to as an *amount-perturbation upscaling model*. Note that  $\boldsymbol{\lambda}(t) = [\lambda_1(t), \lambda_2(t), \dots, \lambda_D(t)]$  is an *amount vector* of sensitivities. As functions of time, these sensitivities are reported to show very complex and varying patterns at different time scales (Cheatham, Sangrey and White, 1993).

Between the full interaction model (Eq. 2) and the amount-perturbation upscaling model (Eq. 3), an intermediate model can be considered. Here we consider the total sensitivity of all channels to ions of type  $j$  as varying along time, denoted as  $\lambda_j(t)$ ; however, the way these counts are split among the  $P$  channels is considered time-independent, and denoted as  $\Lambda_{ij}^*$ . Thus,  $\Lambda_{ij}(t) = \Lambda_{ij}^* \cdot \lambda_j(t)$ .

The resulting model

$$X_i(t) \sim \mathcal{Po} \left( \omega_{0i} \left[ \lambda_{bi} + \sum_{j=1}^D \Lambda_{ij}^* \cdot \lambda_j(t) \cdot (\dot{m}(t) \cdot z_j) \right] \right), \quad t \in [t_M, t_N], \quad (4)$$

is called (*constant*) *matrix-interaction amount-perturbation model*.

## 4 Methods

### 4.1 Notation and common assumptions

Let us assume one particular model  $\Lambda(\mathbf{z}(t_i), \boldsymbol{\lambda}_b, \boldsymbol{\theta}; t_i)$  from those mentioned before, with  $\boldsymbol{\lambda}_b$  denoting expected counts from the background level and  $\boldsymbol{\theta}$  including the rest of model parameters (dwell times  $\omega_{0i}$ , sensitivities  $\lambda_i$ , interactions  $\lambda_{ij}$ , eventually including their own trend parameters). A set of readings of count vectors  $\{\mathbf{x}(t_i), t_i \in T\}$  is available, obtained along a session  $T$  split in  $K$  intervals  $T_1, T_2, \dots, T_K$ . Each interval contains two non-overlapping windows, the background window  $B_k \subset T_k$  and the measurement window  $M_k \subset T_k$ . All readings during the background window are obtained with no material being analysed, i.e. after Eq. (1). During each measuring window  $M_k$  an analyte of different composition was analysed, i.e.  $\mathbf{Z}(t_i) = \mathbf{Z}_k$ . For some of these compositions, very good estimates  $\mathbf{z}_k$  are available (those corresponding to standards): each  $\mathbf{z}_k$  is called a *nominal composition*, to distinguish them from the true composition  $\mathbf{Z}_k$ . Some other of these compositions are totally unknown, and they constitute the actual target of this problem (the samples). Let the set of indices  $\mathcal{K} = \{1, 2, \dots, K\}$  be partitioned in two disjoint subsets  $\mathcal{K}_s$  (corresponding to the time intervals when a standard was analysed) and  $\mathcal{K}_m$  (corresponding to the intervals when a sample of unknown composition was analysed). The goal is thus to estimate all  $\mathbf{z}_k$  for  $k \in \mathcal{K}_m$ , given the set of all observations  $\{\mathbf{x}(t_i), t_i \in T\}$  and the nominal composition of the standards  $\mathbf{z}_k$  for  $k \in \mathcal{K}_s$ . The set of data can also be split in measurements corresponding to all background periods  $B = \bigcup_k B_k$  and measurement of standards, the calibration set  $\mathbf{X}^s = \{\mathbf{x}(t_i), t_i \in B \cup \bigcup_{k \in \mathcal{K}_s} M_k\}$ ; and data correspond to measurement windows of samples of unknown composition, the prediction set  $\mathbf{X}^m = \{\mathbf{x}(t_i), t_i \in \bigcup_{k \in \mathcal{K}_m} M_k\}$ . Given this setting, we will follow the multi-Poisson assumption for the data  $\{\mathbf{x}(t_i), t_i \in T\}$ , with intensity model  $\Lambda(\mathbf{Z}(t_i), \boldsymbol{\lambda}_b, \boldsymbol{\theta}; t_i)$ .

### 4.2 Generalized linear model (GLM)

To apply the formalism of generalized linear models (Nelder and Wedderburn, 1972) we need to further assume that the composition of the standards is perfectly known and homogeneous, i.e. that the nominal and actual values of the standards are the same,  $\mathbf{Z}_k = \mathbf{z}_k$ . In this case we can consider the problem divided in two steps:

1. Calibration phase. In this phase we consider only the data available from all background windows and the measurement windows of the standard.
2. Prediction phase. In this phase the model is used to estimate the composition of the unknown

samples, potentially with confidence intervals (instead of the classical error). As extra results of this step, one can derive the largest component  $z_{ki}$  which cannot be distinguished from 0 with a 99% confidence interval (the DL for variable  $i$  at observation  $k$ ).

#### 4.2.1 Amount-perturbation model with linear time drift

This case is the easiest to understand, as the lack of any form of interaction allows to estimate each component separately. We first reparametrize Eq. (3) using  $\dot{m}(t)\lambda_j(t) := \theta_{0j} + \theta_j t_i$ , i.e. assumed a linear function of time. In this case we have

$$X_j^s(t_i) \sim \mathcal{Po}(\omega_{0j}[(\theta_{0j} + \theta_j t_i)z_j(t_i) + \lambda_{bj}]).$$

To fit the parameters of this model, the GLM formalism establishes that a transformation  $\eta(\cdot)$  of the expected value of  $X_j^s(t_i)$  should be predicted with a linear combination of explanatory variables, i.e. in the calibration phase, given that  $z_j(t_i)$  and  $t_i$  are known everywhere:

$$\eta(\mathbb{E}[X_j^s(t_i)]) = \eta(\omega_{0j}[(\theta_{0j} + \theta_j t_i)z_j(t_i) + \lambda_{bj}]) = a_j + b_j z_j(t_i) + c_j z_j(t_i) t_i.$$

This model can be readily solved if we choose an identity link, i.e.  $\eta(x) = x$ . This is a non-canonical choice, but allows to trivially identify  $a_j = \omega_{0j}\lambda_{bj}$ ,  $b_j = \omega_{0j}\theta_{0j}$  and  $c_j = \omega_{0j}\theta_j$  as the statistical vs. physical parameters to be estimated. Moreover, in the prediction phase, with estimates  $\tilde{a}_j$ ,  $\tilde{b}_j$ ,  $\tilde{c}_j$  set, the unknown is  $z_j(t_i)$ , i.e. the model becomes again a generalized linear model

$$X_j^m(t_i) \sim \mathcal{Po}(\tilde{a}_j + [\tilde{b}_j + \tilde{c}_j t_i]z_j(t_i)),$$

where  $\tilde{a}_j$  is an offset and the predictor variable  $[\tilde{b}_j + \tilde{c}_j t_i]$  is known everywhere. This model does not allow an intercept. These models can be estimated in both steps with the GLM maximization likelihood procedures (Nelder and Wedderburn, 1972).

Note that the canonical choice of the Poisson family (the logarithmic link,  $\eta(x) = \log(x)$ ) would give rise to a different model, called *multiplicative perturbation-scaling model*, and explained in the appendix. It would easily allow the inclusion of the exponential decay observed in  $\dot{m}(t)$ , but then there would be no way to identify the additive effect of the background with any parameter of the model.

#### 4.2.2 matrix-interaction amount-perturbation model with linear time drift

In this case, given the interaction between components it is not possible to consider them totally independently any more. Considering Eq. (2) with  $\dot{m}(t)\Lambda(\mathbf{t}) =: \Theta^0 + \Theta t_i$ , the joint model is

$$\mathbf{X}(t_i) \sim \mathcal{Po}(\omega_0 \oplus_+ [(\Theta^0 + \Theta t_i) \oplus_+ \mathbf{z}(t_i) + \lambda_b]) = \mathcal{Po}(\text{diag}[\omega_0] \cdot [(\Theta^0 + \Theta t_i) \cdot \mathbf{z}(t_i) + \lambda_b]),$$

or distributing

$$\begin{aligned}\mathbf{X}(t_i) &\sim \mathcal{P}o(\text{diag}[\boldsymbol{\omega}_0] \cdot \boldsymbol{\Theta}^0 \cdot \mathbf{z}(t_i) + \text{diag}[\boldsymbol{\omega}_0] \cdot \boldsymbol{\Theta} \cdot [t_i \mathbf{z}(t_i)] + \text{diag}[\boldsymbol{\omega}_0] \cdot \boldsymbol{\lambda}_b) \\ &\sim \mathcal{P}o(\mathbf{a} + \mathbf{B} \cdot \mathbf{z}(t_i) + \mathbf{C} \cdot [t_i \mathbf{z}(t_i)]).\end{aligned}$$

with intercept vector  $\mathbf{a} = \text{diag}[\boldsymbol{\omega}_0] \cdot \boldsymbol{\lambda}_b = \boldsymbol{\omega}_0 \odot \boldsymbol{\lambda}_b$  and two matrices  $\mathbf{B} = \text{diag}[\boldsymbol{\omega}_0] \cdot \boldsymbol{\Theta}^0$  and  $\mathbf{C} = \text{diag}[\boldsymbol{\omega}_0] \cdot \boldsymbol{\Theta}$ , respectively a  $K$ -amounts vector and two  $(K \times D)$ -matrices of coefficients. Arrived at this point, it is possible to split the problem in the several components of the response, i.e.

$$X_j(t_i) \sim \mathcal{P}o(a_j + \mathbf{b}'_j \cdot \mathbf{z}(t_i) + \mathbf{c}'_j \cdot [t_i \mathbf{z}(t_i)]).$$

where  $\mathbf{a} = [a_1, a_2, \dots, a_D]$  and where  $\mathbf{b}'_j$  and  $\mathbf{c}'_j$  are the  $j$ -th rows of the matrices  $\mathbf{B}$  and  $\mathbf{C}$  respectively. They can be then arranged together in one single vector of counts obtained on different channels, and a single GLM with a response of the Poisson family and an identity link fitted. In this way, one single vector model is fitted,

$$\mathbf{X}(t_i) \sim \mathcal{P}o(\mathbf{a} + \mathbf{B} \cdot \mathbf{z}(t_i) + \mathbf{C} \cdot [t_i \mathbf{z}(t_i)]).$$

As we did in the preceding case, in the calibration step we use the data  $\mathbf{X}^s$  and the known predictors  $\{t_i \in T\}$  and  $\{\mathbf{z}_k, k \in \mathcal{K}_s\}$  to obtain estimates  $\tilde{\mathbf{a}}$ ,  $\tilde{\mathbf{B}}$  and  $\tilde{\mathbf{C}}$  of the model parameters. In the prediction step, by fixing these parameters on their estimated values, the model

$$\mathbf{X}(t_i) \sim \mathcal{P}o(\tilde{\mathbf{a}} + \mathbf{Z}(t_i) \cdot [\tilde{\mathbf{B}} + t_i \tilde{\mathbf{C}}])$$

can be fitted with an offset  $\tilde{\mathbf{a}}$  and known predictors  $[\tilde{\mathbf{B}} + t_i \tilde{\mathbf{C}}]$  to obtain estimates of the coefficient vector  $\mathbf{Z}(t_i)$ . Note that all channels of  $\mathbf{X}$  are considered jointly.

### 4.3 Caveats

All methods mentioned before share several main limitations:

- perfectly known standard compositions are required;
- it is not possible to model inhomogeneities of the materials considered (standards or samples);
- no solution exists for other more realistic and flexible models, like a multiplicative interaction-perturbation model with additive error;
- if the hypothesis of Poisson distribution is verified to be inappropriate, the likelihood cannot be computed exactly, and GLM fitting procedures might fail.

All these issues can be tackled with Bayesian estimation techniques. Though they are not much more complex than the methods presented so far, these fall beyond the scope of this contribution and are left for future research. Interested readers can consider the work of van den Boogaart et al. (2015), dealing with a model that can accommodate some of these effects.

## 5 Simulation

This section uses simulations from the several models and methods specified before to illustrate their ranges of validity, as well as to show their behaviour when the underlying hypotheses are not satisfied. In particular, it will be shown that estimation of (log)ratios of components is more stable than estimation of the absolute values of these components.

For this task, several scenarios follow. In each scenario, one parameter is left to vary, and for each value of this parameter, we simulate 300 sets of 200 readings of a three-channel instrument ( $P = 3$ ) depending on a four-part composition ( $D = 4$ ), split in  $B = 100$  background window readings and  $M = 100$  measurement window readings. Only  $K = 2$  periods are considered, one for the standard and one for the measurement. The standard is considered to have a nominal composition of  $\mathbf{z} = [1, 2, 3, 94]\%$ , while the sample has an unknown real composition of  $\mathbf{z} = [3, 2, 1, 94]\%$ . As general settings, we consider a sensitivity of  $\lambda_i = 2$  counts per unit of mass, a background of  $\lambda_i = 2$  counts per dwell time, and a mass of 1000 units ablated. Only results of the three first components are reported. The following cases are considered: varying length of the measurement window; varying total sensitivities and background levels; different masses ablated per unit of time between standard and sample; a heterogeneous standard composition; the presence of a (non-modelled) fractionation effect. These scenarios are built with an amount-perturbation upscaling model. Finally, a constant amount-matrix upscaling-interaction model is considered. In each case, we show boxplots of the (absolute) enrichment/depletion factors  $z_i^m/z_i^s$ , i.e. the number with which one should multiply the nominal value of component  $i$  on the standard to obtain an estimate of the absolute abundance of that element on the sample of unknown composition. If results are not satisfactory, we also report results of the corresponding perturbation,  $\mathcal{C}[z_1^m/z_1^s, z_2^m/z_2^s, z_3^m/z_3^s]$  or some of their logratios, i.e. with either one or the other we can estimate the subcomposition of the sample from the subcomposition of the standard, but not the absolute abundances of these 3 components.

Figure 1 shows that varying width of the measuring window or varying the sensitivity of one channel have the same influence on enrichment factors as the estimation of an average with varying sample size: the larger the total number of counts registered, the less variance the enrichment factor shows. On the other hand, varying the level of noise (i.e. the counts per unit of time of the background) does not affect the variability of results significantly, at least as long as the background represents less than 1:5 parts of the total signal. In any of the cases presented, estimates of the enrichment factors appear to be unbiased. The same can be said of the associated perturbations (results not shown).

A different picture is obtained if the ablated mass per unit of time of standard and of sample are not equal (Fig. 2). If the ablation rate is lower (higher) in the sample than in the standard, less (more) counts of all elements will be produced per unit of time, and the method will consequently estimate a lower (higher) absolute abundance of each element of the subcomposition. Absolute enrichment factors will therefore be strongly biased as long as the ratio of ablated masses differs from 1. On the contrary, perturbations remain fairly stable on a very wide range of this ratio (results shown between 1:2 and 2:1). Note that the reference levels are in this case  $\mathcal{C}[3, 1, 1/3] = [9, 3, 1]/13$ . These facts are well known to geochronologists, who typically work with ratios, but not so common in

other geochemical communities using LA-ICP-MS data.

Figures 3 and 4 show the obtained enrichment factors in absolute or relative scales (as logratios) in the case that the real composition of the standard is assumed to vary along the spot, i.e. that masses with different compositions are sent to the detector during the measuring window of the standard. The actual composition is taken under assumption of additive logistic normal distribution with a diagonal ilr-variance matrix with variances  $\sigma$  within the subcomposition of the first three components, and  $10^{-10}$  on the balance of the fourth component against the other three. For the parameters given, this distribution is undistinguishable from a univariate lognormal distribution on each of the first three components. The two figures differ in which has been considered to be the nominal value: [1, 2, 3, 94] is taken as the compositional mean (closed *geometric* mean) of the standard in Figure 3, while in Figure 4 the *arithmetic* mean is forced to be the nominal value. Results show that a calibration equating nominal value to the geometric mean is not appropriate for moderate to high levels of standard heterogeneity (i.e. high values of  $\sigma$ ) as both enrichment factors and logratios show notable biases in this case: enrichment factors larger than one tend to be underestimated, while those smaller than one tend to be overestimated. On the other hand, a calibration equating nominal value to the arithmetic mean shows almost no bias even in cases of high variance, though ultimately the same overestimation/underestimation bias patterns occur. This highlights the importance of having homogeneous standards. Interestingly, this arithmetic specification of the nominal values delivers better (in the sense of less bias) results than geometric specifications even regarding the estimation of logratios.

Figure 5 (right) shows the results from the case that the ablated mass show an exponential decay fraction  $\dot{m}(t) = \dot{m}_0 \exp(\ln(1 - \theta_t)t/\Delta t)$  of  $100\theta_t\%$  of the original mass during the time period  $\Delta t$  (which is considered equal to the measuring window length). Note that the fitting GLM algorithm suggested cannot take this effect into account, thus we are dealing here with a model misspecification effect. In spite of this, final estimates show no degradation: the same variability and no bias can be observed, independently of the level of fractionation.

Finally, Figure 5 (left) shows a case in which some interaction between components exist. Simulations are obtained with a constant matrix-interaction amount-perturbation model (Eq. 4), using the matrix

$$\mathbf{\Lambda}^* = \begin{pmatrix} 1 & \frac{\alpha}{1+\alpha} & \frac{\alpha}{1+2\alpha} \\ 0 & \frac{1}{1+\alpha} & \frac{\alpha}{1+2\alpha} \\ 0 & 0 & \frac{1}{1+2\alpha} \end{pmatrix}$$

for several values of  $\alpha$  between  $\alpha = 0$  (no interaction) to  $\alpha = 0.2$  (notable interaction). In the calibration phase, the model parameters cannot be estimated with the classical setting (one single standard measured several times) because of collinearity. It is required to measure several standards: in this case, we have assumed nominal values in the subcomposition of interest [1, 2, 3], [1, 4, 9], [9, 4, 1] and [2, 3, 4], each of which was shot for a period of  $N = 25$  readings. Moreover, the model fitted is rather the complete model, i.e a matrix-interaction amount-perturbation model after Eq. (2). Results show no bias and a constant variability with increasing interaction parameter  $\alpha$ .

In summary, the GLM methods proposed are applicable in the presence of moderate variability in

the standard composition (the nominal composition should be the arithmetic expected value of the distribution of the composition of the standard), and do very well filter out exponentially decaying fractionation effects. However, if one cannot ensure that both sample and standard behave in the same way (similar masses per second ablated, i.e. the “matrices match”) then only (log)ratios of the components are reasonably estimated.

## 6 Application

To illustrate the presented concepts, models and solving techniques we use a data set of geochronologically relevant isotopes. These are obtained ablating 35 zircons, including 9 standards, analysed for 6 isotopes: Hg202, Pb204, Pb206, Pb207, Pb208, Th232, U235, U238 (Fig. 6). The analysed sequences is composed of 3 blocks of 10 samples bracketed by 4 blocks of 5 standards (3 types of standards were used, coded GJ1, FC1 and 915, see Figure 9). It is also possible to see that, except for some outliers, the isotopes are ordered in decreasing order of abundance as U238, then Th232 or Pb206, and finally U235, Pb208 and Pb207 in varying orders. Dwell times considered are reported in Table 1.

**Table 1: Dwell times for the several elements considered in ms**

Hg202	Pb204	Pb206	Pb207	Pb208	Th232	U235	U238
4	4	2	8	4	2	2	2

A closer look at the last three measurement periods (Fig. 7), including one standard and two samples, shows several remarkable aspects:

- First, Hg202 and Pb204 do not show up in the samples or in the standard in a way significantly different from the background. These elements are used for analytical quality control, and are not relevant for the problem itself.
- Second, several isotopes show clear exponentially decreasing trends while being shot (linear in log scale), an effect of fractionation of the gas cloud while the laser penetrates deeper in the material.
- Finally, a more relevant aspect for modelling, the variability of these time series is neither additive nor multiplicative, because the background and the signal do not have comparable levels of variability neither in raw nor in log scales. This is the reason why purely multiplicative models (like those presented in section A.2.1) or purely additive models (like using linear regression directly to link counts with composition, as done sometimes in the device calibration literature) are not realistic. This gives a diagnostic tool to decide whether a problem might be attacked with linear regression on raw variables (raw  $Y$ -axis plots show similar variability in the background and signal windows), with linear regression on log-transformed variables (log  $Y$ -axis plots show similar variability in the background and signal windows), or Poisson regression (neither one nor the other are satisfied).

Other aspects and relevant concepts are shown as well in Figure 8. Beside the clear fractionation effects occurring while the laser is shot (increasing trend, followed by a shallow exponential

**Table 2:** Nominal values of some isotopic ratios in standard GJ1 ()

	$^{206}\text{Pb}/^{238}\text{U}$	$^{207}\text{Pb}/^{235}\text{U}$	$^{208}\text{Pb}/^{232}\text{Th}$	$^{206}\text{Pb}/^{207}\text{Pb}$	age (Ma)
GJ1	0.09761	0.8093	0.03011	0.06014	602

decrease) or right after switching it off (pronounced decrease), we see as well the presence of zonations of different composition, thus potentially of different age. The figure also shows that readings have much stronger drifts and variabilities than the relevant isotopic logratios in the measuring windows.

A further assessment on the possible structure of the variability of this series is displayed in Figure 9, which makes use of the property that Poisson distributed variates should show dispersion coefficients of  $\hat{D} = 1$ . For the background windows, overdispersion is clear to see in the heavy ions (U238, Th232). Dispersion coefficients for the measurement period are only reported for the sake of completeness, as they are difficult to estimate due to the presence of the irregularities mentioned before: the values reported in this case are obtained using the residual variance with respect to a fitted exponential trend, which might deliver reasonable estimates for well-behaved samples, but is completely inappropriate for zoned samples. Nevertheless, it shows roughly a similar distribution with a certain bias towards underdispersion.

Given the considerations of this preliminary descriptive analysis, the natural conclusion must be that appropriate models for this dataset should be flexible enough to consider at least the following three effects:

1. natural variability on the composition of the materials (i.e., lack of homogeneity), in particular including zonation of samples and random inhomogeneities of standards and samples;
2. downhole fractionation (typically appearing as local exponentially decreasing trends);
3. Poisson overdispersion (something which is included in state-of-the-art GLM fitting models);
4. additive background variability.

Unfortunately, none of the models presently implemented within the framework of generalized linear models can deal with all these effects, the main limitation being that one cannot simultaneously treat as linear the additive background and the exponential trends. Nevertheless, as we have opted for leaving Bayesian methods for future work, the following is an approximate set of results ignoring the first and third shortcomings using a Poisson GLM regression with identity link (Section 4.2.1).

Results (Fig. 10) show several interesting patterns. The enrichment factors for the standards (Fig. 10 lower plot) should be all 1. The standards show a remarkably constant composition, suggesting that no long-range time drift is necessary. The method proposed is quite robust, for instance showing no to minor influences of the outliers of Pb208 in samples 915-005 or Z-025. If we order the elements by decreasing abundance on the estimated composition and on the observed counts, results are the same, which given the several shortcomings we had to take is a good result.

Finally, using constant nominal values for the standard GJ1 as reported in Table 2, the geochronologically relevant ratios (radiogenic  $^{207}\text{Pb}$  vs.  $^{235}\text{U}$ , and radiogenic  $^{206}\text{Pb}$  vs.  $^{238}\text{U}$ ) of all 35 samples were computed. Given that no evidence of the presence of non-radiogenic  $^{204}\text{Pb}$  is found (except perhaps in some zoned Zircons, like Z-025, Fig. 8), all  $^{207}\text{Pb}$  and  $^{206}\text{Pb}$  detected were considered to be radiogenic. Thus, the direct ratios  $^{207}\text{Pb}/^{235}\text{U}$  and  $^{206}\text{Pb}/^{238}\text{U}$  were used to compute an age,

$$\frac{x\text{Pb}}{y\text{U}} + 1 = \exp(\lambda_y t),$$

which for  $^{207}\text{Pb}/^{235}\text{U}$  uses the decay rate  $\lambda_{235} = 9.8485 \cdot 10^{-10}$ , and for  $^{206}\text{Pb}/^{238}\text{U}$  the rate  $\lambda_{238} = 1.55125 \cdot 10^{-10}$ . Figure 11 shows the so called concordia diagram (Wetherill, 1956), a graphical display of the agreement between the two ages. The plot shows notable variability in the ages provided to the standards (especially, 915), and two samples that are notably far from the concordia curve, corresponding to zoned Zircons. But in general terms, the pair of ages show a satisfactory agreement for the whole data set. The obtention of some confidence regions on these calculated ratios within the scope of the GLMs used here remains to be done, and is left for future research.

## 7 Discussion

The most obvious implication of these results is the fact that models without additive error cannot be accepted as reasonable descriptions of the physical LA-ICP-MS measurement process. This is especially true when the target value is small because then the contribution of the multiplicative error to the measurement uncertainty becomes irrelevant in comparison with the contribution of the additive error. This was visible on the reduction of the estimated concentration of the minor isotopes of the standards in the real case study. This mixed additive-multiplicative nature could also be seen in the simulation studies with heterogeneous standard compositions: there an arithmetic mean specification of the standard delivered better results even with regard to the estimation of logratios. Though it is premature to extract conclusions of this single study, the additive nature of this kind of data perhaps comes from the fact that the ICP-MS does actually count individual atoms/ions/isotopes, and matter (or mass) is an additive property. Building a composition out of the obtained measurements of the several isotopes is then a choice of the analyst, and not intrinsically demanded by this kind of data.

A second family of implications relates to the fact that (log)ratios seem to be much more robustly estimated than their numerator and denominator elements separately, in particular if one cannot ensure that similar amounts of mass are ablated per second in the standard and the samples, i.e. if no good matrix matching is possible. This might be one of the reasons why many ICP-MS users do actually work with ratios, instead of with the absolute abundances provided by the device.

Third, results have implications in the treatment of values below the detection limit. Values below detection limit occur when the number of counts/sec. in the signal window cannot be statistically distinguished from counts/sec. in the background window, i.e. when the number of counts coming from the analysed mass is “low”. In the authors’ opinions, this strongly suggests to avoid additive logistic normal (ALN) models for BDLs, because the low number of counts does not allow the

assumption of a central limit related normal approximation of the distribution. Moreover, the ALN totally disregards the additive effect of the background, which happens to be dominant precisely for the low values. An appropriate alternative should take somehow the counting nature of the problem into account.

Finally, to obtain a full joint compositional calibration model remains a difficult task because of the need of several standards of different composition, which should be each homogeneous, perfectly known and of the same kind of matrix than the samples to analyse. Given the practical problems of fulfilling these conditions in real-world applications even for a single standard, it is foreseeable that ICP-MS calibration will remain univariate. Nevertheless, the fact that perturbation enrichment factors (hence logratios) are notably robust to matrix mismatch opens the door to the possibility to of calibrating devices with several standards of relatively different matrices, and use the results for a full joint compositional calibration along the multi Poisson model shown here.

## References

- Aitchison. J. (1986) *The Statistical Analysis of Compositional Data*, Chapman Ltd.
- van den Boogaart, K. G.; Tolosana-Delgado, R. and Templ, M. and Filzmoser, P. (2015) Regression with compositional response having unobserved components or below detection limit values. *Statistical Modelling*, 15(2), pp. 191-213.
- van den Boogaart, K. G. and Tolosana-Delgado, R. (2008) “Compositions”: a unified R package to analyze compositional data. *Computers and Geosciences*, 34(4), pp. 320-338.
- Cheatham, M. M., Sangrey, W. F. and White, W. M. (1993) Sources of error in external calibration ICP-MS analysis of geological samples and an improved non-linear drift correction procedure. *Spectrochimica Acta*, 48B(3), pp. E467-E506.
- Donevska, S.; Fišerová, E. and Hron, K. (2016) Calibration of compositional measurements. *Communications in Statistics - Theory and Methods*, accepted; doi: 10.1080/03610926.2013.839039
- Jackson S., Pearson N., Griffin W. and Belousova E. (2004) The application of laser ablation-inductively coupled plasma-mass spectrometry to in situ U-Pb zircon geochronology. *Chemical Geology* 211, pp. 47-69.
- Nelder, J. and Wedderburn, R. (1972) Generalized Linear Models. *Journal of the Royal Statistical Society, Series A* 135(3), pp. 370-384.
- Pawlowsky-Glahn, V. and Egozcue, J.J. (2001) Geometric approach to statistical analysis on the simplex. *Stochastic Environmental Research and Risk Assessment*, 15(5), pp. 384-398.
- Pawlowsky-Glahn, V., Egozcue, J.J. and Tolosana-Delgado, R. (2015) *Modelling and Analysis of Compositional Data*, Chichester: Wiley.
- Skellam, J. G. (1946) The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society, Series A* 109(3), pp. 296.

- Sylvester, P. J. (2008) Matrix effects in laser ablation-ICP-MS. In: Sylvester, P.J. (ed.) *Laser ablation ICP-MS in the earth sciences: Current practices and outstanding issues*. Mineralogical Association of Canada, Short Course 40, pp. 67-78.
- Wetherill, G. (1953) Discordant uranium-lead ages, I, transactions. *JAmerican Geophysical Union* 37, pp. 320-326.

## A Appendix

### A.1 Three competing geometries

In this paper, three of the compositional geometries gathered by van den Boogaart and Tolosana-Delgado (2008) are used, two on vectors of positive amounts ( $\mathbf{x} \in \mathbb{R}_+^D$ ), either based on an interval or on a ratio scale; and a relative ratio scale on compositions *s.s.* ( $\mathbf{x} \in \mathcal{S}^D$ , the simplex). This section just summarizes their geometries.

The interval scale on  $\mathbb{R}_+^D$  is captured by a geometry inherited from embedding  $\mathbb{R}_+^D$  on  $\mathbb{R}^D$  with its Euclidean space operations, namely the classical vector sum  $+$  and multiplication by scalars  $\cdot$ . The fundamental isometric operation is the identity, thus this geometry is optimally reproduced by linear models with an identity link function.

The ratio scale on  $\mathbb{R}_+^D$  is captured by equipping this set with the Abelian group operation (*amount*)-*perturbation*  $\oplus_+$  and (*amount*)-*powering*  $\odot_+$ , respectively the component-wise product of the components of two vectors and the component-wise powering of the components of one vector to the same scalar (Pawlowsky-Glahn and Egozcue, 2001) . The fundamental isometric operation is the component-wise log-transformation, thus this geometry is optimally reproduced by linear models with a logarithmic link function.

The relative scale on  $\mathcal{S}^D$  is captured by equipping this set with the Abelian group operation *perturbation*  $\oplus_+$  and *powering*  $\odot_+$ , respectively the closed component-wise product of the components of two vectors (Aitchison, 1986) and the component-wise powering of the components of one vector to the same scalar. The fundamental isometric operation is the centered log-ratio transformation, thus this geometry is optimally reproduced by linear models with a logratio link function.

In general terms, a raw scale should be preferred for one variable which absolute differences are meaningful; a ratio scale for variable which meaningful differences are relative; and a relative scale on the simplex should be the choice if several variables show a ratio scale at the same time and their sum is either an artifact or a meaningless constant. Nevertheless, sometimes the same variables can be studied in one way or another, depending on the question that should be answered, i.e. the scale should be chosen depending on the question and not only on the data.

## A.2 Alternative joint models

### A.2.1 Models ignoring the background or with multiplicative effects

From the point of view of the relative scale on  $\mathcal{S}^D$ , all models presented in this paper are particularly complicated by the presence of the additive background. If this could be ignored (e.g. because it is very small with regard to the signal), and the sum of the components of  $\mathbf{Z}$  equals one (i.e. a whole composition is analysed), then the following models are derived:

- scalar upscaling:  $[\mathbf{X}(t_k)|n] \sim \mathcal{M}u(\mathbf{Z}; n)$  and  $n \sim \mathcal{P}o(\omega_0 \lambda(t_k))$ ;
- perturbation:  $[\mathbf{X}(t_k)|n] \sim \mathcal{M}u(\mathbf{Z} \oplus \boldsymbol{\lambda}(t_k); n)$  and  $n \sim \mathcal{P}o(\omega_0 \sum_i^D \lambda_i(t_k))$ ;
- interaction-perturbation:  $[\mathbf{X}(t_k)|n] \sim \mathcal{M}u(\mathcal{C}[\boldsymbol{\Lambda}^* \cdot (\mathbf{Z} \oplus \boldsymbol{\lambda}(t_k))]; n)$  and  $n \sim \mathcal{P}o(\omega_0 \mathbf{1}' \cdot (\boldsymbol{\Lambda}^* \cdot (\mathbf{Z} \oplus_+ \boldsymbol{\lambda}(t_k))))$ .

This last model is still a mixture of additive and multiplicative geometries. The following models are purely compositional alternatives, using in the compositional part only operations on the simplex

- scalar upscaling:  $[\mathbf{X}(t_k)|n] \sim \mathcal{M}u(\mathbf{Z} \oplus \boldsymbol{\lambda}_b; n)$  and  $n \sim \mathcal{P}o(\omega_0 \lambda(t_k) \sum_i^D \lambda_{bi})$ ;
- perturbation:  $[\mathbf{X}(t_k)|n] \sim \mathcal{M}u(\mathbf{Z} \oplus \boldsymbol{\lambda}(t_k) \oplus \boldsymbol{\lambda}_b; n)$  and  $n \sim \mathcal{P}o(\omega_0 \sum_i^D \lambda_{bi} \lambda_i(t_k))$ ;
- interaction-perturbation:  $[\mathbf{X}(t_k)|n] \sim \mathcal{M}u(\boldsymbol{\Lambda}^* \square (\mathbf{Z} \oplus \boldsymbol{\lambda}(t_k)) \oplus \boldsymbol{\lambda}_b; n)$  and  $n \sim \mathcal{P}o(\omega_0 \mathbf{1}' \cdot (\boldsymbol{\Lambda}^* \square (\mathbf{Z} \oplus \boldsymbol{\lambda}(t_k)) \oplus \boldsymbol{\lambda}_b))$ , in this case with  $\oplus$  the perturbation on the simplex.

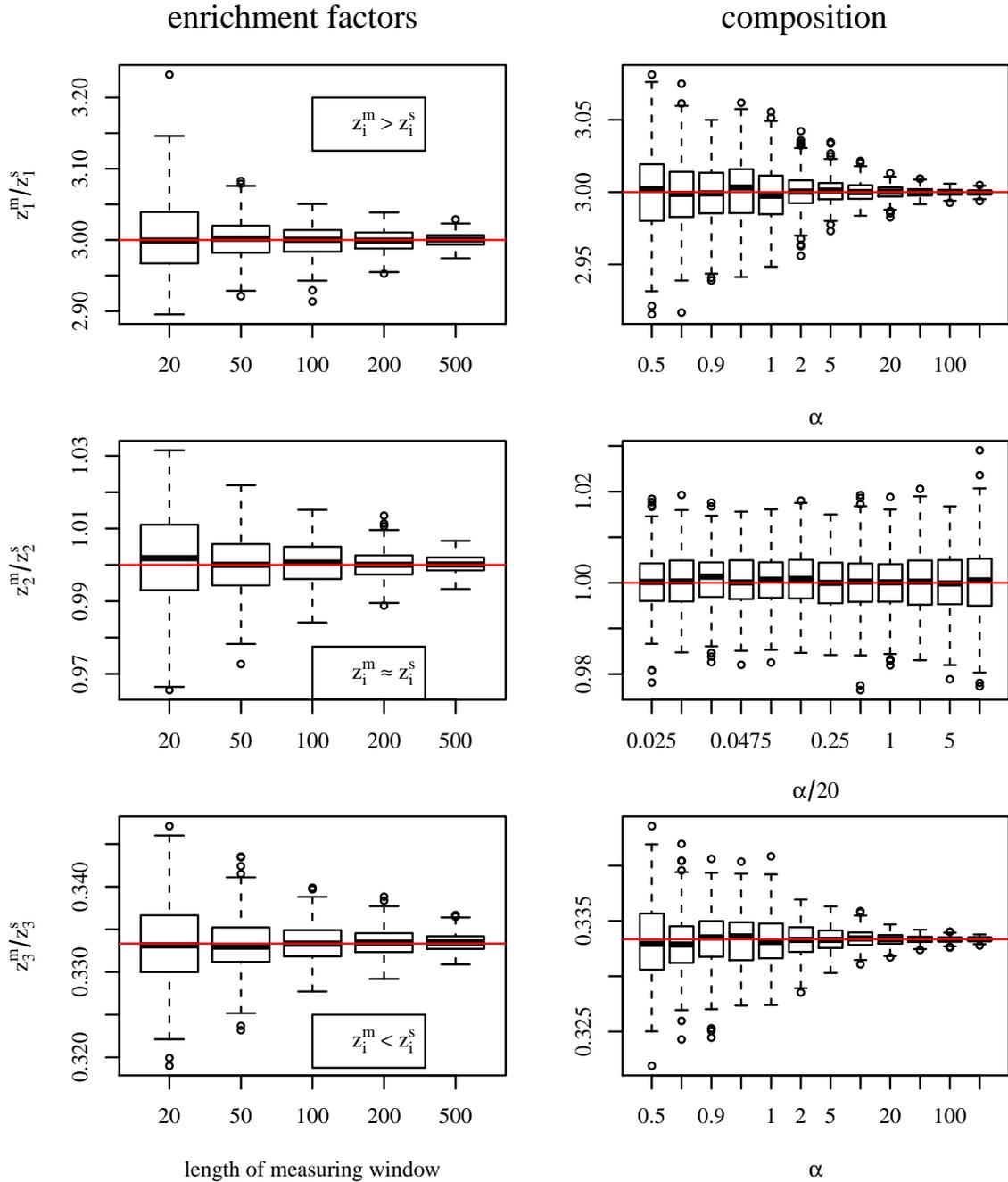
In these expressions, we have used the notation  $\square$  after Pawlowsky-Glahn, Egozcue and Tolosana-Delgado (2015, Chap. 4) to denote a simplicial endomorphism operation, i.e. one such that once expressed in any basis of the simplex becomes a simple matrix-vector product. Note that these purely compositional models imply, among other effects, that the noise induced by the background upscales with the signal, i.e. larger signal should show more background variability.

In any of the cases presented, we finally have distributions for the number of counts on each element class that belong to the multi-Poisson family, with an intensity vector model  $\omega_0 \boldsymbol{\Lambda}(\mathbf{Z}, \boldsymbol{\lambda}_b, \boldsymbol{\theta}; t_k)$  capturing the relationship between the expected partial counts and the composition of the analyte. Hence, we can always consider that the total number of counts  $n(t_k)$  follows a Poisson distribution with  $\lambda_k^T = \omega_0 \mathbf{1}' \cdot \boldsymbol{\Lambda}(\mathbf{Z}, \boldsymbol{\lambda}_b, \boldsymbol{\theta}; t_k)$ ; and, conditional on that total, the vector of counts for each element follows a multinomial distribution with probability parameter vector  $\mathbf{p}_k = \mathcal{C}[\boldsymbol{\Lambda}(\mathbf{Z}, \boldsymbol{\lambda}_b, \boldsymbol{\theta}; t_k)]$ .

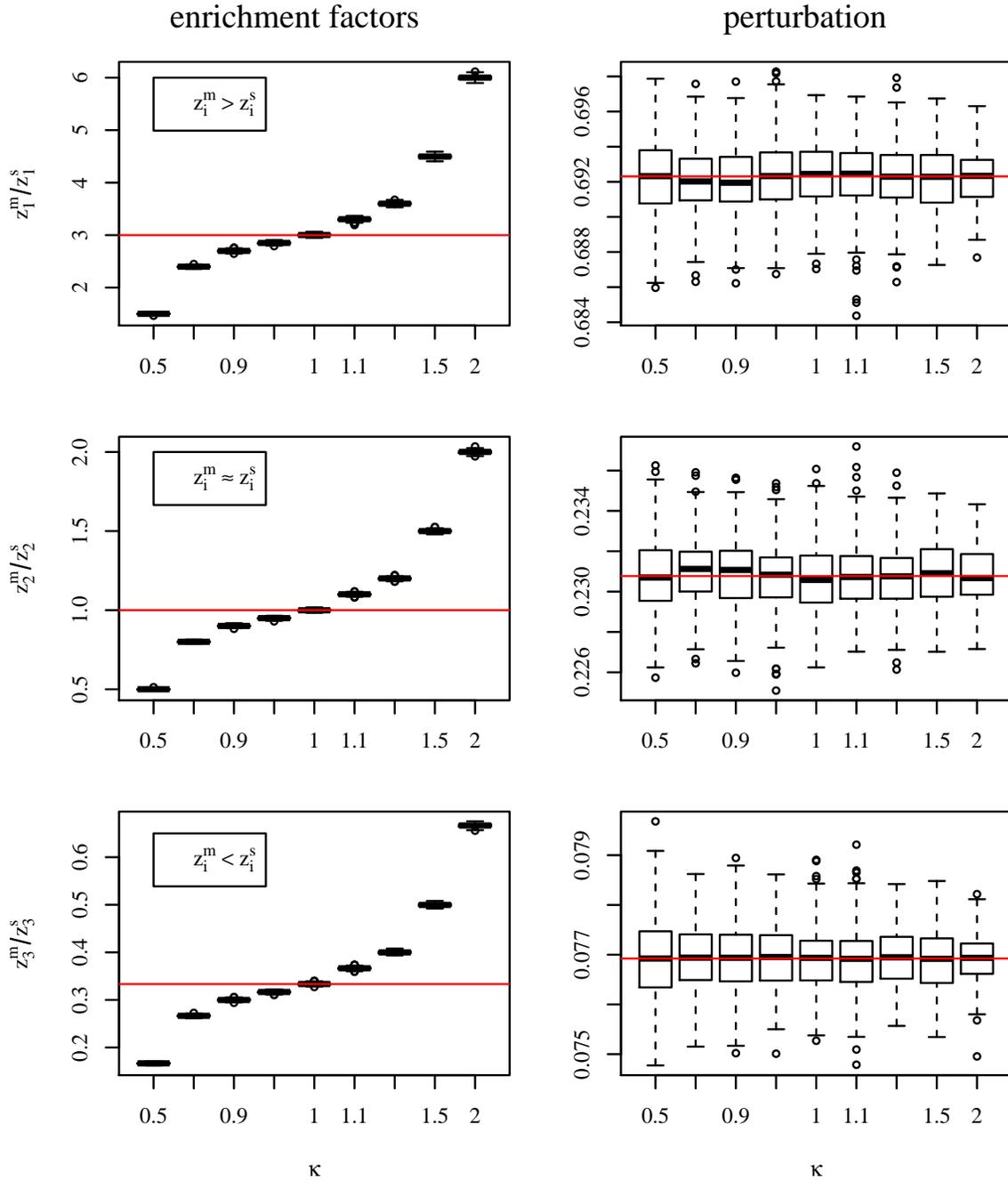
A relevant minor modification for the examples presented in this paper (Sections 5 and 6) consists of the case that the dwell times are not equal for all isotopes. If we denote the vector of dwell times as  $\boldsymbol{\omega}_0 \in \mathbb{R}_+^D$ , then it is immediate to show that the number of counts on each element class belong to the multi-Poisson family, with a perturbed intensity vector model  $\boldsymbol{\omega}_0 \oplus_+ \boldsymbol{\Lambda}(\mathbf{Z}, \boldsymbol{\lambda}_b, \boldsymbol{\theta}; t_k)$ , with perturbation on  $\mathbb{R}_+^D$ . Again, the total number of counts  $n(t_k)$  will follow a Poisson distribution, albeit with total expected counts  $\lambda_k^T = \boldsymbol{\omega}'_0 \cdot \boldsymbol{\Lambda}(\mathbf{Z}, \boldsymbol{\lambda}_b, \boldsymbol{\theta}; t_k)$ ; and the vector of counts for each element conditionally follows a multinomial distribution with probability parameter vector  $\mathbf{p}_k = \mathcal{C}[\boldsymbol{\omega}_0 \oplus_+ \boldsymbol{\Lambda}(\mathbf{Z}, \boldsymbol{\lambda}_b, \boldsymbol{\theta}; t_k)]$ .

### **A.2.2 Implications for the estimation of GLMs**

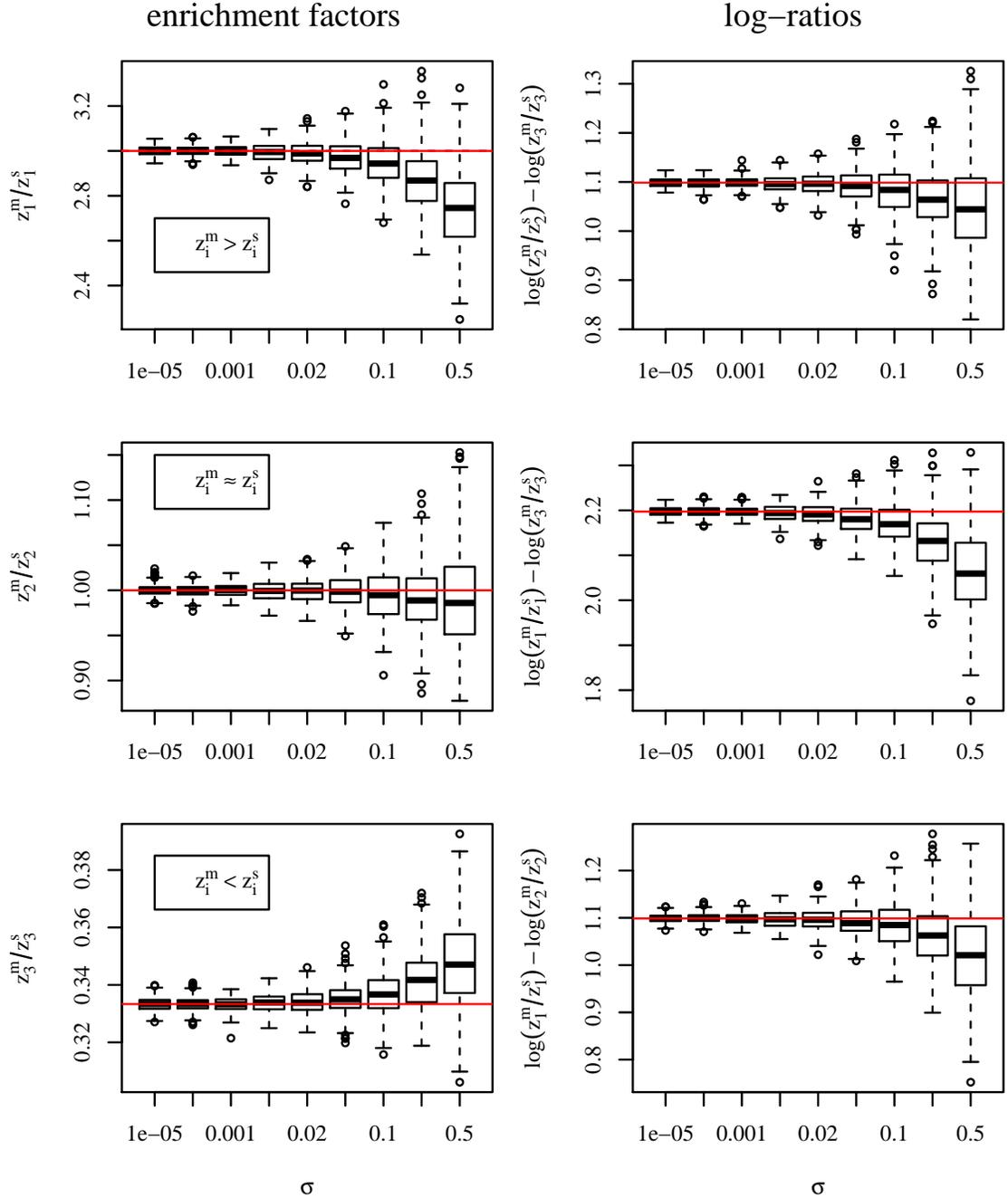
The methods presented in the main part of this contribution were characterized by an identity link function, a requirement of the additive nature of the background and the signal. If this condition is removed, then the class of models can be extended to models with logarithmic link function (actually, the canonical choice of Poisson GLMs). This setting would be suitable to consider the multiplicative models mentioned in section A.2.1. The same structure as in Section 4 would then be used, namely a calibration phase in which all parameters would be estimated with a GLM; and a prediction phase in which another GLM with an offset (equal to the multiplicative background) would be used to estimate the unknown composition of the samples.



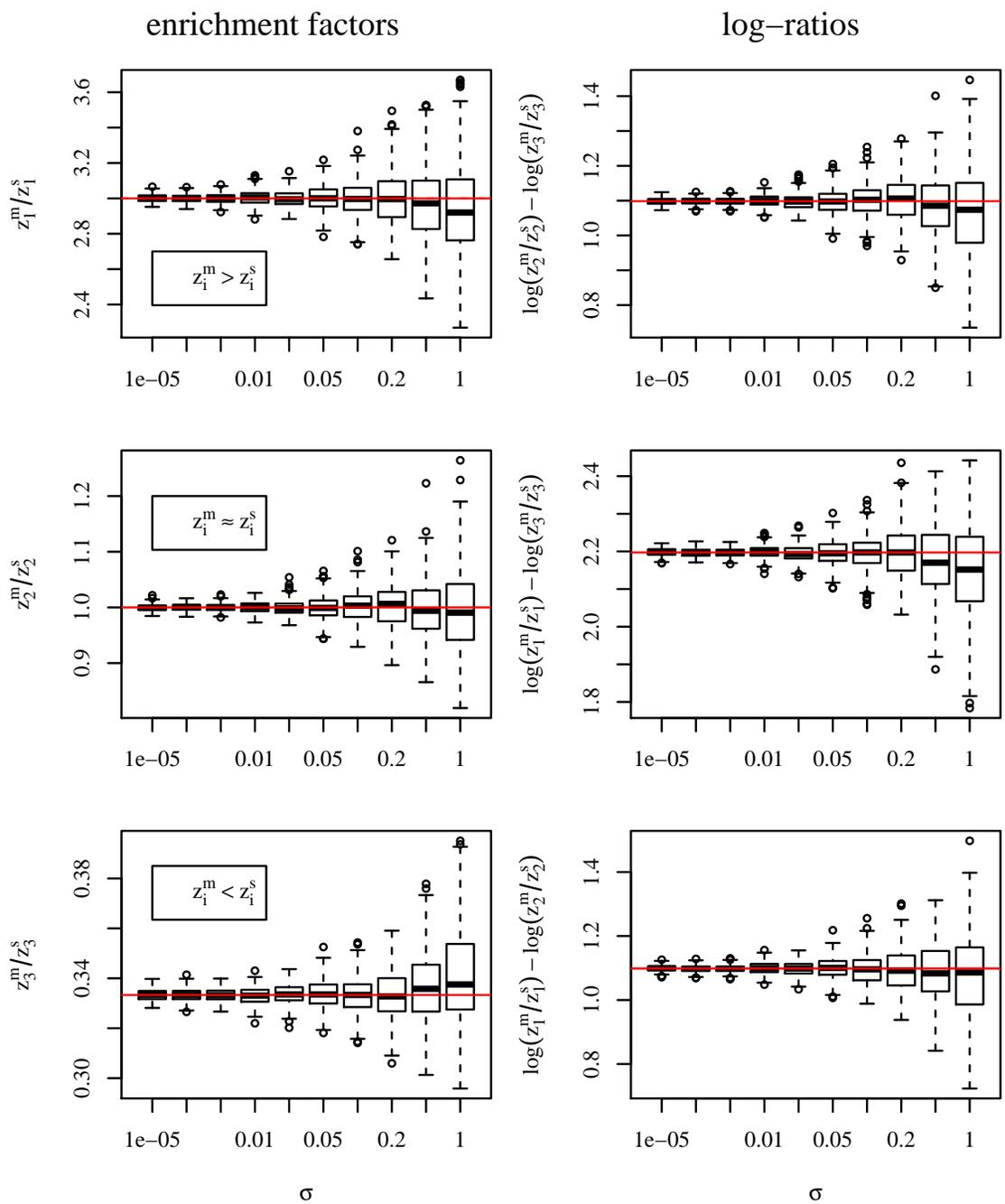
**Figure 1:** Boxplots of enrichment factors obtained for: (left) varying lengths of the measuring window; (right) several values of  $\alpha$ , with background level  $\lambda_b = [1, \alpha, \alpha]$  and sensitivity  $\lambda = 100[\alpha, 1, \alpha]$ ; in the middle right plot the X axis reports  $\alpha/20$  which corresponds to the percentage of expected counts coming from the background compared with the total expected counts; on the contrary, the X axis upper and lower right plots directly reports  $\alpha$ .



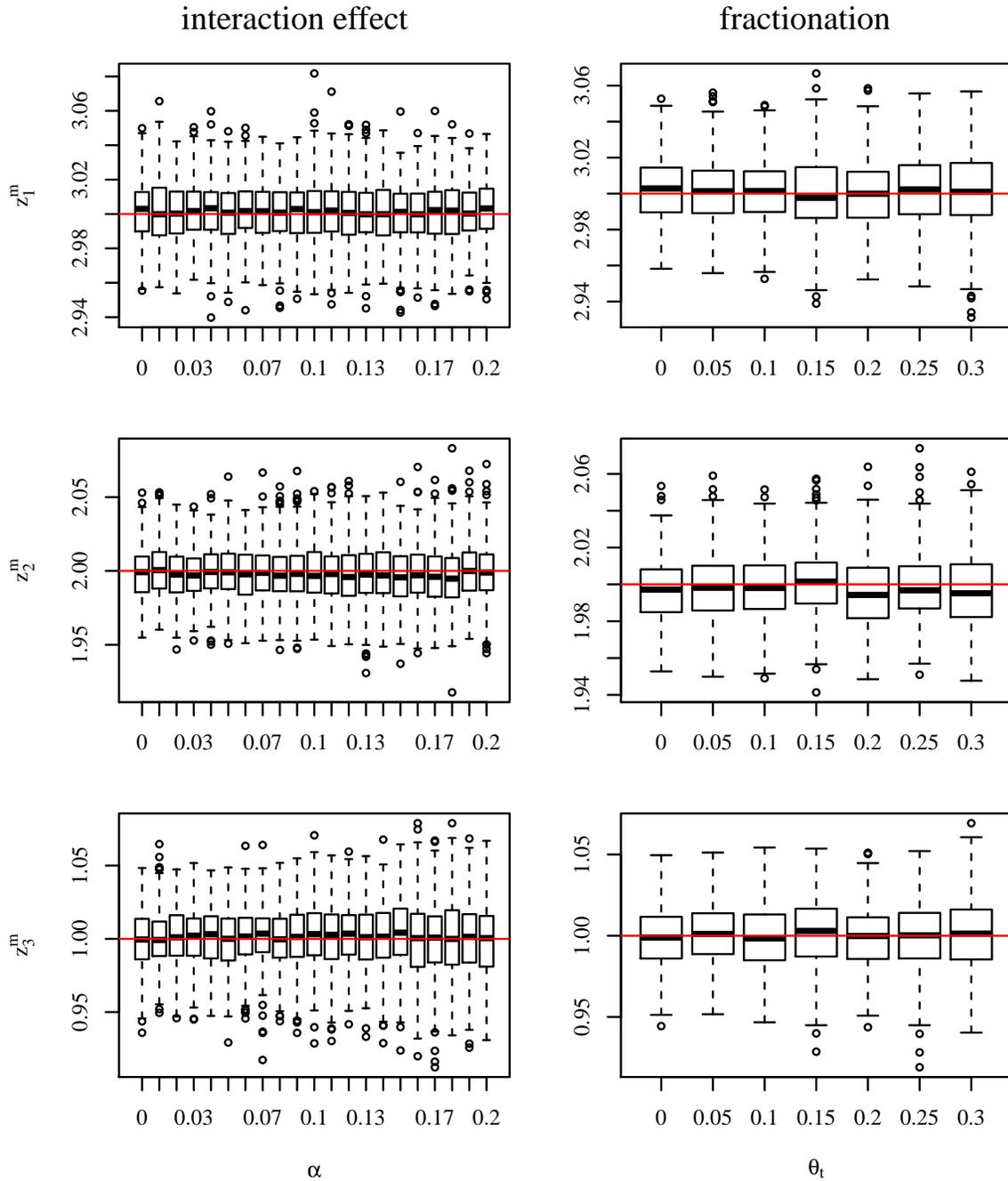
**Figure 2:** Boxplots of enrichment factors and perturbations obtained for several ratios of ablated mass per unit of time from the sample and from the standard, denoted  $\kappa$ .



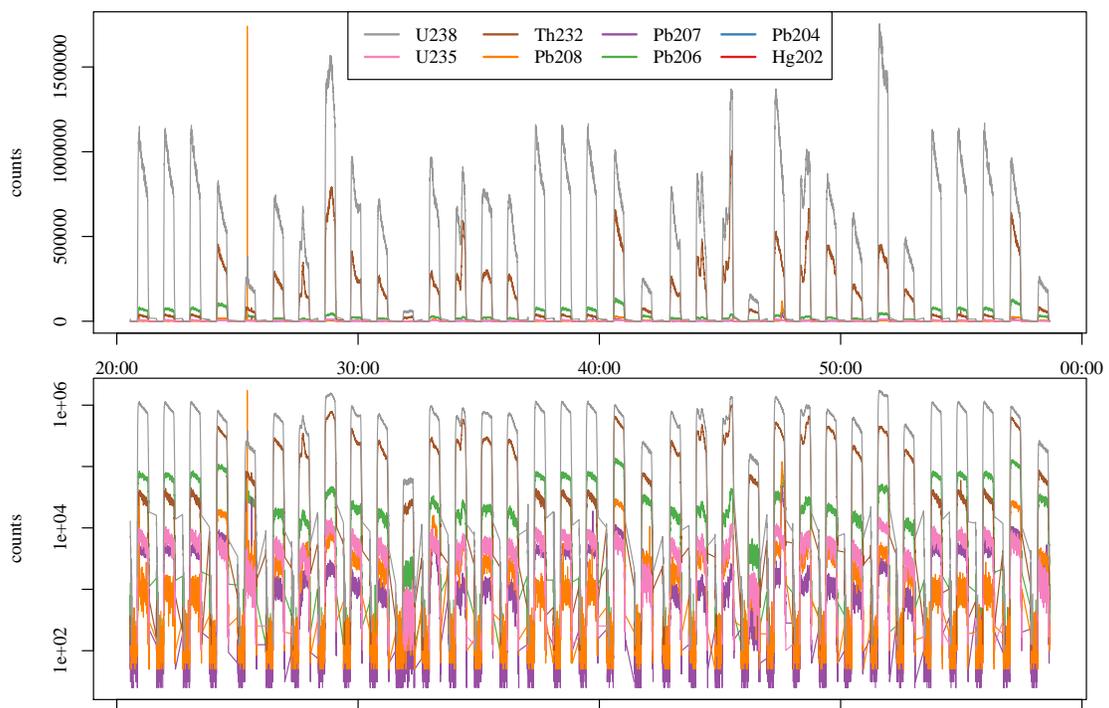
**Figure 3:** Boxplots of enrichment factors and of their logratios obtained for several variability levels  $\sigma$  of the composition of the standard, which nominal value equal is taken as the closed geometric mean.



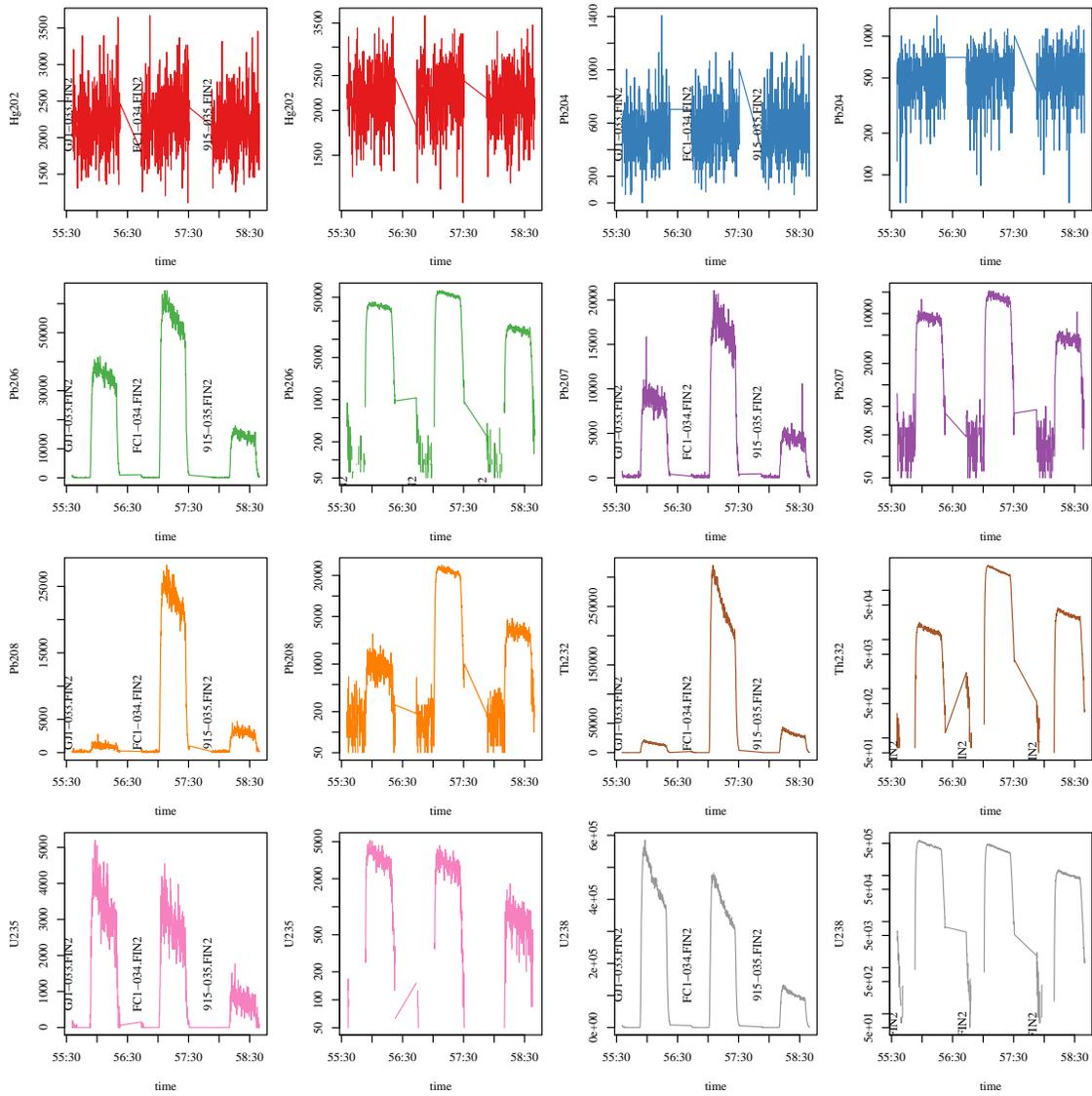
**Figure 4:** Boxplots of enrichment factors and of their logratios obtained for several variability levels  $\sigma$  of the composition of the standard, which nominal value equal is taken as the arithmetic mean.



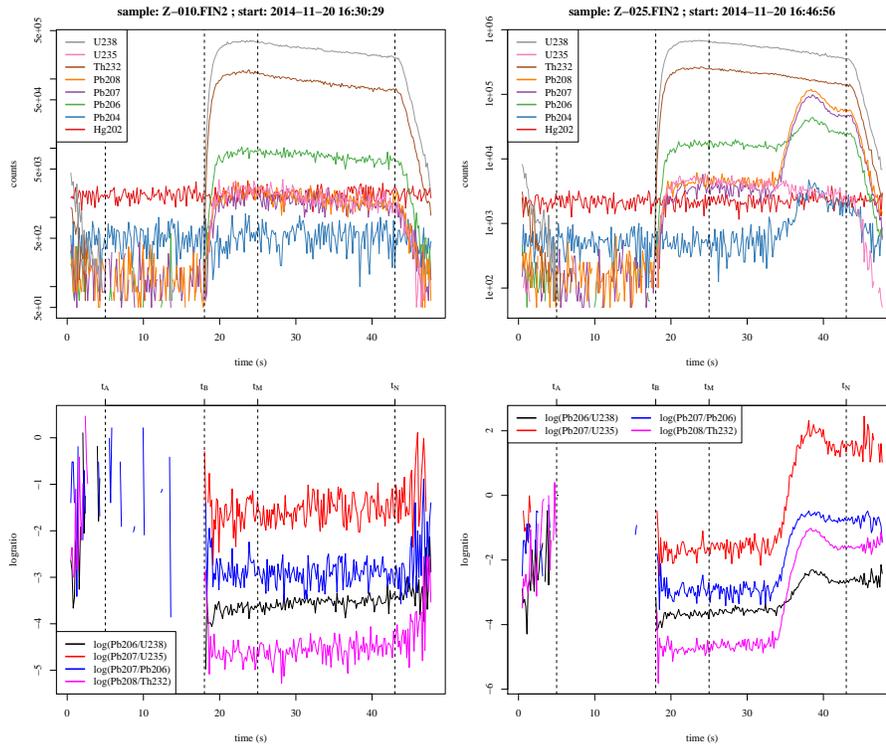
**Figure 5:** Boxplots of estimated compositions obtained: (left) with different levels of interaction; (right) with different intensities of fractionation. See text for details.



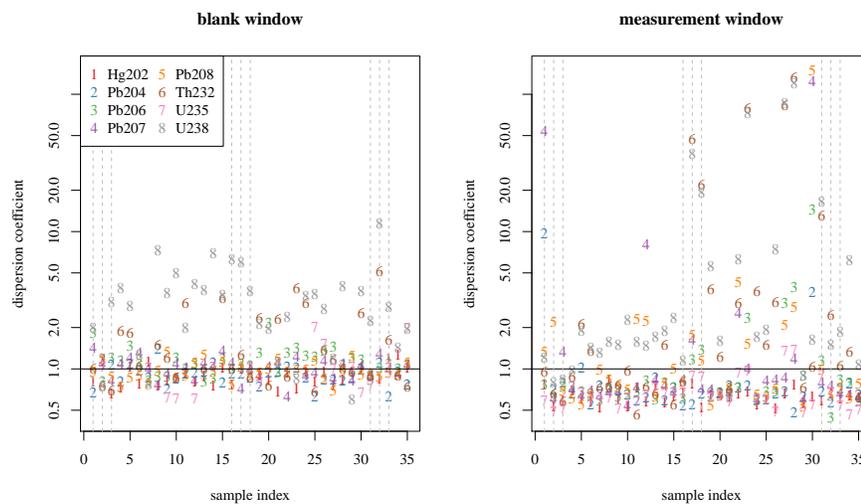
**Figure 6:** Time series of counts obtained for each component considered, in raw scale (top) and in log scale (bottom), for 35 analytes. The X axis reports the time at which each measurement was obtained.



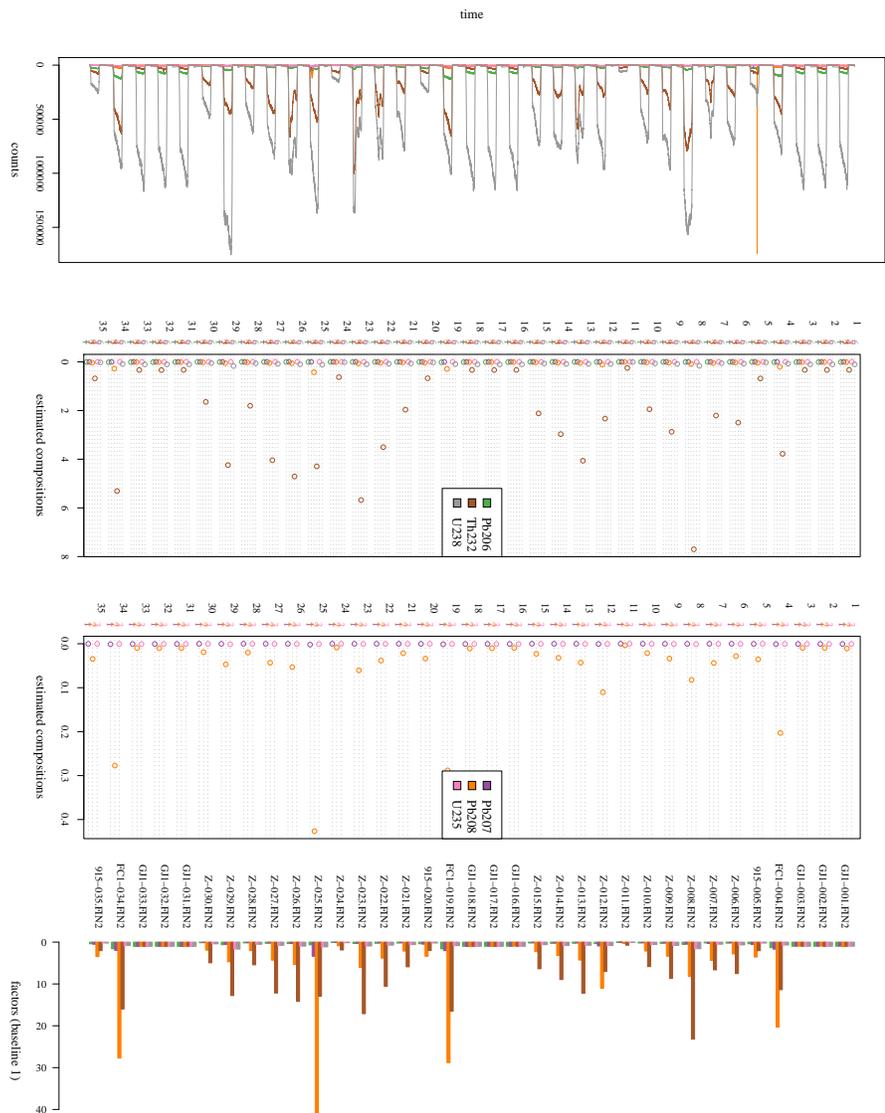
**Figure 7:** Time series of counts obtained for each component considered, in raw scale and in log scale, for the last three samples. In particular, note that the variability of the data has neither a pure additive nor a pure multiplicative structure.



**Figure 8:** Time series of counts obtained for each component considered, for a well-behaved sample (left) and for a sample showing clear zonation (right). Upper plots show the counts (in log scale) while lower plots show the naive logratios relevant for geochronological calculations. Note as well the dashed vertical lines, showing the windows for background (between the first two lines) and measurement (between the last two lines).



**Figure 9:** Dispersion coefficients of the background and measurement windows for each sample. Vertical dashed lines mark the standards.



**Figure 10:** Results of the analysis of counts for each isotope: (1) original data, (2) results for the major components, (3) results for the minor components, and (4) enrichment factors with respect to standard GJ1.

