Analysis of lead isotopic ratios of glass objects with the aim of comparing them for forensic purposes

Agnieszka Martyna [a], Knut-Endre Sjastad [b], Grzegorz Zadora [c,*], Daniel Ramos [d]

[a] Faculty of Chemistry, Jagiellonian University, Krakow, Poland
[b] National Criminal Investigation Service, Oslo, Norway
[c] Institute of Forensic Research, Westerplatte 9, 31-033 Krakow, Poland
[d] Institute of Forensic Science and Security (ICFS), and ATVS-Biometric Recognition Group (ATVS), Universidad Autonoma de Madrid, Spain

ARTICLE INFO

ABSTRACT

This paper presents the possibilities of applying the likelihood ratio (LR) approach for the comparison problem to the data collected as a result of the Isotope Ratio Mass Spectrometry (IRMS) analysis targeted at lead (Pb)-isotope ratios. The assessment of the applied LR models performance was conducted by an Empirical Cross Entropy (ECE) approach. 35 glass samples were subjected to IRMS analysis and were described by Pb-isotope ratios: $^{208}Pb/^{204}Pb$, $^{207}Pb/^{204}Pb$, $^{206}Pb/^{204}Pb$, $^{208}Pb/^{206}Pb$, and $^{207}Pb/^{206}Pb$. Univariate and bivariate LR computations were performed, assuming normally distributed data subjected or not to a logarithmic transformation. Principal Component Analysis (PCA) was employed for creating orthogonal variables to propose an alternative LR model. It was found that the application of variable $^{208}Pb/^{204}Pb$ seems to be promising as it delivers one of the lowest percentages of false positive and false negative rates as well as being the only variable for which an ECE plot gave satisfactory results.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Glass fragments are a commonly encountered type of evidence in the forensic field. They occur in such events as vehicle collisions, burglaries, robberies and fights. The feature that makes glass valuable for forensic experts is that it can disintegrate into small fragments, which may be unnoticeably carried on clothes and transferred by the event participants [1].

Due to the fact that glass traces are usually of linear dimensions less than 0.5 mm, it is necessary to apply some analytical methods appropriate for the determination of physicochemical data of traces. These include the widely used methods among the forensic experts: Glass Refractive Index Measurement (GRIM) [2–4] and Scanning Electron Microscopy coupled with an Energy Dispersive X-ray detector (SEM–EDX) [3,5–8]. Some other techniques include Laser Ablation Inductively Coupled Plasma Mass Spectrometry (LA–ICP-MS) [9,10], μ-X-ray Fluorescence [11] and Laser Induced Breakdown Spectroscopy (LIBS) [12–14].

Lead is a trace element in non-leaded glass with concentrations varying from 0 to 5000 ppm [15]. The lead isotope ratio (IR) varies over geographical areas due to different biological and geological processes. Owing to geographical variation of isotopic ratio, proper use of IR-determination with a high degree of accuracy pinpoints the source of a given object (e.g. window panes, containers, etc.). The natural variation of lead isotopic ratios over different regions may also be useful for solving the comparison problem of glass fragments for forensic purposes.

In this study, the lead isotope ratios [15] in glass were determined by MC–ICP-MS, which is a multi-collector mass spectrometer combined with inductively coupled plasma ion source. It is a newly applied technique to forensic glass analysis. In this technique the ions generated in the ICP source are transferred to the mass spectrometer and separated according to their mass to charge ratio. Such ion beams are directed into a set of collectors, which generate voltages according to the ion energies. The ratios of the isotopes are then computed based on the generated voltages. So the use of Isotope Ratio Mass Spectrometry (IRMS) analysis has developed to be one of the most promising methods for the determination of an object's origin [16].

The comparison is one of the most commonly encountered problems in the analysis of glass objects for forensic purposes [17]. It involves the comparison of physicochemical data obtained as a result of the glass analysis (such as refractive index and/or elemental composition) performed on recovered glass fragments (e.g. from suspect's clothes) and on control glass fragments (e.g. collected from a broken window at a scene of crime).

The question of interest from the forensic point of view within the comparison problem is: What is the value of the evidence of

---

these measurements in relation to the propositions that the two samples of glass fragments did, or did not, come from the same source? The answer requires knowledge about:

a) similarity of the data obtained for compared glass fragments,
b) the possible sources of uncertainty, which include:
   (i) the variation of measurements within recovered and control glass fragments,
   (ii) the variation of measurements between objects in the glass population,
c) information about the rarity of measured physicochemical data. For instance, one would expect refractive index (RI) values from different locations of the same glass object to be very similar. However, equally similar RI values could also be observed from different glass items. Without a wider context, it is not possible to ascribe meaning to the observed similarity. Therefore, inferences about the source of glass fragments made purely on the basis of similarity of measurements are incomplete. Information about the rarity of a determined RI value has to be taken into account. Intuition suggests that the value of the evidence in support of the proposition that the recovered glass fragments and the control sample have a common origin is greater when the determined RI values are similar and rare in the relevant population, than when the RI values are equally similar but common in the same population,
d) existing correlation between variables in the case of multi-dimensional data.

The evidential value of physicochemical data, taking into account all the mentioned requirements stemming from the forensic practice of glass fragments analysis, could be assessed by the application of the likelihood ratio approach (LR), a well-documented measure of evidential value in the forensic sciences [2,5,8,17–21]. It provides the possibility of comparing the data describing the compared glass fragments, being the evidence (E), in the context of two contrasting hypotheses. The first one, referred to as the so-called prosecutor's hypothesis, θ_p, is the proposition that the compared glass fragments come from the same object, while the second one, termed the defence hypothesis, θ_d, is the proposition that the glass fragments have different origins. The LR is defined by the following equation:

$$LR = \frac{\Pr(E|\theta_p)}{\Pr(E|\theta_d)} \tag{1}$$

In the case of continuous type data $\Pr(\cdot)$ are substituted by suitable probability density functions $f(\cdot)$. Values of LR above 1 support the prosecutor's hypothesis, while values of LR below 1 support the defence hypothesis. The values equal to 1 support neither of the hypotheses. The higher (lower) the value of LR, the stronger the support for the prosecutor's (defence) hypothesis.

The likelihood ratio approach is based on *Bayes' theorem*.

$$\frac{\Pr(\theta_p)}{\Pr(\theta_d)}\frac{\Pr(E|\theta_p)}{\Pr(E|\theta_d)} = \frac{\Pr(\theta_p)}{\Pr(\theta_d)}LR = \frac{\Pr(\theta_p|E)}{\Pr(\theta_d|E)} \tag{2}$$

$\Pr(\theta_p)$ and $\Pr(\theta_d)$ are called *a priori* probabilities and their quotient is called *a priori odds*. Their estimation lies within the competence of the fact finder (judge, prosecutor or police) expressing their opinions about the considered hypotheses before the evidence is analysed, thus without having any further information in this matter. It is the duty of a fact finder, police or court to determine whether the objects are deemed to stem from the same or different sources and this decision is based on the results expressed in the form of conditional probabilities—$\Pr(\theta_p|E)$ and $\Pr(\theta_d|E)$, namely *a posteriori* probabilities and their quotient is called *a posteriori odds*. These could be estimated by taking into account *a priori odds* and the information delivered by the forensic expert in the form of LR.

Therefore, it is important that the method used for the evidence evaluation delivers strong support for the correct hypothesis, i.e. LR $\gg$ 1 when θ_p is correct and LR $\ll$ 1 when θ_d is correct. Additionally, it is desired that if an incorrect hypothesis is supported by LR value (i.e. LR > 1 when θ_d is correct and LR < 1 when θ_p is correct) then LR value should be close to 1 as it allows to deliver only weak misleading evidence. Roughly speaking, according to Eq. (2) it seems to be of great importance to obtain LR values that do not provide misleading information for the court or police. This implies the need for evaluation of performance of the applied methodology of data evaluation, which could be made by the application of an Empirical Cross Entropy (ECE) approach [6,21,22].

The study focuses on the application of the likelihood ratio approach for the comparison problem to the data collected as a result of the IRMS aimed at Pb-isotope ratios obtainment as described in [15]. LR models for uni- and multi-dimensional data were computed, differing in the way of data preparation. The scope of the paper is also targeted at the assessment of the applied models performance by an ECE approach [6,21,22].

2. Materials and methods

2.1. Glass database

35 glass fragments from the National Criminal Investigation Services reference collection of street samples were selected. These samples are of unknown origin, but they are all of different sources. The glass fragments were subjected to the Isotope Ratio Mass Spectrometry analysis. For each glass fragment three measurements were performed according to normal protocols described in detail in [15] delivering the information on the following lead isotope ratios: $^{208}Pb/^{204}Pb$, $^{207}Pb/^{204}Pb$, $^{206}Pb/^{204}Pb$, $^{208}Pb/^{206}Pb$ and $^{207}Pb/^{206}Pb$.

2.2. Instrumental

Analyses of lead-isotope ratios in glass were performed using a Nu Plasma magnetic sector, multi-collector inductively coupled plasma source mass spectrometer (MC–ICP-MS), Nu Instruments, Wrexham, UK, at the Department of Earth Sciences, University of Oslo. Ion beams of mass 208–202 were measured by static multicollection in Faraday cups.

The MC–ICP-MS technique requires some sample preparation. 35 glass fragments were crushed to a fine powder in an agate mortar. The powder was subsequently washed in 2% HNO_3 for half an hour in an ultrasonic bath and finally rinsed with water. Then the glass powder was dried in filtered air for a minimum of 2 h in 60 °C. Three samples of 5–100 mg were created from the glass powder originating from a single object. Each sample was dissolved in 5–7 ml of HF, HCl, and HNO_3 acid mixture (2:1:1, v/v/v) for 2 h. After adding portions of HCl and HBr and evaporating to hard dryness, the samples were centrifuged with 1 ml of 0.8 M HBr and introduced into the ion-exchange column (a brominated resin, type AG1-X8, BioRad) to remove matrix elements (including those giving isobaric overlap with lead). Stripping with 1 and 2 ml of subsequent portions of HNO_3 acid released only lead. The details of the sample preparation, column preparation, and conditioning are provided in [15]. Dissolved samples containing lead ions were then transferred into the MC–ICP-MS system and were analysed for the lead isotopic ratios.

2.3. Likelihood ratio

Details of the distributional assumptions used for the likelihood ratio computation and the required variance estimates as well as LR expressions are given in this section.

The prosecution proposition, θ_p, states that the control and recovered means, $\overline{\mathbf{y}}_1$ and $\overline{\mathbf{y}}_2$ respectively, come from the same object, while the defence proposition, θ_d, states that they come from different objects.

Aitken and Lucy [18] gave various expressions for the numerator and denominator of the likelihood ratio (Eq. (1)) in the case of evaluating the continuous type data when between-object distribution could be assumed normal or not. The LR model presented below was applied in calculations based on the assumption that between-object distribution of physicochemical data of 35 measured glass samples is normal, as observed in Q–Q plots in Section 3.1. It should be mentioned that if between-object distribution could not be estimated by a normal distribution, then a probability density function could be estimated using Gaussian kernels (for details see [18]).

The numerator and the denominator of LR formulae, when between-object distribution is assumed normal, are respectively given as [18]

$$(2\pi)^{-p} \left| \frac{\mathbf{U}}{n_1} + \frac{\mathbf{U}}{n_2} \right|^{-1/2} \exp\left\{ -\frac{1}{2} (\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2)^T \left(\frac{\mathbf{U}}{n_1} + \frac{\mathbf{U}}{n_2} \right)^{-1} (\overline{\mathbf{y}}_1 - \overline{\mathbf{y}}_2) \right\}$$
$$\times \left| \frac{\mathbf{U}}{n_1 + n_2} + \mathbf{C} \right|^{-1/2} \exp\left\{ -\frac{1}{2} (\overline{\mathbf{y}}^* - \boldsymbol{\mu})^T \left(\frac{\mathbf{U}}{n_1 + n_2} + \mathbf{C} \right)^{-1} (\overline{\mathbf{y}}^* - \boldsymbol{\mu}) \right\} \quad (3)$$

$$(2\pi)^{-p} \left| \frac{\mathbf{U}}{n_1} + \mathbf{C} \right|^{-1/2} \exp\left\{ -\frac{1}{2} (\overline{\mathbf{y}}_1 - \boldsymbol{\mu})^T \left(\frac{\mathbf{U}}{n_1} + \mathbf{C} \right)^{-1} (\overline{\mathbf{y}}_1 - \boldsymbol{\mu}) \right\}$$
$$\times \left| \frac{\mathbf{U}}{n_2} + \mathbf{C} \right|^{-1/2} \exp\left\{ -\frac{1}{2} (\overline{\mathbf{y}}_2 - \boldsymbol{\mu})^T \left(\frac{\mathbf{U}}{n_2} + \mathbf{C} \right)^{-1} (\overline{\mathbf{y}}_2 - \boldsymbol{\mu}) \right\} \quad (4)$$

Within-object variance–covariane estimate $\mathbf{U}$ in the case of multivariate data is expressed as

$$\mathbf{U} = \frac{\mathbf{S_w}}{m(n-1)},$$

with $\mathbf{S_w}$ expressed as

$$\mathbf{S_w} = \sum_{i=1}^{m} \sum_{j=1}^{n} (\mathbf{x_{ij}} - \overline{\mathbf{x}}_i)(\mathbf{x_{ij}} - \overline{\mathbf{x}}_i)^T$$

where

$\mathbf{x_{ij}}$—a vector of values of p variables obtained in j-th measurement for the i-th object,

$\overline{\mathbf{x}}_i$—a vector of means of p variables calculated using n measurements for the i-th object in the database: $\overline{\mathbf{x}}_i = \frac{1}{n} \sum_{j=1}^{n} \mathbf{x_{ij}}$.

Between-object variance–covariance estimate $\mathbf{C}$ in the case of multivariate data can be expressed as follows:

$$\mathbf{C} = \frac{\mathbf{S}^*}{m-1} - \frac{\mathbf{S_w}}{nm(n-1)},$$

with $\mathbf{S}^*$ expressed as

$$\mathbf{S}^* = \sum_{i=1}^{m} (\overline{\mathbf{x}}_i - \overline{\mathbf{x}})(\overline{\mathbf{x}}_i - \overline{\mathbf{x}})^T$$

$\overline{\mathbf{x}}$—a vector of means of p variables calculated using n measurements for m objects in the database $\overline{\mathbf{x}} = 1/mn \sum_{i=1}^{m} \sum_{j=1}^{n} \mathbf{x_{ij}}$.

Moreover in Eqs. (3) and (4) $\overline{\mathbf{y}}_1$, $\overline{\mathbf{y}}_2$, $\boldsymbol{\mu}$ and $\overline{\mathbf{y}}^*$ are defined as

$\boldsymbol{\mu}$—a vector of overall means of p variables estimated using n measurements for m objects from the database $\boldsymbol{\mu} = \overline{\mathbf{x}}$,

$\overline{\mathbf{y}}_1$—a vector of means of p variables calculated using n_1 measurements performed on the control object $\overline{\mathbf{y}}_1 = 1/n_1 \sum_{j=1}^{n_1} \mathbf{y_{1j}}$,

$\overline{\mathbf{y}}_2$—a vector of means of p variables calculated using n_2 measurements performed on the recovered object $\overline{\mathbf{y}}_2 = 1/n_2 \sum_{j=1}^{n_2} \mathbf{y_{2j}}$, and

$$\overline{\mathbf{y}}^* = \frac{n_1 \overline{\mathbf{y}}_1 + n_2 \overline{\mathbf{y}}_2}{n_1 + n_2}.$$

In the case of analysis of univariate data ($p=1$) vectors and matrices become suitable scalars.

2.4. Experimental protocol

A number of experiments were undertaken in order to evaluate the performance of the proposed methods. These included so-called false positive and false negative rates. These performance metrics are typically defined as the percentage of likelihood ratio values that would lead to an incorrect decision if the decision threshold is set to LR=1. These include the rates of LR values lower than 1 under the prosecution hypothesis, and the rates of LR values greater than 1 under the defence hypothesis. These values as defined here have also been called *rates of misleading evidence*, since LR values presenting such behaviour will provide misleading results. The need for controlling the levels of misleading information arises from the forensic scientist's practice. False positive answers occur when fragments coming from different glass objects are thought to be from the same source. False negative answers appear when the calculations indicate different origins of the compared objects which, in fact, stem from one glass object.

Two kinds of experiments were performed with the aim of establishing the percentage of false positive and false negative answers:

a) the rates of false positive answers were evaluated by conducting LR calculations by comparing results obtained for two different objects, which gives $N_d = \binom{35}{2} = 35!/2!(35-2)! = (34 \times 35/2) = 595$ comparisons. The desirable answer was LR < 1; hence each value of LR > 1 was considered a false positive answer. Rates of false positive answers should be especially observed in the forensic sphere due to the fact that they may lead to serious legal consequences. Imagine a situation when two compared glass samples are reported to come from the same object on the basis of obtained analytical results followed by LR calculations, whereas in fact they come from different objects. The accused person can have serious legal consequences, despite being innocent.

b) the rates of false negative answers were estimated by forming two samples, control and recovered, from observations for a single object in a way that the recovered sample was created from one of the three available measurements ($n_2=1$), hence the remaining two formed a control sample ($n_1=2$). The number of performed comparisons was $N_p=35$. The reason for creating the recovered sample always consisting of less measurements than the control one underlies the assumption that the forensic experts in most cases has at their disposal more control material than recovered. The desirable response was LR > 1. Each value of LR < 1 was considered a false negative answer.

2.5. Empirical Cross Entropy

False positive and false negative rates present limitations as measures of performance. They only consider values that are misleading according to the threshold at LR=1, but they do not consider the magnitude of a misleading LR value. For instance, a

LR value computed under the defence hypothesis would be much worse if its value was LR=1000 than if it was LR=2. However, these will compute as a single false positive answer in both cases, and no distinction will be made.

Empirical Cross Entropy (ECE) overcomes all those problems. ECE was proposed as an assessment metric of the performance of the evidence evaluation methods such as the likelihood ratio model, which is thoroughly described in this study [23,24]. ECE is a framework derived from information theory firstly presented in the 1950s.

ECE, being a measure of information, is aimed at assessing the performance of a statistic, such as the aforementioned likelihood ratio, with respect to correctness of decision making. It was mentioned that the higher (lower) the LR values, the greater the support for the θ_p (θ_d). Thus for a forensic expert the best method for evidence evaluation is the one delivering the extreme values supporting the correct hypothesis. Roughly speaking, according to Eq. (2) it seems to be of great importance to obtain such LR values as they do not provide misleading information for the court or police. This implies the need for measuring the performance of the applied LR methodology of data evaluation.

The Empirical Cross Entropy approach is related to the strictly proper scoring rules. In the study, the strictly proper scoring rules are expressed as logarithmic scoring rules (LS) in the following way:

a) if θ_p is true: $-\log_2(\Pr(\theta_p|E))$,
b) if θ_d is true: $-\log_2(\Pr(\theta_d|E))$.

The logarithmic scoring rule is illustrated in Fig. 1. In [24] the overall measure of goodness of a forecaster is defined as the average value of a strictly proper scoring rule over many different forecasts, which are expressed by posterior probabilities. For instance, for the logarithmic scoring rule, this mean value could be expressed by

$$LS = -\frac{1}{N_p}\sum_{i \in p}\log_2(\Pr(\theta_p|e_i)) - \frac{1}{N_d}\sum_{j \in d}\log_2(\Pr(\theta_d|e_j))$$

where p and d refer to the comparisons of the objects having the same and different origins respectively, N_p and N_d refer to the number of the comparisons made under each of the considered propositions θ_p and θ_d (see Section 2.4). This average value, LS, can be viewed as an overall loss. The ECE, is the proposed measure of goodness as a variant of LS, and is expressed as follows:

$$ECE = -\frac{\Pr(\theta_p)}{N_p}\sum_{i \in p}\log_2(\Pr(\theta_p|e_i)) - \frac{\Pr(\theta_d)}{N_d}\sum_{j \in d}\log_2(\Pr(\theta_d|e_j)) \quad (6)$$

Taking into account Eq. (2) it can be seen that ECE could be expressed as

$$ECE = \frac{\Pr(\theta_p)}{N_p}\sum_{i \in p}\log_2\left(1 + \frac{\Pr(\theta_d)}{LR_i\Pr(\theta_p)}\right) + \frac{\Pr(\theta_d)}{N_d}\sum_{j \in d}\log_2\left(1 + \frac{LR_j\Pr(\theta_p)}{\Pr(\theta_d)}\right)$$
$$(7)$$

The *a priori* probabilities $\Pr(\theta_p)$ and $\Pr(\theta_d)$ are not generally known in the forensic evaluation of the evidence, because they depend on various information sources: witnesses, police investigations, other evidence, etc. Because ECE cannot be computed if *a priori* probabilities are not known, the adopted solution is to plot ECE for a set of all possible *a priori* probability quotients, further referred to as *a priori* odds and expressed as its logarithm $\log_{10}$ Odds(θ). The details about the derivation and interpretation
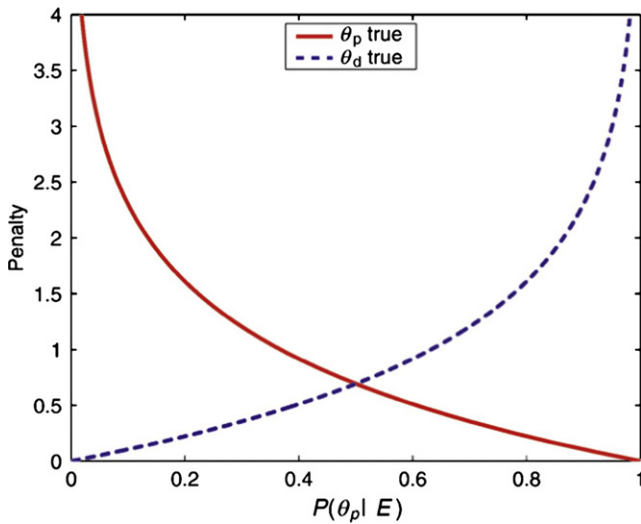

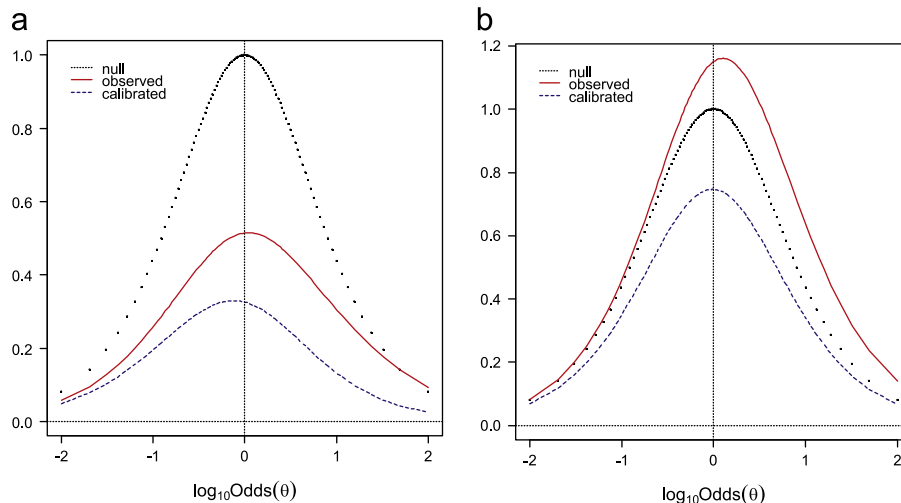**Fig. 1.** Logarithmic scoring rule (for details, see Section 2.5).

Fig. 2. Examples of ECE plots (for details see Section 2.5): (a) for a reasonably working LR model, (b) for an incorrectly working LR model (here, a model is assumed to be a combination of a particular LR equation and variables used for evidence evaluation). (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

of ECE can be found in [22]. That leads to the so-called ECE plot, which can be seen in Fig. 2. The ECE plot consists of 3 curves [21,22]:

a) the solid (red) curve (named *observed* in Fig. 2)—represents the ECE (average information loss) values calculated using the statistic evidence evaluation method under analysis (see Eq. (7)).
b) the dashed (blue) curve (named *calibrated* in Fig. 2)—represents the calibrated ECE values obtained from computing ECE for the experimental LR values (Eqs. (3) and (4)) transformed using Pool Adjacent Violators algorithm (PAV) [23]. The discriminating power of the calibrated method is unaltered, which means that it represents the LR values set of the best performance of all other LR sets offering the same discriminating power. Therefore, the observed differences between the calibrated method curve and the ECE curve for the experimental LR set are due to the problems with the calibration of the applied evidence evaluation method.
c) the dotted (black) curve (named *null* in Fig. 2)—represents the performance of a method always providing LR=1. Therefore within this method (referred to as a null method) a curve is always the same for different sets of experimental LR values. This method is equivalent to assigning no value to the evidence, and will be used as a reference.

The interpretation of the relative location of the ECE curve (Fig. 2) in relation to the remaining two mentioned earlier illustrates the performance of the method of evidence evaluation. If the LR values of the evidence evaluation process are misleading to the fact finder, then the ECE will grow, and more information on average will be needed in order to know the true values of the hypotheses. In other words, the higher the curve (Fig. 2a), the more uncertainty remains and therefore the worse the method of choice is for interpretation of the evidence under analysis. If the curve appears to have greater values than the ones (Fig. 2b) in the neutral method, the evidence evaluation introduces more misleading information than when not evaluating the evidence at all and therefore the method within the range of occurrence of such a situation is treated as a great misuse.

3. Results and discussion

3.1. Descriptive statistics

The distributions of ten considered variables (5 lead isotope ratios and their logarithms) are shown in Fig. 3 in the form of box-plots. All the variables were subjected to a logarithmic transformation aimed at reducing differences in the orders of magnitude and, what is of greater importance, bringing the data closer to normality. Therefore, the LR calculations (Section 3.2) were conducted using both the data before and after taking the base 10 logarithms.

Q–Q plots were drawn with the aim of checking whether or not data could be estimated by normal distribution. The assumption that they are normally distributed is not unreasonably stated for most of the data (Fig. 4) and therefore such an assumption was made and consequently the authors focused on obtaining LR values with the use of the expressions assuming normality of the data (see Section 2.3).

Further data inspection indicates that a strong correlation between variables exists, which is presented in the partial correlation coefficient matrix in Table 1. The correlation coefficients are quite high, reaching almost 1 for $^{206}Pb/^{204}Pb$ and $^{207}Pb/^{206}Pb$, $^{208}Pb/^{206}Pb$ and $^{207}Pb/^{206}Pb$ as well as for $^{206}Pb/^{204}Pb$ and $^{208}Pb/^{206}Pb$. It means that information contained in one of the variables $^{206}Pb/^{204}Pb$, $^{208}Pb/^{206}Pb$ and $^{207}Pb/^{206}Pb$ provides similar information about the two remaining practically without any loss. The same values of
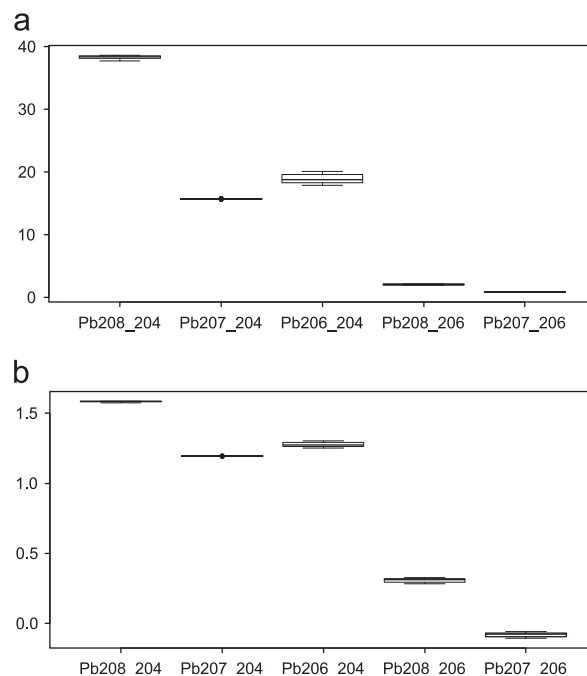

Fig. 3. Box-plots for the variables before (a) and after (b) the logarithmic transformation. Note: Boxplot lines refer to the lower quartile (Q1) and the upper quartile (Q3) and the line inside a box indicates the median (Q2). The whiskers are drawn in a way that they do not exceed 1.5 IQR (which stands for interquartile range being a measure of statistical dispersion of data expressed as Q3–Q1) and are moved to the nearest data points.

correlation coefficients (when up to two decimal places are considered) were obtained for logarithmically transformed data.

3.2. Likelihood ratio models performance

The LR computations involved univariate and multivariate problems. Univariate calculations were based on each of the considered variables: $^{208}Pb/^{204}Pb$, $^{207}Pb/^{204}Pb$, $^{206}Pb/^{204}Pb$, $^{208}Pb/^{206}Pb$ and $^{207}Pb/^{206}Pb$ and the multivariate LR calculations were in fact conducted for 10 bivariate combinations. The calculation of LR using more than 2 variables is possible but the number of samples in the database is too small to reliably estimate all parameters considered in applied LR Eqs. (3) and (4), e.g. when $p=5$ then five means five variances and ten covariances should be reliably estimated.

It is well known that tuning and test sets should be created when a particular model for data interpretation is analysed. In this study only 35 glass objects were available in the database and it was difficult to create such sets. Therefore, a jack-knife procedure was applied in order to optimally use the available database. The jack-knife procedure implies that the database, used for the estimation of parameters in the applied LR model, consists of all objects except two being actually compared in the between-objects comparisons (estimation of false positive answers) or one object in the within-object comparisons (estimation of false negative answers) (see the Section 2.4).

The results of the performed analyses, aimed at estimating false positive and negative answers levels, are presented in Table 2. Application of univariate or bivariate models allowed one to obtain, in most of cases, less than 10% of false negative answers (1–3 incorrect answers). Only a bivariate model based on variables $^{206}Pb/^{204}Pb$ and $^{208}Pb/^{206}Pb$ delivered 4 false negative answers (11.4%). Quite significant is the fact that the bivariate combinations do not lead to lower levels of false negative answers, but even make them higher. It could be easily observed from Table 2 that
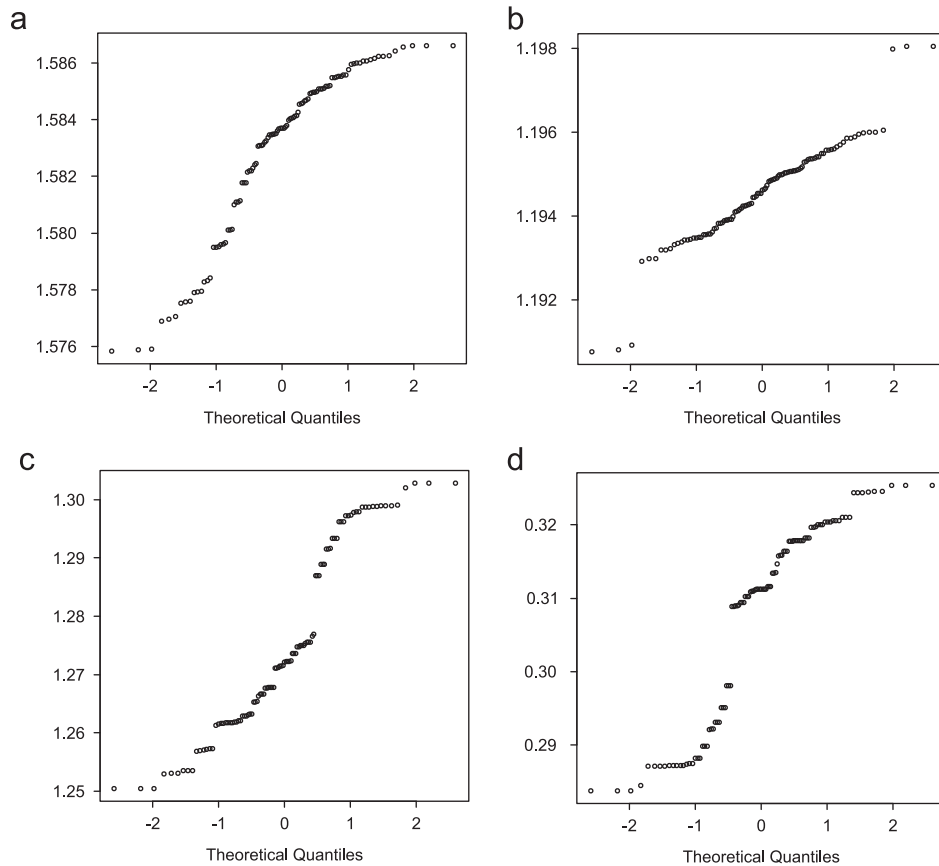
Fig. 4. Q–Q plots for the variables before and after the logarithmic transformation; (a) $\log_{10}(^{208}\text{Pb}/^{204}\text{Pb})$, (b) $\log_{10}(^{207}\text{Pb}/^{204}\text{Pb})$, (c) $\log_{10}(^{206}\text{Pb}/^{204}\text{Pb})$, (d) $\log_{10}(^{208}\text{Pb}/^{206}\text{Pb})$ (also for $\log_{10}(^{207}\text{Pb}/^{206}\text{Pb})$ similar shape of Q–Q plot was observed).

Table 1
Partial correlation coefficients matrix calculated for the variables before and after the logarithmic transformation.

	$^{208}\text{Pb}/^{204}\text{Pb}$	$^{207}\text{Pb}/^{204}\text{Pb}$	$^{206}\text{Pb}/^{204}\text{Pb}$	$^{208}\text{Pb}/^{206}\text{Pb}$	$^{207}\text{Pb}/^{206}\text{Pb}$
$^{208}\text{Pb}/^{204}\text{Pb}$	1.00	0.78	0.85	−0.79	−0.84
$^{207}\text{Pb}/^{204}\text{Pb}$		1.00	0.69	−0.64	−0.65
$^{206}\text{Pb}/^{204}\text{Pb}$			1.00	−0.99	−1.00
$^{208}\text{Pb}/^{206}\text{Pb}$				1.00	0.99
$^{207}\text{Pb}/^{206}\text{Pb}$					1.00

Table 2
Results of experiments which aim was to establish the number of false positive and false negative answers.

Variable(s)	Raw data		Log transformed data	
	False positive	False negative	False positive	False negative
$^{208}\text{Pb}/^{204}\text{Pb}$	16.13[a] (96)[b]	5.71 (2)	16.13 (96)	5.71 (2)
$^{207}\text{Pb}/^{204}\text{Pb}$	15.97 (95)	5.71 (2)	15.97 (95)	5.71 (2)
$^{206}\text{Pb}/^{204}\text{Pb}$	21.01 (125)	2.86 (1)	20.67 (123)	2.86 (1)
$^{208}\text{Pb}/^{206}\text{Pb}$	22.69 (135)	2.86 (1)	23.87 (142)	2.86 (1)
$^{207}\text{Pb}/^{206}\text{Pb}$	20.34 (121)	2.86 (1)	20.67 (123)	2.86 (1)
$^{208}\text{Pb}/^{204}\text{Pb}$ and $^{207}\text{Pb}/^{204}\text{Pb}$	7.23 (43)	8.57 (3)	7.23 (43)	8.57 (3)
$^{208}\text{Pb}/^{204}\text{Pb}$ and $^{206}\text{Pb}/^{204}\text{Pb}$	7.56 (45)	8.57 (3)	7.56 (45)	8.57 (3)
$^{208}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{206}\text{Pb}$	7.39 (44)	8.57 (3)	7.56 (45)	8.57 (3)
$^{208}\text{Pb}/^{204}\text{Pb}$ and $^{207}\text{Pb}/^{206}\text{Pb}$	7.06 (42)	8.57 (3)	7.06 (42)	8.57 (3)
$^{207}\text{Pb}/^{204}\text{Pb}$ and $^{206}\text{Pb}/^{204}\text{Pb}$	10.25 (61)	5.71 (2)	10.08 (60)	5.71 (2)
$^{207}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{206}\text{Pb}$	10.92 (65)	5.71 (2)	11.26 (67)	5.71 (2)
$^{207}\text{Pb}/^{204}\text{Pb}$ and $^{207}\text{Pb}/^{206}\text{Pb}$	10.25 (61)	5.71 (2)	10.08 (60)	5.71 (2)
$^{206}\text{Pb}/^{204}\text{Pb}$ and $^{208}\text{Pb}/^{206}\text{Pb}$	7.90 (47)	11.43 (4)	7.56 (45)	5.71 (2)
$^{206}\text{Pb}/^{204}\text{Pb}$ and $^{207}\text{Pb}/^{206}\text{Pb}$	9.75 (58)	8.57 (3)	10.76 (64)	8.57 (3)
$^{208}\text{Pb}/^{206}\text{Pb}$ and $^{207}\text{Pb}/^{206}\text{Pb}$	8.24 (49)	5.71 (2)	8.24 (49)	5.71 (2)

[a] Results in [%].
[b] Number of incorrect answers.

the number of false positive answers for bivariate models is about half that for univariate ones. Therefore, based on the levels of false responses, there are no such models that are simultaneously providing the most reliable results. Moreover, in most cases, the percentage of false positive answers exceeds the percentage of false negative answers.

The results showed that slight differences, occurring in the rates of false answers for models concerning data subjected or not to the logarithmic transformation, are negligible, as they refer to only a few incorrect model responses. All the emerging differences seem to be rather a matter of chance.

It was mentioned that the used database is not large enough to reliably estimate all parameters considered in the applied LR Eqs. (3) and (4). An approach based on graph theory [3,9] could be used to factorise the joint density function into the product of several density functions of lower dimensions which allows estimation of the parameters even for smaller sample sizes but it could not be used in this case as it was noticed that the variables are highly correlated (see Table 1). Therefore, Principal

Component Analysis (PCA) was used for creating uncorrelated orthogonal variables. In other words it was not the PCA aim to reduce the dimensionality, but to deliver some orthogonal variables which can subsequently be used for the model involving the multiplication of the univariate LR model results. Nevertheless, the PCA performed for the data under analysis showed that due to the high correlation between the considered variables, just the first 3 components explain almost the whole variance, i.e. more than 99.9%. Such variables can therefore be used in LR calculations taking into account all the information about the objects under analysis.

All the newly formed uncorrelated variables, being the principal components, were subjected to the LR calculations in the form of $LR_f = \Pi_{i=1} LR_{PCi}$, where LR_{PCi} is LR value calculated using i-th Principal Component by applying the LR formula for the univariate problem (3) and (4). The LR_f model, based on all PC-based variables, reduces the error levels for different glass object comparisons to less than 5% when the data was not transformed and to 6.5% when the data was logarithmically transformed. The presented error levels are still lower than any other obtained taking into account other considered variables. Contrastingly, when comparing the objects coming from a common source, the rates of false negative answers exceed even 11%, which was only encountered for the worst bivariate combination ($^{206}Pb/^{204}Pb$ and $^{208}Pb/^{206}Pb$ variables) in the case of raw data. However, the advantage of results obtained on the basis of PC is that the rates of false positive answers, indisputably more concerning from the forensic point of view, are significantly lowered.

According to [25], LR models concerning SEM-EDX data deliver ca. 5% of false positive answers and 3.5% of false negative, which is lower than for the proposed LR models based on isotopic ratios. Similarly, for LR models based on ICP-MS data [26] no false negative answers were obtained and the false positive rate was 3.3%. The LR models effectiveness (assessed by Empirical Cross Entropy) is still comparable for the methods routinely used in forensic glass analysis (such as SEM-EDX) and the IRMS method (the loss of information reaches ca. 50–60%) [27]. However, this comparison of methods is only for a brief illustration, as it is quite difficult to compare the performance of the LR models concerning lead isotopic ratios with the error levels typical for other methods, commonly applied for glass analysis for forensic purposes. This is due to the fact that the analyses were carried out for different glass samples and the LR models differed in the assumptions (e.g. whether the between-object distribution is modelled by normal distribution or Kernel Density Estimation is used for its estimation).

For the method evaluation on the basis of the LR results, it is not only the information on the rates of false answers that is essential, but also the most crucial seems to be the LR values distribution. The histograms presented in Fig. 5 depict the examples of the LR values distributions calculated on the basis of one of the variables having the best performance e.g. $^{208}Pb/^{204}Pb$ (Fig. 5a and b for different and the same source object
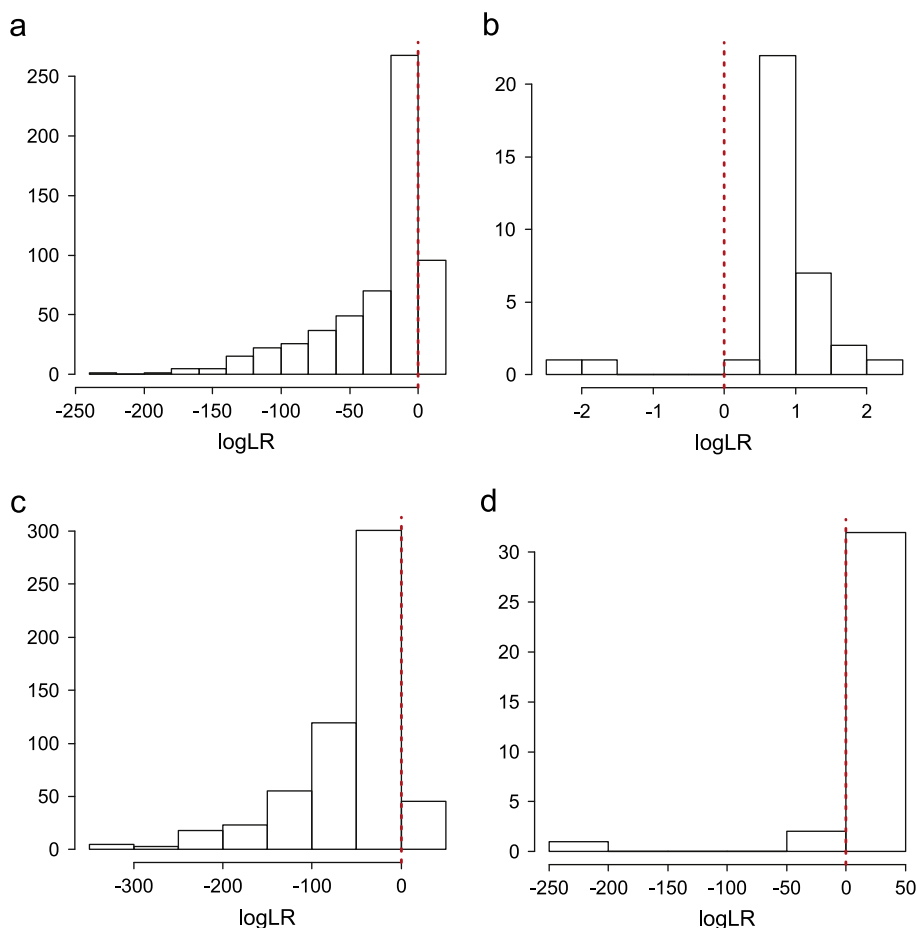

Fig. 5. Histograms depicting the LR values distributions: (a) and (c) comparison of glass fragments from different sources (log (LR) < 0 are desirable) for the variables with satisfying performance (e.g. $^{208}Pb/^{204}Pb$) and poor performance (e.g. log ($^{208}Pb/^{204}Pb$) and log ($^{206}Pb/^{204}Pb$)) respectively; (b) and (d) comparison of glass fragments from the same source (log (LR) > 0 are desirable) for the variables with satisfying performance and poor performance respectively. Note that log (LR)=0, which stands for LR=1, is marked by a dashed (green) line. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

comparisons respectively) and one revealing the poorest performance e.g. log (^{208}Pb/^{204}Pb) and log (^{206}Pb/^{204}Pb) (Fig. 5c and d for different and the same source object comparisons respectively). A brief inspection of Fig. 5a and b suggests that a great deal of LR values are far from 1 in support of the proper hypotheses. Contrastingly, for poorly behaving variables (Fig. 5c and d), more numerous and extreme are the LR values supporting the incorrect hypotheses.

The models efficiency illustrated by the structure of the distributions for each of the considered models (here, a model is assumed to be a combination of a particular LR equation and variables used for evidence evaluation) is reflected in the ECE plots, which were prepared using Empirical Cross Entropy approach as an assessment metric of the evidence evaluation method performance (see Section 2.5).

The performed ECE plots (Fig. 6) show that the loss of information is reduced by the evidence evaluation method from 100% to 50–70% (observed for $\log_{10}(\text{Odds}(\theta)=0$) only for ^{208}Pb/^{204}Pb and ^{207}Pb/^{204}Pb variables with respect to not evaluating the evidence. For the variables ^{206}Pb/^{204}Pb, ^{208}Pb/^{206}Pb and ^{207}Pb/^{206}Pb for the whole range of the *a priori* odds the evaluation of the evidence seems to be pointless, as information provided by

a priori odds is more reliable than information gained from the analysis (the solid (red) line above the *null* (black) line).

This situation does not seem to improve for the data subjected to the logarithmic transformation as the ECE plots for the LR sets obtained from the calculations using the data subjected and not to the logarithmic transformation practically do not differ. When the ECE procedure was used for bivariate data (Fig. 7), only for a combination of variables: ^{208}Pb/^{204}Pb and ^{207}Pb/^{204}Pb or log (^{208}Pb/^{204}Pb) and log (^{207}Pb/^{204}Pb) the reduction of uncertainty occurs for at least some part of the *a priori* odds range ($\log_{10}(\text{Odds}(\theta)) > -0.3$), whereas for the remaining variables not evaluating the evidence at all proves to deliver more adequate results regardless of the *a priori* odds. Similar observations could be made for a LR model based on the PC-variables. This indicates that by creating a model based on totally uncorrelated PC-variables the performance of the method under assessment does not appear to improve.

The performed ECE plots suggest that the performance of the LR models is rather poor as only for some variables the information gained by analysing the evidence reduces the loss of information from 100% to 50–70%. In some cases even tens of percents of the differences between the ECE (the solid (red) line)
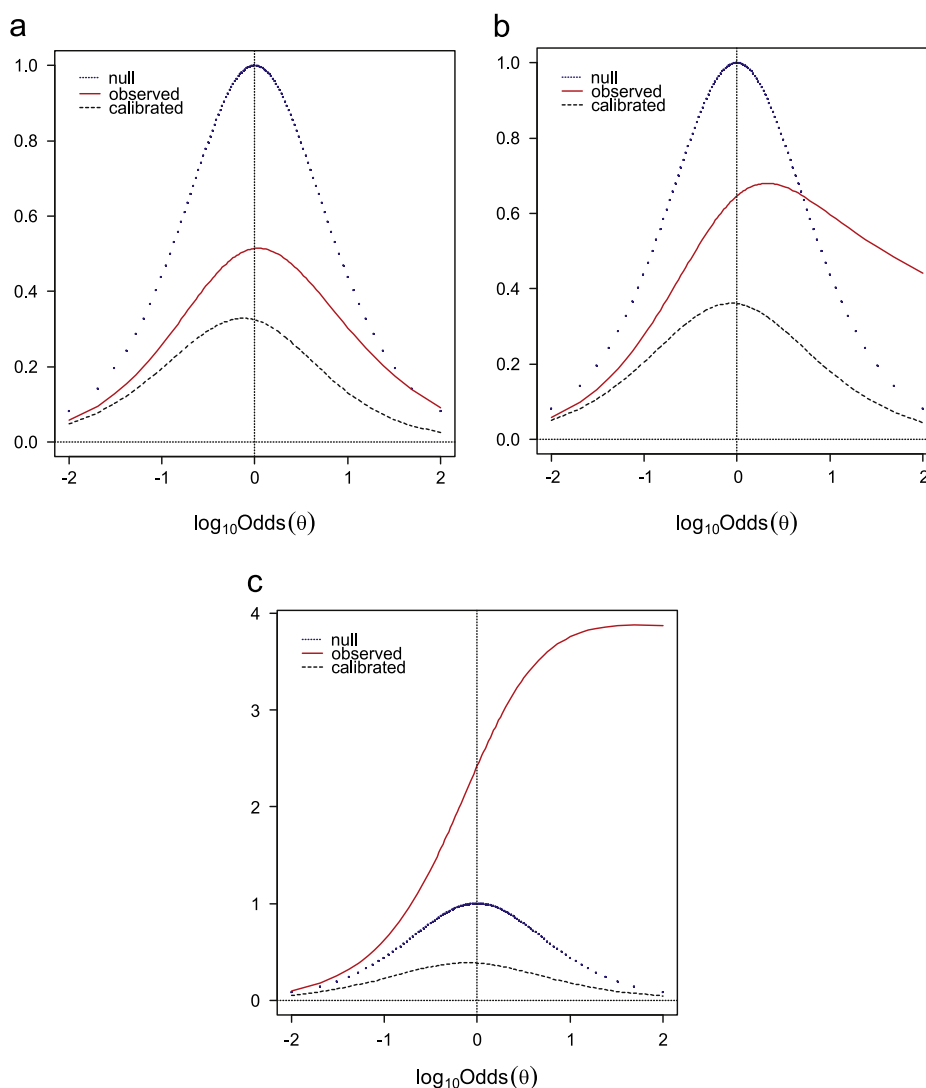

Fig. 6. ECE plots for single variables; (a) log (^{208}Pb/^{204}Pb), (b) log (^{207}Pb/^{204}Pb), (c) log (^{206}Pb/^{204}Pb) (also for log (^{208}Pb/^{206}Pb) and log (^{207}Pb/^{206}Pb) similar shapes of ECE plots were observed).

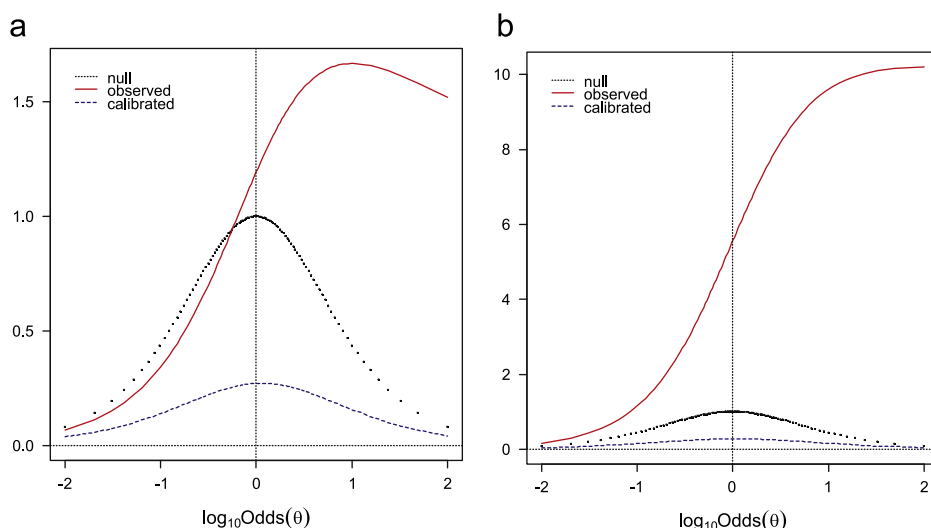**Fig. 7.** ECE plots for bivariate combinations of variables; (a) the ECE plot typical for a set $^{208}Pb/^{204}Pb$ and $^{207}Pb/^{204}Pb$ as well as for a set of $\log(^{208}Pb/^{204}Pb)$ and $\log(^{207}Pb/^{204}Pb)$ variables, (b) the ECE plot typical for the remaining variables.

for experimental LR set and the calibrated method (the dashed (blue) line) indicate the problems with the calibration of all proposed LR models under assessment. Nevertheless, these results should be confirmed when more data is available.

4. Conclusions

It could be concluded that any set of variable(s) cannot be selected which gave the best results taking into account results of false positive and false negative answers as well as ECE plots analysis. The application of the $^{208}Pb/^{204}Pb$ variable seems to be promising as it delivers one of the lowest percentages of false positive and false negative answers as well as being the only variable for which the ECE plot gave satisfactory results. The $^{207}Pb/^{204}Pb$ delivered a slightly lower number of false positive answers than $^{208}Pb/^{204}Pb$, but the ECE is not as good as for $^{208}Pb/^{204}Pb$. Also, a combination of these variables delivers good results in terms of false positive and false negative answers as well as the results of ECE.

What is more, IRMS data is rather easy to present in a descriptive context, and therefore easier to explain to the layman than more complex datasets. Development of more advanced instrumentation will move towards faster analysis with higher sensitivity and precision. As such, IRMS determination moves towards a promising future.

References

[1] J.M. Curran, T.N. Hicks, J.S. Buckleton, Forensic Interpretation of glass evidenceCRC Press LLC, Boca Raton, USA, 2000.
[2] G. Zadora, Anal. Chim. Acta 642 (2009) 279–290.
[3] G. Zadora, J. Forensic Sci. 54 (2009) 49–59.
[4] G. Zadora, D. Wilk, Probl. Forensic Sci. 80 (2009) 365–377.
[5] T. Neocleous, C.G.G. Aitken, G. Zadora, Chemom. Intell. Lab. Syst. 109 (2011) 77–85.
[6] D. Ramos, G. Zadora, Anal. Chim. Acta 705 (2011) 207–217.
[7] G. Zadora, Z. Brożek-Mucha, Mater. Chem. Phys. 81 (2003) 345–348.
[8] G. Zadora, T. Neocleous, C.G.G. Aitken, J. Forensic Sci. 55 (2010) 371–384.
[9] T. Trejos, J.R. Almirall, Talanta 67 (2005) 388–395.
[10] T. Trejos, J.R. Almirall, Talanta 67 (2005) 396–401.
[11] T. Hicks, F. Monard-Sermier, T. Goldmann, A. Brunelle, C. Champod, P. Margot, Forensic Sci. Int. 137 (2003) 107–118.
[12] C.M. Bridge, J. Powell, K.L. Steele, M.E. Sigman, Spectrochim. Acta B 62 (2007) 1419–1425.
[13] N. Carmona, M. Oujja, S. Gaspard, M. Garcia-Heras, M.A. Villegas, M. Castillejo, Spectrochim. Acta B 62 (2007) 94–100.
[14] E.M. Rodriguez-Celis, I.B. Gornushkin, U.M. Heitmann, J.R. Almirall, B.W. Smith, J.D. Winefordner, N. Omenetto, Anal. Bioanal. Chem. 391 (2008) 1961–1968.
[15] K.E. Sjastad, S.L. Simonsen, T. Anderson, J. Anal. At. Spectrom. 20 (2010) 325–333.
[16] C.J. Allegre, Isotope Geology, Cambridge University Press, Cambridge, New York, Melbourne, 2008.
[17] C.G.G. Aitken, F. Taroni, Statistics and the Evaluation of Evidence for Forensic ScientistsJohn Wiley & Sons, Chichester, 2004.
[18] C.G.G. Aitken, D. Lucy, Appl. Stat. 53 (2004) 109–122. (with corrigendum 665–666).
[19] C.G.G. Aitken, G. Zadora, D. Lucy, J. Forensic Sci. 52 (2007) 412–419.
[20] G. Zadora, J. Chemom. 24 (2010) 346–366.
[21] G. Zadora, D. Ramos, Chemom. Intell. Lab. Syst. 102 (2010) 63–83.
[22] D. Ramos, Forensic Evaluation of the Evidence Using Automatic Speaker Recognition Systems, Ph.D. Thesis, Depto. De Ingenierai Informatica, Escuela Politecnica Superior, Universidad Autonoma de Madrid, Madrid, Spain, 2007. Available from : http://atvs.ii.uam.es.
[23] N. Brümmer, J. du Preez, Compt. Speech Lang. 20 (2006) 230–275.
[24] M.H. deGroot, S.E. Fienberg, Statistician 32 (1982) 12–22.
[25] T. Neocleous, C.G.G. Aitken, G. Zadora, Chemom. Intell. Lab. Syst. 109 (2011) 77–85.
[26] C.G.G. Aitken, D. Lucy, Appl. Stat.—J. R. Stat. Soc. C 53 (1) (2004) 109–122.
[27] G. Zadora, D. Ramos, Chemom. Intell. Lab. Syst 102 (2010) 63–83.