



**Repositorio Institucional de la Universidad Autónoma de Madrid**

<https://repositorio.uam.es>

Esta es la **versión de autor** del artículo publicado en:

This is an **author produced version** of a paper published in:

Food Chemistry 150 (2014): 287 – 295

**DOI:** <http://dx.doi.org/10.1016/j.foodchem.2013.10.111>

**Copyright:** © 2014 Elsevier B.V.

El acceso a la versión del editor puede requerir la suscripción del recurso

Access to the published version may require subscription

**Wine authenticity verification as a forensic problem. An application of likelihood ratio test to label verification.**

*Agnieszka Martyna<sup>a</sup>, Grzegorz Zadora<sup>b,\*</sup>, Ivana Stanimirova<sup>c</sup>, Daniel Ramos<sup>d</sup>*

<sup>a</sup> Faculty of Chemistry, Jagiellonian University, Ingardena 3, 30-060 Krakow, Poland, e-mail address: rzepecka@chemia.uj.edu.pl

<sup>b</sup> Institute of Forensic Research, Westerplatte 9, 31-033, Krakow, Poland, e-mail address: gzadora@ies.krakow.pl

<sup>c</sup> Department of Analytical Chemistry, Chemometric Research Group, Institute of Chemistry, The University of Silesia, Szkolna 9, 40-006 Katowice, Poland, e-mail address: ivana.stanimirova-daszykowska@us.edu.pl

<sup>d</sup> Institute of Forensic Science and Security and ATVS - Biometric Recognition Group (ATVS), Universidad Autonoma de Madrid, C/ Francisco Tomas y Valiente 11, 28049 Madrid, Spain, e-mail address: daniel.ramos@uam.es

**Abstract**

The aim of the study was to investigate the applicability of the likelihood ratio (LR) approach for verifying the authenticity of 178 samples of 3 Italian wine brands: Barolo, Barbera, and Grignolino described by 27 parameters describing their chemical compositions. Since the problem of products authenticity may be of forensic interest, the likelihood ratio approach, expressing the role of the forensic expert, was proposed for determining the true origin of wines. It allows us to analyse the evidence in the context of two hypotheses, that the object belongs to 1° or 2° wine brand. Various LR models were the subject of the research and their correctness was evaluated by the Empirical Cross Entropy (ECE) approach. The rates of

correct classifications for the proposed models were higher than 90% and their performance evaluated by ECE was satisfactory.

# **Key words**

evaluation of forensic evidence, food products authenticity, likelihood ratio, empirical cross entropy, classification problem

\* Corresponding author: Tel.: +48 12 422 87 55, Fax: +48 12 422 38 50, e-mail address: gzadora@ies.krakow.pl

## 1. Introduction

Verifying the authenticity of food products is one of the most important issues in food quality control aiming to guarantee the safety and to protect the rights of consumers and producers. A chemical approach to inferring the properties of food products is based on analysis of chemical composition of a particular food product as a unique combination of constituents. Then, either a classification or a discriminant chemometric method can be used to predict the assignment of an unknown food sample described by its chemical features to a group of similar samples. The classification/discriminant rules are first created for samples grouped according to their geographical origins, years of production, producers or brands, *etc.* (Charlton, Wrobel, Stanimirova, Daszykowski, Grundy & Walczak, 2010; Stanimirova *et al.*, 2010) and then these rules are used for prediction purposes. Even though such an approach seems straightforward, it requires a delivery of new food quality specifications for different authentic food commodities and a selection of a classification/discriminant chemometric model with a relatively high efficiency, sensitivity and specificity for the problem studied. Therefore, the development of cost-effective procedures for identification of fraudulent products by checking the compliance with the food quality specifications is highly valued. This was essentially the goal of the EU-funded project TRACE - Tracing food commodities in Europe.

The authenticity of food products may be an issue of forensic interest, especially when it involves economic consequences or causes negative health effects. Then, representatives of the administration of justice are interested in answering the question of *what is the value of the evidence of the measurements in relation to the propositions that the analysed sample came from either category 1 or 2?* This problem is known in the forensic field as a classification problem.

A situation in the court is that the prosecutor and the defence have opposite hypotheses e.g.  $\theta_1$ : a wine is not from Grignolino brand and  $\theta_2$ : a wine is from Grignolino brand. In general,

the prosecutor and the defence think in a sense of the following conditional probabilities –  $\Pr(\theta_1|E)$  and  $\Pr(\theta_2|E)$ , where  $E$  describes the evidence (e.g. physicochemical data obtained during analysis of a wine sample, quality specifications). The role of the forensic expert is to evaluate an evidence ( $E$ ) in the context of these hypotheses. It requires estimation of the following conditional probabilities  $\Pr(E|\theta_1)$  and  $\Pr(E|\theta_2)$ .

The evaluation of physicochemical data (quality specifications) from a forensic point of view requires some knowledge about the rarity of the measured physicochemical properties (quality specifications) in a population representative for the analysed casework - called the relevant population (e.g. the population of wines of a particular type). For instance, similar values of particular wine characteristics could be observed in different brand of wines. Therefore, information about the rarity of a determined value of wine characteristics has to be taken into account. For example, the value of the evidence in support of the proposition that the wine sample originated from category 1 is greater when the determined value of these characteristics is rare in the relevant population of category 1, than when this value is common in the relevant population of category 2. In the aim to obtain information about the rarity of the physicochemical data suitable databases should be available. Moreover, it should be pointed out that information about the rarity is not included in most of the discriminant methods, e.g. LDA.

Moreover, it is important that the results of the physicochemical analysis (quality specifications data) of products subjected to authenticity verification made by forensic experts should be evaluated by methods which also allow for including information about the possible sources of uncertainty (e.g. the variation of measurements within the analysed objects, the variation of measurements between objects in the relevant population) and existing correlation between variables in the case of multi-dimensional data.

The evidential value of physicochemical data (quality specifications), taking into account all the mentioned requirements stemming from forensic practice, could be assessed by the

application of the likelihood ratio approach (LR), a well-documented measure of evidential value in the forensic sciences. An extensive body of literature exists on the applications of LR in the forensic field (Aitken & Taroni, 2004). The likelihood ratio approach is widely used in the interpretation of data collected in the analysis of glass fragments (Zadora, 2009, Zadora & Neocleous, 2009, Zadora & Ramos, 2010) and in genetics for DNA profiling (e.g. Aitken *et al.*, 2004; Evett & Weir, 1998). It allows for analysis of the evidence (E) in the context of two hypotheses, that the object belongs to either 1° category ( $\theta_1$ ) or the 2° one ( $\theta_2$ ). The LR is defined by the following equation:

$$LR = \frac{\Pr(E | \theta_1)}{\Pr(E | \theta_2)} \quad \{1\}$$

In the case of continuous type data,  $\Pr(\cdot)$  are substituted by suitable probability density functions  $f(\cdot)$ . Values of LR above 1 support  $\theta_1$ , while values of LR below 1 support the  $\theta_2$  hypothesis. The values equal to 1 support neither of them. The higher (lower) the value of LR, the stronger the support for the relevant hypothesis is.

The likelihood ratio approach is a part of the *Bayes' theorem* expressed in Eq. 2.

$$\frac{\Pr(\theta_1)}{\Pr(\theta_2)} \cdot \frac{\Pr(E | \theta_1)}{\Pr(E | \theta_2)} = \frac{\Pr(\theta_1)}{\Pr(\theta_2)} \cdot LR = \frac{\Pr(\theta_1 | E)}{\Pr(\theta_2 | E)} \quad \{2\}$$

$\Pr(\theta_1)$  and  $\Pr(\theta_2)$  are called *a priori* probabilities and their quotient is called the prior odds. Their estimation lies within the competence of the fact finder (judge, prosecutor, or police) expressing their opinion about the considered hypotheses before the evidence is analysed, thus without having any further information in this matter. This opinion may be modified by accounting LR values supporting one of the propositions and delivered by an expert after the analysis of evidence. It is the duty of a fact finder, police, or court to determine whether the objects are deemed to belong to one of the considered categories and this decision is taken based, as mentioned previously, on the results expressed in the form of conditional

probabilities -  $\Pr(\theta_1 | E)$  and  $\Pr(\theta_2 | E)$ , namely posterior probabilities, whose quotient is called the posterior odds.

For every evidence evaluation method it is crucial that it delivers strong support for the correct hypothesis (i.e.  $LR \gg 1$  when  $\theta_1$  is correct and  $LR \ll 1$  when  $\theta_2$  is correct). Additionally, it is desired that if an incorrect hypothesis is supported by LR value (i.e.  $LR < 1$  for true  $\theta_1$  and  $LR > 1$  for true  $\theta_2$ ), then the LR value should concentrate close to 1 delivering 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 of evaluating the performance of the applied methodology for data evaluation, which could be made by the application of the Empirical Cross Entropy (ECE) approach (Brümmer & du Preez, 2006; Ramos, Gonzalez-Rodriguez, Aitken & Zadora, 2013; Ramos & Zadora, 2011; Zadora *et al.*, 2010).

The aim of this study is to investigate the applicability of the likelihood ratio approach for verifying the authenticity of samples for forensic purposes. For illustration purposes, a set of authentic wine samples described by physicochemical features that belong to three production brands (Grignolino, Barolo, and Barbera) was considered. The assessment of the performance of the applied models was conducted by the Empirical Cross Entropy approach (Brümmer *et al.*, 2006; Ramos *et al.*, 2013).

The aim of the paper is to present LR approach, which could be used when the authenticity of food products is an issue of forensic interest and to show the performance of LDA when the method was applied for the same forensic purpose.

## 2. Methods

### 2.1 Wines database

The data subjected to the evaluation process was taken from Forina, Armanino, Castino & Ubiegli (1986). They were obtained from the analysis of 178 wine samples from 3 brands of

Italian wines (59 samples of Barolo (further denoted as BAR), 71 samples of Grignolino (GRI), and 48 samples of Barbera (BRB)). Each sample represented a single bottle of wine. Samples were collected and pretreated in a way conditioned on the type of the subsequent analysis. The applied methods of the analysis were mostly specific for wines analysis such as a group of methods known under common name *wet chemical analysis*. The rest of the methods involved HPLC, GC, and enzymatic analysis. Schlesier *et al.* (2009) discuss these issues.

For each sample, 27 parameters were determined and listed in Table 1. All of them represent the commonly determined characteristics of wines for commercial and scientific purposes.

## 2.2 Likelihood ratio

In this research LR values were calculated for each of the 178 analysed objects (wine samples). Therefore, the data matrix consisted of 178 rows (each corresponding to one of the analysed samples) and 27 columns (each describing one of the determined parameters for the wine samples). Therefore, the data for the sample under classification were in the form of a  $\bar{y}$  vector with the length of 27. A so-called *one-level* LR model (firstly introduced in Zadora (2009)) was applied since there were only single measurements made for each parameter within an object, thus the within-object variability was not available (Zadora *et al.*, 2009). For the purpose of this study a likelihood ratio (LR) was computed for logarithmically transformed data (i.e. for example  $\log_{10}(\text{Alc})$ , where Alc stands for the original data describing the alcohol content in the samples). A kernel density estimation procedure (KDE) using Gaussian kernels was applied for the estimation of between-object distribution as some of the variables could not be described by normal distribution (see section 3.1).

The jack-knife procedure was applied for the estimation of suitable population parameters, which implies excluding the object already classified from the total population of the analysed objects. Following this procedure, LR values obtained for different objects are based on



slightly differing information derived from the database. However, the presented approach ensures that all of the available data are exploited at once to the limits of possibilities and the proposed model is not over-fitted.

When the  $\theta_1$  hypothesis states that the object with mean vector  $\bar{\mathbf{y}}$  belongs to category 1, the  $\theta_2$  states that it belongs to the second category, and the between-object distribution is estimated by KDE, then the *one-level* LR is calculated according to Eq. 3:

$$LR = \frac{\left| h_1^2 \mathbf{C}_1 \right|^{-1/2} \frac{1}{m_1} \sum_{i=1}^{m_1} \exp \left\{ -\frac{1}{2} (\bar{\mathbf{y}} - \bar{\mathbf{x}}_{1i})^T (h_1^2 \mathbf{C}_1)^{-1} (\bar{\mathbf{y}} - \bar{\mathbf{x}}_{1i}) \right\}}{\left| h_2^2 \mathbf{C}_2 \right|^{-1/2} \frac{1}{m_2} \sum_{i=1}^{m_2} \exp \left\{ -\frac{1}{2} (\bar{\mathbf{y}} - \bar{\mathbf{x}}_{2i})^T (h_2^2 \mathbf{C}_2)^{-1} (\bar{\mathbf{y}} - \bar{\mathbf{x}}_{2i}) \right\}} \quad \{3\}$$

The between-object variance-covariance estimate ( $\mathbf{C}$ ) in the case of multivariate data in *one-level* LR model can be expressed as follows:

$$\mathbf{C}_g = \frac{\mathbf{S}_g^*}{m_g - 1}, \quad \{4\}$$

where:

$$\mathbf{S}_g^* = \sum_{i=1}^m (\bar{\mathbf{x}}_{gi} - \bar{\mathbf{x}}_g)(\bar{\mathbf{x}}_{gi} - \bar{\mathbf{x}}_g)^T,$$

$\bar{\mathbf{x}}_{gi}$  - a vector of means of  $p$  variables calculated using  $n$  measurements (here:  $n=1$ ) for the  $i$ -th object coming from  $g$  ( $g=1,2$ ) objects category:  $\bar{\mathbf{x}}_{gi} = \frac{1}{n} \sum_{j=1}^n \mathbf{x}_{gij}$  (here:  $\bar{\mathbf{x}}_{gi} = \mathbf{x}_{gi}$ ),

$\bar{\mathbf{x}}_g$  - a vector of means of  $p$  variables calculated using  $n$  measurements performed for  $m_g$

objects coming from  $g$  ( $g=1,2$ ) objects category:  $\bar{\mathbf{x}}_g = \frac{1}{m_g n} \sum_{i=1}^{m_g} \sum_{j=1}^n \mathbf{x}_{gij},$

$h_g$  - smoothing parameter calculated according to the expression (Silverman, 1986):

$h_g = h_{opt} = \left( \frac{4}{m_g(2p+1)} \right)^{\frac{1}{p+4}}.$  The smoothing parameter helps in fitting the probability density

curve to the analysed data using Kernel Density Estimation procedure with Gaussian kernels (Silverman, 1986).

In the case of the analysis of univariate data ( $p=1$ ), vectors and matrices become suitable scalars (e.g.  $\mathbf{C}$  is replaced by  $c^2$ ).

The evaluation of the correct model answers (correct classifications) acts as one of the performance measures that are typically defined as the percentage of likelihood ratio values that would lead to a correct decision if the decision threshold is set at  $LR=1$ . In the aim of evaluating the levels of correctly classified objects, a number of experiments were conducted. Therefore, each object was classified into one of two categories (see section 3.2) based on LR values obtained in the course of the univariate LR calculations and other alternative models (see section 3.3).

### 2.3 Empirical Cross Entropy

However, the rates of correct and incorrect LR model answers are limited measures of performance. They only provide information about supported proposition according to the threshold at  $LR=1$ , but they ignore the strength of such support carried by the magnitude of the LR value. For instance, a LR value computed under the  $\theta_1$  hypothesis would be much worse if its value is  $LR=1000$  than if it is  $LR=2$  in the case of true-  $\theta_2$  hypothesis. In the *misleading evidence rates* approach, these two LR values are treated equally and no distinction between their strength is being made, contrastingly to the Empirical Cross Entropy (ECE), which takes into account both the support and strength of the hypothesis.

ECE was proposed as an assessment technique for evaluation methods such as the likelihood ratio model (Brümmer *et al.*, 2006; Ramos *et al.*, 2013). ECE is a framework derived from information theory firstly presented in 1950's.

ECE, being a measure of information, is aimed at assessing the performance of a statistic (such as the likelihood ratio) with respect to correctness of decision making (Lucy & Zadora, 2011; Ramos *et al.*, 2011; Zadora *et al.*, 2010). It was mentioned that the higher (lower) the LR values, the greater the support for the  $\theta_1$  ( $\theta_2$ ). 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 that 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*. Commonly, the strictly proper scoring rules are expressed as logarithmic scoring rules (LS) in the following way:

a) if  $\theta_1$  is true:  $-\log_2(\Pr(\theta_1 | E))$ ,

b) if  $\theta_2$  is true:  $-\log_2(\Pr(\theta_2 | E))$ .

In Brümmer *et al.* (2006) the overall measure of the goodness of a forecaster is defined as the average value of a strictly proper scoring rule over many different forecasts ( $e_i$  referring to evidence information considered under  $\theta_1$  hypothesis and  $e_j$  referring to evidence information considered under  $\theta_2$  hypothesis), which are expressed by posterior probabilities. For instance, for the logarithmic scoring rule, this mean value could be expressed by:

$$LS = -\frac{1}{N_1} \sum_{i \in cat.1} \log_2(\Pr(\theta_1 | e_i)) - \frac{1}{N_2} \sum_{j \in cat.2} \log_2(\Pr(\theta_2 | e_j)) \quad \{5\}$$

where  $N_1, N_2$  refer to the number of the objects originally belonging to each of the considered categories. This average value (LS) can be viewed as an overall loss. The ECE, is

the proposed measure of goodness as a variant of LS, weighted by the prior probabilities  $\Pr(\theta_1)$  and  $\Pr(\theta_2)$ , and is expressed as follows:

$$ECE = -\frac{\Pr(\theta_1)}{N_1} \sum_{i \in \text{cat.1}} \log_2(\Pr(\theta_1 | e_i)) - \frac{\Pr(\theta_2)}{N_2} \sum_{j \in \text{cat.2}} \log_2(\Pr(\theta_2 | e_j)) \quad \{6\}$$

Taking into account equation {2} it can be concluded that ECE could be expressed as:

$$ECE = \frac{\Pr(\theta_1)}{N_1} \sum_{i \in \text{cat.1}} \log_2 \left( 1 + \frac{\Pr(\theta_2)}{LR_i \Pr(\theta_1)} \right) + \frac{\Pr(\theta_2)}{N_2} \sum_{j \in \text{cat.2}} \log_2 \left( 1 + \frac{LR_j \Pr(\theta_1)}{\Pr(\theta_2)} \right) \quad \{7\}$$

The *a priori* probabilities  $\Pr(\theta_1)$  and  $\Pr(\theta_2)$  are not generally known in the forensic evaluation of evidence, because they depend on various information sources (witnesses, police investigations, other evidence, *etc.*). Because ECE cannot be computed if prior probabilities are not known, the adopted solution is to plot ECE for a set of all possible prior probability quotients, further referred to as prior odds and expressed as its logarithm  $\log_{10}\text{Odds}(\theta)$ . The details about the derivation and interpretation of ECE can be found in Brümmer *et al.* (2006), Ramos *et al.* (2013). That leads to the so-called ECE plot consisting of 3 curves (also referred to in sections 3.2 and 3.3, and Figure 2):

a) the solid (red) curve (named *observed*) – represents the ECE (average information loss) values calculated using the evidence evaluation method under analysis (see Eq. {7}).

b) the dashed (blue) curve (named *calibrated*) – represents the calibrated ECE values obtained from computing ECE for the experimental LR values transformed using Pool Adjacent Violators algorithm (PAV) (Ayer, Brunk, Ewing, Reid & Silverman, 1955; Best & Chakravarti, 1990). 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*) – 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 for the experimental set of LR values (solid, red line) in relation to the remaining two (dashed and dotted lines) 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 will be needed in order to know the true values of the hypotheses. In other words, the higher the curve, the more uncertainty remains and therefore the worse the method of choice is for the interpretation of the evidence under analysis. If the curve appears to have greater values than the ones in the neutral method, the evidence evaluation introduces more misleading information than when not evaluating the evidence at all.

For the purposes of this paper, the information about the reduction of information loss due to the analysis of evidence always refers to the point of  $\log_{10}\text{Odds}(\theta)=0$ .

### **3. Results and discussion**

#### *3.1 Descriptive statistics and experimental protocol*

Descriptive statistics in the form of box-plots (Figure 1) for each of the variables within a single category suggest, that most of the parameters ranges do not distinctively differ between the considered wine brands. Transmittance of flavonoids (T\_flav), transmittance for diluted samples (T\_diluted), and flavonoid (Flav) content seem to have the highest classification

power as their data overlap the least. The descriptive statistics also successfully identify the variables that should be avoided in the classification analysis. Such variables include Bu\_diol, Ca, P, TN, pH, Ash, K, Cl, and Meth.

Performed statistical analysis (e.g. Q-Q plots) proved that some variables are not normally distributed, and therefore, the kernel density estimation procedure (KDE) was used for their distribution modelling.

There were 6 classification problems considered; three were concerned with classifying objects into 2 categories formed from single brands such as Barolo *vs.* Grignolino (denoted further as BAR *vs.* GRI), Barolo *vs.* Barbera (BAR *vs.* BRB), and Grignolino *vs.* Barbera (GRI *vs.* BRB). The next three took into account all the analysed samples by classifying them into categories (out of which one was the single brand category and the second was formed from the two remaining classes joined together). The classification included classification into Grignolino wine class *vs.* combined Barolo and Barbera classes (denoted further as GRI *vs.* BARBRB), Barbera *vs.* combined Barolo and Grignolino classes (BRB *vs.* BARGRI), and Barolo *vs.* combined Grignolino and Barbera classes (BAR *vs.* GRIBRB). There were proposed 27 univariate LR models based on each of the 27 variables as well as 3 multivariate LR models. The first one took into account all of the variables assuming their independency (a naïve LR model denoted as  $LR_{27}$ ). The second naïve model involved variables selected by F-test and ECE curve shapes ( $LR_F$ ). The last one eliminated the independency assumption by employing PCA for creating orthogonal variables ( $LR_{PCA}$ ). ECE plots were generated in order to assess the performance of LR models in the evidence evaluation process for each LR model.

### *3.2 Univariate LR models*

The performed LR calculations as well as ECE plots proved that for each of the considered 6 classification problems, the rates of correct models responses significantly differed (Table 2 and Table 3). Thus, within each classification problem, there are different sets of variables that have the best performance as well as those delivering the most misleading information. For solving the classification problem within categories BAR *vs.* BRB the best variables are Flav (100% of correct classifications), T\_diluted (*ca.* 99%), T\_flav (*ca.* 97%), TPh (*ca.* 95%), Hue (*ca.* 94%), and Proline (*ca.* 89%). For Flav, ECE analysis showed that the reduction of information loss reaches 100% (Figure 2). For both transmittance variables (T\_diluted and T\_flav), the evidence evaluation makes the loss of information reduce by *ca.* 90% (all percentage values of information loss reduction refer to the  $\log_{10}\text{Odds}(\Theta)=0$ ). Hue, Proline, and TPh decrease the information loss from 100% to less than 30% in relation to the situation of not evaluating the evidence.

In the case of the classification into BAR and GRI categories the only variables for which the rate of correct classification exceeds 90% are Alc and Proline. These reduced the information loss to 28% and 24% in respect to 100% for the neutral method (dotted line in Figure 2). Within the GRI *vs.* BRB class, the most suitable variables are Flav, T\_diluted, Col\_int (nearly 90%), Hue, and T\_flav (both *ca.* 86%). The information gained by analysing the evidence has decreased the loss of information from 100% to 29-45% with respect to not evaluating the evidence at all.

For the classification into BAR and GRIBRB categories, variables such as Proline (*ca.* 92%) and Flav (*ca.* 88%) seem to deliver the most satisfying results. They reduce the information loss from 100% to 26% and 40% respectively.

For the classification into BRB and BARGRI categories, the best performance is achieved relating to the Flav and T\_diluted (both *ca.* 93%), Hue, and T\_flav (both *ca.* 91%). For these variables the loss of information is greatly reduced by the evidence evaluation method from 100% to approximately 26-34% with respect to not evaluating the evidence at all.

The results of ECE interpretation of LR values obtained in classification to GRI and BARBRB categories indicate that for the most effective variables (Alc and Col\_int with nearly 90% of correct classifications), the loss of information is the most noticeable, although still high and reaches the level of 50% and 37%. Great attention should be paid to all the remaining variables. They hardly reduce or even increase the information loss in comparison to the neutral method, which acts as if the evidence had not been analysed and does not support any of the propositions (LR=1).

As already proved, the analysis of ECE plots confirms that the variables with the best performance chosen on the basis of correct answers rates generate the most efficient and reliable LR models. This demonstrates that the LR values obtained within the most efficient univariate models strongly support the correct hypothesis and provide weak support for the incorrect one. However, there is no universal set of the variables with the best classification power for all the classification problems that were considered. The variables should be chosen for each problem individually. The reason for obtaining different sets of variables for distinctive classification problems lies in the nature of each of the brands of wines. These brands are described by many parameters, sometimes some of them overlap in two brands. However, the most effective variables gave the correct answers above 85% and effectively reduced the uncertainty about the evidence supporting the correct hypothesis. This is satisfactory and therefore they were mainly used for the creation of alternative LR models.

### *3.3 Multivariate LR models*

#### *3.3.1 A naïve multivariate LR model involving all the variables*

A naïve multivariate LR model accounting for all the variables was created by multiplication of 27 univariate problems concerning each of the 27 variables ( $LR_{27} = \prod_{i=1}^{p=27} LR_i$ ). Such a model may only be used in instances when the variables are independent. However, due to having a



limited number of objects in the database (and with only a single observation for each of the 27 variables describing them) such a naïve approach was justifiable.

The rates of correct classifications for each of 6 considered classification problems (obtained on the basis of the proposed LR models) exceeded 98% and in the case of classification into BAR vs. BRB were equal to 100%.

This model reduced the information loss due to the evidence analysis from 100% to less than 6%. This means that using this model, there remains less than 6% of information loss to improve the effectiveness of the model.

### *3.3.2 A naïve multivariate LR model involving the variables selected by F-test and ECE plots analysis*

Two criteria were proposed, based on the observation that some variables may introduce misleading support for either of hypotheses, with the aim to reduce the dimensionality of the data. This is done by removing variables which do not provide any additional, reliable information.

The procedure embedded within the  $F$ -test (Otto, 1999) was proposed for selecting a series of variables with the best classification power within the set of the 27 variables. Its objective is to compare the variance estimate of the data between classes with the variance estimate of observations within each of the class. High values of the  $F$ -statistic imply that the data variation between classes is much more significant than the variation within each class. The higher the  $F$ -statistic, the most powerful the variable for solving the classification problem, since such a variable represents relatively separate sets of observation values describing the objects belonging to different classes. The variables for which the  $F$ -statistic was lower than 3.90 ( $F_{1,176} = 3.90$ ) were removed from the variables set. Under the second criterion the removed variables were those for which the ECE curve representing the experimental LR

values (solid, red line in Figure 2) exceeds the neutral curve (dotted, black line in Figure 2) for any range of the prior odds. The selected variables for each of the classification problems are marked in Table 1. They were further used for an alternative model proposition. This was a naïve approach based on multiplication of the LR values for univariate problems concerning all the variables ( $p$ ) except those which were removed from the original dataset

$$(LR_F = \prod_{i=1}^p LR_i).$$

Such LR models delivered more than 98% of correct classification rates and reduced the information loss from 100% to less than 10%. Such a good performance of the model proclaims that naïve models based on reduced number of variables are still effective. However,  $LR_F$  is still a naïve approach and does not concern the existing correlation between the variables.

### *3.3.3 A multivariate LR model involving the variables selected by F-test and and ECE plots analysis and orthogonalised by PCA*

With the aim of removing the variables mutual dependency, PCA was applied (taking into account only the variables selected based on F-test and ECE plots), and delivered new orthogonal variables. All the computed principal components were used due to the assumption that in the forensic sciences all information about the evidence may be relevant and should be taken into account. The model that was proposed involved the multiplication of LR values obtained for univariate problems, but based on principal components

$$(LR_{PCA} = \prod_{i=1}^p LR_i).$$

The rates of correct classifications within the 6 considered classification problems were satisfactory. Only for classification BAR vs. GRI was the rate of correct classifications slightly lower (*ca.* 95%), but on the other hand it was 100% for BAR vs. BRB. The

information loss was significantly reduced from 100% to less than 25% (the poorest reduction for BAR vs. GRI and GRI vs. BARBRB), which was slightly worse than for previously proposed models. However, this was still satisfactory and better than for univariate models. The performance of each of the proposed alternative LR models (as well as for the most effective univariate model) is illustrated in Figure 3.

### 3.4 Comparison of the likelihood ratio model with linear discriminate analysis (LDA)

One of the most widely used discriminant methods in food authenticity testing is linear discriminant analysis, LDA. For a comparative purpose, the results obtained from LDA for the studied data are presented below. A cross-model validation methodology (Westerhuis *et al.*, 2008) was adopted avoiding the presentation of either over-optimistic or over-pessimistic predictions that could be obtained from a uniform selection (by the Kennard and Stone or Duplex algorithms) of samples in the training set. This methodology consists of multiple creations of discriminant rules with LDA using a set of samples randomly selected from each class and use of the created rules to predict the class membership of the remaining test samples from each class. This procedure was repeated 1000 times and efficiency (correct classification rate) of the LDA model was calculated as the average of the 1000 efficiency values obtained for the test sets. Because of the LDA assumption for equal class-covariances, the number of class samples selected in the training set is critical. This number was selected as 75% of the number of samples of the least numerous class (i.e. Barbera wine with 48 samples) guarantying a balanced training set. In contrast to the LR approach, when variables are in different units, an autoscaling procedure combining centering and scaling to unitary standard deviation of variables in the training set is required. The test set is preprocessed using the mean and standard deviation of the training set. The multiple LDA models using two discriminant functions each were built for the three-class problem. The average efficiency of the LDA model is 97.62%. The sensitivities (the percentage of correct predictions of class

membership) of the Barolo, Grignolino, and Barbera wines are of 99.87%, 95.53%, and 99.41%, respectively, indicating that the probability of recognising Barolo or Barbera wine as the other tested wine using 27 parameters is rather negligible, while there is a larger probability that a Grignolino wine could be labelled as a Barolo or Barbera wine.

The LR and LDA approaches are hard modelling techniques. In contrast to soft modelling tools, with hard modelling techniques, a sample described by a set of parameters will always be assigned to one of the modelled classes. In contrast to the LR approach, LDA does not provide any information on the probability given evidence (the probability with which the membership of an unknown sample is predicted) and does not include information about the rarity of a determined value of wine characteristics as is required in forensic science. In this context, the LR models are preferred in solving authenticity problems for wine samples when it is an issue of forensic interest.

#### **4. Conclusions**

This research was aimed at investigating if the LR models used in forensic science (for example in glass analysis (Zadora, 2009)) could be successfully adapted to verify the authenticity of food products and constitute reliable evidence for the court.

The results of this study clearly demonstrate that the classification problem of wines can be successfully solved by means of LR models. The percentage of correctly classified objects in some univariate LR models as well as for all proposed alternative models exceeded 90%. Application of these models to the forensic evaluation of evidence significantly reduces the information loss by tens of percent. The problems with calibration of LR values are also minimal (the inconsiderable distance between the solid, red and dashed, blue curve in ECE plots).

It can be concluded that there are no such universal sets of variables that yield the best results for each of the considered classification problems. However, the LR should be calculated

mainly based on Flav, spectral parameters such as T\_diluted and T\_flav, Hue, Col\_int, Alc, and Proline. Using these variables guarantees that the reduction of information loss will be most significant after the evidence analysis under the method of choice in relation to the neutral method, which delivers the performance of the method without investigating the evidence at all. The easiest classification seems to be the classification into categories BAR vs. BRB, as it is faultless when based on Flav and all alternative multivariate models.

A connection between the model performance and the number of the considered variables was firmly established. The more univariate problems were taken into account and used for calculating overall LR values, the better performance the model had and consequently the classification power proved to be greater. This corresponds to the fact that each variable should be treated as separate evidence and the more variables are taken into account, the more complete and improved the model is. Therefore, it seems reasonable to use multivariate LR models including all the information about the evidence.

## **Acknowledgments**

The authors wish to thank Dr. Christopher Rogers (Natural History Museum, London, UK) for his helpful comments and language support.

## References

1. Aitken, C. G. G & Taroni, F. (2004). Statistics and the Evaluation of Evidence for Forensic Scientists. (2nd ed.). Chichester: Wiley, UK.
2. Ayer, M., Brunk, H. D., Ewing, G. M., Reid, W. T. & Silverman, E. (1955). An Empirical Distribution Function for Sampling with Incomplete Information. *Annals of Mathematical Statistics*, 26, 641-647.
3. Best, M. J. & Chakravarti N. (1990). Active set algorithms for isotonic regression: A unifying framework. *Mathematical Programming*, 47, 425-439.
4. Brümmer, N. & du Preez, J. (2006). Application-independent evaluation of speaker detection. *Computer Speech and Language*, 20, 230-275.
5. Charlton, A. J., Wrobel, M. S., Stanimirova, I., Daszkowski, M., Grundy, H. H. & Walczak, B. (2010). Multivariate discrimination of wines with respect to their grape varieties and vintages. *European Food Research and Technology*, 231, 733–743.
6. Evett, I. W. & Weir, B. S., (1998). *Interpreting DNA evidence*. Sunderland: Sinauer Associates, USA.
7. Forina, M., Armanino, C., Castino, M. & Ubiegli, M. (1986). Multivariate data analysis as a discriminating method of the origin of wines. *Vitis*, 25, 189–201.
8. Lucy, D. & Zadora, G. (2011). Mixed effects modelling for glass category estimation from glass refractive indices. *Forensic Science International*, 212, 189-197.
9. Otto, M. (1999). *Chemometrics: statistics and computer application in analytical chemistry*. Weinheim: Wiley-VCH, Germany.
10. Ramos, D., Gonzalez-Rodriguez, J., Aitken, C. G. G. & Zadora, G. (2013). Information theoretical assessment of the performance of likelihood ratio computation methods. In press.
11. Ramos, D. & Zadora, G. (2011). Information-theoretical feature selection using data obtained by Scanning Electron Microscopy coupled with an Energy Dispersive X-ray

- spectrometer for the classification of glass traces. *Analytica Chimica Acta*, 705, 207-217.
12. Schlesier, K., Fauhl-Hassek, C., Forina, M., Cotea, V., Kocsi, E., Schoula, R., van Jaarsveld, F. & Wittkowski, R. (2009). Characterisation and determination of the geographical origin of wines. Part I: overview. *European Food Research and Technology*, 230, 1-13.
13. Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis. London: Chapman & Hall, UK.
14. Stanimirova, I., Üstün, B., Cajka, T., Riddellova, K., Hajslova, J., Buydens, L. M. C. & Walczak, B. (2010). Tracing the geographical origin of honeys using the GCxGC-MS and pattern recognition techniques. *Food Chemistry*, 118, 171-176.
15. Westerhuis, J. A., Hoefsloot, H. C. J., Smit, S., Vis, D. J., Smilde, A. K., van Velzen, E. J. J., van Duijnhoven, J. P. M. & van Dorsten, F. A. (2008). Assessment of PLS-DA cross validation. *Metabolomics*, 4, 81–89.
16. Zadora, G. (2009). Classification of glass fragments based on elemental composition and refractive index. *Journal of Forensic Sciences*, 54(1), 49-59.
17. Zadora, G. & Neocleous, T. (2009). Likelihood ratio model for classification of forensic evidence. *Analytica Chimica Acta*, 642, 266-278.
18. Zadora, G. & Ramos, D. (2010). Evaluation of glass samples for forensic purposes-an application of likelihood ratios and an information theoretical approach. *Chemometrics and Intelligent Laboratory Systems*, 102, 63-83.

## Figure captions

Fig. 1. Descriptive statistics (in the form of box-plots) presented for the variables mostly differing between the categories (Flav, T\_diluted, T\_flav) and one of the variables overlapping between categories (Bu\_diol).

Fig. 2. The ECE plots presenting the performance of 3 univariate LR models within BAR vs. BRB classification: a) LR model with the best performance based on Flav, b) LR model with a satisfactory performance based on Hue, c) LR model with a poor performance based on Meth.

Fig. 3. The values of ECE (at the  $\log_{10}\text{Odds}(\Theta)=0$ ) for the univariate model with the best performance and 3 multivariate LR models (LR<sub>27</sub> – a naïve LR model based on all variables, LR<sub>F</sub> – a naïve LR model based on variables chosen by F-test and ECE plots analysis, LR<sub>PCA</sub> – LR model based on data from PCA). Black bars illustrate the ECE for a calibrated set of LR values and white bars (with the black ones) present ECE for experimental LR values.



555 **Table captions**

556 Table 1. The parameters describing the analysed wine samples. The numbers indicate the  
557 variables taken into account in the  $LR_F$  model (1-BAR *vs.* BRB, 2-BAR *vs.* GRI, 3-GRI *vs.*  
558 BRB, 4-BAR *vs.* GRIBRB, 5-BRB *vs.* BARGRI, 6-GRI *vs.* BARBRB).

559

560 Table 2. The rates of correct classifications [%] within each classification problem obtained  
561 for the considered univariate and multivariate LR models and the corresponding ECE values  
562 for  $\log_{10}Odds(\Theta)=0$ .

563

564 Table 3. The rates of correct classifications [%] to each of the categories within each  
565 classification problem obtained for these considered univariate and multivariate LR models.

**Figure captions**

Fig. 1. Descriptive statistics (in the form of box-plots) presented for the variables mostly differing between the categories (Flav, T\_diluted, T\_flav) and one of the variables overlapping between categories (Bu\_diol).

Fig. 2. The ECE plots presenting the performance of 3 univariate LR models within BAR vs. BRB classification: a) LR model with the best performance based on Flav, b) LR model with a satisfactory performance based on Hue, c) LR model with a poor performance based on Meth.

Fig. 3. The values of Cllr (values of ECE at the  $\log_{10}\text{Odds}(\Theta)=0$ ) for the univariate model with the best performance and 3 multivariate LR models (LR<sub>27</sub> – a naïve LR model based on all variables, LR<sub>F</sub> – a naïve LR model based on variables chosen by F-test and ECE plots analysis, LR<sub>PCA</sub> – LR model based on data from PCA). Black bars illustrate the ECE for a calibrated set of LR values and white bars (with the black ones) present ECE for experimental LR values.

Table 1. The parameters describing the analysed wine samples. The numbers indicate the variables taken into account in the  $LR_F^{(a)}$  model (1-BAR<sup>b</sup>) vs. BRB, 2-BAR vs. GRI, 3-GRI vs. BRB, 4-BAR vs. GRI\_BRB, 5-BRB vs. BAR\_GRI, 6-GRI vs. BAR\_BRB).

no	abbreviations	explanation
1	Alc <sup>(1,2,3,4,6)</sup>	Alcohol
2	Sfe <sup>(2,4,6)</sup>	Sugar-free extract
3	Fa <sup>(1,3,4,5,6)</sup>	Fixed acidity
4	Ta <sup>(1,2,3,4,5)</sup>	Tartaric acid
5	Ma <sup>(1,3,5,6)</sup>	Malic acid
6	Ua <sup>(1,3,4,5,6)</sup>	Uronic acid
7	pH <sup>(1,5)</sup>	pH
8	Ash <sup>(2,3,4,5,6)</sup>	Ash
9	Aash <sup>(1,3,5)</sup>	Alkalinity of Ash
10	K <sup>(2,6)</sup>	Potassium
11	Ca <sup>(2,3,4,6)</sup>	Calcium
12	Mg <sup>(1,2,3,4,6)</sup>	Magnesium
13	P <sup>(1,2,4,6)</sup>	Phosphate
14	Cl <sup>(1,5)</sup>	Chloride
15	TPh <sup>(1,2,3,4,5)</sup>	Total phenols
16	Flav <sup>(1,2,4)</sup>	Flavonoids
17	NFPh <sup>(1,2,3,4,5)</sup>	Nonflavonoid phenols
18	Proanth <sup>(1,2,3,4,5)</sup>	Proanthocyanins
19	Col_int <sup>(1,2,3,4,5,6)</sup>	Color intensity
20	Hue <sup>(1,3,4,5,6)</sup>	Hue
21	T_diluted <sup>(1,2,3,4,5,6)</sup>	Transmittance ratio of diluted samples of wines measured by 280 and 315 nm
22	T_flav <sup>(1,2,3,4,5,6)</sup>	Transmittance ratio of flavonoids measured by 280 and 315 nm
23	Gly <sup>(1,4)</sup>	Glycerol
24	Bu_diol <sup>(2,3,5,6)</sup>	2,3-butanediol
25	TN <sup>(5)</sup>	Total nitrogen
26	Proline <sup>(1,2,3,4,5,6)</sup>	Proline
27	Meth	Methanol

a) a naïve multivariate likelihood ratio model accounting for variables selected with application of the  $F$ -test,

b) BAR – Barolo wine brand, BRB – Barbera wine brand, GRI – Grignolino wine brand.

Table 2. The rates of correct classifications [%] within each classification problem obtained for the considered univariate and multivariate LR models and the corresponding Cllr values, i.e. ECE values for  $\log_{10}\text{Odds}(\Theta)=0$ .

Problem of classification into categories:												
	BAR vs. BRB		BAR vs. GRI		GRI vs. BRB		BRB vs. BAR_GRI		BAR vs. GRI_BRB		GRI vs. BAR_BRB	
LR models	% corr. class.	Cllr	% corr. class.	Cllr	% corr. class.	Cllr	% corr. class.	Cllr	% corr. class.	Cllr	% corr. class.	Cllr
Alc	68.2	0.80	91.5	0.28	79.0	0.66	59.6	0.87	77.0	0.54	88.2	0.50
Sfe	72.0	0.88	75.4	0.73	48.7	1.05	52.8	1.08	71.4	0.79	66.9	0.90
Fa	84.1	0.55	67.7	0.91	68.1	0.74	74.2	0.66	71.4	0.82	61.8	0.95
Ta	86.0	0.58	57.7	0.95	78.2	0.76	83.7	0.70	62.4	0.82	58.4	0.96
Ma	86.0	0.57	68.5	0.81	80.7	0.72	81.5	0.67	73.6	0.75	72.5	0.87
Ua	75.7	0.64	52.3	1.02	76.5	0.67	80.9	0.65	51.7	0.91	55.6	0.92
pH	62.6	0.95	50.8	1.02	56.3	1.02	62.4	0.98	55.1	0.99	32.6	1.03
Ash	44.9	1.01	63.1	0.92	58.8	0.84	44.4	0.92	57.9	0.98	69.7	0.89
Aash	80.4	0.56	69.2	0.93	56.3	0.92	62.9	0.78	74.2	0.85	51.7	1.09
K	58.9	1.00	54.6	0.95	60.5	0.94	59.0	0.97	51.1	0.97	65.7	0.94
Ca	61.7	0.96	73.9	0.79	66.4	0.91	64.0	0.98	62.4	0.87	71.9	0.85
Mg	64.5	0.93	77.7	0.66	66.4	0.88	37.6	0.99	65.2	0.79	74.2	0.79
P	76.6	0.82	75.4	0.76	42.9	1.01	56.7	0.99	73.0	0.77	66.3	0.93
Cl	69.2	0.83	51.5	1.11	76.5	0.81	80.3	0.77	53.4	1.01	55.1	0.98
TPh	95.3	0.21	74.6	0.68	74.0	0.81	80.3	0.63	80.9	0.53	64.0	0.97
Flav	100.0	0.00	83.9	0.54	89.1	0.38	93.3	0.27	87.6	0.40	72.5	0.81
NFPh	82.2	0.66	69.2	0.88	65.6	0.94	72.5	0.85	71.9	0.80	56.2	1.00
Proanth	83.2	0.52	60.8	0.90	74.0	0.81	79.8	0.71	64.6	0.78	50.6	0.98
Col_int	72.0	0.84	87.7	0.42	89.9	0.29	69.7	0.62	71.9	0.68	89.9	0.37
Hue	93.5	0.22	50.0	0.90	86.6	0.41	90.5	0.34	66.9	0.72	57.3	0.87
T_diluted	99.1	0.04	63.1	0.84	89.1	0.36	93.3	0.26	72.5	0.64	61.2	0.87
T_flav	97.2	0.10	65.4	0.90	85.7	0.45	91.0	0.34	68.5	0.70	56.2	0.95
Gly	64.5	0.87	84.6	0.67	74.0	0.94	55.6	0.99	74.2	0.76	76.4	0.84
Bu_diol	61.7	0.98	70.0	0.89	74.8	0.83	74.2	0.93	58.4	0.97	68.5	0.86
TN	68.2	0.81	71.5	0.87	60.5	0.91	52.3	0.89	63.5	0.87	65.2	0.94
Proline	88.8	0.27	94.6	0.24	67.2	0.84	69.7	0.69	91.6	0.26	75.3	0.64
Meth	56.1	1.01	42.3	0.98	54.6	1.03	60.7	1.04	52.3	0.98	52.3	1.01
LR <sub>27</sub>	100.0	0.00	99.2	0.03	99.2	0.03	98.9	0.04	99.4	0.04	98.3	0.06
LR <sub>F</sub>	100.0	0.00	99.2	0.10	98.3	0.07	98.9	0.05	99.4	0.04	98.3	0.06
LR <sub>PCA</sub>	100.0	0.00	95.4	0.16	99.2	0.01	97.2	0.09	98.3	0.07	96.6	0.25

There are 59 samples in BAR, 71 in GRI, 48 in BRB, 130 in BAR\_GRI, 119 in GRI\_BRB, and 107 in BAR\_BRB categories.

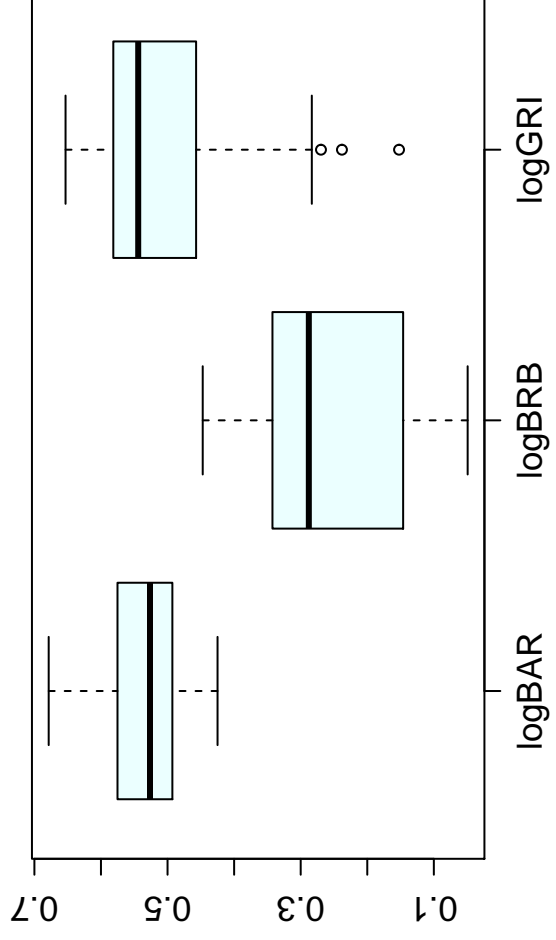
Table 3. The rates of correct classifications [%] to each of the categories within each classification problem obtained for the considered univariate and multivariate LR models.

Problem of classification into categories:												
	BAR vs. BRB		BAR vs.GRI		GRI vs. BRB		BRB vs. BAR_GRI		BAR vs. GRI_BRB		GRI vs. BAR_BRB	
LR models	BAR	BRB	BAR	GRI	GRI	BRB	BRB	BAR_GRI	BAR	GRI_BRB	GRI	BAR_BRB
Alc	72.9	62.5	96.6	87.3	77.5	81.3	81.3	51.5	89.8	70.6	87.3	88.8
Sfe	81.4	60.4	81.4	70.4	54.9	39.6	62.5	49.2	81.4	66.4	70.4	64.5
Fa	84.8	83.3	81.4	56.3	60.6	79.2	83.3	70.8	84.8	64.7	45.1	72.9
Ta	94.9	75.0	59.3	56.3	87.3	64.6	64.6	90.8	89.8	48.7	87.3	39.3
Ma	84.8	87.5	91.5	49.3	76.1	87.5	85.4	80.0	83.1	68.9	43.7	91.6
Ua	84.8	64.6	40.7	62.0	84.5	64.6	64.6	86.9	76.3	39.5	88.7	33.6
pH	67.8	56.3	66.1	38.0	63.4	45.8	54.2	65.4	66.1	49.6	16.9	43.0
Ash	39.0	52.1	78.0	50.7	39.4	87.5	87.5	28.5	66.1	53.8	45.1	86.0
Aash	74.6	87.5	71.2	67.6	36.6	85.4	87.5	53.9	71.2	75.6	71.8	38.3
K	54.2	64.6	96.6	19.7	53.5	70.8	68.8	55.4	57.6	47.9	18.3	97.2
Ca	66.1	56.3	83.1	66.2	71.8	58.3	39.6	73.1	84.8	51.3	66.2	75.7
Mg	78.0	47.9	88.1	69.0	62.0	72.9	83.3	20.8	84.8	55.5	66.2	79.4
P	84.8	66.7	86.4	66.2	18.3	79.2	62.5	54.6	86.4	66.4	62.0	69.2
Cl	81.4	54.2	66.1	39.4	88.7	58.3	58.3	88.5	76.3	42.0	87.3	33.6
TPh	98.3	91.7	93.2	59.2	71.8	77.1	87.5	77.7	98.3	72.3	57.8	68.2
Flav	100.0	100.0	94.9	74.7	90.1	87.5	97.9	91.5	98.3	82.4	81.7	66.4
NFPh	86.4	77.1	86.4	54.9	64.8	66.7	75.0	71.5	86.4	64.7	45.1	63.6
Proanth	88.1	77.1	76.3	47.9	74.7	72.9	77.1	80.8	79.7	57.1	63.4	42.1
Col_int	86.4	54.2	91.5	84.5	87.3	93.8	77.1	66.9	91.5	62.2	87.3	91.6
Hue	98.3	87.5	67.8	35.2	85.9	87.5	87.5	91.5	94.9	52.9	87.3	37.4
T_diluted	100.0	97.9	94.9	36.6	87.3	91.7	95.8	92.3	96.6	60.5	88.7	43.0
T_flav	98.3	95.8	91.5	43.7	81.7	91.7	95.8	89.2	93.2	56.3	73.2	44.9
Gly	81.4	43.8	83.1	85.9	87.3	54.2	52.1	56.9	83.1	69.8	85.9	70.1
Bu_diol	76.3	43.8	64.4	74.7	85.9	58.3	54.2	81.5	62.7	56.3	77.5	62.6
TN	66.1	70.8	91.5	54.9	39.4	91.7	91.7	37.7	86.4	52.1	54.9	72.0
Proline	88.1	89.6	94.9	94.4	53.5	87.5	91.7	61.5	89.8	92.4	78.9	72.9
Meth	55.9	56.3	47.5	38.0	56.3	52.1	47.9	65.4	39.0	58.8	39.4	60.8
LR <sub>27</sub>	100.0	100.0	100.0	98.6	98.6	100.0	100.0	98.5	100.0	99.2	95.8	100.0
LR <sub>F</sub>	100.0	100.0	100.0	98.6	97.2	100.0	100.0	98.5	100.0	99.2	97.2	99.1
LR <sub>PCA</sub>	100.0	100.0	98.3	93.0	98.6	100	95.8	97.7	96.6	99.2	93.0	99.1

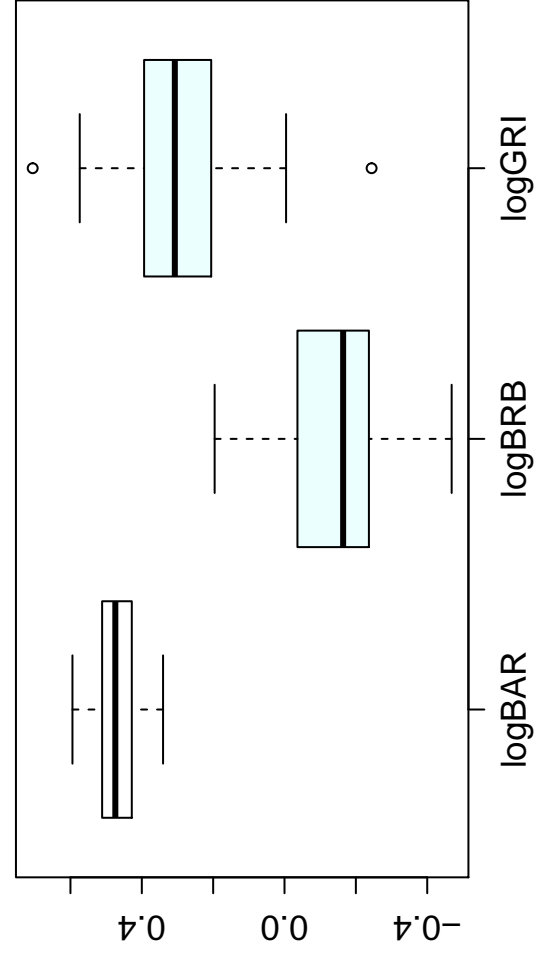
There are 59 samples in BAR, 71 in GRI, 48 in BRB, 130 in BAR\_GRI, 119 in GRI\_BRB, and 107 in BAR\_BRB categories.

Figure 1

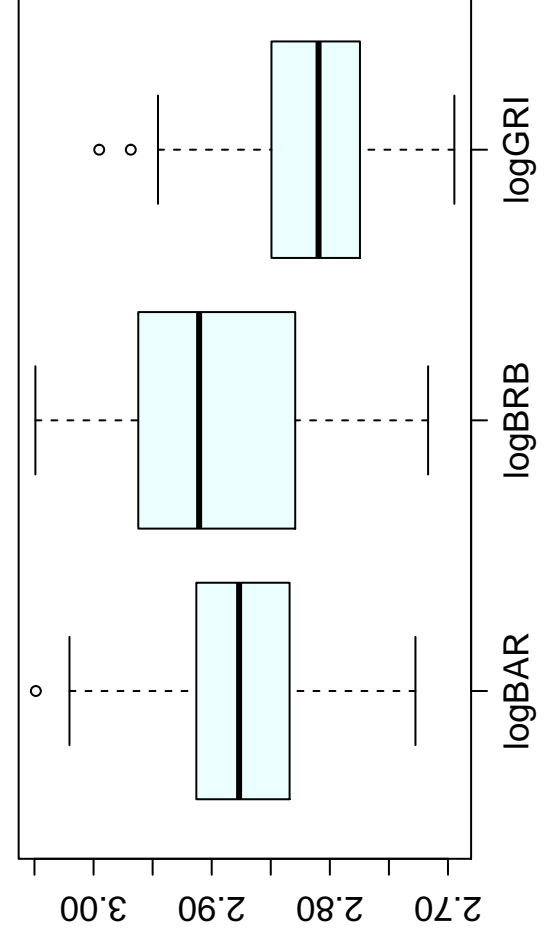
**T\_flav**



**Flav**



**Bu\_diol**



**T\_diluted**

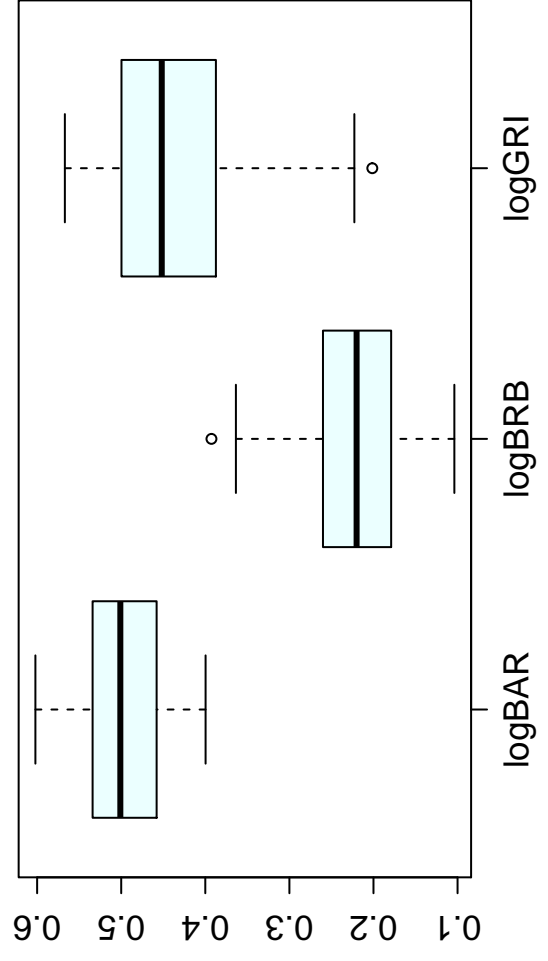


Figure 2a

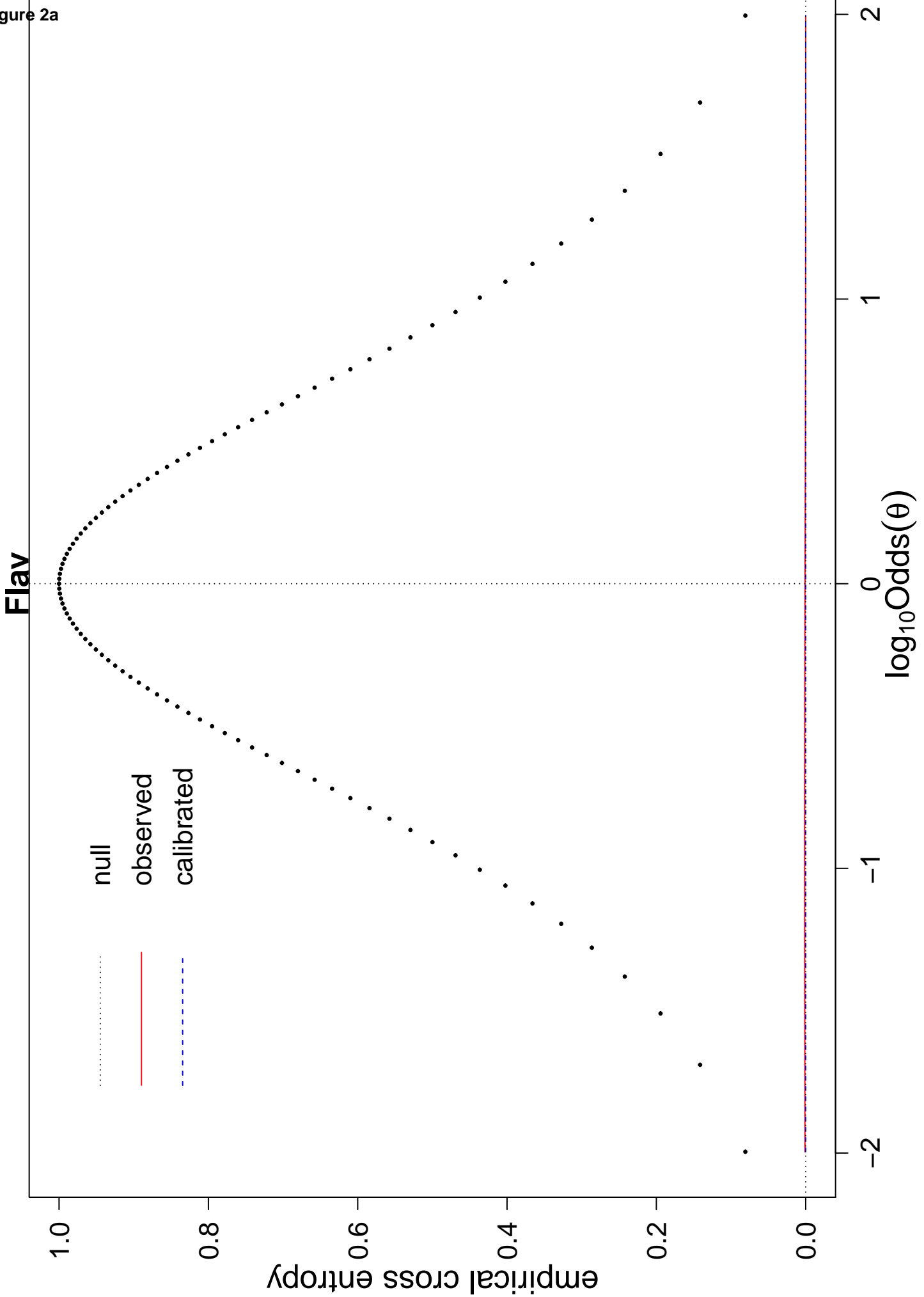


Figure 2b

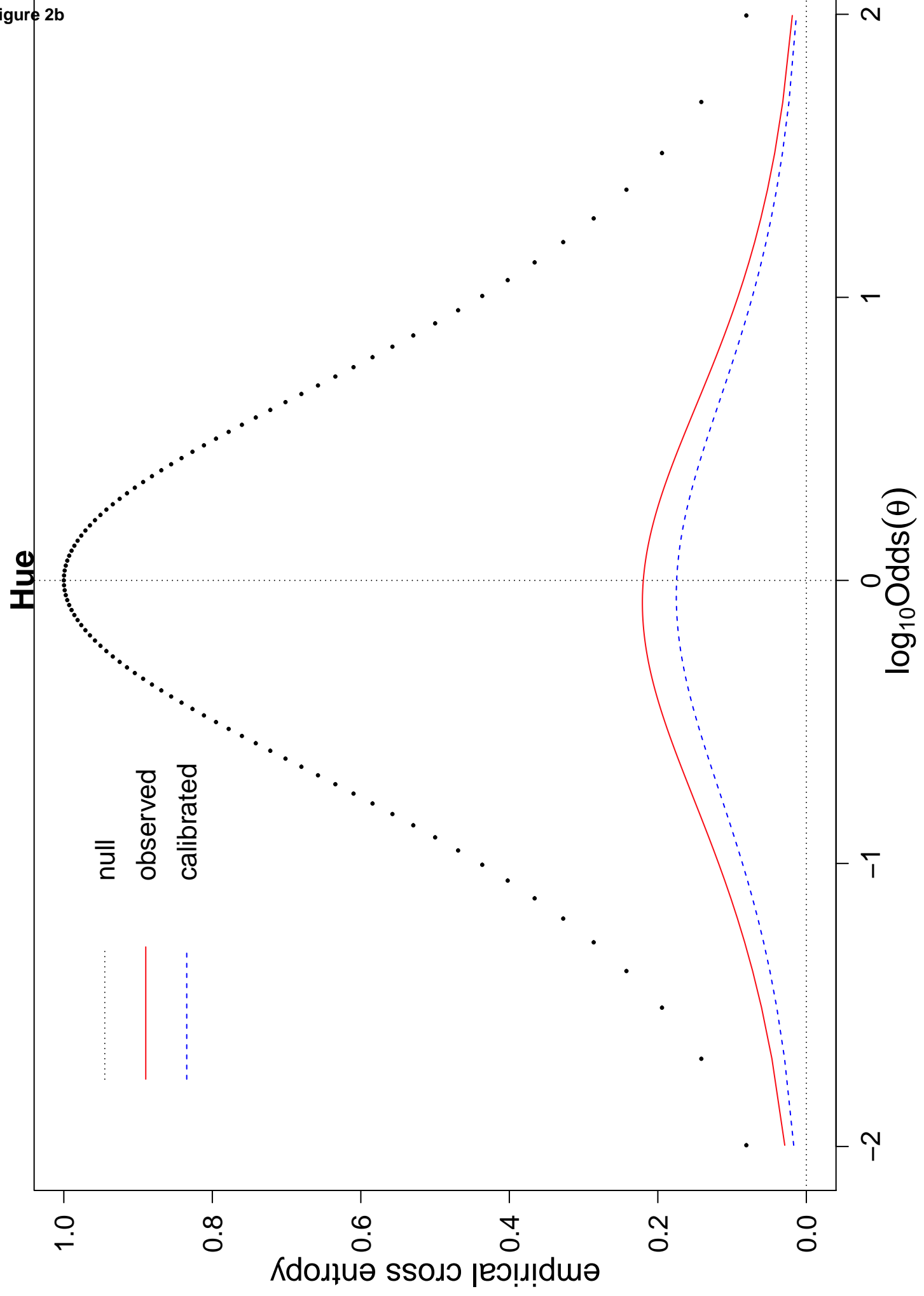




Figure 2c

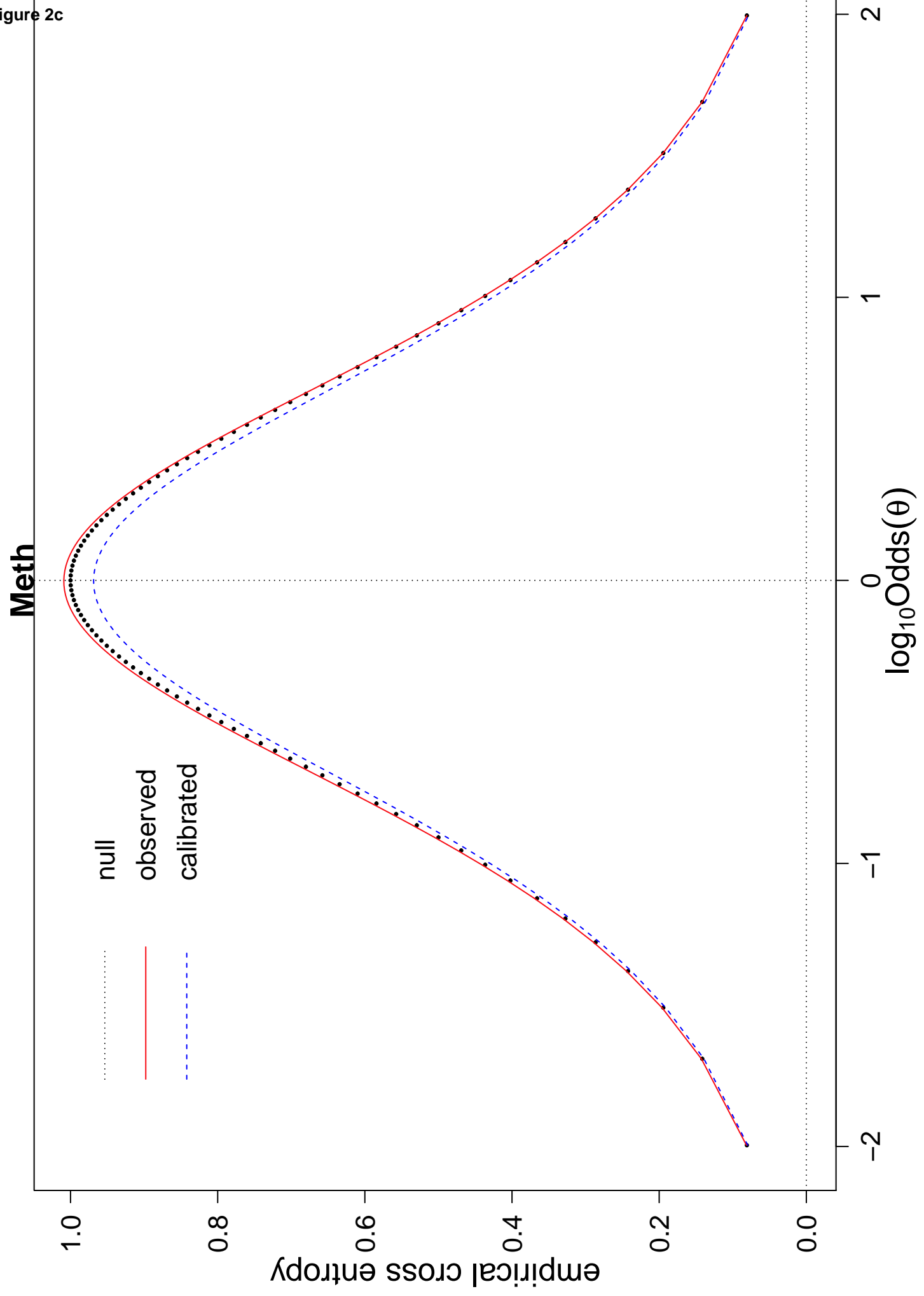


Figure 3

