Article:
Napier, Gary, Neocleous, Tereza and Nobile, Agostino orcid.org/0000-0002-5344-8525 (2015) A composite Bayesian hierarchical model of compositional data with zeros. Journal of Chemometrics. 96 - 108. ISSN: 1099-128X

https://doi.org/10.1002/cem.2681

A composite Bayesian hierarchical model of compositional data with zeros

Gary Napier* Tereza Neocleous* Agostino Nobile†

September 26, 2014

Abstract

We present an effective approach for modelling compositional data with large concentrations of zeros and several levels of variation, applied to a database of elemental compositions of forensic glass of various use types. The procedure consists of: (i) partitioning the data set in subsets characterised by the same pattern of presence/absence of chemical elements and (ii) fitting a Bayesian hierarchical model to the transformed compositions in each data subset. We derive expressions for the posterior predictive probability that newly observed fragments of glass are of a certain use type and for computing the evidential value of glass fragments relating to two competing propositions about their source. The model is assessed using cross-validation, and it performs well in both the classification and evidence evaluation tasks.

Keywords: Bayes factor; classification; evidence evaluation; forensic glass; Markov chain Monte Carlo

*School of Mathematics and Statistics, University of Glasgow

†Department of Mathematics, University of York

1 Introduction

Chemical compositions often contain large concentrations of zeros, which require special consideration during the statistical modelling process. We focus on the elemental composition of glass and construct a model that deals with this data complexity in a practical way. The elemental composition data, described in detail in Section 2, consist of the percentage weights (wt%) of each chemical element comprising a glass fragment. They contain many zeros indicating that an element is either below limits of detection or absent from the composition of a fragment. Our model accounts for the presence (above limits of detection) or absence (below limits of detection) of particular elements, and seems to improve performance in tasks related to the statistical analysis of glass fragments in a forensic context.

Analysis of glass fragments for forensic purposes usually focuses on evidence evaluation, which relates to the comparison of two sets of fragments under competing propositions, or, at the investigation stage, on the classification of a fragment into a use-type category (type of glass object from which the fragments could have originated). Most glass fragments analysed by forensic experts are too small for their use type to be determined by their thickness or colour [1], so measurements of physico-chemical features of the fragments, such as the refractive index or elemental composition, are obtained. Such measurements are also used for computing a numerical measure of the evidential value of glass fragments transferred to or from a crime scene.

In this paper we present a model that allows us to address both evidence evaluation and classification of glass fragments, while dealing with the issue of large concentrations of zeros in an effective way. Our composite model

combines models on lower-dimensional subsets of the data, which are determined by presence/absence patterns of the elements iron and potassium, and performs well in simulation studies to assess classification and evidence evaluation performance.

The paper is organised as follows: the glass data set and data transformations applied to it are described in Section 2. Section 3 presents the approach to handling compositional zeros and Section 4 describes the Bayesian hierarchical model for the forensic glass data. Section 5 gives details of how the composite model is put together and describes how the model is used to classify glass items into use-type categories. Section 6 discusses the evidence evaluation procedure. Concluding remarks are provided in Section 7.

2 Glass data

The data were provided by the Institute of Forensic Research, Krakow, and were collected in an experimental setting. Glass fragments from 320 glass objects of five use types (26 bulbs, 94 car windows, 16 headlamps, 79 containers and 105 building windows) were analysed. Their elemental content was measured using a scanning electron microscope with an energy-dispersive X-ray (SEM-EDX) spectrometer [1]. SEM-EDX produces measurements on the percentage weight (wt%) of the main elements making up the composition of the glass items. These are oxygen (O), sodium (Na), magnesium (Mg), aluminium (Al), silicon (Si), potassium (K), calcium (Ca) and iron (Fe). Three replicate measurements were taken on four glass fragments from each of the 320 items, for a total of 3840 measurements in the database.

The data are compositional: the percentage weights of each fragment are non-negative and sum to 100%. Some of the percentage weights are zero; the frequency of zeros for each element is shown in Table 1.

Table 1 about here.

Denoting the number of elements in the composition by D, the percentage weights $\mathbf{w} = (w_1, \ldots, w_D)$ satisfy $\sum_{d=1}^{D} w_d = 100$ and $w_d \geq 0$ and can be transformed by taking the ratio of $D-1$ elements to the remaining one. This removes the issue of the constrained sample space and reduces the dimension of the data vector to $D - 1$. The transformed vector is

$$\mathbf{w}^* = \left(\frac{w_1}{w_D}, \ldots, \frac{w_{D-1}}{w_D} \right), \tag{1}$$

where oxygen (O) was chosen as the common divisor, w_D, because it is always present in glass and has the highest weight percentage.

While the Dirichlet distribution would seem a natural fit to modelling compositional data, it is restrictive in a way that prevents it from detecting correlation between subcompositions from the same full composition; this is referred to as complete subcompositional independence [2, Chapter 10]. Instead, a data transformation is typically applied to (1) to achieve variance stability and normality. The most common choice of transformation for compositional data is the additive log-ratio (ALR) of Aitchison [2] which takes the logarithm of (1). Other transformations that have been applied to compositional data include the Box-Cox [3], isometric log-ratio [4], hyperspherical [5, 6, 7], centred log-ratio [8], multiplicative log-ratio [9] and complementary log-log [10] transformations.

Some members of the Box-Cox family of transformations were examined, with improvements in variance stability and normality of the data obtained by applying the square root transformation to (1). A comparison of the ALR and square root transformations can be seen in Figure 1 and shows that the square root transformation is more effective at stabilising the variability in the data. Furthermore, the square root transformation can be applied directly to compositional zeros, while logarithmic transformations require replacing them by a small constant (see Section 3). For these reasons, the square root transformation was considered the appropriate choice for the analysis of these data.

Figure 1 about here.

3 Compositional zeros

There are two types of compositional zeros: rounded zeros, indicating that if present a component is below some detection limit, and essential zeros, denoting the absolute absence of a component from an observation [12]. Compositional zeros in glass are most often treated as rounded zeros under the assumption that traces of certain elements are present but below detection. The simplest strategy then is to replace rounded zeros by some small constant equal to or below the detection limit. Techniques for doing this include the additive replacement strategy of Aitchison [2] and the multiplicative replacement strategy of Martín-Fernández et al. [13]. Palarea-Albaladejo et al. [14] introduced a parametric approach that reduces artificial correlation induced by such strategies. For the case of essential zeros Butler and Glasbey [15] introduced a latent Gaussian random variable that creates a point

mass at zero; see also [16].

Here, we partition the glass data depending on whether the elements iron and potassium are present (above the detection limit) or absent (below the detection limit) from each composition. Any $(D-1)$ dimensional composition (1) with Z zero elements is reduced to a $(D-Z-1)$ dimensional subcomposition, by simply removing the zeros. A separate model is then estimated for each resulting subset of the data. In fact, Stewart and Field [9] handled zeros in a similar way by proposing a mixture model that splits the data according to the presence or absence of components.

Observing the presence or absence of an element from the composition of a glass item can help determine its use type. For example, none of the bulbs and headlamps in our database contain iron; therefore, a composition containing iron is thought as being unlikely to be of either of these types. Other techniques of obtaining the elemental composition of glass fragments, such as laser ablation inductively coupled plasma mass spectrometry, may detect traces of such elements in these glass types [17]. Oxygen, silicon and sodium are always present in glass. The remaining five elements could be present or absent, giving 32 possible presence/absence configurations. Only 10 of these configurations were found in the glass database, with most accounting for very few items. In fact, as can be seen in Table 1, iron and potassium are responsible for 87.9% of zero measurements, thus only focusing on the presence or absence of those two elements allows for the majority of zeros to be removed from the data. We therefore consider only four configurations as shown in Table 2.

Table 2 about here.

Typically, the presence or absence of an element in a glass item is unambiguous: out of 320 items, only eight have a chemical element with its 12 measurements not all positive or all zero. In general, we assumed that an element is present in an item if at least one of its 12 measurements is positive.

Modelling of nonzero subcompositions reduces the biasing influence of zeros on the distribution of the data. This is shown in Figures 2 and 3, containing plots of the item means for the whole data set, and for the subset having configuration 2 ($\overline{\text{Fe}}$, K) from Table 2. In Figure 3, an improvement can be seen in the symmetry and concentration of points once the large mass at zero for iron is removed.

The next section discusses a Bayesian hierarchical model for the glass data. A separate model is estimated for each subset of the data with a given pattern of presence/absence of chemical elements, which we call elemental configuration. We consider four models, one for each configuration $m = 1, \ldots, M = 4$ reported in Table 2.

Figure 2 about here.

Figure 3 about here.

4 Bayesian hierarchical model

Aitken and Lucy [18] and Neocleous et al. [10] used frequentist approaches to modelling the elemental composition of glass fragments using random effects models with two levels of variation: between-item and within-item. Here we take a Bayesian approach and model the hierarchical structure of

the data using a mixed effects model. The model contains a fixed effect for the mean by glass type, and three random effects: at item level, at fragment level and at measurement level.

For each data set with a given elemental configuration m (often not explicitly indicated), we denote by $\mathbf{z}_{tijk}$ the p-vector of square roots of the compositional ratios from the k-th measurement on the j-th fragment from the i-th item of use type t, and assume that

$$\begin{gathered} \mathbf{z}_{tijk} = \boldsymbol{\theta}_t + \boldsymbol{b}_{ti} + \boldsymbol{c}_{tij} + \boldsymbol{\epsilon}_{tijk}, \\ \boldsymbol{b}_{ti} \overset{\text{iid}}{\sim} N_p(\mathbf{0}, \Omega_t^{-1}), \quad \boldsymbol{c}_{tij} \overset{\text{iid}}{\sim} N_p(\mathbf{0}, \Psi^{-1}), \quad \boldsymbol{\epsilon}_{tijk} \overset{\text{iid}}{\sim} N_p(\mathbf{0}, \Lambda^{-1}). \end{gathered} \tag{2}$$

The parameter $\boldsymbol{\theta}_t$ is the mean vector for use type t; $\boldsymbol{b}_{ti}$ is the item-level random effect; $\boldsymbol{c}_{tij}$ is the fragment within item random effect; and $\boldsymbol{\epsilon}_{tijk}$ denotes the measurement error. Multivariate normal distributions are assumed for all random effects, with unknown precision matrices, Ω_t, Ψ and Λ. The dimension p may differ across elemental configurations. The parameters in the model are collectively designated as $\xi_m = \{\boldsymbol{\theta}, \Omega, \Psi, \Lambda\}$, where $\boldsymbol{\theta} = \{\boldsymbol{\theta}_t\}_{t=1}^{T}$ and $\Omega = \{\Omega_t\}_{t=1}^{T}$; for the random effects we use the shorthands $\boldsymbol{b} = \{\boldsymbol{b}_{ti}\}_{i=1\ t=1}^{I_t\ \ T}$ and $\boldsymbol{c} = \{\boldsymbol{c}_{tij}\}_{j=1\ i=1\ t=1}^{J\ \ I_t\ \ T}$. The symbol $T = 5$ denotes the number of use types; I_t is the number of glass items of use type t ($I_1 = 26$, $I_2 = 94$, $I_3 = 16$, $I_4 = 79$, $I_5 = 105$); $J = 4$ is the number of fragments from each item; and $K = 3$ is the number of repeated measurements on each fragment. If we denote by $\mathbf{z}$ the JK measurements on an item of use type $\mathcal{T}_{\mathbf{z}} = t$, then model (2) implies that the distribution of $\mathbf{z}$, without conditioning on the random effects, is

$$\mathbf{z} | \mathcal{T}_{\mathbf{z}} = t, \xi_m \sim N_{JKp}(\mathbf{1}_{JK} \otimes \boldsymbol{\theta}_t,\, \Sigma_t), \tag{3}$$

where the covariance matrix Σ_t is given by

$$\Sigma_t = (\mathbf{1}_{JK}\mathbf{1}'_{JK}) \otimes \Omega_t^{-1} + \left[\mathbb{I}_J \otimes (\mathbf{1}_K\mathbf{1}'_K)\right] \otimes \Psi^{-1} + \mathbb{I}_{JK} \otimes \Lambda^{-1}, \qquad (4)$$

$\mathbf{1}_d$ denotes a column vector of d 1's, and $\mathbb{I}_d$ is the $d \times d$ identity matrix.

The prior distributions on the fixed effects $\boldsymbol{\theta}_t$ are independent multivariate normals truncated to the positive orthant, to ensure that the means for the square-root-transformed data are non-negative:

$$\boldsymbol{\theta}_t \overset{\text{iid}}{\sim} N_p(\mathbf{0}, \Phi^{-1}), \quad \boldsymbol{\theta}_t > \mathbf{0}, \quad t = 1, \ldots, T.$$

The covariance matrix Φ^{-1} is fixed and equal to $s \cdot \mathbb{I}_p$, with s a relatively large constant, we used $s = 1000$. Conjugate Wishart priors are placed on the precision matrices of the random effects:

$$\Omega_t \sim W_p(d_{1t}, A_t), \quad \Psi \sim W_p(d_2, B), \quad \Lambda \sim W_p(d_3, C),$$

where the degrees of freedom d_{1t}, d_2 and d_3 are all set equal to p, and the precision matrices A_t, B and C are set to $(1/1000) \cdot \mathbb{I}_p$. It was necessary to introduce separate precision matrices, Ω_t, for each glass type, due to the random variation between items having different properties across use types, as can be seen from the results in Section 4.2.

4.1 MCMC implementation

Markov chain Monte Carlo (MCMC) methods are used to sample from the joint posterior distribution of the parameters in the model, $\xi_m = \{\boldsymbol{\theta}, \Omega, \Psi, \Lambda\}$ and the random effects $\{\boldsymbol{b}, \boldsymbol{c}\}$. The full conditional distributions of all these

quantities are standard distributions and are reported in Appendix A.1. These are used to update $\boldsymbol{b}$, $\boldsymbol{c}$, Ω, Ψ and Λ, using Gibbs sampling moves; see [19] for an introduction to Gibbs sampling. The update of $\boldsymbol{\theta}_t$ is performed by means of a Metropolis-Hastings (M-H) move, with proposal distribution equal to the multivariate normal in the full conditional distribution of $\boldsymbol{\theta}_t$, disregarding the restriction to the positive orthant. The acceptance probability is either 1 or 0, depending on whether the candidate vector falls in the positive orthant or not.

We also use two additional M-H moves to update the $\boldsymbol{\theta}_t$'s. The first one is performed on $\boldsymbol{\theta}_1$ only, as the samples for this vector displayed appreciable positive autocorrelation. It is a random walk M-H move, performed on each element of $\boldsymbol{\theta}_1$ separately, with uniform proposal over an interval centred at the current value, and with interval widths determined from a preliminary run.

The second move changes at once $\boldsymbol{\theta}$ and $\boldsymbol{b}$, with the candidate state chosen in a way that leaves the likelihood unchanged. This move is a special case of the M-H algorithm as described in [20]. It is discussed in detail in appendix A.2 and, in our experiments, it substantially reduced autocorrelation of the samples.

4.2 Posterior samples for configuration $m = 2$ ($\overline{\text{Fe}}$, K)

The results shown are those for items with configuration $m = 2$ ($\overline{\text{Fe}}$, K) from Table 2. Similar results for the three other configurations are not reported here. All of the analysis was carried out using the statistical programming language R [21]. The time taken to obtain the model simulation results was

approximately 10 hours, which included a burn-in period of $10,000$, and also thinning of the Markov chain, where every 200^{th} draw was stored and the rest discarded. The acceptance rate for the M-H move performed on $\boldsymbol{\theta}_1$ only was 31%, and the rate for the joint move on $\boldsymbol{\theta}$ and $\boldsymbol{b}$ was 54%. Time series plots of the sampled fixed effect $\boldsymbol{\theta}_t$ are shown in Figure 4. Scatterplots of the draws of $\boldsymbol{\theta}_t$ are displayed in Figure 5 and show clear separation in the means between the five use-type categories.

As can be seen in Table 3, the variability at item level is shown to be rather different between use types, which is why the model accommodates for these differences by allowing the covariance matrix at item level, Ω_t^{-1}, to change by use type. When we compare the variability at fragment level, Ψ^{-1}, with that for the measurement error, Λ^{-1}, there is little difference observed, with the variability at fragment level being slightly greater, as would be expected. As expected the variability between glass items is much greater than that found within items.

Table 3 about here.

Figure 4 about here.

Figure 5 about here.

5 Composite model

In the previous section we specified multivariate mixed effects models for the square-root-transformed compositions $\mathbf{z}$, one model for each configuration $\mathcal{C}_{\mathbf{z}}$ in Table 2, and conditional on the known use types $\mathcal{T}_{\mathbf{z}}$. In this section

we show how these configuration-specific models can be pulled together in a single model. The model is then used in section 5.1 to compute the use-type probabilities for a newly observed item $\mathbf{y}$ of unknown use type. We begin with some definitions.

Let $D = \{\mathbf{z}_{ti}, i = 1, \ldots, I_t, t = 1, \ldots, T\}$ be the reference data, where the numbers I_t of items of use type t are under control of the experimenter. Also let $D_m = \{\mathbf{z} \in D : \mathcal{C}_\mathbf{z} = m\}$ be the subset of D with elemental configuration m. For any given $\mathbf{z} \in D$, the hierarchical model (2) for configuration m supplies the distribution

$$p(\mathbf{z}|\mathcal{T}_\mathbf{z} = t, \mathcal{C}_\mathbf{z} = m, \xi) = p(\mathbf{z}|\mathcal{T}_\mathbf{z} = t, \mathcal{C}_\mathbf{z} = m, \xi_m)$$

where $\xi = \{\xi_m\}_{m=1}^M$ denotes the collection of parameters across all configurations. More specifically, $p(\mathbf{z}|\mathcal{T}_\mathbf{z} = t, \mathcal{C}_\mathbf{z} = m, \xi_m)$ is given by formulae (3) and (4).

Let $\varphi_t = (\varphi_{t1}, \ldots, \varphi_{tM})$ be an unknown vector of configuration probabilities for an item $\mathbf{z}$ of use type t:

$$\varphi_{tm} \;=\; p(\mathcal{C}_\mathbf{z} = m|\mathcal{T}_\mathbf{z} = t, \varphi, \xi) \;=\; p(\mathcal{C}_\mathbf{z} = m|\mathcal{T}_\mathbf{z} = t, \varphi_t). \tag{5}$$

We assume that a priori the configuration probabilities $\varphi = \{\varphi_t\}_{t=1}^T$ are independent of ξ and have independent Dirichlet prior distributions:

$$\varphi_t|\xi \;\sim\; Dir(\alpha_{t1}, \ldots, \alpha_{tM}), \qquad t = 1, \ldots, T, \tag{6}$$

where the α_{tm}'s are suitable constants reflecting any prior information about which configurations are likely for each use type. See Section 5.1 for more

details on the choice of the α's.

Next we derive the likelihood function $\mathcal{L}(\xi, \varphi)$. The distribution of a single item $\mathbf{z} \in D$, given its type $\mathcal{T}_{\mathbf{z}}$ and the parameters ξ and φ, is

$$\begin{aligned} p(\mathbf{z}|\mathcal{T}_{\mathbf{z}} = t, \xi, \varphi) &= \sum_{r=1}^{M} p(\mathcal{C}_{\mathbf{z}} = r|\mathcal{T}_{\mathbf{z}} = t, \xi, \varphi)\, p(\mathbf{z}|\mathcal{T}_{\mathbf{z}} = t, \mathcal{C}_{\mathbf{z}} = r, \xi, \varphi) \\ &= \sum_{r=1}^{M} \varphi_{tr}\, p(\mathbf{z}|\mathcal{T}_{\mathbf{z}} = t, \mathcal{C}_{\mathbf{z}} = r, \xi_r). \end{aligned} \tag{7}$$

Therefore, the distribution of the reference data D, given ξ and φ (and the items use types, fixed by design), is

$$p(D|\xi, \varphi) = \prod_{t=1}^{T} \prod_{i=1}^{I_t} \left\{ \sum_{r=1}^{M} \varphi_{tr}\, p(\mathbf{z}_{ti}|\mathcal{T}_{\mathbf{z}_{ti}} = t, \mathcal{C}_{\mathbf{z}_{ti}} = r, \xi_r) \right\} \tag{8}$$

The likelihood $\mathcal{L}(\xi, \varphi)$ is given by (8), regarded as a function of ξ and φ, with D fixed at the observed data. This means that the configurations $\mathcal{C}_{\mathbf{z}_{ti}}$ are all known, which implies that the $\sum_r$ contains only one term corresponding to the observed configuration of $\mathbf{z}_{ti}$. Thus the likelihood can be written as

$$\begin{aligned} \mathcal{L}(\xi, \varphi) &= \prod_{t=1}^{T} \prod_{i=1}^{I_t} \varphi_{tm}\, p(\mathbf{z}_{ti}|\mathcal{T}_{\mathbf{z}_{ti}} = t, \mathcal{C}_{\mathbf{z}_{ti}} = m, \xi_m) \\ &= \left\{ \prod_{t=1}^{T} \prod_{m=1}^{M} \varphi_{tm}^{N_{tm}} \right\} \cdot \left\{ \prod_{m=1}^{M} \prod_{t=1}^{T} \prod_{i \in E_{tm}} p(\mathbf{z}_{ti}|\mathcal{T}_{\mathbf{z}_{ti}} = t, \mathcal{C}_{\mathbf{z}_{ti}} = m, \xi_m) \right\}, \end{aligned} \tag{9}$$

where $E_{tm} = \{i : \mathcal{T}_{\mathbf{z}_{ti}} = t, \mathcal{C}_{\mathbf{z}_{ti}} = m\}$ and $N_{tm} = \#E_{tm}$. In words, the N_{tm}'s are the counts in Table 2: the number of items in D that are of use type t and configuration m.

Because ξ and φ are assumed a priori independent, the preceding factori-

sation of the likelihood implies that they are also a posteriori independent. Moreover, combining the first term on the right-hand side of (9) with the prior distribution of φ in (6), yields independent Dirichlet posterior distributions for the φ_t's:

$$\varphi_t|\xi, D \sim Dir(\alpha_{t1} + N_{t1}, \ldots, \alpha_{tM} + N_{tM}), \qquad t = 1, \ldots, T. \tag{10}$$

Posterior independence of φ and ξ implies that a sample from their joint posterior distribution can be obtained in two stages: (i) sample φ from the independent posterior distributions in (10) and (ii) sample ξ_m, for each configuration m, using the MCMC procedure described in Section 4.1.

We conclude this section by remarking that formula (7) provides a mixture representation for the density of $\mathbf{z}$. This has been already recognised by Stewart and Field [9], see their formula (3.1). Because the mixture component that has generated $\mathbf{z}$ is immediately known on inspection of the item's measurements, we prefer the use of the term "composite", rather than "mixture", model.

5.1 Glass classification

The object of interest is $p(\mathcal{T}_\mathbf{y}|\mathbf{y}, D)$, that is, the posterior distribution of the use type $\mathcal{T}_\mathbf{y}$ of a newly observed glass item $\mathbf{y}$, conditional on the reference data D and the new item $\mathbf{y}$. Let the elemental configuration of $\mathbf{y}$ be $\mathcal{C}_\mathbf{y} = m$, which is known if $\mathbf{y}$ is conditioned upon. Then, using Bayes theorem,

$$\begin{aligned} p(\mathcal{T}_\mathbf{y} = t|\mathbf{y}, D) &= p(\mathcal{T}_\mathbf{y} = t|\mathbf{y}, \mathcal{C}_\mathbf{y} = m, D) \\ &\propto p(\mathcal{T}_\mathbf{y} = t|\mathcal{C}_\mathbf{y} = m, D)\, p(\mathbf{y}|\mathcal{T}_\mathbf{y} = t, \mathcal{C}_\mathbf{y} = m, D). \end{aligned} \tag{11}$$

Next, we derive expressions for the two terms in the right-hand side of (11). Beginning with the first quantity and using again Bayes theorem, one has

$$p(\mathcal{T}_{\mathbf{y}} = t|\mathcal{C}_{\mathbf{y}} = m, D) \;\propto\; p(\mathcal{T}_{\mathbf{y}} = t|D)\, p(\mathcal{C}_{\mathbf{y}} = m|\mathcal{T}_{\mathbf{y}} = t, D). \tag{12}$$

On its own, the reference data set D is not informative about $\mathcal{T}_{\mathbf{y}}$, as the use types of the items in D were under the control of the experimenter. Therefore, $p(\mathcal{T}_{\mathbf{y}} = t|D) = p(\mathcal{T}_{\mathbf{y}} = t)$ and (12) becomes

$$p(\mathcal{T}_{\mathbf{y}} = t|\mathcal{C}_{\mathbf{y}} = m, D) \;\propto\; p(\mathcal{T}_{\mathbf{y}} = t)\, p(\mathcal{C}_{\mathbf{y}} = m|\mathcal{T}_{\mathbf{y}} = t, D). \tag{13}$$

The prior distribution $p(\mathcal{T}_{\mathbf{y}} = t)$ should be chosen to reflect any available information about the prevalence of use types as forensic samples; if no such information is available, it may be set equal to a discrete uniform distribution. The second term in the right-hand side of (13) can be computed as follows:

$$\begin{aligned} p(\mathcal{C}_{\mathbf{y}} = m|\mathcal{T}_{\mathbf{y}} = t, D) &= \int p(\mathcal{C}_{\mathbf{y}} = m|\mathcal{T}_{\mathbf{y}} = t, \varphi_t, D)\, p(\varphi_t|\mathcal{T}_{\mathbf{y}} = t, D)\, \mathrm{d}\varphi_t \\ &= \int \varphi_{tm}\, p(\varphi_t|D)\, \mathrm{d}\varphi_t \\ &= \frac{\alpha_{tm} + N_{tm}}{\sum_{r=1}^{M}(\alpha_{tr} + N_{tr})}, \end{aligned} \tag{14}$$

where we used the definition of the φ's in (5) and the posterior distribution of φ_t in (10). Substituting in (13) yields the posterior distribution of the use type $\mathcal{T}_{\mathbf{y}}$ conditional only on D and the configuration $\mathcal{C}_{\mathbf{y}}$, but without conditioning on the actual new item $\mathbf{y}$:

$$p(\mathcal{T}_{\mathbf{y}} = t|\mathcal{C}_{\mathbf{y}} = m, D) \;\propto\; p(\mathcal{T}_{\mathbf{y}} = t)\, \frac{\alpha_{tm} + N_{tm}}{\sum_{r=1}^{M}(\alpha_{tr} + N_{tr})}. \tag{15}$$

Table 4 reports the values of $p(\mathcal{T}_{\mathbf{y}} = t|\mathcal{C}_{\mathbf{y}} = m, D)$ computed using $p(\mathcal{T}_{\mathbf{y}} = t) = 1/T$ and the hyperparameters $\alpha_{tm} = 0.1$ for all t and m.

Table 4 about here.

Incidentally, we remark that the choice of 0.1 as the value for the α's did not seem to matter much: we repeated five times the classification exercise reported in Section 5.2, setting the α's to 0, 0.2, 0.3, 0.4 and 0.5, and in all cases, the classifications given in Table 5 remained unchanged. Similarly, the evidence evaluation error rates reported in Section 6.1 were unaffected by changing the α's from 0.1 to 0.5.

As for the second term in (11), the posterior predictive distribution of $\mathbf{y}$ conditional on its use type, configuration and the reference data, one has

$$\begin{aligned} p(\mathbf{y}|\mathcal{T}_{\mathbf{y}} = t, \mathcal{C}_{\mathbf{y}} = m, D) = & \\ &= \int p(\mathbf{y}|\mathcal{T}_{\mathbf{y}} = t, \mathcal{C}_{\mathbf{y}} = m, \xi_m, D)\, p(\xi_m|\mathcal{T}_{\mathbf{y}} = t, \mathcal{C}_{\mathbf{y}} = m, D)\, \mathrm{d}\xi_m \\ &= \int p(\mathbf{y}|\mathcal{T}_{\mathbf{y}} = t, \mathcal{C}_{\mathbf{y}} = m, \xi_m)\, p(\xi_m|D_m)\, \mathrm{d}\xi_m \\ &= E_{\xi_m|D_m}[p(\mathbf{y}|\mathcal{T}_{\mathbf{y}} = t, \mathcal{C}_{\mathbf{y}} = m, \xi_m)], \end{aligned} \tag{16}$$

where $E_{\xi_m|D_m}$ denotes expectation with respect to the posterior distribution of ξ_m. The density $p(\mathbf{y}|\mathcal{T}_{\mathbf{y}} = t, \mathcal{C}_{\mathbf{y}} = m, \xi_m)$ is provided by the Bayesian hierarchical model (2) for elemental configuration m. To be more specific, assuming that $\mathbf{y}$ is a vector consisting of $\tilde{K}$ measurements on each of $\tilde{J}$ fragments from the same item of use type t and configuration m, then the distribution of $\mathbf{y}|\mathcal{T}_{\mathbf{y}} = t, \xi_m$ is given by formulae (3) and (4), after replacing J with $\tilde{J}$ and K with $\tilde{K}$:

$$\mathbf{y}|\mathcal{T}_{\mathbf{y}} = t, \xi_m \sim N_{\tilde{J}\tilde{K}p}(\mathbf{1}_{\tilde{J}\tilde{K}} \otimes \boldsymbol{\theta}_t, \Sigma_t)$$

$$\Sigma_t = (\mathbf{1}_{\tilde{J}\tilde{K}}\mathbf{1}'_{\tilde{J}\tilde{K}}) \otimes \Omega_t^{-1} + \left[\mathbb{I}_{\tilde{J}} \otimes (\mathbf{1}_{\tilde{K}}\mathbf{1}'_{\tilde{K}})\right] \otimes \Psi^{-1} + \mathbb{I}_{\tilde{J}\tilde{K}} \otimes \Lambda^{-1}. \quad (17)$$

Plugging (15) and (16) into (11) gives the expression for the use-type probability of a future item $\mathbf{y}$:

$$p(\mathcal{T}_\mathbf{y} = t|\mathbf{y}, D) \propto p(\mathcal{T}_\mathbf{y} = t) \frac{\alpha_{tm} + N_{tm}}{\sum_{r=1}^{M}(\alpha_{tr} + N_{tr})} E_{\xi_m|D_m}[p(\mathbf{y}|\mathcal{T}_\mathbf{y} = t, \mathcal{C}_\mathbf{y} = m, \xi_m)], \quad (18)$$

where the expectation on the right-hand side can be estimated by averaging the densities $p(\mathbf{y}|\mathcal{T}_\mathbf{y} = t, \mathcal{C}_\mathbf{y} = m, \xi_m)$ over an MCMC sample from the posterior of ξ_m, obtained as detailed in Section 4.1.

5.2 Classification results

A simulation study was conducted in order to assess the performance of the composite model in classifying glass fragments into one of the five use types (bulb, headlamp, container, car window or building window). Probabilities that each set of fragments was from an item of a certain use type were estimated using expression (18) with $p(\mathcal{T}_\mathbf{y} = t) = 1/T$ and $\alpha_{tm} = 0.1$ for all t and m. Fragments were classified into the use-type category with the highest probability.

The simulation study used five-fold cross-validation with the reference data D randomly divided into five parts, each consisting of 64 items. One part was kept as test data consisting of unobserved glass items $\mathbf{y}$. The remaining four parts were considered as training data containing reference glass items $\mathbf{z}$, from which model parameters were estimated. This was repeated five times in order to classify all 320 items.

The classification results are shown in Table 5, with Figure 6 giving an indication of the uncertainty about the classification of each item. The overall misclassification rate was 20.6%, reflecting good classification performance for bulbs, containers and headlamps, and poorer performance for car and building windows. From Table 5, it is clear that misclassification of a window type is most often to the other window type. This is because car and building windows have a very similar elemental composition, thus making it difficult to correctly distinguish between them based on this alone. Zadora [1] reports improved classification rates for car and building windows when, in addition to the elemental composition, the refractive index before and after annealing is used.

Table 5 about here.

Comparing our classification results with those obtained using a hierarchical model that does not take into account the configurations, shows that the composite model leads to a reduction in the number of items misclassified (misclassification rates of 20.6% for the composite model compared with 22.8% for the model without configurations). In addition, the composite model achieves a lower misclassification rate (20.6%) than support vector machines (SVM) [22] (22.8%). The composite model also outperforms SVM across two other classification performance measures: Cohen's kappa ($\kappa = 0.721$ for the composite model compared with 0.688 for SVM) and Brier score ($BS = 0.319$ for the composite model, 0.447 for SVM). Cohen's kappa [23] is a measure of agreement ranging from 0 to 1 with $\kappa = 1$ indicating perfect agreement. The Brier score [24] is a measure of prediction strength with $BS = 0$ implying perfect predictions. For comparison, the corresponding performance measures for the hierarchical model

without considering configurations are $\kappa = 0.693$ and $BS = 0.338$. In both comparisons, the benefit of modelling the configurations is clear.

Figure 6 about here.

6 Evidence evaluation

Glass fragments obtained from a suspect can be used as evidence in support of (or against) the proposition that the suspect was at the scene of the crime. The statistical approach to evaluating the strength or value V of such evidence stems from [25]; a recent overview is provided by [26, Chapter 10], whose terminology we adopt. Let E be the evidence, H_p the prosecution proposition, H_d the defence proposition and I additional background information related to the case. The value of the evidence for H_p, also known in the forensic literature as the *likelihood ratio*, is defined as the factor by which to multiply the prior odds on H_p, as a result of observing E: $V = \Pr(E|H_p, I)/\Pr(E|H_d, I)$. Typically, the probability statements in V are obtained by integrating over the posterior distributions of unknown parameters, then the appropriate term is *Bayes factor* [27] for H_p and against H_d, on evidence E.

More specifically, denote by $\mathbf{x}$ the measurements collected from a sample of glass fragments found at the crime scene (source evidence) and by $\mathbf{y}$ the measurements obtained from fragments found on the suspect (receptor object), under the assumption that all fragments in $\mathbf{y}$ are from the same item. The glass evidence comprises both control and recovered samples: $E = (\mathbf{x}, \mathbf{y})$. The prosecution proposition, H_p, is that $\mathbf{y}$ is from the same

item as $\mathbf{x}$, that is, the fragments found on the suspect originated from the broken item found at the crime scene. The defence proposition, H_d, is that $\mathbf{y}$ is not from the same item as $\mathbf{x}$, that is, the fragments found on the suspect originated from some source outwith the crime scene. The value of the glass evidence is then

$$V = \frac{p(\mathbf{x}, \mathbf{y}|H_p, I)}{p(\mathbf{x}, \mathbf{y}|H_d, I)}. \tag{19}$$

Here, we assume that two sets of fragments $\mathbf{x}$ and $\mathbf{y}$ do not come from the same item if their elemental configurations do not match; that is, $\mathcal{C}_\mathbf{x} \neq \mathcal{C}_\mathbf{y}$, yielding $V = 0$. This assumption may not always hold in practice as it is possible for two sets of fragments from the same item to have non-matching configurations, although for the glass data, this occurs very rarely (less than 1% of the time). In the following, we assume that $\mathcal{C}_\mathbf{x} = \mathcal{C}_\mathbf{y} = \mathcal{C} = m$. Let $\mathcal{T}_\mathbf{x} = t$ be the known use type of the glass recovered at the crime scene. Under H_p, one has that $\mathcal{T}_\mathbf{y} = \mathcal{T}_\mathbf{x}$, while under H_d, the use type of $\mathbf{y}$ is uncertain. Dropping in (19) the explicit conditioning on I, save for the reference data set D used to assess the competing propositions, the known elemental configuration $\mathcal{C}$, and the known use type $\mathcal{T}_\mathbf{x}$ of $\mathbf{x}$, one has:

$$V = \frac{p(\mathbf{x}, \mathbf{y}|\mathcal{T}_\mathbf{x} = t, \mathcal{C} = m, D, H_p)}{p(\mathbf{x}, \mathbf{y}|\mathcal{T}_\mathbf{x} = t, \mathcal{C} = m, D, H_d)}. \tag{20}$$

We show in Appendix A.3 that V can be rewritten as

$$V = \frac{E_{\xi_m|D_m}\left[p(\mathbf{x}, \mathbf{y}|\mathcal{T}_{(\mathbf{x},\mathbf{y})} = t, \mathcal{C} = m, \xi_m)\right]}{\sum_{s=1}^{T} p(\mathcal{T}_\mathbf{y} = s|\mathcal{C} = m, D)\, E_{\xi_m|D_m}[p(\mathbf{x}|\mathcal{T}_\mathbf{x} = t, \mathcal{C} = m, \xi_m)\, p(\mathbf{y}|\mathcal{T}_\mathbf{y} = s, \mathcal{C} = m, \xi_m)]}, \tag{21}$$

where $p(\mathcal{T}_\mathbf{y} = s|\mathcal{C} = m, D)$ is given in (15). The density $p(\mathbf{x}, \mathbf{y}|\mathcal{T}_{(\mathbf{x},\mathbf{y})} = t, \mathcal{C} = m, \xi_m)$ in the numerator is that of a $N_{\tilde{J}\tilde{K}p}(\mathbf{1}_{\tilde{J}\tilde{K}} \otimes \boldsymbol{\theta}_t, \Sigma_t)$ distribution, where

$\tilde{J} = \tilde{J}_{\mathbf{x}} + \tilde{J}_{\mathbf{y}}$ is the total number of fragments obtained, $\tilde{K}$ is the number of measurements taken on each fragment, and the covariance matrix Σ_t has the expression given in (17). The densities $p(\mathbf{x}|\mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m)$ and $p(\mathbf{y}|\mathcal{T}_{\mathbf{y}} = s, \mathcal{C} = m, \xi_m)$ in the denominator are of $N_{\tilde{J}_{\mathbf{x}}\tilde{K}p}(\mathbf{1}_{\tilde{J}_{\mathbf{x}}\tilde{K}} \otimes \boldsymbol{\theta}_t, \Sigma_{t\mathbf{x}})$ and $N_{\tilde{J}_{\mathbf{y}}\tilde{K}p}(\mathbf{1}_{\tilde{J}_{\mathbf{y}}\tilde{K}} \otimes \boldsymbol{\theta}_t, \Sigma_{t\mathbf{y}})$ respectively, where $\Sigma_{t\mathbf{x}}$ and $\Sigma_{t\mathbf{y}}$ are given by formula (17), with $\tilde{J}$ replaced by $\tilde{J}_{\mathbf{x}}$ and $\tilde{J}_{\mathbf{y}}$.

6.1 Evidence evaluation results

The performance of the composite model in the evidence evaluation task was assessed in terms of the percentage of false negative (FN) and false positive (FP) answers produced in a simulation study. An FN occurs when glass fragments from the same item are evaluated as originating from different items, a decision that is made whenever $V \leq v$, for some threshold value v. An FP happens when glass fragments are from different items, but $V > v$ so that they are evaluated as coming from the same item. Here, V is obtained using (21), with parameter values and resulting $p(\mathcal{T}_{\mathbf{y}} = s|\mathcal{C} = m, D)$ estimates as in Table 4.

Five-fold cross-validation was used in the simulation study to estimate the percentages of FN and FP answers. Each test set consisted of 64 items with the percentage of FN answers obtained by randomly choosing two fragments from each item as the source evidence, $\mathbf{x}$, and comparing them with the remaining two fragments from the same item as the receptor sample, $\mathbf{y}$. This was repeated for each of the five test sets yielding a total of 320 same-source comparisons. The percentage of FP answers was obtained by taking all 12 measurements from an item as $\mathbf{x}$ and all 12 measurements from another item as $\mathbf{y}$, and comparing all $\binom{64}{2} = 2016$ possible item pairs in a test set. This

was repeated for each of the five test sets giving a total of $5 \times \binom{64}{2} = 10,080$ comparisons. Note that because many more comparisons were made for different-source pairs than for same-source pairs, estimates of FN rates are more uncertain than those of FP rates.

Using as threshold $v = 1$, the rates of FNs and FPs produced by cross-validation were 4.4% and 1.4%, respectively, which are improvements on previous publications with similar glass databases; see [10]. However, the two types of error are of different seriousness: because incorrectly evaluating two sets of fragments as originating from the same item may contribute to the conviction of an innocent person, emphasis should be placed on avoiding false positives. This is readily achieved by varying the threshold v. To each value of v there corresponds a pair of error rates and these are represented in the receiver operating characteristic (ROC) curve displayed in Figure 7, where as usual the true positive (TP) rate (TP rate $= 1-$FN rate) is plotted against the FP rate. The ROC curve is very steep in the region of FP rates close to 0: small reductions in the FP rate, say below 1%, can be achieved only at the cost of noticeably increasing the FN rate to about 10% or more. The value of the area under the curve is 0.99, a value of 1 is achieved by an "ideal" procedure with zero FP rate and unit TP rate.

Figure 7 about here.

7 Conclusion

We have presented a composite model to deal with a large point mass at zero for various components of glass elemental compositions. The glass data

set was partitioned according to the presence or absence of the elements iron and potassium, and a Bayesian hierarchical model was fit to each resulting subset of the data. While this approach allows for the majority of compositional zeros to be accounted for, a small proportion of zeros persists, as seen from Figure 3, mainly occurring when the element magnesium is absent from a composition. To check whether accounting for these additional zeros would improve results, we split configuration 2 in Table 2, $(\overline{\text{Fe}}, \text{K})$, into two configurations based on the presence or absence of magnesium, $(\overline{\text{Fe}}, \text{K}, \text{Mg})$ and $(\overline{\text{Fe}}, \text{K}, \overline{\text{Mg}})$. Repeating the analysis with the resulting five configurations did not change or improve upon the classification and evidence evaluation results obtained using the original four configurations.

Before proceeding with the analysis, we have applied a square root transformation to the ratios of chemical elements' contents to that of oxygen. This is a departure from the more commonly used ALR transformation. We have found that, in addition to being more effective at stabilising the variability in these data, use of the square root also meant that any remaining zeros in the data did not require further special treatment such as replacement by a small amount.

Our hierarchical model is more general than previous random effects models for similar data [10, 18] as it contains a fixed effect for use type of glass and three levels of variability (item, fragment and measurement) and allows for different variances for each use type of glass. A normality assumption was made for the distribution of all random effects, which may not be ideal for the between-item distribution in particular. An alternative would be to use mixture models – this is an area of future work. However, the simplicity of a normal linear mixed model is rather appealing, especially when it produces

satisfactory results, as is the case here in both classification and evidence evaluation tasks.

The composite model outperforms SVM as well as a hierarchical model without configurations in the classification task, with very good results obtained for the classification of glass items of use types bulb, headlamp and container. The relatively high overall misclassification rate of 20.6% is due to the difficulty in distinguishing between car and building windows, which are manufactured in a similar way and have similar elemental compositions. However, whenever a window is misclassified, it is most often misclassified as the other window type. Perhaps different glass measurements such as the refractive index would be more useful in distinguishing between window types.

The performance of the composite model in the evidence evaluation task is also good. The FP and FN rates, obtained using cross-validation and giving equal importance to the two types of error, were 1.4% and 4.4%, respectively. More generally, the ROC curve (area under the curve = 0.99) shows that very low FP rates can be achieved, if one is willing to accept moderate FN rates.

Acknowledgements

The authors would like to thank Prof. Grzegorz Zadora for providing us with access to the glass database used in the analysis. Gary Napier acknowledges financial support under a doctoral training grant from the U.K. Engineering and Physical Sciences Research Council.

A Appendix

A.1 Full conditional distributions

The full conditional distribution of all the unknown quantities in the model are reported below. We use "$|\cdots$" to mean "conditionally on all the other variables".

- $\boldsymbol{\theta}_t \mid \cdots \sim N_p(\tilde{\boldsymbol{\phi}}_t, \tilde{\Phi}_t^{-1}), \quad \boldsymbol{\theta}_t > \mathbf{0}$,

  where $\tilde{\boldsymbol{\phi}}_t = \tilde{\Phi}_t^{-1}\Big[JKI_t(\bar{\mathbf{z}}_{t\cdots} - \bar{\boldsymbol{b}}_{t\cdot} - \bar{\boldsymbol{c}}_{t\cdot\cdot})\Big]$, and $\tilde{\Phi}_t = JKI_t\Lambda + \Phi$.

- $\boldsymbol{b}_{ti} \mid \cdots \sim N_p(\tilde{\boldsymbol{\omega}}_{ti}, \tilde{\Omega}_t^{-1})$,

  where $\tilde{\boldsymbol{\omega}}_{ti} = \tilde{\Omega}_t^{-1}\Big[JK\Lambda(\bar{\mathbf{z}}_{ti\cdot\cdot} - \boldsymbol{\theta}_t - \bar{\boldsymbol{c}}_{ti\cdot})\Big]$, and $\tilde{\Omega}_t = JK\Lambda + \Omega_t$.

- $\boldsymbol{c}_{tij} \mid \cdots \sim N_p(\tilde{\boldsymbol{\psi}}_{tij}, \tilde{\Psi}^{-1})$,

  where $\tilde{\boldsymbol{\psi}}_{tij} = \tilde{\Psi}^{-1}\Big[K\Lambda\big(\bar{\mathbf{z}}_{tij\cdot} - \boldsymbol{\theta}_t - \boldsymbol{b}_{ti}\big)\Big]$, and $\tilde{\Psi} = K\Lambda + \Psi$.

- $\Omega_t \mid \cdots \sim W_p(\tilde{d}_{1t}, \tilde{A}_t)$,

  where $\tilde{d}_{1t} = d_{1t} + I_t$, and $\tilde{A}_t = A_t + \sum_{i=1}^{I_t} \boldsymbol{b}_{ti}\boldsymbol{b}_{ti}'$.

- $\Psi \mid \cdots \sim W_p(\tilde{d}_2, \tilde{B})$,

  where $\tilde{d}_2 = d_2 + J\sum_{t=1}^{T} I_t$, and $\tilde{B} = B + \sum_{t=1}^{T}\sum_{i=1}^{I_t}\sum_{j=1}^{J} \boldsymbol{c}_{tij}\boldsymbol{c}_{tij}'$.

- $\Lambda \mid \cdots \sim W_p(\tilde{d}_3, \tilde{C})$,

  where $\tilde{d}_3 = d_3 + JK\sum_{t=1}^{T} I_t$, and $\tilde{C} = C + \sum_{t=1}^{T}\sum_{i=1}^{I_t}\sum_{j=1}^{J}\sum_{k=1}^{K}(\mathbf{z}_{tijk} - (\boldsymbol{\theta}_t + \boldsymbol{b}_{ti} + \boldsymbol{c}_{tij}))(\mathbf{z}_{tijk} - (\boldsymbol{\theta}_t + \boldsymbol{b}_{ti} + \boldsymbol{c}_{tij}))'$.

A.2 MH move on $\boldsymbol{\theta}$ and $\boldsymbol{b}$ jointly

This move updates simultaneously the fixed effect $\boldsymbol{\theta}_t$ and the random effects at item level $\boldsymbol{b}_{ti}$, separately for each use type t. The candidate state is chosen to leave $\boldsymbol{\theta}_t + \boldsymbol{b}_{ti}$ unchanged, so that it shares with the current state the same value of the likelihood. The candidate for the fixed effect is obtained as

$$\tilde{\boldsymbol{\theta}}_t = \boldsymbol{\theta}_t + \boldsymbol{v},$$

where $\boldsymbol{v} = (v_1, \ldots, v_p)'$, with components $v_l \sim Unif(-\delta_{tl}, \delta_{tl})$ independently, with interval widths determined from a preliminary run. The candidates for the random effects are then set to

$$\tilde{\boldsymbol{b}}_{ti} = \boldsymbol{b}_{ti} - \boldsymbol{v}, \qquad i = 1, \ldots, I_t.$$

Since the likelihood is left unchanged, the ratio of target densities reduces to the ratio of prior densities evaluated at the candidate and current state. Then, following the approach in §2.2 of [20], the acceptance probability can be computed as

$$\alpha = \min\left(1, \frac{p(\tilde{\boldsymbol{\theta}}_t)\, p(\tilde{\boldsymbol{b}}_t|\Omega_t)}{p(\boldsymbol{\theta}_t)\, p(\boldsymbol{b}_t|\Omega_t)} \frac{\tilde{f}(\tilde{\boldsymbol{v}})}{f(\boldsymbol{v})} \left|\frac{\partial(\tilde{\boldsymbol{\theta}}_t, \tilde{\boldsymbol{b}}_t, \tilde{\boldsymbol{v}})}{\partial(\boldsymbol{\theta}_t, \boldsymbol{b}_t, \boldsymbol{v})}\right|\right)$$

where f is the density, uniform on a p-hyperrectangle, of the random numbers $\boldsymbol{v}$, and $\tilde{f} = f$ is the density of the random numbers $\tilde{\boldsymbol{v}} = -\boldsymbol{v}$ in the

reverse move. The absolute value of the determinant of Jacobian matrix is

$$\left|\frac{\partial(\tilde{\boldsymbol{\theta}}_t, \tilde{\boldsymbol{b}}_t, \tilde{\boldsymbol{v}})}{\partial(\boldsymbol{\theta}_t, \boldsymbol{b}_t, \boldsymbol{v})}\right| = \begin{vmatrix} \mathbb{I}_p & 0_{p,pI_t} & \mathbb{I}_p \\ 0_{pI_t,p} & \mathbb{I}_{pI_t} & -\mathbf{1}_{I_t} \otimes \mathbb{I}_p \\ 0_{p,p} & 0_{p,pI_t} & -\mathbb{I}_p \end{vmatrix} = 1.$$

Therefore the acceptance probability only involves the prior density at the current and candidate values:

$$\alpha = \min\left(1, \frac{p(\tilde{\boldsymbol{\theta}}_t)\, p(\tilde{\boldsymbol{b}}_t|\Omega_t)}{p(\boldsymbol{\theta}_t)\, p(\boldsymbol{b}_t|\Omega_t)}\right).$$

A.3 Computing the evidence value V

Here we show how to derive the expression (21) for V. Starting with (20),

$$\begin{aligned} V &= \frac{p(\mathbf{x}, \mathbf{y}|\mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, D, H_p)}{p(\mathbf{x}, \mathbf{y}|\mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, D, H_d)} \\ &= \frac{\int p(\mathbf{x}, \mathbf{y}|\mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m, D, H_p)\, p(\xi_m|\mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, D, H_p)\, \mathrm{d}\xi_m}{\int p(\mathbf{x}, \mathbf{y}|\mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m, D, H_d)\, p(\xi_m|\mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, D, H_d)\, \mathrm{d}\xi_m}. \end{aligned} \tag{22}$$

For the first term of the integrand in the numerator one has

$$\begin{aligned} p(\mathbf{x}, \mathbf{y}|\mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m, D, H_p) &= \sum_{s=1}^{T} p(\mathbf{x}, \mathbf{y}|\mathcal{T}_{\mathbf{x}} = t, \mathcal{T}_{\mathbf{y}} = s, \mathcal{C} = m, \xi_m, D, H_p) \\ &\qquad \cdot p(\mathcal{T}_{\mathbf{y}} = s|\mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m, D, H_p) \\ &= p(\mathbf{x}, \mathbf{y}|\mathcal{T}_{\mathbf{x}} = t, \mathcal{T}_{\mathbf{y}} = t, \mathcal{C} = m, \xi_m, D) \\ &= p(\mathbf{x}, \mathbf{y}|\mathcal{T}_{(\mathbf{x},\mathbf{y})} = t, \mathcal{C} = m, \xi_m) \end{aligned}$$

since H_p implies that $\mathcal{T}_{\mathbf{y}} = \mathcal{T}_{\mathbf{x}} = t$, the known use type of $\mathbf{x}$, and we use the notation $\mathcal{T}_{(\mathbf{x},\mathbf{y})}$ to emphasize that, under H_p, $\mathbf{x}$ and $\mathbf{y}$ are measurements from the same glass item. Then the numerator of (22) can be written as

$$\int p(\mathbf{x},\mathbf{y}|\mathcal{T}_{(\mathbf{x},\mathbf{y})} = t, \mathcal{C} = m, \xi_m)\, p(\xi_m|D_m)\, \mathrm{d}\xi_m = E_{\xi_m|D_m}\left[p(\mathbf{x},\mathbf{y}|\mathcal{T}_{(\mathbf{x},\mathbf{y})} = t, \mathcal{C} = m, \xi_m)\right]. \quad (23)$$

Consider next the denominator of (22). Under H_d and conditional on the parameters ξ_m, $\mathbf{x}$ is independent of $\mathbf{y}$ so that

$$\begin{aligned} p(\mathbf{x},\mathbf{y}|\mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m, D, H_d) \;&= \\ &= p(\mathbf{x}|\mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m, D, H_d)\, p(\mathbf{y}|\mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m, D, H_d) \\ &= p(\mathbf{x}|\mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m)\, p(\mathbf{y}|\mathcal{C} = m, \xi_m, D). \end{aligned} \quad (24)$$

The second term on the right hand side of (24) can be written as

$$\begin{aligned} p(\mathbf{y}|\mathcal{C} = m, \xi_m, D) &= \sum_{s=1}^{T} p(\mathbf{y}|\mathcal{T}_{\mathbf{y}} = s, \mathcal{C} = m, \xi_m, D)\, p(\mathcal{T}_{\mathbf{y}} = s|\mathcal{C} = m, \xi_m, D) \\ &= \sum_{s=1}^{T} p(\mathbf{y}|\mathcal{T}_{\mathbf{y}} = s, \mathcal{C} = m, \xi_m)\, p(\mathcal{T}_{\mathbf{y}} = s|\mathcal{C} = m, D), \end{aligned}$$

where $p(\mathcal{T}_{\mathbf{y}} = s|\mathcal{C} = m, D)$ is given in (15). It then follows that the denominator of (22) is

$$\begin{aligned} &\int p(\mathbf{x}|\mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m) \\ &\qquad \left[\sum_{s=1}^{T} p(\mathbf{y}|\mathcal{T}_{\mathbf{y}} = s, \mathcal{C} = m, \xi_m)\, p(\mathcal{T}_{\mathbf{y}} = s|\mathcal{C} = m, D)\right] p(\xi_m|D_m)\, \mathrm{d}\xi_m = \\ &= \sum_{s=1}^{T} p(\mathcal{T}_{\mathbf{y}} = s|\mathcal{C} = m, D) \\ &\qquad E_{\xi_m|D_m}[p(\mathbf{x}|\mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m)\, p(\mathbf{y}|\mathcal{T}_{\mathbf{y}} = s, \mathcal{C} = m, \xi_m)]. \end{aligned} \quad (25)$$

Substituting (23) and (25) into (22) yields (21).

References

1. Zadora G. Classification of glass fragments based on elemental composition and refractive index. *J. Forensic Sci.* 2009; 54(1):49–59.

2. Aitchison J. *The statistical analysis of compositional data*. Chapman & Hall, 1986.

3. Rayens WS, Srinivasan C. Box-Cox transformations in the analysis of compositional data. *J. Chemometr.* 1991; 5(3):227–239.

4. Egozcue JJ, Pawlowsky-Glahn V, Mateu-Figueras G, Barceló-Vidal C. Isometric logratio transformations for compositional data analysis. *Math. Geol.* 2003; 35(3):279–300.

5. Scealy JL, Welsh AH. Regression for compositional data by using distributions defined on the hypersphere. *J. R. Stat. Soc. B* 2011; 73(3): 351–375.

6. Scealy JL, Welsh AH. Fitting Kent models to compositional data with small concentration. *Stat. Comput.* 2014; 24(2): 165–179.

7. Wang H, Liu Q, Mok HMK, Fu L, Tse WM. A hyperspherical transformation forecasting model for compositional data. *Eur. J. Oper. Res.* 2007; 179(2):459–468.

8. Campbell GP, Curran JM, Miskelly GM, Coulson S, Yaxley GM, Grunsky EC, Cox SC. Compositional data analysis for elemental data in forensic science. *Forensic Sci. Int.* 2009; 188(1–3):81–90.

9. Stewart C, Field C. Managing the essential zeros in quantitative fatty acid signature analysis. *J. Agr. Bio. Envir. St.* 2011; 16(1):45–69.

10. Neocleous T, Aitken CGG, Zadora G. Transformations for compositional data with zeros with an application to forensic evidence evaluation. *Chemometr. Intell. Lab.* 2011; 109(1):77–85.

11. Box GEP, Cox DR. An analysis of transformations. *J. R. Stat. Soc. B* 1964; 26(2):211–252.

12. Martín-Fernández JA, Barceló-Vidal C, Pawlowsky-Glahn V. Dealing with zeros and missing values in compositional data sets using nonparametric imputation. *Math. Geol.* 2003; 35(3):253–277.

13. Martín-Fernández JA, Barceló-Vidal C, Pawlowsky-Glahn V. Zero replacement in compositional data sets. In Kiers H, Rasson J, Groenen P, Shader M, editors, *Proceedings of the 7th Conference of the International Federation of Classification Societies, (University of Namur (Belgium))*, pages 155–160. Springer-Verlag, Berlin, 2000.

14. Palarea-Albaladejo J, Martín-Fernández JA, Gómez-García JG. A parametric approach for dealing with compositional rounded zeros. *Math. Geol.* 2007; 39(7):625–645.

15. Butler A, Glasbey C. A latent Gaussian model for compositional data with zeros. *J. R. Stat. Soc. C* 2008; 57(5):505–520.

16. Leininger TL, Gelfand AE, Allen JM, Silander Jr JA. Spatial regression modeling for compositional data with many zeros. *J. Agr. Bio. Envir. St.* 2013; 18(3):314–334.

17. Trejos T, Almirall JR. Sampling strategies for the analysis of glass fragments by LA-ICP-MS Part 1: micro-homogeneity study of glass and its application to the interpretation of forensic evidence. *Talanta* 2005; 67(2):396–401.

18. Aitken CGG, Lucy D. Evaluation of trace evidence in the form of multivariate data. *J. R. Stat. Soc. C* 2004; 53(1):109–122.

19. Casella G, George EI. Explaining the Gibbs Sampler. *Am. Stat.* 1992; 46(3):167–174.

20. Green PJ. Trans-dimensional Markov chain Monte Carlo. In Green PJ, Hjort NL, Richardson S, editors, *Highly structured stochastic systems*, pages 179–196. Oxford: Oxford University Press, 2003.

21. R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. URL http://www.R-project.org.

22. Cortes C, Vapnik VN. Support-vector networks. *Mach. Learn.* 1995; 20 (3):273–297.

23. Cohen J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* 1960; 20(1):37–46.

24. Brier GW. Verification of forecasts expressed in terms of probability. *Mon. Weather Rev.* 1950; 78(1):1–3.

25. Lindley DV. A problem in forensic science. *Biometrika* 1977; 64(2): 207–213.

26. Aitken CGG, Taroni F. *Statistics and the evaluation of evidence for forensic scientists*. John Wiley and Sons, 2004.

27. Kass RE, Raftery AE. Bayes factors. *J. Am. Statist. Ass.* 1995; 90(430): 773–795.

Table 1: Frequency of zero measurements by chemical element.

Element	O	Si	Na	Ca	Al	Mg	K	Fe
Frequency	0	0	0	108	205	265	1168	3036
Percentage	0.0	0.0	0.0	2.8	5.3	6.9	30.4	79.1

Table 2: Presence (Fe, K) and absence ($\overline{\text{Fe}}, \overline{\text{K}}$) at item level by use type.

Glass type	**Configuration** m				**Total**
	1: Fe, K	2: $\overline{\text{Fe}}$, K	3: Fe, $\overline{\text{K}}$	4: $\overline{\text{Fe}}$, $\overline{\text{K}}$	
bulb	0	25	0	1	26
car window	23	40	11	20	94
headlamp	0	14	0	2	16
container	12	48	0	19	79
building window	7	55	15	28	105
	42	182	26	70	320

Table 3: Standard deviations (multiplied by 10) from covariance matrices Ω_t^{-1}, Ψ^{-1} and Λ^{-1}. For Ω_t^{-1}, $t = 1, \ldots, 5$ correspond to use types: bulb, car window, headlamp, container and building window.

	Na	Mg	Al	Si	K	Ca
Ω_1^{-1}	0.95	0.94	0.28	0.65	1.12	1.34
Ω_2^{-1}	0.10	0.25	0.26	0.30	0.15	0.26
Ω_3^{-1}	0.18	0.91	0.52	0.51	0.31	0.78
Ω_4^{-1}	0.12	0.51	0.15	0.44	0.19	0.37
Ω_5^{-1}	0.10	0.13	0.13	0.26	0.13	0.22
Ψ^{-1}	0.07	0.05	0.09	0.30	0.11	0.25
Λ^{-1}	0.04	0.03	0.04	0.15	0.06	0.11

Table 4: Use type probabilities $p(\mathcal{T}_{\mathbf{y}} = t|\mathcal{C}_{\mathbf{y}} = m, D)$, with $\alpha_{tm} = 0.1$ for all t and m and $p(\mathcal{T}_{\mathbf{y}} = t) = 1/T$.

Glass type	$\mathcal{C}_{\mathbf{y}} = m$			
	1	2	3	4
bulb	0.008	0.283	0.014	0.047
car window	0.516	0.126	0.432	0.239
headlamp	0.013	0.256	0.022	0.144
container	0.321	0.180	0.005	0.270
building window	0.142	0.155	0.527	0.300

Table 5: Classification of each glass item into one of five use type categories.

Classification	Glass type					
	bulb	car window	headlamp	container	building window	**Total**
bulb	25	0	1	0	1	27
car window	1	74	0	4	29	108
headlamp	0	1	15	1	1	18
container	0	2	0	72	6	80
building window	0	17	0	2	68	87
Total	26 (96.2%)	94 (78.7%)	16 (93.8%)	79 (91.1%)	105 (64.8%)	320

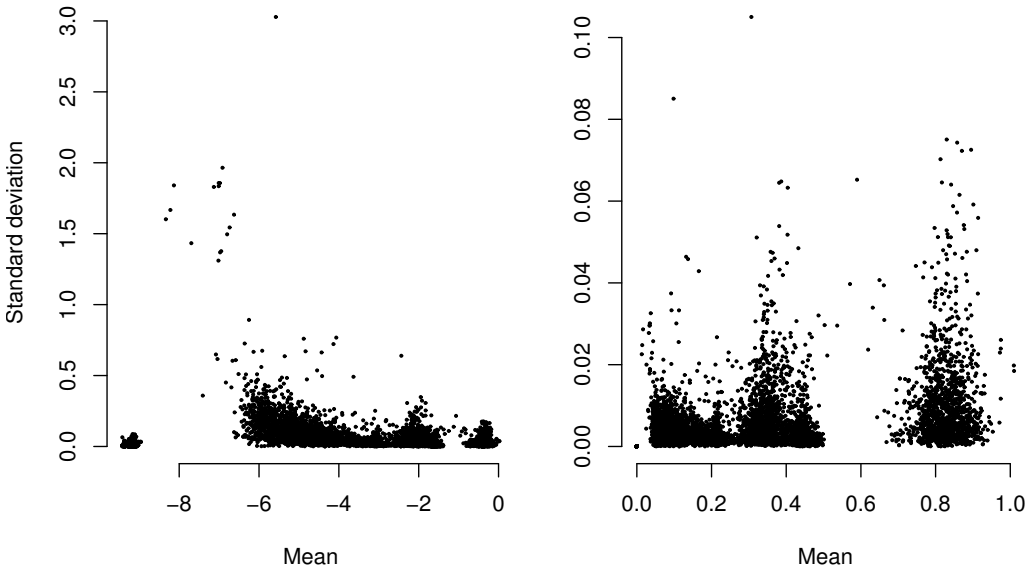

Figure 1: Plots of fragments' standard deviations against corresponding means, using the ALR (left panel) and the square root (right panel) transformations. For the ALR, 0.005 was added to all compositional zeros. Seven pairs (mean, sd) are plotted for each fragment, one for each element of $\mathbf{w}^*$ in (1), computed using the fragment's three repeated measurements. While the variability of the square root transformed data is roughly the same across the range of mean levels, for the ALR transformed data the range of sd's changes considerably across mean levels.

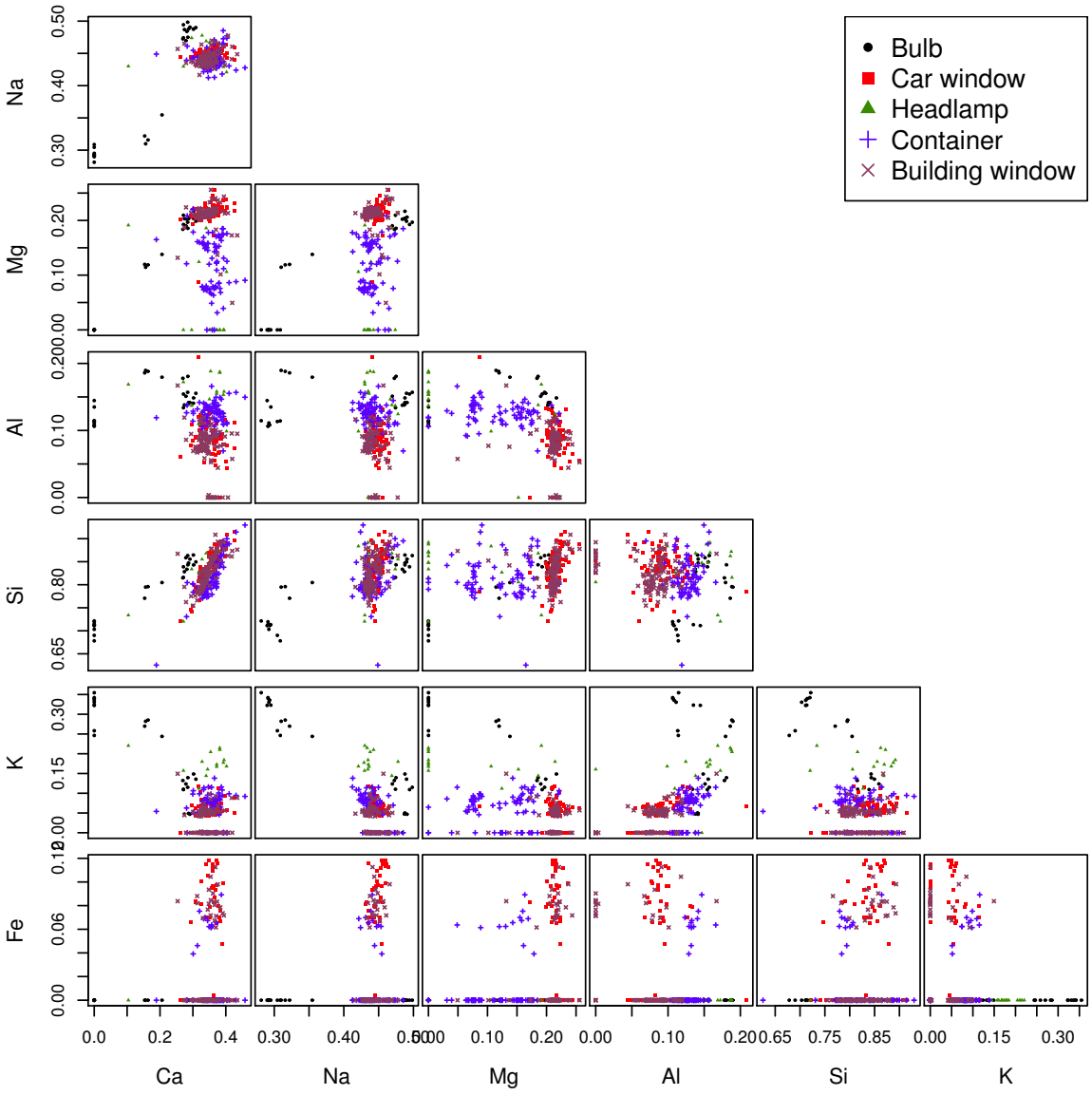

Figure 2: Scatterplots of all item means from the database. The mean of each item is taken across fragments, i.e. obtained from all 12 measurements.

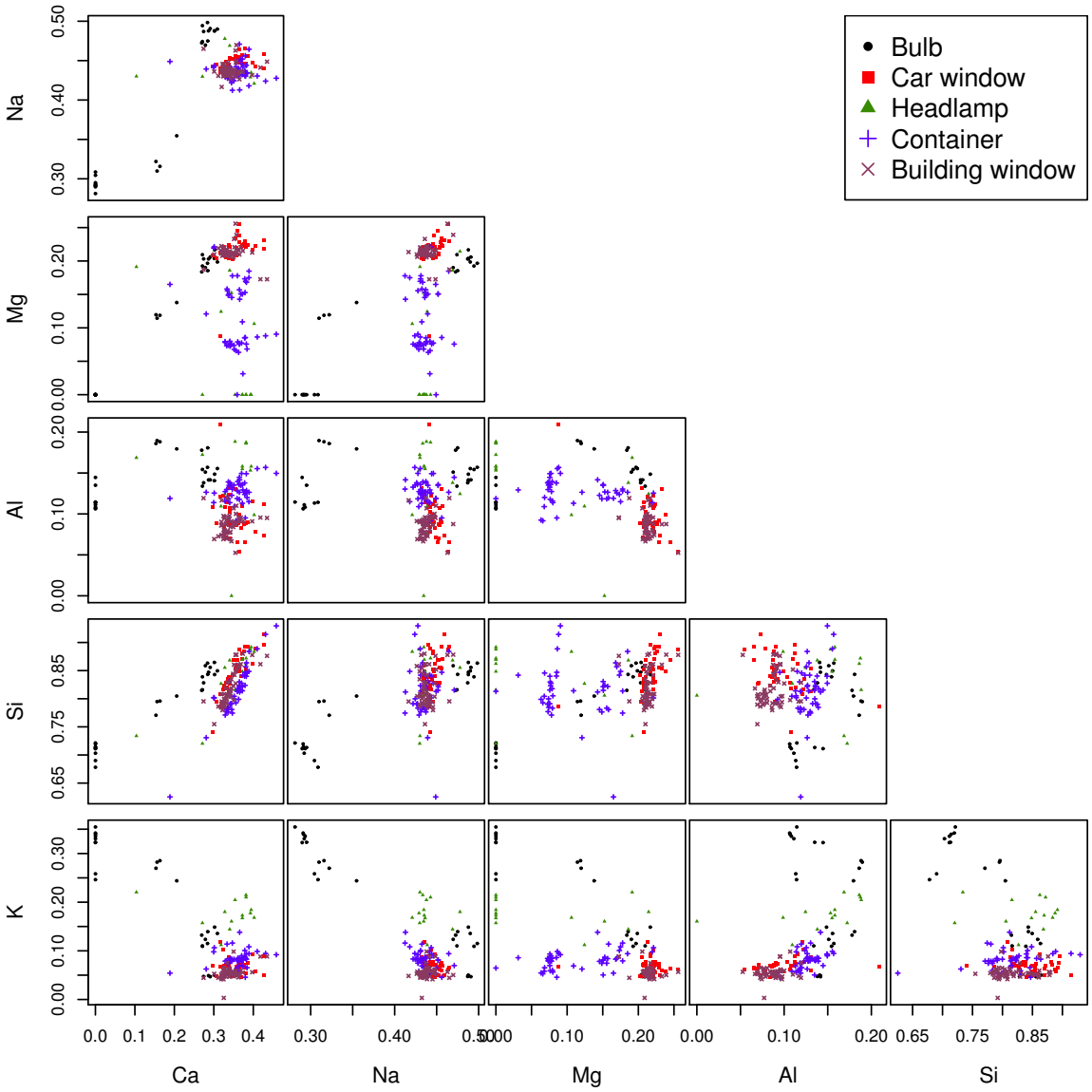

Figure 3: Scatterplots of the item means for items with configuration 2 from Table 2. The mean of each item is taken across fragments, i.e. obtained from all 12 measurements.

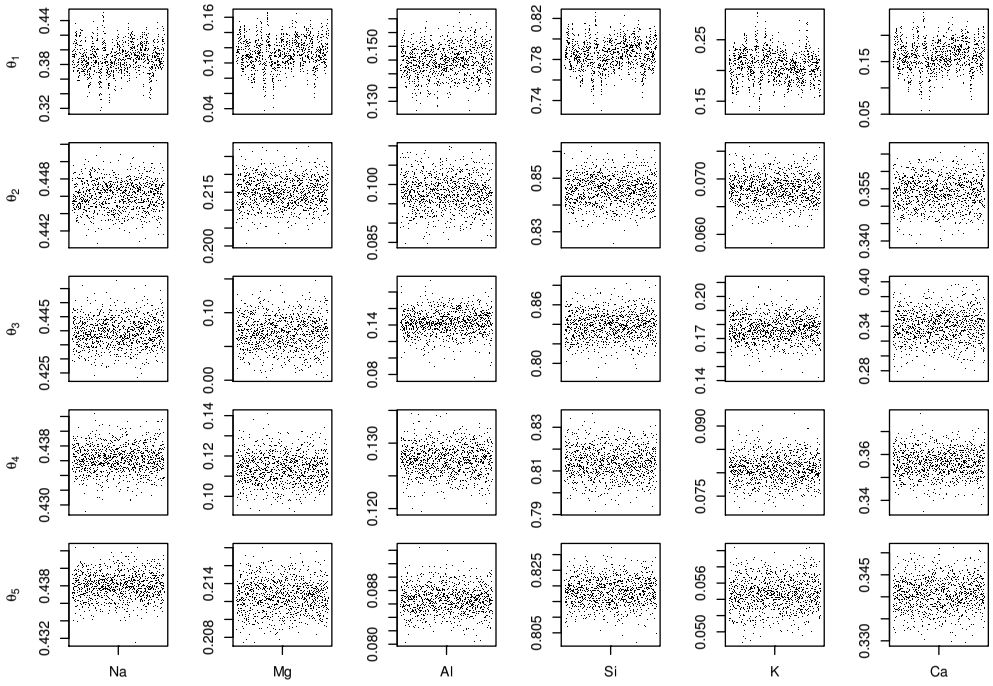

Figure 4: Trace plots of the mean $\boldsymbol{\theta}_t$. A burn-in period of $10,000$ was used, and thinning of the Markov chain with every 200^{th} draw being stored.

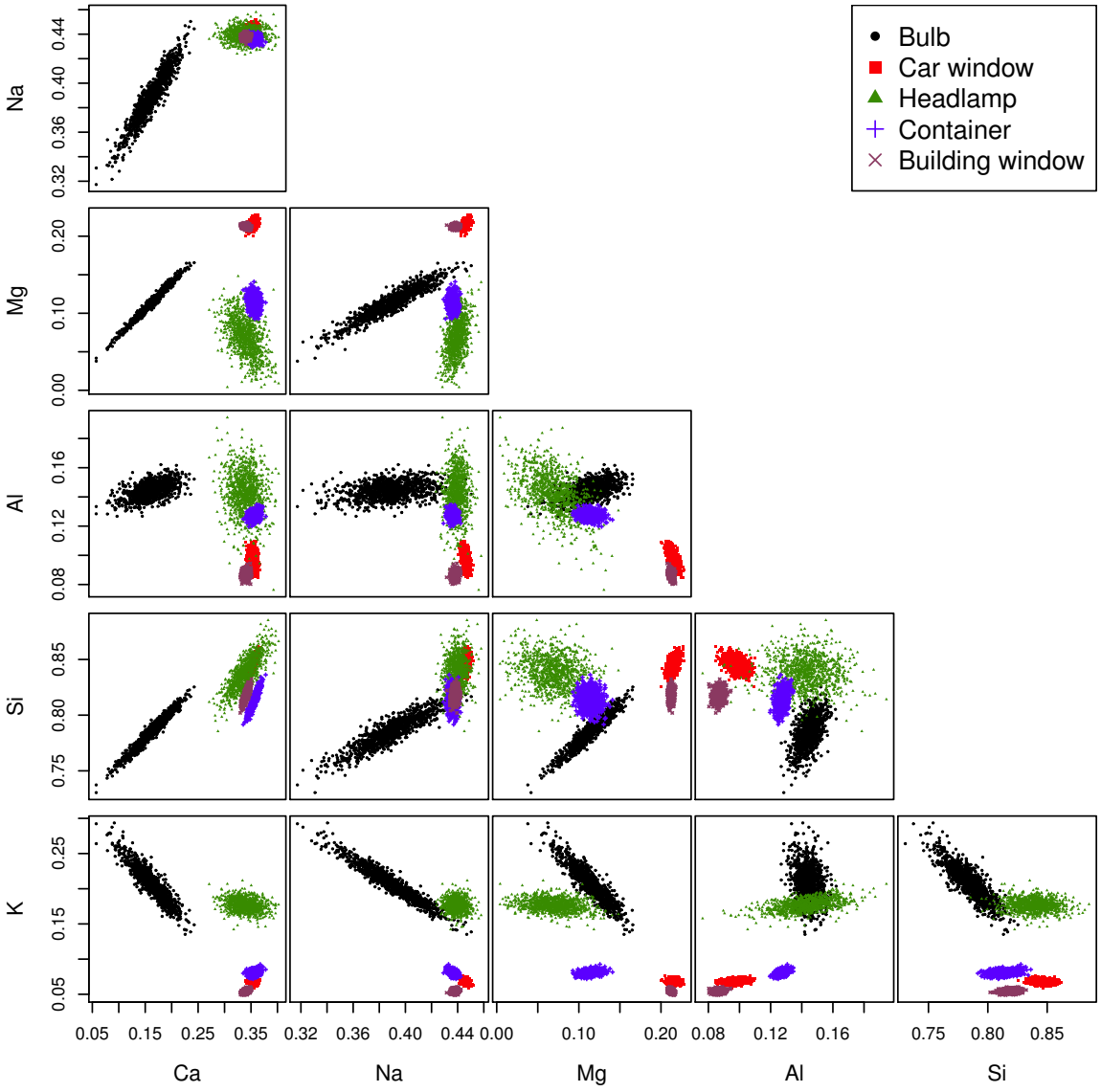

Figure 5: Scatterplots of the draws from the posterior distribution of $\boldsymbol{\theta}_t$ in the model for configuration $m = 2$ ($\overline{\text{Fe}}$, K).

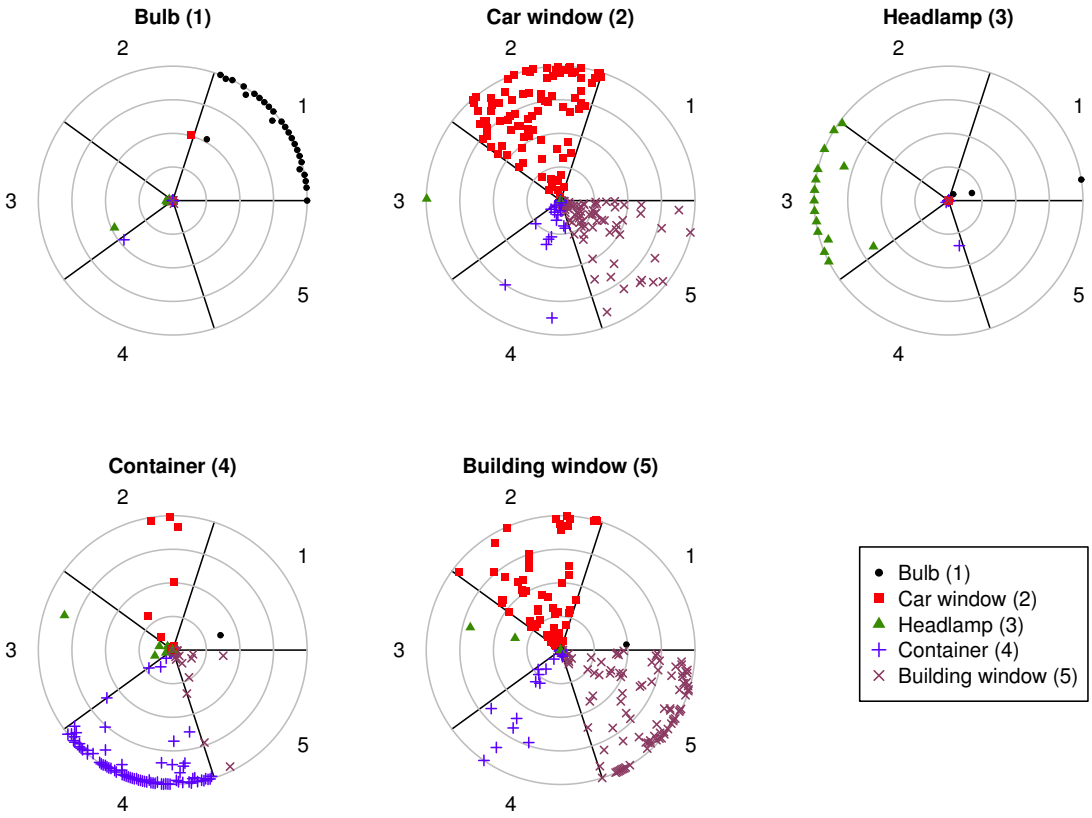

Figure 6: Plot of the classification probabilities in five-fold cross validation. Each panel shows the classification probabilities of items of a certain use type, with five probabilities plotted for each item, one in each circular sector corresponding to the five classes labelled 1 (bulb) to 5 (building window). The four grey circles correspond to probability values of 0.25, 0.5, 0.75 and 1, with the center of the circles having a value of 0. In each circular sector the order of items is the same. The ideal situation, where items of type t (in panel t) are correctly classified with probability one, produces a display where all points in sector t are on the outer circle, while points in the other sectors lie at the center: this is nearly achieved in panel 1 (except for two items) and, to a lesser extent, in panels 3 and 4.

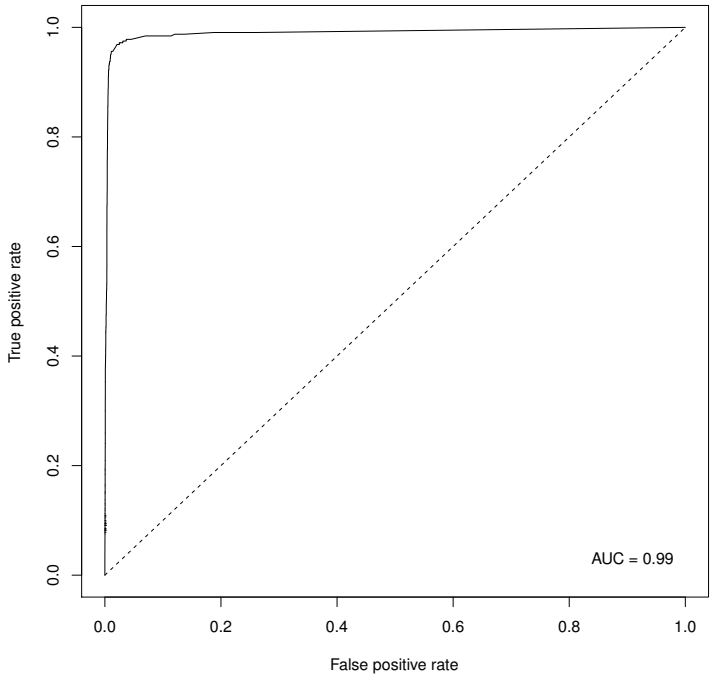

Figure 7: ROC curve for different thresholds of V.