



University
of Glasgow

Napier, Gary (2014) *A Bayesian hierarchical model of compositional data with zeros: classification and evidence evaluation of forensic glass*. PhD thesis.

<http://theses.gla.ac.uk/5793/>

Copyright and moral rights for this thesis are retained by the author

A copy can be downloaded for personal non-commercial research or study, without prior permission or charge

This thesis cannot be reproduced or quoted extensively from without first obtaining permission in writing from the Author

The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the Author

When referring to this work, full bibliographic details including the author, title, awarding institution and date of the thesis must be given



A Bayesian hierarchical model of compositional data
with zeros: classification and evidence evaluation of
forensic glass

Gary Napier

A Dissertation Submitted to the

University of Glasgow

for the degree of

Doctor of Philosophy

School of Mathematics & Statistics

November 2014

© Gary Napier, November 2014

Abstract

A Bayesian hierarchical model is proposed for modelling compositional data containing large concentrations of zeros. Two data transformations were used and compared: the commonly used additive log-ratio (alr) transformation for compositional data, and the square root of the compositional ratios. For this data the square root transformation was found to stabilise variability in the data better. The square root transformation also had no issues dealing with the large concentrations of zeros. To deal with the zeros, two different approaches have been implemented: the data augmentation approach and the composite model approach. The data augmentation approach treats any zero values as rounded zeros, i.e. traces of components below limits of detection, and updates those zero values with non-zero values. This is better than the simple approach of adding constant values to zeros as it reduces any artificial correlation produced by updating the zeros as part of the modelling procedure. However, due to the small detection limit it does not necessarily alleviate the problems of having a point mass very close to zero. The composite model approach treats any zero components as being absent from a composition. This is done by splitting the data into subsets according to the presence or absence of certain components to produce different data con-

figurations that are then modelled separately. The models are applied to a database consisting of the elemental configurations of forensic glass fragments with many levels of variability and of various use types. The main purposes of the model are (i) to derive expressions for the posterior predictive probabilities of newly observed glass fragments to infer their use type (classification) and (ii) to compute the evidential value of glass fragments under two complementary propositions about their source (forensic evidence evaluation). Simulation studies using cross-validation are carried out to assess both model approaches, with both performing well at classifying glass fragments of use types bulb, headlamp and container, but less well so when classifying car and building windows. The composite model approach marginally outperforms the data augmentation approach at the classification task; both approaches have the edge over support vector machines (SVM). Both model approaches also perform well when evaluating the evidential value of glass fragments, with false negative and false positive error rates below 5%. The results from glass classification and evidence evaluation are an improvement over existing methods. Assessment of the models as part of the evidence evaluation simulation study also leads to a restriction being placed upon the reported strength of the value of this type of evidence. To prevent strong support in favour of the wrong proposition it is recommended that this glass evidence should provide, at most, moderately strong support in favour of a proposition. The classification and evidence evaluation procedures are implemented into an online web application, which outputs the corresponding results for a given set of elemental composition measurements. The web application contributes a quick and easy-to-use tool for forensic scientists that deal with this type of forensic evidence in real-life casework.

Keywords: Bayes factor; compositional data; compositional zeros; classification; evidence evaluation; forensic glass; hierarchical model; Markov chain Monte Carlo

Acknowledgements

First and foremost, I would like to thank my supervisors Dr Tereza Neocleous and Dr Agostino Nobile for their support throughout the project. I would like to thank my fellow officemates for keeping me sane over the past few years. I would also like to thank the U.K. Engineering and Physical Sciences Research Council for funding my research.

Declaration

I have prepared this thesis myself; no section of it has been submitted previously as part of any application for a degree. I carried out the work reported in it, except where otherwise stated.

Contents

1	Introduction	1
1.1	Overview of thesis	6
2	Compositional data	8
2.1	Transformations	11
2.1.1	Additive log-ratio	11
2.1.2	Centred log-ratio	12
2.1.3	Isometric log-ratio	13
2.1.4	Box-Cox family	14
2.1.5	Hypersphere	15
2.2	Compositional zeros	16
2.2.1	Rounded zeros	16

2.2.2	Essential zeros	19
2.3	Glass database	20
3	Models	31
3.1	Bayesian hierarchical model	33
3.1.1	Markov Chain Monte Carlo implementation	35
3.1.2	Posterior samples from the Bayesian hierarchical model	43
3.2	Modelling compositional zeros	49
3.2.1	Data augmentation approach	49
3.2.2	Posterior samples from data augmentation	51
3.2.3	Composite model approach	56
3.2.4	Posterior samples for configuration $m = 2$ ($\overline{\text{Fe}}$, K)	60
3.3	Log-ratio transformation	65
3.3.1	Log-ratio data augmentation posterior samples	67
3.3.2	Log-ratio composite model posterior samples	72
3.4	Model diagnostics	77
3.4.1	Results of model diagnostics	78
3.4.2	Simulating data from the models	78

4	Class classification	99
4.1	Classification simulation study	103
4.1.1	Classification results	105
4.2	Classification performance measures	118
4.2.1	Goodman and Kruskal's tau	119
4.2.2	Theil's U	119
4.2.3	Cohen's kappa	120
4.2.4	Brier score	120
4.2.5	Classification performance results	121
5	Evidence evaluation	123
5.1	Evidence evaluation performance measures	128
5.1.1	Measuring performance	128
5.2	Evidence evaluation simulation study	143
5.2.1	Evidence evaluation simulation results	145
6	Web application	156
6.1	Classification	159

6.2	Evidence evaluation	162
7	Discussion and conclusions	167
A	Elemental configuration scatterplots	175
B	MCMC	179
B.1	Full conditional distributions	179
B.2	M-H moves interval widths	180
C	Posterior samples from composite model	181
C.1	Posterior samples for configuration $m = 1$	182
C.2	Posterior samples for configuration $m = 3$	185
C.3	Posterior samples for configuration $m = 4$	188
D	Log-ratio composite model samples	193
D.1	Log-ratio posterior samples for configuration $m = 1$	194
D.2	Log-ratio posterior samples for configuration $m = 3$	197
D.3	Log-ratio posterior samples for configuration $m = 4$	200
E	Simulated data from the composite model	204

List of Tables

2.1	Frequency of zero measurements by chemical element.	22
2.2	Presence and absence of elements at item level by use type . .	26
2.3	Presence and absence of Fe and K at item level by use type . .	26
2.4	Items containing elements with both zero and non-zero values	27
3.1	Effective sample size from BHM	45
3.2	Standard deviations of covariance matrices from BHM	48
3.3	Effective sample size from data augmentation	52
3.4	Standard deviations of covariance matrices from data aug. . .	55
3.5	Effective sample size from configuration 2	61
3.6	Standard deviations of covariance matrices from config. 2 . . .	64
3.7	Effective sample size from DA using log-ratios	68
3.8	Standard deviations of covariance matrices from log-ratios . .	71

3.9	Effective sample size from config. 2 using log-ratios	74
3.10	Log-ratio standard deviations for configuration 2	76
4.1	Use type probabilities	102
4.2	Composite model classification results	106
4.3	Data augmentation classification results	106
4.4	SVM.sub classification results	107
4.5	SVM.full classification results	107
4.6	Composite model expected counts	109
4.7	Data augmentation expected counts	109
4.8	SVM.sub expected counts	110
4.9	SVM.full expected counts	110
4.10	Composite model expected probabilities	112
4.11	Data augmentation expected probabilities	112
4.12	SVM.sub expected probabilities	113
4.13	SVM.full expected probabilities	113
4.14	Classification performance measures	122
5.1	Verbal interpretation of V	144

5.2	SKL scale for interpreting V	145
5.3	Percentage of FN and FP answers	146
6.1	Example application data file	157
B.1	δ_{it} 's for Metropolis-Hastings moves M-H 2 and M-H 3	180
C.1	Effective sample size from configuration 1	183
C.2	Standard deviations of covariance matrices from config. 1	185
C.3	Effective sample size from configuration 3	186
C.4	Standard deviations of covariance matrices from config. 3	188
C.5	Effective sample size from configuration 4	189
C.6	Standard deviations of covariance matrices from config. 4	192
D.1	Effective sample size from config. 1 using log-ratios	194
D.2	Log-ratio standard deviations for configuration 1	197
D.3	Effective sample size from config. 3 using log-ratios	198
D.4	Log-ratio standard deviations for configuration 3	200
D.5	Effective sample size from config. 4 using log-ratios	201
D.6	Log-ratio standard deviations for configuration 4	203

E.1 Elemental configurations across use types for simulated data . 205

List of Figures

2.1	Scatterplots of item means for the ratios to oxygen	23
2.2	Standard deviation of each fragment against its mean	24
2.3	Scatterplots of all item means from the database	29
2.4	Scatterplots of the item means with configuration 2	30
3.1	Trace plots of θ_t with the zeros unaltered	46
3.2	Scatterplots of θ_t with the zeros unaltered	47
3.3	Trace plots of θ_t using data augmentation	53
3.4	Scatterplots of θ_t for items using data augmentation	54
3.5	Trace plots of θ_t for configuration 2	62
3.6	Scatterplots of θ_t for items with configuration 2	63
3.7	Trace plots of θ_t using log-ratios	69
3.8	Scatterplots of θ_t using log-ratios	70

3.9	Log-transformed configuration 2 θ_t trace plots	74
3.10	Log-transformed configuration 2 θ_t scatterplots	75
3.11	Boxplots of simulated θ_t from data augmentation	81
3.12	Boxplots of simulated Ω_1^{-1} from data augmentation	82
3.13	Boxplots of simulated Ω_2^{-1} from data augmentation	83
3.14	Boxplots of simulated Ω_3^{-1} from data augmentation	84
3.15	Boxplots of simulated Ω_4^{-1} from data augmentation	85
3.16	Boxplots of simulated Ω_5^{-1} from data augmentation	86
3.17	Boxplots of simulated Ψ^{-1} from data augmentation	87
3.18	Boxplots of simulated Λ^{-1} from data augmentation	88
3.19	Boxplots of simulated θ_t for configuration 2	91
3.20	Boxplots of simulated Ω_1^{-1} for configuration 2	92
3.21	Boxplots of simulated Ω_2^{-1} for configuration 2	93
3.22	Boxplots of simulated Ω_3^{-1} for configuration 2	94
3.23	Boxplots of simulated Ω_4^{-1} for configuration 2	95
3.24	Boxplots of simulated Ω_5^{-1} for configuration 2	96
3.25	Boxplots of simulated Ψ^{-1} for configuration 2	97

3.26	Boxplots of simulated Λ^{-1} for configuration 2	98
4.1	Composite model posterior probabilities	114
4.2	Data augmentation posterior probabilities	115
4.3	SVM.sub posterior probabilities	116
4.4	SVM.full posterior probabilities	117
5.1	Example histograms of simulated V values	131
5.2	Example Tippett plots	134
5.3	Example DET curve	135
5.4	Example DET curves and Tippett plots	137
5.5	SPSR plot	139
5.6	ECE example plot	142
5.7	Histogram of V values from composite model	148
5.8	Histogram of V values from data augmentation	148
5.9	ROC curve from the composite model	149
5.10	ROC curve from data augmentation	149
5.11	Tippett plots from the composite model	150
5.12	Tippets plots from data augmentation	151

5.13	DET curves from the composite model	152
5.14	DET curves from data augmentation	152
5.15	ECE plots from the composite model	154
5.16	ECE plots from data augmentation	155
6.1	Snapshot of default application screen	158
6.2	Snapshot of classification application screen	161
6.3	Snapshot of evidence evaluation application screen	164
6.4	Snapshot of evidence evaluation results screen	165
6.5	Application screen for differing configurations	166
A.1	Scatterplots of the item means with configuration 1	176
A.2	Scatterplots of the item means with configuration 3	177
A.3	Scatterplots of the item means with configuration 4	178
C.1	Trace plots of θ_t for configuration 1	183
C.2	Scatterplots of θ_t for items with configuration 1	184
C.3	Trace plots of θ_t for configuration 3	186
C.4	Scatterplots of θ_t for items with configuration 3	187
C.5	Trace plots of θ_t for configuration 4	190

C.6	Scatterplots of θ_t for items with configuration 4	191
D.1	Log-transformed configuration 1 θ_t trace plots	195
D.2	Log-transformed configuration 1 θ_t scatterplots	196
D.3	Log-transformed configuration 3 θ_t trace plots	198
D.4	Log-transformed configuration 3 θ_t scatterplots	199
D.5	Log-transformed configuration 4 θ_t trace plots	201
D.6	Log-transformed configuration 4 θ_t scatterplots	202
E.1	Boxplots of simulated θ_t for configuration 1	206
E.2	Boxplots of simulated Ω_1^{-1} for configuration 1	207
E.3	Boxplots of simulated Ω_2^{-1} for configuration 1	208
E.4	Boxplots of simulated Ω_3^{-1} for configuration 1	209
E.5	Boxplots of simulated Ω_4^{-1} for configuration 1	210
E.6	Boxplots of simulated Ω_5^{-1} for configuration 1	211
E.7	Boxplots of simulated Ψ^{-1} for configuration 1	212
E.8	Boxplots of simulated Λ^{-1} for configuration 1	213
E.9	Boxplots of simulated θ_t for configuration 3	214
E.10	Boxplots of simulated Ω_1^{-1} for configuration 3	215

E.11	Boxplots of simulated Ω_2^{-1} for configuration 3	216
E.12	Boxplots of simulated Ω_3^{-1} for configuration 3	217
E.13	Boxplots of simulated Ω_4^{-1} for configuration 3	218
E.14	Boxplots of simulated Ω_5^{-1} for configuration 3	219
E.15	Boxplots of simulated Ψ^{-1} for configuration 3	220
E.16	Boxplots of simulated Λ^{-1} for configuration 3	221
E.17	Boxplots of simulated θ_t for configuration 4	222
E.18	Boxplots of simulated Ω_1^{-1} for configuration 4	223
E.19	Boxplots of simulated Ω_2^{-1} for configuration 4	224
E.20	Boxplots of simulated Ω_3^{-1} for configuration 4	225
E.21	Boxplots of simulated Ω_4^{-1} for configuration 4	226
E.22	Boxplots of simulated Ω_5^{-1} for configuration 4	227
E.23	Boxplots of simulated Ψ^{-1} for configuration 4	228
E.24	Boxplots of simulated Λ^{-1} for configuration 4	229

Chapter 1

Introduction

Forensic science is the application of existing sciences to help determine the outcome of an investigation based on the evidence collected. The objective of the forensic scientist, or expert witness, is to quantify the strength of the accumulated evidence found at a crime scene or accident. Most often the evidence gathered in such cases is recovered in the form of traces and is therefore referred to as *trace evidence*. All trace evidence collected by the forensic expert is brought to a forensic laboratory where various measurements are recorded for analysis. Common forms of trace evidence found at crime scenes are body fluids and glass fragments. The class of trace evidence that is of most interest is known as *transfer evidence*.

Transfer evidence is evidence that has been transferred to or from a crime scene. This can include blood from an assault victim and glass fragments collected from an individual's clothing, hair or shoes. There are two types of transfer evidence to be considered: that of unknown (questioned) origin and

of known origin. For example, glass fragments obtained from a suspect are considered to be of unknown origin as the source of the fragments is questionable; they are referred to as the *recovered sample*. Fragments found at the scene of a crime would usually be of known origin as they are collected from an on-scene broken source; they are referred to as the *control sample* (Aitken and Taroni, 2004). However, control and recovered samples of transfer evidence are not always associated with crime scenes and suspects, respectively. For example, a footprint found at a crime scene would be a recovered sample and not a control sample as the footwear that made the print is unknown, and therefore the footprint is of unknown (questioned) origin. The control sample in this case would come from the footwear of a suspect as it would then be of known origin.

A measure of the evidential value or strength of transfer evidence can be computed relevant to two complementary propositions: the prosecution proposition and the defence proposition. Three levels of propositions comprise a hierarchy of propositions (Aitken and Taroni, 2004). These are the *source level*, *activity level* and *crime level* propositions. Source level propositions involve the analysis of control and recovered samples. Here the prosecution proposition would be that the control and recovered samples come from the same source, while the defence proposition would be that the control and recovered samples come from different sources. Activity level propositions include some form of action. For example, the prosecution proposition could be that the suspect assaulted the victim, while the defence proposition would be that the suspect did not assault the victim. Crime level propositions involve non-scientific information of interest to the jury, which can include such

information as the validity or reliability of eyewitness reports. This thesis will focus on source level propositions involving glass fragments. Chapter 10 of [Aitken and Taroni \(2004\)](#) details how the evidence under evaluation is computed under these two complementary propositions, with their terminology adopted in this thesis. Let E denote the evidence, H_p and H_d denote the prosecution and defence propositions, and I be additional background information connected to the case under investigation. The value V of the evidence forms the likelihood ratio or Bayes factor in favour of the prosecution proposition H_p , given the evidence E :

$$V = \frac{\Pr(E|H_p, I)}{\Pr(E|H_d, I)}. \quad (1.1)$$

The value V in (1.1) can be derived from Bayes' Theorem which is used to convert prior beliefs about a proposition, say H_p , to posterior beliefs after observing some evidence E :

$$\Pr(H_p|E, I) = \frac{\Pr(E|H_p, I) \times \Pr(H_p|I)}{\Pr(E|I)}, \quad (1.2)$$

where $\Pr(H_p|E, I)$ is the posterior probability of H_p given the evidence E , and $\Pr(H_p|I)$ is the prior probability of H_p before observing E . The value of the evidence V in (1.1) then arises by taking the ratio of (1.2) for H_p against H_d to acquire the posterior odds:

$$\underbrace{\frac{\Pr(H_p|E, I)}{\Pr(H_d|E, I)}}_{\text{Posterior odds}} = \underbrace{\frac{\Pr(E|H_p, I)}{\Pr(E|H_d, I)}}_{\substack{\text{Likelihood ratio} \\ \text{(Bayes factor)}}} \times \underbrace{\frac{\Pr(H_p|I)}{\Pr(H_d|I)}}_{\text{Prior odds}} = V \times \underbrace{\frac{\Pr(H_p|I)}{\Pr(H_d|I)}}_{\text{Prior odds}}. \quad (1.3)$$

Prior to receiving E the odds in favour of H_p are given by the prior odds, or ratio of initial beliefs between H_p and H_d . Formula (1.3) converts these prior

beliefs in favour of the prosecution proposition, H_p , into posterior beliefs in favour of H_p , relative to the defence proposition, H_d , by multiplying the prior odds by V . Hence, a value of $V > 1$ provides support for H_p , whereas $V < 1$ provides support for H_d . The way in which V is computed allows for the evidence to be considered under two complementary propositions, i.e. H_p and H_d are mutually exclusive and exhaustive propositions, while allowing for other possible factors to be considered during the evaluation of the evidence. The expert witness receives less information about a case than the judge and jury, with their sole purpose to quantify the strength of the evidence, and so should not be concerned with trying to determine the prior odds ([Lucy, 2005](#)). For more information on the evaluation of evidence see [Aitken and Taroni \(2004\)](#).

The type of forensic evidence that will be analysed within this thesis is on fragments of glass from an experimental database by the Institute of Forensic Research, Krakow, which will be referred to as the glass database for the remainder of the thesis. Glass fragments are a common source of trace evidence in forensic investigations, with the strength of the evidence evaluated under two complementary propositions as described earlier. Here the prosecution proposition would be that the control and recovered samples come from the same glass object, whereas the defence proposition would be that they originated from different glass objects. The main question of interest is then: do the fragments obtained from the suspect come from the glass object found at the crime scene?

As glass fragments collected from a suspect are of unknown origin it would be useful to infer their use type. For example, if a person is involved in an

incident such as a car crash or assault where fragments of glass are recovered, being able to determine the use type of those fragments could aid a police investigation. Most glass fragments collected for analysis by forensic experts are too small for their use type to be determined by physical properties such as their thickness or colour (Zadora, 2009), so physico-chemical measurements, such as the refractive index or elemental composition, are acquired.

Lindley (1977) proposed an approach for continuous measurements in the form of refractive indices from glass fragments. Here focus is on the elemental composition of glass. The elemental compositions consist of the percentage weights associated with each of the main chemical elements comprising a glass fragment. As is common with compositional data, many of the recorded weight percentages are zero, indicating those elements' absence from a fragments composition. Standard statistical procedures developed for analysing compositional data require careful consideration of any zeros present within a composition. The most common approaches to analysing compositional data include transformations involving logarithmic terms, such as the additive log-ratio (alr) and centred log-ratio (clr) transformations of Aitchison (1986), and the more recently introduced isometric log-ratio (ilr) transformation of Egozcue *et al.* (2003). This means that any zeros present within a composition require special treatment before such transformations can be computed.

This thesis considers two different approaches to handling the compositional zeros. The first method is the most common approach to dealing with zeros as it treats them as rounded zeros. Rounded zeros indicate that an element

is present within a fragment but that traces of that element are below the detection limit of the measuring equipment. This method essentially updates any zeros with non-zero values that are below limits of detection. The second method treats the zeros as essential zeros, indicating the absolute absence of that element from a fragment's composition (Martín-Fernández *et al.*, 2003). This method is similar to that of Stewart and Field (2011) in that the glass data is partitioned depending on whether elements are present or absent from each glass composition. This can lead to a reduction in the dimension of the data, and to the modelling of subcompositions.

The methods of dealing with zeros mentioned above will be incorporated into a Bayesian hierarchical model. The model takes into consideration the glass use type and the hierarchical structure of the glass data by including three levels of variability. The model is then used to classify glass items into one of five use type categories and for computing the evidential value of glass fragments relating to two competing propositions.

1.1 Overview of thesis

Chapter 2 provides a review of current methods used in the analysis of compositional data, as well as a detailed description and exploratory analysis of the glass database. Chapter 3 details the Bayesian hierarchical model and the Markov Chain Monte Carlo (MCMC) implementation, and also describes the two different approaches taken to handling the many compositional zeros present in the glass database. Chapter 3 also presents the posterior draws obtained from the hierarchical model as well as model diagnostics. The clas-

sification procedure and results of classifying all glass items in the database are detailed in Chapter 4, while the evaluation of glass fragments as evidence is described, including the results, in Chapter 5. Chapter 6 describes a web application of the classification and evidence evaluation approaches along with instructions on how to load the data into the application. Discussion and conclusions are contained in Chapter 7.

Chapter 2

Compositional data

As already mentioned in the introduction, the glass database contains measurements on percentage weights and so is compositional in nature. This chapter will begin with an overview of statistical methods for compositional data and the techniques used to analyse it before going on to later detail the glass database itself.

Compositional data commonly occurs in scientific disciplines such as chemistry, geology, economics, and many others. Compositional data are vectors comprised of non-negative parts of some whole. The sample space or simplex as defined by [Aitchison \(1986\)](#) is

$$S^{D-1} = \left\{ \mathbf{w} = (w_1, \dots, w_D) : w_d \geq 0; \sum_{d=1}^D w_d = c \right\}, \quad (2.1)$$

where c is the constant sum constraint and value of the full composition, i.e. $c = 1, 100, \dots$ for proportions, percentages etc. The simplex has the vector operations of perturbation and power transformation which are analagous to

translation and scalar multiplication in \mathbb{R}^D .

The perturbation operation can be used to record changes in a composition. If \mathbf{x} is a D -part composition consisting of decaying components and \mathbf{u} is a D -part vector containing the decay rates of \mathbf{x} , then the newly perturbed composition \mathbf{y} can be written as

$$\mathbf{y} = \mathbf{x} \oplus \mathbf{u} = \mathcal{C}(x_1 u_1, \dots, x_D u_D). \quad (2.2)$$

The closure operator \mathcal{C} used in (2.2) is defined as

$$\mathcal{C}(\mathbf{x}) = \frac{c}{\sum_{d=1}^D x_d} \mathbf{x}, \quad (2.3)$$

and is used to convert raw composition measurements into proportions, percentages etc. depending on the choice of c . The vector \mathbf{u} containing any change or perturbation in a composition can be found from the inverse operation:

$$\mathbf{u} = \mathbf{y} \ominus \mathbf{x} = \mathcal{C}(y_1/x_1, \dots, y_D/x_D), \quad (2.4)$$

with $\mathbf{u} = \mathbf{1}_D$ the neutral element corresponding to no change from \mathbf{x} to \mathbf{y} . The perturbation operation allows for compositional measurements collected from a source, that may be represented in different units, to produce the same qualitative results under statistical analyses if there exists a perturbing vector \mathbf{u} that allows for easy transformations between units. The power transformation can also be used to rescale a composition given some constant a :

$$\mathbf{y} = a \otimes \mathbf{x} = \mathcal{C}(x_1^a, \dots, x_D^a), \quad a > 0. \quad (2.5)$$

An analogue to the Euclidean distance in the simplex, known as Aitchison distance ([Aitchison, 1986](#)), is defined by

$$d_a(\mathbf{x}, \mathbf{y}) = \left[\frac{1}{D} \sum_{i < j} \left(\ln \frac{x_i}{x_j} - \ln \frac{y_i}{y_j} \right)^2 \right]^{1/2} = \left[\sum_{i=1}^D \left(\ln \frac{x_i}{g(\mathbf{x})} - \ln \frac{y_i}{g(\mathbf{y})} \right)^2 \right]^{1/2}, \quad (2.6)$$

where $g(\mathbf{x}) = (x_1 \cdots x_D)^{1/D}$ is the geometric mean of \mathbf{x} . The Aitchison distance satisfies the property of subcompositional dominance ([Aitchison, 1992](#)). Subcompositional dominance is the property whereby the distance between two full compositions \mathbf{x} and \mathbf{y} is at least as large as the distance between any two subcompositions \mathbf{x}^* and \mathbf{y}^* , i.e. $d_a(\mathbf{x}, \mathbf{y}) \geq d_a(\mathbf{x}^*, \mathbf{y}^*)$. There are two principles of compositional data to consider before carrying out any analysis. These are scale invariance and subcompositional coherence ([Aitchison and Egozcue, 2005](#)).

Compositional vectors \mathbf{w}_1 and \mathbf{w}_2 are considered compositionally equivalent if $\mathcal{C}(\mathbf{w}_1) = \mathcal{C}(\mathbf{w}_2)$, which implies the existence of a proportionality constant, λ , such that $\mathbf{w}_1 = \lambda \cdot \mathbf{w}_2$. A function f is said to be scale invariant if

$$f(\mathbf{w}_1) = f(\lambda \cdot \mathbf{w}_1). \quad (2.7)$$

Since scale invariance is satisfied by taking the ratio of $D - 1$ components to the D^{th} component of a composition, and as compositions carry relative information, compositional data are most often expressed in terms of ratios of the component parts. A function is also said to be permutation invariant if it is unaffected by changes to the ordering of the component parts.

For any subset S of the indices $(1, \dots, D)$ of a D -part composition such that $\mathbf{w}_S = \mathbf{s}$ is a subcomposition of \mathbf{w} , the ratio of any two components in \mathbf{s}

should be equal to the corresponding ratio in the full composition \mathbf{w} :

$$\frac{s_i}{s_j} = \frac{w_i}{w_j} \quad \forall (i, j \in S). \quad (2.8)$$

This is referred to as subcompositional coherence, and along with scale invariance, is the reason why analyses of compositional data regularly involve the use of component ratios. The use of ratios also removes the constant sum constraint of the simplex allowing for the use of standard multivariate techniques, and for the dimension of the data to be reduced by one. Typically data transformations are then applied to the ratios to improve variance stability and normality.

2.1 Transformations

2.1.1 Additive log-ratio

The most common transformation applied to compositional data was introduced by [Aitchison \(1982\)](#) and involves taking the logarithm of $D - 1$ components to the remaining one. This is referred to as the additive log-ratio (alr) transformation:

$$\text{alr}(\mathbf{w}) = \left(\ln \left(\frac{w_1}{w_D} \right), \dots, \ln \left(\frac{w_{D-1}}{w_D} \right) \right). \quad (2.9)$$

The alr transformation removes the sum constraint allowing for the application of standard multivariate techniques, such as assuming the data to be multivariate normally distributed, where the variance-covariance matrix, Σ ,

would be of dimension $(D - 1) \times (D - 1)$ and given as

$$\Sigma = \left[\text{cov} \left\{ \ln \left(\frac{w_i}{w_D} \right), \ln \left(\frac{w_j}{w_D} \right) \right\} : i, j = 1, \dots, D - 1 \right]. \quad (2.10)$$

The alr transformation satisfies the principles of compositional data described earlier, but it is not without its issues. One potential issue is that it does not treat the parts of \mathbf{w} symmetrically with w_D taking on the role of the common divisor in the log-ratios. Also, despite the invariant property, the choice of the common divisor w_D in the ratio still has an important role to play in terms of the goodness of fit to normality of the log-transformed data (Rayens and Srinivasan, 1991a). The component chosen as the divisor needs to be greater than zero within a composition in order to avoid further complications involving compositional zeros.

Baxter *et al.* (2005) point out that performing a principal component analysis (PCA) on the standardised ratios recovers the structure of the data much better than using the alr transformed data. This is due to component parts with very low absolute values and high variability strongly influencing the structure of the log-transformed data; see also Wang *et al.* (2008). The alr transformation also requires that all component parts be strictly positive, but zero measurements occur frequently in compositional data, leading to the development of strategies to deal with the presence of such values. These strategies are detailed in Section 2.2.

2.1.2 Centred log-ratio

Another transformation proposed by Aitchison (1986) that satisfies the principles of compositional data is the centred log-ratio (clr) transformation. The

clr transformation is the logarithm of the ratio of all D -parts to their geometric mean:

$$\text{clr}(\mathbf{w}) = \left(\ln \left(\frac{w_1}{g(\mathbf{w})} \right), \dots, \ln \left(\frac{w_D}{g(\mathbf{w})} \right) \right), \quad (2.11)$$

where $g(\mathbf{w})$ is the geometric mean. Unlike the alr, the clr treats compositions symmetrically by not having a component singled out as the common divisor. For the clr, the $D \times D$ variance-covariance matrix, Γ , of a D -part composition \mathbf{w} is given as

$$\Gamma = \left[\text{cov} \left\{ \ln \left(\frac{w_i}{g(\mathbf{w})} \right), \ln \left(\frac{w_j}{g(\mathbf{w})} \right) \right\} : i, j = 1, \dots, D \right]. \quad (2.12)$$

An advantage the clr has over the alr is being able to visualise all D -parts of a composition when performing an exploratory analysis of the data (Campbell *et al.*, 2009). However, as (2.12) is singular, it is simpler to revert back to the alr transformation when using standard methods requiring the inverse of the variance-covariance matrix, similarly to how Campbell *et al.* (2009) analysed data on New Zealand nephrite.

2.1.3 Isometric log-ratio

Egozcue *et al.* (2003) introduced the isometric log-ratio (ilr) transformation for compositional data which allows for angles and distances in the simplex to be associated with angles and distances in real space. The ilr transformation is given as $\text{ilr}(\mathbf{w}) = \mathbf{x} = (x_1, \dots, x_{D-1}) \in \mathbb{R}^{D-1}$ where

$$x_i = \frac{1}{\sqrt{i(i+1)}} \ln \left(\frac{\prod_{j=1}^i w_j}{(w_{i+1})^i} \right). \quad (2.13)$$

The relationships between the ilr and the alr and clr transformations are described in detail in Egozcue *et al.* (2003).

Other logarithmic transformations that have been applied to compositional data include the complementary log-log transformation (Neocleous *et al.*, 2011) and the multiplicative log-ratio (mlr) transformation (Stewart and Field, 2011). However, as previously stated, in order to compute logarithmic transformations all component parts must be strictly positive. This has led to the proposal of non-logarithmic transformations for compositional data.

2.1.4 Box-Cox family

Rayens and Srinivasan (1991a) propose using the family of transformations considered by Box and Cox (1964) to improve the fit to normality of compositional data. Rayens and Srinivasan believe that the ratio transformation alone alleviates most of the issues concerning the statistical analysis of compositional data, with the additional logarithmic transformation used mainly to improve normality. As the log transformation is part of the Box-Cox family of transformations, a better fit to normality, and thus a better use of normal based multivariate techniques, should be achievable by considering the whole family of transformations.

Let \mathbf{x} be the ratios of $D - 1$ components to the remaining one. The Box-Cox transformation is then applied to each component of \mathbf{x} as follows

$$y_i = \begin{cases} \frac{x_i^{\lambda_i} - 1}{\lambda_i}, & \text{if } \lambda_i \neq 0, \\ \ln(x_i), & \text{if } \lambda_i = 0, \end{cases} \quad (2.14)$$

where $\lambda = (\lambda_1, \dots, \lambda_{D-1})$ are the power parameters of the transformation.

Rayens and Srinivasan (1991a) recognise the potential issue pertaining to

subcompositional analysis with the use of non-logarithmic transformations. However, they believe this concern can be resolved in practice if an ever present component is commonly accepted by practitioners to be the common divisor.

Application of the Box-Cox and alr transformations to real data in [Rayens and Srinivasan \(1991a\)](#) show improvements in goodness of fit to normality in favour of the Box-Cox transformation. Rayens and Srinivasan extend their application of the Box-Cox transformation in parametric and non-parametric approaches to modelling compositional data in [Rayens and Srinivasan \(1991b\)](#).

2.1.5 Hypersphere

[Wang *et al.* \(2007\)](#) and [Neocleous *et al.* \(2011\)](#) avoid the complication of zero components by applying a hyperspherical transformation to compositional data. First the square root is applied to all D -parts of a composition \mathbf{w} :

$$\text{sqrt}(\mathbf{w}) = (\sqrt{w_1}, \dots, \sqrt{w_D}), \quad (2.15)$$

transforming it onto the surface of the $(D-1)$ -dimensional hypersphere. The Cartesian coordinates of (2.15) are then mapped to polar coordinates using a recursive relationship. [Scealy and Welsh \(2011\)](#) used (2.15) to allow for directional data distributions - such as the Kent distribution - to be used when modelling compositional data; see also [Scealy and Welsh \(2014\)](#).

Regardless of which transformation is applied to compositional data, there still lies the problem pertaining to zeros, especially in datasets where they

are abundant. However, zeros are still notably more problematic for any of the logarithmic transformations, with even a small number of zeros causing computational issues; they also have a stronger influence on the data distribution, as will be demonstrated in Section 2.3.

2.2 Compositional zeros

Zero measurements occur frequently in compositional data causing problems with the application of any of the logarithmic transformations mentioned in Section 2.1. There are two types of compositional zeros: rounded zeros, indicating the presence of a component but below some detection limit; and essential zeros, denoting the absolute absence of a component from an observation (Martín-Fernández *et al.*, 2003). Different approaches have been adopted for both types of zeros.

2.2.1 Rounded zeros

Non-zero traces of components that are below limits of detection for some measuring equipment are reported as zero concentrations and referred to as rounded zeros. The simplest strategy then is to replace rounded zeros by some small constant equal to or below the detection limit (Neocleous *et al.*, 2011). To satisfy the constant sum constraint the non-zero components would then have to be adjusted accordingly, which can be done using one of several techniques.

Simple replacement strategy

The easiest of such zero replacement techniques is known as the simple replacement strategy (Martín-Fernández and Thió-Henestrosa, 2006). A D -part composition \mathbf{w} containing rounded zeros is replaced by an updated non-zero composition \mathbf{u} :

$$u_d = \begin{cases} \frac{c}{c + \sum_{\{k:w_k=0\}} \delta_k} \delta_d, & \text{if } w_d = 0, \\ \frac{c}{c + \sum_{\{k:w_k=0\}} \delta_k} w_d, & \text{if } w_d > 0, \end{cases} \quad (2.16)$$

where $c = \sum w_d$ is the constant sum constraint, and δ is a value below limits of detection.

Additive replacement strategy

Aitchison (1986) proposed an additive replacement strategy where any D -part composition \mathbf{w} , containing Z zeros is replaced by an updated non-zero composition \mathbf{u} :

$$u_d = \begin{cases} \frac{\delta(Z+1)(D-Z)}{D^2}, & \text{if } w_d = 0, \\ w_d - \frac{\delta(Z+1)Z}{D^2}, & \text{if } w_d > 0, \end{cases} \quad (2.17)$$

where δ is a value smaller than the given threshold of the measuring equipment. A problem with this strategy is that it fails to preserve the ratios between components of the original composition \mathbf{w} and the updated non-zero composition \mathbf{u} , leading to subcompositional incoherence. This led to other replacement strategies being proposed with Fry *et al.* (2000) adopting a modified version of the additive replacement strategy of Aitchison when

dealing with share ratios in economic data, and to the introduction of the multiplicative replacement strategy of [Martín-Fernández *et al.* \(2000\)](#).

Multiplicative replacement strategy

The zero replacement technique of [Martín-Fernández *et al.* \(2000\)](#) does not rely on the number of components D , or the number of zeros Z , but only on the given threshold value δ :

$$u_d = \begin{cases} \delta_d, & \text{if } w_d = 0, \\ w_d - \frac{w_d}{c} \sum_{\{d:w_d=0\}} \delta_d, & \text{if } w_d > 0, \end{cases} \quad (2.18)$$

where $c = \sum w_d$ is the constant sum constraint. This proposed replacement strategy preserves the ratios between the old observations \mathbf{w} , and the newly adjusted observations \mathbf{u} . [Sanford *et al.* \(1993\)](#) suggest values equal to 0.55 times the threshold δ be used when imputing values below limits of detection. A comparison of the performance of the additive and multiplicative replacement strategies can be found in [Martín-Fernández *et al.* \(2003\)](#).

Although the multiplicative replacement strategy preserves ratios it can run into problems with compositional datasets containing a large number of zeros (more than 10% ([Sanford *et al.*, 1993](#))). The issue of artificial correlation induced by the strategy for datasets containing a large proportion of zeros is raised in [Palarea-Albaladejo *et al.* \(2007\)](#).

[Palarea-Albaladejo *et al.* \(2007\)](#) propose a parametric approach to replacing components below the limits of detection in the presence of many zero values. [Palarea-Albaladejo *et al.*](#) use a modified version of the EM algo-

rithm that takes into account that rounded zeros are seen as missing values (Martín-Fernández and Thió-Henestrosa, 2006), however they are not considered missing at random (MAR), but below limits of detection, so are therefore treated as not missing at random (NMAR). Their method also takes into account the covariance structure of the data ensuring imputed values for zeros of the same component differ from one another, thus reducing the artificial correlation induced by the multiplicative replacement strategy.

2.2.2 Essential zeros

Essential zeros are different to rounded zeros in that they are not zero due to limits of detection, but considered to be actual zeros. Essential zeros appear in compositions in disciplines such as economics. Most often zeros in compositional data are treated as rounded zeros and so fewer advancements have been made for zeros considered to be essential. Aitchison and Kay (2003) believe that such zero problems may not exist once the precise aims of a study of compositional data containing essential zeros have been obtained.

Zadora *et al.* (2010) modelled non-zero subcompositions using a two stage model approach, with the presence of zeros treated using an independent binary model as suggested by Aitchison and Kay (2003). Bacon-Shone (2003) explores the problem of essential zeros arising in household expenditure data, i.e. households who spend money on alcohol versus those that do not. Bacon-Shone suggests the use of multivariate logistic or probit models for modelling the essential zero structure within the data. Butler and Glasbey (2008) and Leininger *et al.* (2013) approached the issue of zeros by including a latent

Gaussian random variable in their models, that has the effect of creating a point mass at zero. [Stewart and Field \(2011\)](#) developed mixture models for compositional data containing zeros when examining the diets of predators. Stewart and Field partitioned their data depending on whether different species were present or absent from a predator’s diet, thus only modelling the non-zero components. [Stewart \(2013\)](#) adapted the model using the zero-inflated beta distribution, simplifying the previous approach by not having to apply the multiplicative log-ratio transformation to the same data set.

[Daunis-i-Estadella *et al.* \(2008\)](#) introduce the idea of a different type of compositional zeros that occur in count data sets. Referring to them as count zeros, [Daunis-i-Estadella *et al.*](#) propose the use of the multiplicative replacement strategy in order to preserve ratios between prior and posterior counts when the Dirichlet distribution is used as the conjugate prior of the multinomial distribution within a Bayesian framework, thus allowing for the application of the alr transformation. The glass database will now be detailed, where its unique aspects are highlighted during an exploratory analysis that is informative about the approaches used in its analysis.

2.3 Glass database

The glass database being analysed consists of measurements obtained from an experimental setting, and was provided by Prof. G. Zadora of the Institute of Forensic Research, Krakow. The database is comprised of four fragments - each having three replicate measurements – from 320 glass items to give a total of 3840 measurements. The 320 glass items are split across five use types:

26 bulbs, 94 car windows, 16 headlamps, 79 containers and 105 building windows. The elemental content of each fragment was obtained from a scanning electron microscope with an energy dispersive X-ray (SEM-EDX) spectrometer. The fragments chosen for analysis were obtained from an experimental setting by breaking a glass item of a specified use type into smaller fragments. The fragments with the smoothest and flattest surfaces with linear dimension less than 0.5 mm were then selected for analysis (Zadora, 2009). The measurements produced by SEM-EDX are on the percentage weights (wt%) – to two decimal places – of the main elements comprising each fragment’s composition. These main chemical elements are: oxygen (O), sodium (Na), magnesium (Mg), aluminium (Al), silicon (Si), potassium (K), calcium (Ca) and iron (Fe). Other methods of elemental analysis of glass fragments include non-destructive energy dispersive X-ray microfluorescence (microXRF) (Hicks *et al.*, 2003) and laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) (Trejos and Almirall, 2005).

The percentage weights of each fragment are compositional and therefore are non-negative and sum to 100%. Let the number of elements in a fragment’s composition be D . The percentage weights are then denoted by the D dimensional vector $\mathbf{w} = (w_1, \dots, w_D)$, with $w_d \geq 0$ and $\sum_{d=1}^D w_d = 100$. To remove the sum constraint the elements can be transformed into a $(D - 1)$ dimensional vector of ratios by taking the ratio of $(D - 1)$ elements to the D^{th} element. The transformed vector is

$$\mathbf{w}^* = \left(\frac{w_1}{w_D}, \dots, \frac{w_{D-1}}{w_D} \right), \quad (2.19)$$

where oxygen was chosen as the common divisor, w_D , due to having the highest weight percentage and being ever present in glass.

A number of the percentage weights are zero with their prevalence differing across elements, as shown in Table 2.1. Figure 2.1, containing plots of the item means for the ratios to oxygen, illustrates how a large number of compositional zeros can influence the distribution of the data.

Table 2.1: 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

In order to improve variance stability and normality, a transformation was then applied to the ratios in (2.19). In addition to the additive log-ratio transformation, members from the family of transformations of [Box and Cox \(1964\)](#) were examined, with improvements in variance stability and normality of the data obtained by applying the square root to (2.19). Figure 2.2 contrasts the alr and square root transformations, and shows that the square root transformation improves variance stability in the data more so than the alr transformation. This is due to the chemical elements with many zeros and very low weight percentages having a much stronger influence over the log-transformed ratios than those of the square root. Not only is the square root transformation more effective at stabilising the variability in the data, it also does not require any zeros be altered in order to be computed. Due to these reasons the square root transformation has been chosen as the appropriate transformation for the analysis of the glass database.

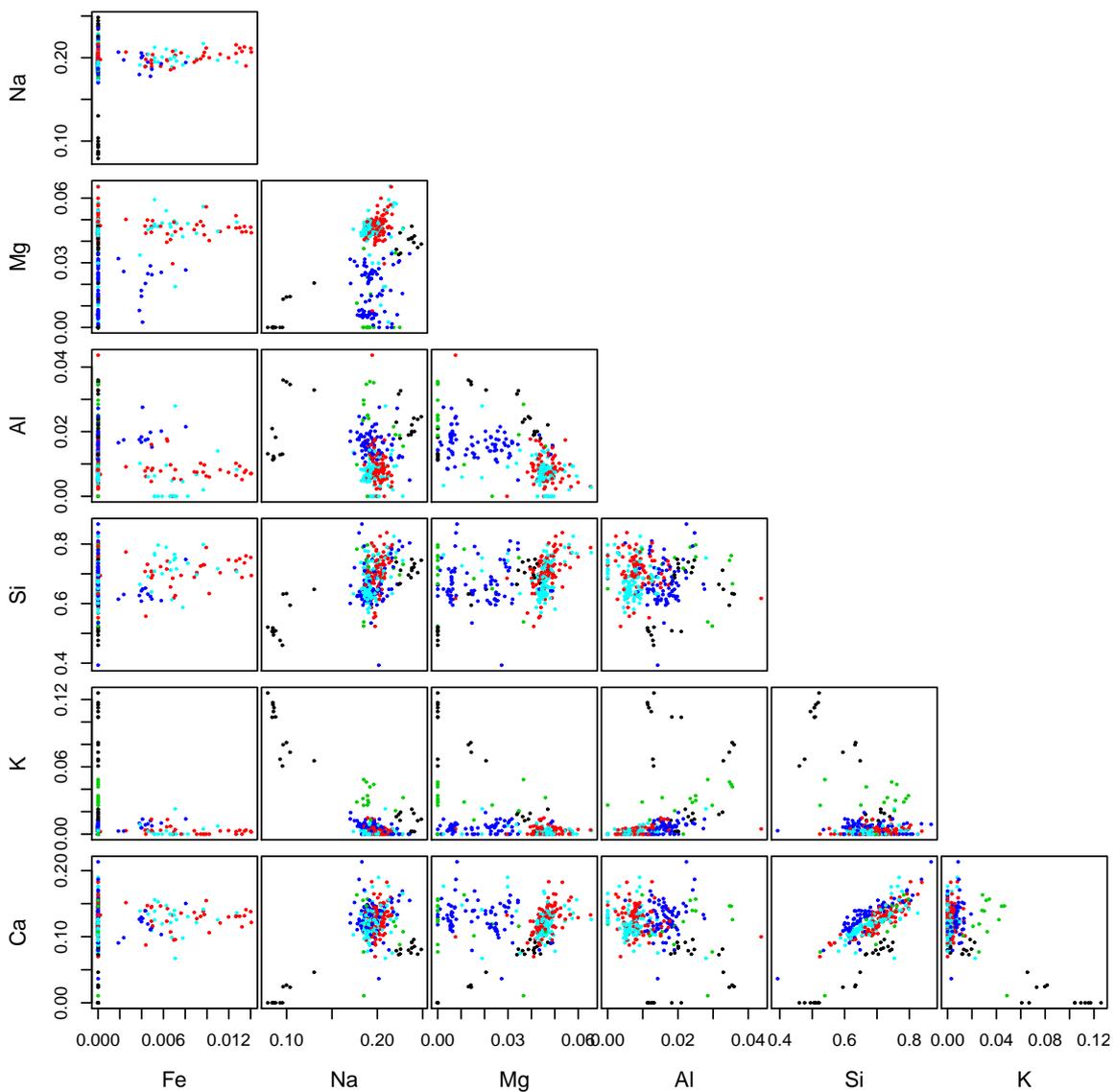


Figure 2.1: Scatterplots of the ratios to oxygen for all item means from the database. The different coloured points correspond to the five use type categories: bulb, car window, headlamp, container and building window.

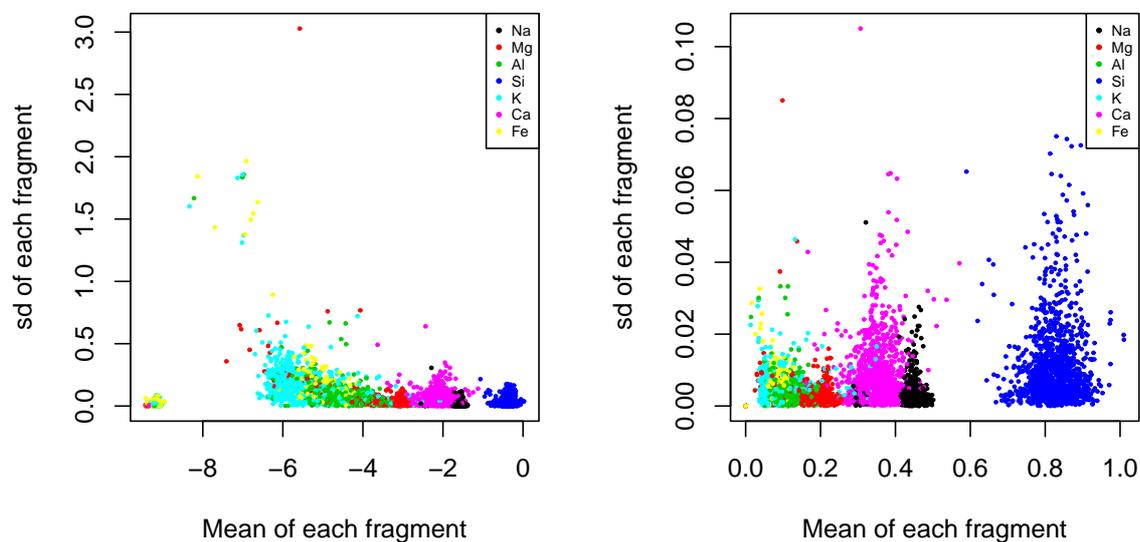


Figure 2.2: Plots of fragments’ standard deviations against corresponding means, using the alr (left panel) and the square root (right panel) transformations. For the alr, 0.005 was added to all compositional zeros. Seven elemental pairs (mean, sd) are plotted for each fragment, computed using the fragment’s three repeated measurements. While the variability is roughly the same across the range of mean levels for the square root transformed data, the range of sd’s changes considerably across mean levels for the alr transformed data.

Despite the square root transformation improving normality and stabilising the variability in the data, the presence of many zeros still has a bearing on the distribution of the data. One way to reduce the influence of the zeros is to partition the glass data depending on whether an element is present or absent from a glass item’s composition. This reduces (2.19) from a $(D - 1)$ dimensional composition to a $(D - Z - 1)$ dimensional subcomposition by eliminating the Z zero elements.

Another benefit to examining whether elements are present or absent from

an item's composition is when it comes to determining that item's use type. For example, the database has no bulbs or headlamps containing iron, and so if a composition containing iron is observed then it is unlikely that it is either of these two use types. Three elements are always present in glass: oxygen, silicon and sodium. The remaining five can be present or absent from the composition of a glass item, which gives $2^5 = 32$ possible configurations of the elements according to their presence or absence. Ten of the possible 32 configurations were observed in the glass database, as shown in Table 2.2. As can be seen from Table 2.2, most of the configurations account for very few of the items in the database. The elements iron and potassium are accountable for 87.9% of the zero measurements in the database, as seen in Table 2.1, therefore most of the zeros present in the database can be removed by only focusing on the presence or absence of these two elements. Therefore only four configurations are later considered as shown in Table 2.3, where some configurations from Table 2.2 are combined together.

Table 2.2: Presence (1) and absence (0) of elements at item level by use type.

	Element					Glass type					Total
	Mg	Al	K	Ca	Fe	bulb	car window	headlamp	container	building window	
1	1	1	1	1	1	0	23	0	12	7	42
2	1	1	1	1	0	16	40	5	47	55	163
3	1	1	0	1	1	0	10	0	0	6	16
4	1	1	0	1	0	1	18	0	17	25	61
5	1	0	1	1	0	0	0	1	0	0	1
6	1	0	0	1	1	0	1	0	0	9	10
7	1	0	0	1	0	0	2	0	0	3	5
8	0	1	1	1	0	0	0	8	1	0	9
9	0	1	1	0	0	9	0	0	0	0	9
10	0	1	0	1	0	0	0	2	2	0	4
						26	94	16	79	105	320

Table 2.3: 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: $\overline{\text{Fe}}, \text{K}$	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

Just from observing an item's composition it is obvious which elements are present and which are absent: only eight of the 320 items have an element where all 12 measurements are neither all positive nor zero, as shown in Ta-

ble 2.4. Here an element is assumed to be present in an item's composition if at least one of its 12 measurements is positive. Of the eight items in Table 2.4 five of them contain measurements on iron or potassium that are neither all positive nor zero. For those five items 28/120 of the total measurements associated with the elements iron and potassium are zero. Under the assumption that an element is present if at least one of its measurements is positive this means that $28/4204 = 0.7\%$ of zero measurements would 'slip by' when observing the presence or absence of these two elements. Also, across all eight of the items in Table 2.4 containing an element with its 12 measurements not all positive or all zero, these zeros account for only 0.9% of zeros in the entire glass database, and so the assumption that an element is present if at least one measurement is positive seems reasonable.

Table 2.4: Frequency of items containing chemical elements where all 12 measurements are neither all positive nor zero.

Glass type	Element					Total
	Mg	Al	K	Ca	Fe	
car window	0	1	0	0	2	3
building window	1	1	3	0	0	5
Total	1	2	3	0	2	8

Taking into account the presence or absence of elements should improve normality assumptions when it comes to modelling the data by decreasing the influence of the zeros on the distribution of the data. This can be seen by comparing Figure 2.3, containing plots of the means at item level for all items in the database, with Figure 2.4, containing plots of the item means for all items with configuration 2 from Table 2.3.

The next chapter discusses a Bayesian hierarchical model for the glass data. It also details the two different approaches taken to dealing with the many compositional zeros. The first approach assumes all of the compositional zeros are due to the concentrations of those elements being below the limits of the detection equipment, and updates the zeros during the modelling procedure with values below this detection limit. The second approach focuses on using the four configurations mentioned above by separating the data into four distinct subsets according to the presence or absence of the elements iron and potassium. This gives four separate models for this approach that are then brought together to form a single model.

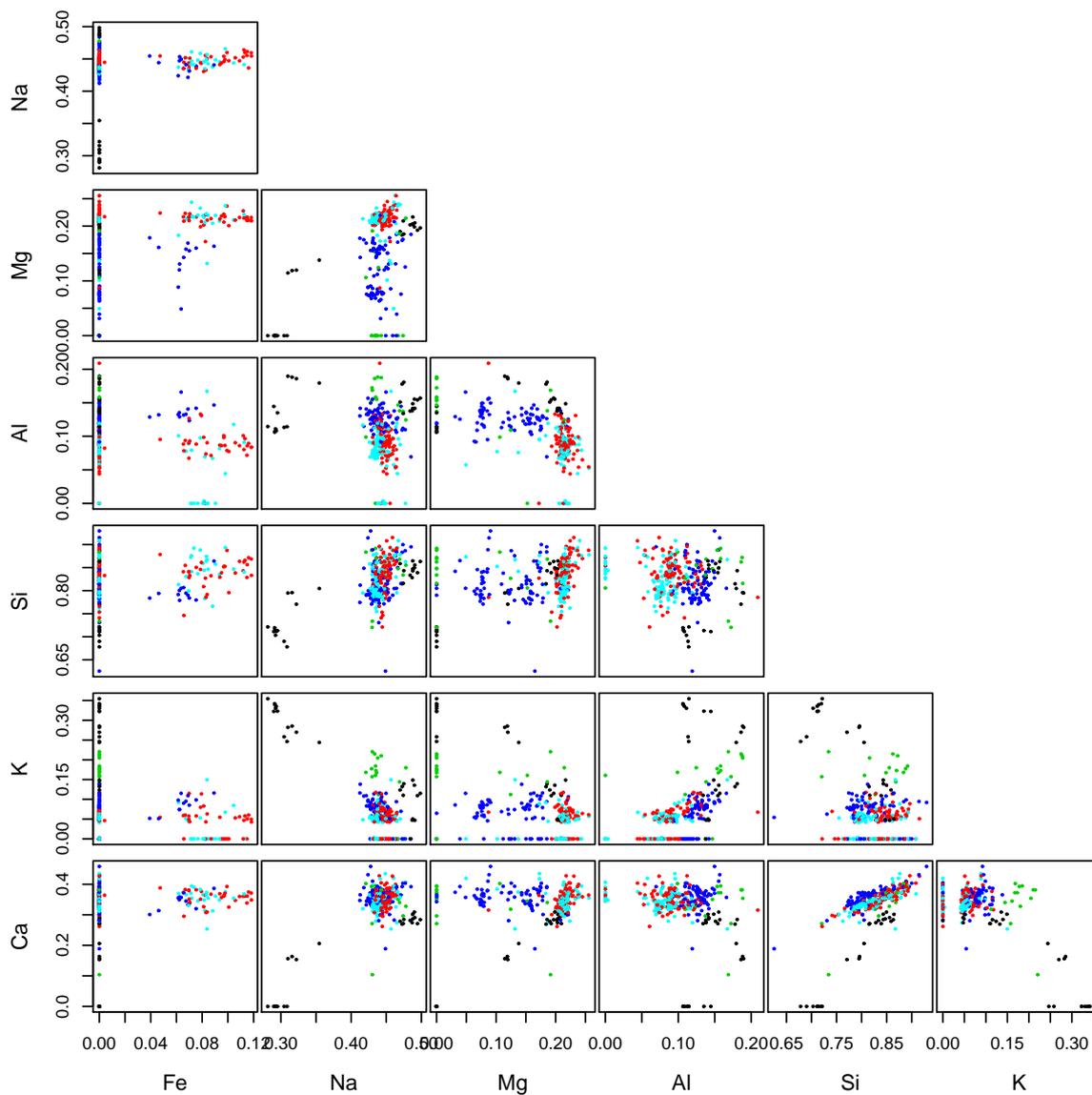


Figure 2.3: Scatterplots of the square root transformed ratios of all item means from the database. The different coloured points correspond to the five use type categories: bulb, car window, headlamp, container and building window.

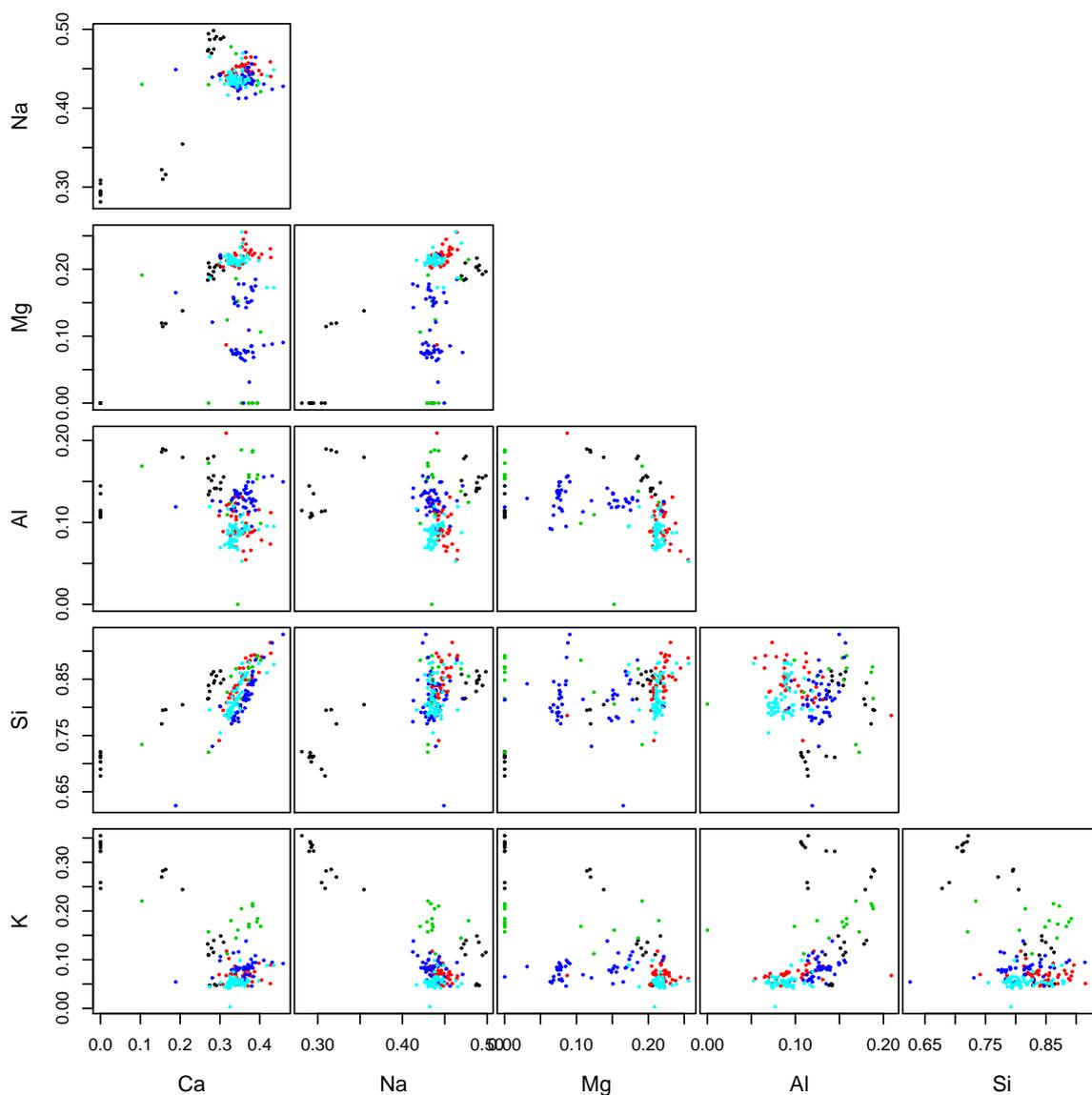


Figure 2.4: Scatterplots of the square root transformed ratios of the item means for items with configuration 2 from Table 2.3. The different coloured points correspond to the five use type categories: bulb, car window, headlamp, container and building window.

Chapter 3

Models

The elemental content of glass fragments has been previously modelled by [Aitken and Lucy \(2004\)](#) and [Neocleous *et al.* \(2011\)](#) from a frequentist perspective. The models proposed were random effects models incorporating two levels of variation: between-item and within-item. The between-item level variability is captured by a random effect associated with individual glass items, and the within-item variability by a random effect associated with individual fragments from the same glass item. [Aitken and Lucy \(2004\)](#) performed their analysis on building windows using the logarithm of three different ratios believed to be the most discriminatory for such data. [Neocleous *et al.* \(2011\)](#) used log-ratios for the entire elemental composition of glass, with oxygen chosen as the common divisor. Along with the log-ratio transformation, [Neocleous *et al.* \(2011\)](#) also used a complementary log-log transformation – logarithm of the negative log-ratio transformation – and a spherical transformation mapping the elemental composition onto the unit

hypersphere. In terms of results, there was no single transformation outperforming the others. For instance under the normal model used by [Neocleous *et al.* \(2011\)](#), the spherical transformation yielded the lowest false positive rate, while the log-ratio transformation yielded the lowest false negative rate.

Here a Bayesian approach is used to model the hierarchical structure of the data using a mixed effects model. As the hierarchical structure of the data contains an additional layer relating to the repeated measurements on each fragment, an additional random effect is placed at this measurement error level on top of the two levels of variability already mentioned. The model is a mixed effects model and not a random effects model as it also includes a fixed effect term for the overall mean for each glass type. For a detailed introduction to mixed effects models see [Pinheiro and Bates \(2000\)](#). See also [Gelman *et al.* \(2004\)](#) for details on hierarchical models in a Bayesian framework. As was seen from the exploratory analysis in Chapter 2 the square roots of the compositional ratios improved normality and stability in the data variability more so than the log-ratios. Therefore the primary transformation used when analysing the glass database will be the square roots of the compositional ratios, with oxygen chosen as the common divisor. Also, unlike the log-ratio transformation, the square root transformation does not require any zeros present in the data to be altered. Results from the model using the log-ratio transformation will also be produced for comparison's sake, and to demonstrate why it does not perform as well as the square root transformation.

3.1 Bayesian hierarchical model

Denote the square root ratios for each glass measurement by \mathbf{z}_{tijk} , where \mathbf{z}_{tijk} is the k -th replicate measurement from the j -th fragment of the i -th glass item of use type t . The dimension p of \mathbf{z}_{tijk} at item level may differ across the elemental configurations detailed in Section 2.3 when the approach to modelling the presence or absence of the elements iron and potassium is taken; see Section 3.2.3. It is then assumed that

$$\begin{aligned} \mathbf{z}_{tijk} &= \boldsymbol{\theta}_t + \mathbf{b}_{ti} + \mathbf{c}_{tij} + \boldsymbol{\epsilon}_{tijk}, \\ \mathbf{b}_{ti} &\stackrel{\text{iid}}{\sim} N_p(\mathbf{0}, \Omega_t^{-1}), \quad \mathbf{c}_{tij} \stackrel{\text{iid}}{\sim} N_p(\mathbf{0}, \Psi^{-1}), \quad \boldsymbol{\epsilon}_{tijk} \stackrel{\text{iid}}{\sim} N_p(\mathbf{0}, \Lambda^{-1}). \end{aligned} \tag{3.1}$$

The fixed effect term for the mean of use type t is denoted by the parameter $\boldsymbol{\theta}_t$; the item-level random effect by \mathbf{b}_{ti} ; the fragment within-item random effect by \mathbf{c}_{tij} ; and the error at measurement level by $\boldsymbol{\epsilon}_{tijk}$. Each of the random effects are assumed to have multivariate normal distributions, with unknown precision matrices Ω_t , Ψ and Λ . The separate covariance matrices, Ω_t^{-1} , were introduced at item level for each use type after observing dissimilar levels of random variability between items of differing use types, which will be seen from the model results. The assumption of normality is questionable and may not hold after looking at scatterplots of the data in Chapter 2, even when the zero concentrations are removed. However, the validity of this assumption did not have any substantial affect on the conclusions drawn from the models used, as seen from the results of classification and evidence evaluation in chapters 4 and 5. Misspecifying random effects distributions does not greatly affect estimations of the random effects variances (McCulloch and Neuhaus, 2011). Shorthand notations for each of the parameters in the model are

$\boldsymbol{\theta} = \{\boldsymbol{\theta}_t\}_{t=1}^T$; $\mathbf{b} = \{\mathbf{b}_{ti}\}_{i=1}^{I_t} \quad T_{t=1}$; $\mathbf{c} = \{\mathbf{c}_{tij}\}_{j=1}^J \quad I_{i=1} \quad T_{t=1}$ and $\Omega = \{\Omega_t\}_{t=1}^T$. Here $T = 5$ denotes the number of use types in the database; I_t denotes the number of glass items of each use type t ($I_1 = 26$, $I_2 = 94$, $I_3 = 16$, $I_4 = 79$, $I_5 = 105$); $J = 4$ denotes the total number of fragments associated with each item; and $K = 3$ is the number of replicate measurements on each fragment. For a glass item \mathbf{z} of use type $\mathcal{T}_{\mathbf{z}} = t$ with JK measurements, model (3.1) implies - without conditioning on the random effects - that the distribution of item \mathbf{z} is

$$\mathbf{z} | \mathcal{T}_{\mathbf{z}} = t, \boldsymbol{\xi} \sim N_{JKp}(\mathbf{1}_{JK} \otimes \boldsymbol{\theta}_t, \Sigma_t), \quad (3.2)$$

where $\boldsymbol{\xi} = \{\boldsymbol{\theta}, \Omega, \Psi, \Lambda\}$ collectively denotes the model parameters. When modelling the presence and absence of iron and potassium the model parameters for the subset of items with configuration m in Table 2.3 will be denoted by ξ_m . The covariance matrix Σ_t is given by

$$\Sigma_t = (\mathbf{1}_{JK} \mathbf{1}'_{JK}) \otimes \Omega_t^{-1} + [\mathbb{I}_J \otimes (\mathbf{1}_K \mathbf{1}'_K)] \otimes \Psi^{-1} + \mathbb{I}_{JK} \otimes \Lambda^{-1}, \quad (3.3)$$

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

The prior distributions placed on the fixed effects $\boldsymbol{\theta}_t$ are also assumed independent multivariate normals, but they are restricted to the positive orthant, in order to ensure that the square root transformed means are non-negative:

$$\boldsymbol{\theta}_t \stackrel{\text{iid}}{\sim} N_p(\mathbf{0}, \Phi^{-1}), \quad \boldsymbol{\theta}_t > \mathbf{0}, \quad t = 1, \dots, T. \quad (3.4)$$

The covariance matrix, Φ^{-1} , of the fixed effects is fixed and set equal to $s \cdot \mathbb{I}_p$, where s is a relatively large constant. All subsequent results used the value $s = 1000$. This gives a priori independent components of $\boldsymbol{\theta}_t$ with very large spread, so that the posterior modes of $\boldsymbol{\theta}_t$ will be very close to the

corresponding sample means. The precision matrices for the random effects have conjugate Wishart hyperpriors placed on each of them:

$$\Omega_t \sim W_p(d_{1t}, A_t), \quad \Psi \sim W_p(d_2, B), \quad \Lambda \sim W_p(d_3, C), \quad (3.5)$$

where d_{1t} , d_2 and d_3 denote the degrees of freedom; and A_t , B and C are precision matrices. The degrees of freedom of the Wishart distribution need to be greater than the data dimension minus one, e.g. $d_2 > p - 1$, so noninformative prior values for the degrees of freedom are set equal to p ; see [DeGroot \(1970\)](#) for details on the Wishart distribution. The precision matrices A_t , B and C are all set equal to $(1/1000) \cdot \mathbb{I}_p$ so that, as mentioned for the θ_t 's above, posterior inferences would be largely driven by the data.

3.1.1 Markov Chain Monte Carlo implementation

In Bayesian inference Markov Chain Monte Carlo (MCMC) methods are used to simulate and draw samples from distributions of interest. Monte Carlo methods are used to approximate integrals and closed-form expressions that are otherwise extremely difficult or impossible to evaluate. This is done by creating a Markov chain which, after reaching a state of equilibrium, is effectively sampling from the desired target distribution. This is an iterative algorithm procedure that after a period of time will have produced a chain consisting of samples drawn from the target distribution.

The two most frequently used MCMC algorithms are the Gibbs sampler and the Metropolis-Hastings algorithm. Depending on the model, only one of the sampling techniques may be required, but it is also possible to have a hybrid

sampling technique that uses both methods. As the core of the sampler used here is Gibbs sampling, the Gibbs sampler will be described first, before then going on to look at the Metropolis-Hastings algorithm and the corresponding moves incorporated into the sampler.

Gibbs sampling

Since the full conditional distributions of the random effects $\{\mathbf{b}, \mathbf{c}\}$ and the parameters of $\xi = \{\boldsymbol{\theta}, \Omega, \Psi, \Lambda\}$ - minus $\boldsymbol{\theta}$ with its applied restriction - are known standard distributions, Gibbs sampling moves can be used to update these parameters. Gibbs sampling moves work by generating values from the full conditional distribution of a variable, i.e. the conditional distribution of a variable given all other variables. For example, let the vector containing all parameters of interest be $\boldsymbol{\zeta} = (\zeta_1, \dots, \zeta_n)'$, then the Gibbs sampler operates as follows:

1. First set initial values for the parameters $\boldsymbol{\zeta}^0 = (\zeta_1^0, \dots, \zeta_n^0)'$
2. Iteratively generate values for $\boldsymbol{\zeta}$ where the first iterate is generated as follows:

$$\zeta_1^1 \sim p(\zeta_1^1 \mid \zeta_2^0, \dots, \zeta_n^0)$$

$$\zeta_2^1 \sim p(\zeta_2^1 \mid \zeta_1^1, \zeta_3^0, \dots, \zeta_n^0)$$

$$\zeta_3^1 \sim p(\zeta_3^1 \mid \zeta_1^1, \zeta_2^1, \zeta_4^0, \dots, \zeta_n^0)$$

$$\vdots$$

$$\zeta_n^1 \sim p(\zeta_n^1 \mid \zeta_1^1, \dots, \zeta_{n-1}^1)$$

3. Repeat step 2 for each iteration of the sampler.

The Gibbs sampler is a special case of the Metropolis-Hastings algorithm where each move is always accepted with probability one. For a detailed explanation of the Gibbs sampler see [Casella and George \(1992\)](#). The full conditional distributions of each of the parameters used in the Gibbs sampler are given in [Appendix B.1](#). A “burn-in” period can be implemented into the sampler where the initial iterations that have not reached a state of equilibrium are discarded, with only the draws made after this period used in the analysis. Any autocorrelation in the draws from the sampler can also be reduced by “thinning” the Markov chain: every m -th draw of the sampler is stored and the rest discarded. Thinning can reduce autocorrelation, but it can also lead to a loss in precision, with an increase in the variance for thinned chains compared to unthinned chains ([Link and Eaton, 2012](#)).

In mixed effects models there may be strong correlations between the parameters, such as relationships between the fixed effects and the different levels of random effects in the hierarchical model. There may also be strong relationships between the random effects and their corresponding variances. This can cause issues with poor mixing using Gibbs samplers, with different MCMC strategies proposed to deal with such problems. [Gelfand *et al.* \(1995\)](#) introduced hierarchical centering that can be used to reduce strong correlation between parameters. [Liu *et al.* \(1998\)](#) proposed a parameter expansion technique which was adapted for the Gibbs sampler ([Liu and Wu, 1999](#)) and involves the inclusion of additional auxiliary parameters to reduce the effects of correlation between parameters. An overview of each method can be found in [Browne \(2004\)](#).

Metropolis-Hastings algorithm

The Metropolis-Hastings algorithm (Hastings, 1970) is a generalisation of the Metropolis algorithm of Metropolis *et al.* (1953) and allows draws to be made from any probability distribution, given that the target distribution of interest can be computed at a specific value. The Metropolis-Hastings algorithm uses a proposal density to generate a candidate state for the Markov chain to move to with some probability, else the chain remains at its current state. Let ζ denote the current state, where $p(\zeta)$ denotes the target distribution at ζ . Let $q(\tilde{\zeta}|\zeta)$ denote the proposal distribution, which proposes moving to the state $\tilde{\zeta}$ when currently at state ζ ; where $q(\zeta|\tilde{\zeta})$ denotes the move in the opposite direction. The Metropolis-Hastings algorithm is given as follows:

1. Specify an initial state, ζ_0 , for the current state ζ .
2. For $i = 1, \dots, n$
 - (a) Sample a new candidate state $\tilde{\zeta}_i$ from $q(\tilde{\zeta}_i|\zeta_i)$.
 - (b) $\tilde{\zeta}_i$ is accepted with probability $\min(1, \alpha)$ with

$$\alpha = \frac{p(\tilde{\zeta}_i) q(\zeta_i|\tilde{\zeta}_i)}{p(\zeta_i) q(\tilde{\zeta}_i|\zeta_i)}. \quad (3.6)$$

Generate a random $u \sim Unif(0, 1)$, accepting $\tilde{\zeta}_i$ if $u < \alpha$, otherwise remain at the current state ζ_i .

If the proposal density $q(\tilde{\zeta}|\zeta)$ is symmetric then $q(\tilde{\zeta}|\zeta) = q(\zeta|\tilde{\zeta})$, thus reducing the acceptance probability α to

$$\alpha = \frac{p(\tilde{\zeta})}{p(\zeta)}, \quad (3.7)$$

the ratio of target probability densities at the candidate state $\tilde{\zeta}$ and the current state ζ .

A symmetrical proposal often used generates candidate values by adding to the current state a value drawn from a $Unif(-\delta, \delta)$ distribution:

$$\tilde{\zeta} = \zeta + v, \quad \text{where } v \sim Unif(-\delta, \delta). \quad (3.8)$$

This is sometimes referred to as a random walk Metropolis-Hastings and is symmetric due to $q(\tilde{\zeta}|\zeta) = q(\zeta|\tilde{\zeta}) = 1/2\delta$. There are various other proposal choices that can also be implemented. For a more detailed description of the Metropolis-Hastings algorithm see [Chib and Greenberg \(1995\)](#). The proposed model implements three separate Metropolis-Hastings moves detailed below.

Metropolis-Hastings move 1

The first Metropolis-Hastings move (M-H 1) is used to update the fixed mean parameter $\boldsymbol{\theta}_t$. As the prior distribution on $\boldsymbol{\theta}_t$ is restricted to the positive orthant, it requires updating using a Metropolis-Hastings move in order to prevent the acceptance of negative $\boldsymbol{\theta}_t$ values. M-H 1 updates $\boldsymbol{\theta}_t$ using a proposal distribution equal to the full conditional distribution of $\boldsymbol{\theta}_t$ minus the positive orthant restriction. The candidate value $\tilde{\boldsymbol{\theta}}_t$ is therefore drawn from a multivariate normal distribution:

$$\tilde{\boldsymbol{\theta}}_t \sim N_p(\tilde{\boldsymbol{\phi}}_t, \tilde{\boldsymbol{\Phi}}_t^{-1}),$$

with acceptance probability α given by

$$\begin{aligned}
\alpha &= \min \left(1, \frac{p(\tilde{\xi}, \mathbf{b}, \mathbf{c}) q(\boldsymbol{\theta}_t | \tilde{\boldsymbol{\theta}}_t)}{p(\xi, \mathbf{b}, \mathbf{c}) q(\tilde{\boldsymbol{\theta}}_t | \boldsymbol{\theta}_t)} \right) \\
&= \min \left(1, \frac{p(\tilde{\boldsymbol{\theta}}_t | \dots) p(\dots) q(\boldsymbol{\theta}_t | \tilde{\boldsymbol{\theta}}_t)}{p(\boldsymbol{\theta}_t | \dots) p(\dots) q(\tilde{\boldsymbol{\theta}}_t | \boldsymbol{\theta}_t)} \right) \\
&= \min \left(1, \frac{N_p(\tilde{\boldsymbol{\theta}}_t | \tilde{\boldsymbol{\phi}}_t, \tilde{\Phi}_t^{-1}) I(\tilde{\boldsymbol{\theta}}_t \in \mathbb{R}_+^p) N_p(\boldsymbol{\theta}_t | \tilde{\boldsymbol{\phi}}_t, \tilde{\Phi}_t^{-1})}{N_p(\boldsymbol{\theta}_t | \tilde{\boldsymbol{\phi}}_t, \tilde{\Phi}_t^{-1}) \underbrace{I(\boldsymbol{\theta}_t \in \mathbb{R}_+^p)}_{=1} N_p(\tilde{\boldsymbol{\theta}}_t | \tilde{\boldsymbol{\phi}}_t, \tilde{\Phi}_t^{-1})} \right) \\
&= I(\tilde{\boldsymbol{\theta}}_t \in \mathbb{R}_+^p), \tag{3.9}
\end{aligned}$$

where the joint posterior distribution, $p(\xi, \mathbf{b}, \mathbf{c})$, of the model is reduced to the expressions only involving $\boldsymbol{\theta}_t$: the full conditional distribution $p(\boldsymbol{\theta}_t | \dots)$, where “ \dots ” denotes all other variables in the model. The full conditional distributions of all the model variables are shown in Section B.1 of Appendix B. The candidate value $\tilde{\boldsymbol{\theta}}_t$ is then accepted with probability 1 or 0 depending on whether it lies in the positive orthant or not.

Metropolis-Hastings move 2

The second Metropolis-Hastings move (M-H 2) is also used to update $\boldsymbol{\theta}_t$, but this time only for glass items of use type bulb, i.e. $\boldsymbol{\theta}_1$. This additional move is performed only on $\boldsymbol{\theta}_1$ and not on glass items of the other four use types as the MCMC samples for bulbs displayed noticeable positive autocorrelation. This may be down to the fact that the percentage weights amongst bulbs are more variable than the other use types, with two separate clusters of bulbs observed; see Figure 2.3. M-H 2 is a random walk move that is performed

on each element of $\boldsymbol{\theta}_1$ separately:

$$\tilde{\boldsymbol{\theta}}_{1l} = \boldsymbol{\theta}_{1l} + v, \quad \text{where } v \sim \text{Unif}(-\delta_{1l}, \delta_{1l}), \quad (3.10)$$

where the interval widths δ_{1l} were determined from preliminary runs of the model; see Section B.2 of Appendix B. The probability of accepting the newly generated candidate is then

$$\begin{aligned} \alpha &= \min \left(1, \frac{p(\tilde{\boldsymbol{\xi}}, \mathbf{b}, \mathbf{c}) q(\boldsymbol{\theta}_{1l} | \tilde{\boldsymbol{\theta}}_{1l})}{p(\boldsymbol{\xi}, \mathbf{b}, \mathbf{c}) q(\tilde{\boldsymbol{\theta}}_{1l} | \boldsymbol{\theta}_{1l})} \right) \\ &= \min \left(1, \frac{p(\tilde{\boldsymbol{\theta}}_1 | \cdots) p(\cdots) q(\boldsymbol{\theta}_{1l} | \tilde{\boldsymbol{\theta}}_{1l})}{p(\boldsymbol{\theta}_1 | \cdots) p(\cdots) q(\tilde{\boldsymbol{\theta}}_{1l} | \boldsymbol{\theta}_{1l})} \right) \\ &= \min \left(1, \frac{p(\tilde{\boldsymbol{\theta}}_1 | \cdots)}{p(\boldsymbol{\theta}_1 | \cdots)} \right), \end{aligned} \quad (3.11)$$

where $q(\boldsymbol{\theta}_{1l} | \tilde{\boldsymbol{\theta}}_{1l}) = q(\tilde{\boldsymbol{\theta}}_{1l} | \boldsymbol{\theta}_{1l}) = 1/2\delta_{1l}$. This is done for each element $l = 1, \dots, p$ where the only difference between the densities $p(\tilde{\boldsymbol{\theta}}_1 | \cdots)$ and $p(\boldsymbol{\theta}_1 | \cdots)$ in (3.11) is the candidate generated value of the l -th element of $\tilde{\boldsymbol{\theta}}_1$.

Metropolis-Hastings move 3

The third Metropolis-Hastings move (M-H 3) is performed on both the fixed effect $\boldsymbol{\theta}_t$ and the item-level random effect \mathbf{b}_{ti} simultaneously. Here the proposed candidate state is chosen in a way that leaves the value of the likelihood unchanged between the current and candidate states. Therefore, the candidate values generated for $\boldsymbol{\theta}_t$ and \mathbf{b}_{ti} are generated such that $\boldsymbol{\theta}_t + \mathbf{b}_{ti} = \tilde{\boldsymbol{\theta}}_t + \tilde{\mathbf{b}}_{ti}$. The candidate for the fixed effect $\boldsymbol{\theta}_t$ is generated as follows:

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

where $\mathbf{v} = (v_1, \dots, v_p)'$, with components $v_l \sim \text{Unif}(-\delta_{tl}, \delta_{tl})$ independently; with the values δ_{tl} for the interval widths determined from a pilot run; see

Section B.2 of Appendix B. In order for the likelihood to remain unchanged and for $\boldsymbol{\theta}_t + \mathbf{b}_{ti} = \tilde{\boldsymbol{\theta}}_t + \tilde{\mathbf{b}}_{ti}$ to hold, the candidate values of the item-level random effects are then set equal to

$$\tilde{\mathbf{b}}_{ti} = \mathbf{b}_{ti} - \mathbf{v}, \quad i = 1, \dots, I_t.$$

As the proposed candidate values do not change the value of the likelihood, the ratio of the target densities in the acceptance probability reduces to the ratio of prior densities for $\boldsymbol{\theta}_t$ and \mathbf{b}_t evaluated at the candidate and current states, along with two additional terms. The acceptance probability can then be computed as follows, using the approach described in Section 2.2 of [Green \(2003\)](#):

$$\begin{aligned} \alpha &= \min \left(1, \frac{p(\tilde{\boldsymbol{\xi}}, \tilde{\mathbf{b}}, \mathbf{c}) \tilde{f}(\tilde{\mathbf{v}})}{p(\tilde{\boldsymbol{\xi}}, \tilde{\mathbf{b}}, \mathbf{c}) f(\mathbf{v})} \left| \frac{\partial(\tilde{\boldsymbol{\theta}}_t, \tilde{\mathbf{b}}_t, \tilde{\mathbf{v}})}{\partial(\boldsymbol{\theta}_t, \mathbf{b}_t, \mathbf{v})} \right| \right) \\ &= \min \left(1, \frac{p(\tilde{\boldsymbol{\theta}}_t) p(\tilde{\mathbf{b}}_t | \Omega_t) \tilde{f}(\tilde{\mathbf{v}})}{p(\boldsymbol{\theta}_t) p(\mathbf{b}_t | \Omega_t) f(\mathbf{v})} \left| \frac{\partial(\tilde{\boldsymbol{\theta}}_t, \tilde{\mathbf{b}}_t, \tilde{\mathbf{v}})}{\partial(\boldsymbol{\theta}_t, \mathbf{b}_t, \mathbf{v})} \right| \right). \end{aligned}$$

The density $f(\cdot)$ is uniform on a hyper-cube across the values of \mathbf{v} , with $\tilde{f}(\cdot) = f(\cdot)$ such that $\tilde{\mathbf{v}} = -\mathbf{v}$ in the reverse move. The absolute value of the determinant of the Jacobian matrix is

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

The acceptance probability for M-H 3 is then shown to only involve the prior densities of $\boldsymbol{\theta}_t$ and \mathbf{b}_t at the candidate and current values:

$$\alpha = \min \left(1, \frac{p(\tilde{\boldsymbol{\theta}}_t) p(\tilde{\mathbf{b}}_t | \Omega_t)}{p(\boldsymbol{\theta}_t) p(\mathbf{b}_t | \Omega_t)} \right). \quad (3.12)$$

The three Metropolis-Hastings moves detailed are integrated into the Gibbs sampler that is used to update the parameters where their full conditional distributions are known.

The above sampler describes how posterior draws are obtained for the parameters in model (3.1) but the issue of compositional zeros still needs to be addressed. However, since the square root of the compositional ratios does not need any of the zeros to be altered in order to be used, the results from the MCMC with all compositional zero values unaltered will be displayed first before introducing the two different approaches to modelling the zeros using model (3.1).

3.1.2 Posterior samples from the Bayesian hierarchical model

The posterior samples shown here were obtained from the Bayesian hierarchical model with all compositional zeros left unaltered using the MCMC procedure detailed in Section 3.1.1. To obtain the posterior draws, the MCMC algorithm was coded in the statistical programming language R (R Development Core Team, 2011). The time taken to obtain the simulated model results was approximately 12 hours. This included a burn-in period of 10,000. Thinning of the Markov chain was also used to aid convergence with every 200th draw from the chain kept and the rest discarded. All subsequent samples from the posterior consist of 1,000 draws. The acceptance rate for Metropolis-Hastings move M-H 1 in Section 3.1.1 is not explicitly stated due to almost all new candidates being accepted due to the mean lying far enough

away from the constraint, with the variance small enough to ensure the constraint is rarely broken. The acceptance rate for Metropolis-Hastings move M-H 2 in Section 3.1.1 that is performed on θ_1 only was 39%, and the acceptance rate for Metropolis-Hastings move M-H 3 in Section 3.1.1 which jointly updates both θ and \mathbf{b} was 37%. Time series plots of the draws obtained for the sampled fixed effect θ_t are shown in Figure 3.1 and indicate good convergence of θ_t for $t = 2, 3, 4, 5$ with the convergence for $t = 1$, i.e. bulbs, less so. This is due to the bulb use type displaying more variability at item level than the other four use types. This can be most easily seen in the second column of plots (for the element sodium) of Figure 2.3 where there appear to be two separate groupings of bulbs containing different levels of sodium. This is the main reason why separate covariance matrices at item level, Ω_t^{-1} , were incorporated into the model, to allow change across use types, which is displayed in Table 3.2. Also seen from Figure 3.1 is the influence of the abundant levels of zero concentrations of iron in bulbs and headlamps, i.e. θ_1 and θ_3 . Table 3.1 contains the effective sample sizes for θ_t and reflects the poorer convergence for bulbs, with much smaller effective sample sizes obtained for the elements which differ between the two separate groups seen in Figure 2.3. The effective sample sizes were obtained using the *effectiveSize* function from the *coda* package in R (R Development Core Team, 2011). Figure 3.2 displays scatterplots of the draws of θ_t and shows clear separation in the means between the five different glass use types. The influence of the zeros can be seen (first column of Figure 3.2) from the point mass close to zero for the use types bulb and headlamp. Figure 3.2 also shows how the mean for bulbs, θ_1 , appears stretched between the two separate clusters of bulbs seen in Figure 2.3 which could be seen as a potential problem. However, with the

introduction of the two approaches used to handle the zeros, the models do not have any problems in correctly classifying or distinguishing glass items of that use type from the others, as seen in Chapter 4.

Table 3.1: Effective sample size from the Bayesian hierarchical model with unaltered compositional zeros for the mean vector θ_t . For θ_t , $t = 1, \dots, 5$ correspond to use types: bulb, car window, headlamp, container and building window.

	Na	Mg	Al	Si	K	Ca	Fe
θ_1	70.1	68.0	231.4	79.3	79.9	71.6	1134.0
θ_2	1000.0	1000.0	1000.0	1000.0	732.4	1000.0	1000.0
θ_3	353.7	616.4	500.93	732.7	189.1	539.8	855.4
θ_4	780.5	1000.0	1454.4	1000.0	821.9	1000.0	991.3
θ_5	1000.0	1000.0	791.4	894.9	802.4	871.6	775.4

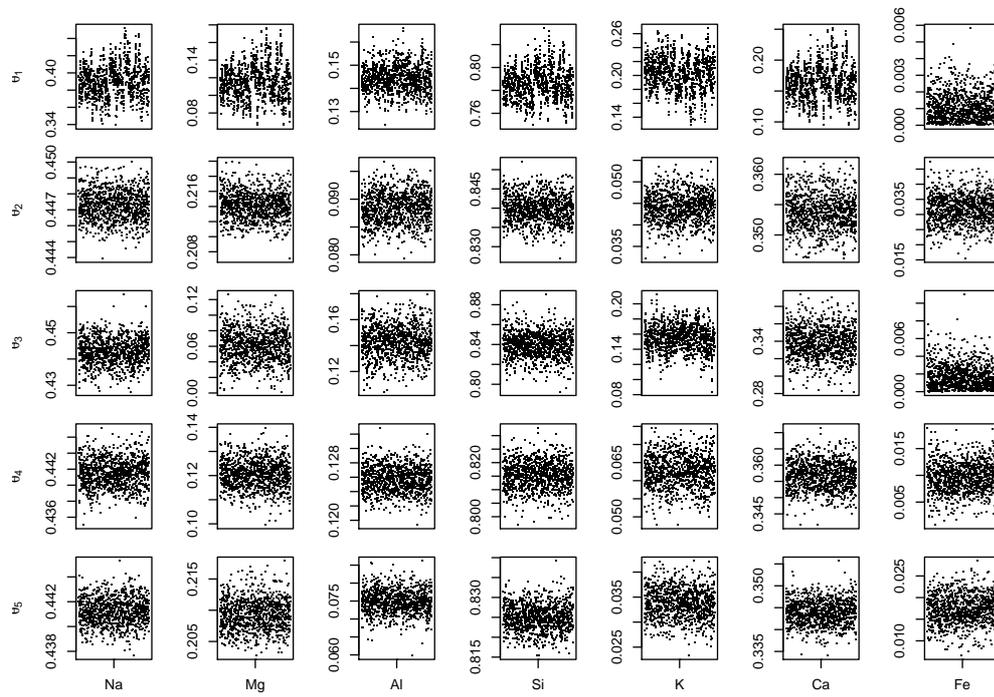


Figure 3.1: Trace plots of the mean θ_t from the Bayesian hierarchical model with all compositional zeros unaltered. A burn-in period of 10,000 was used, and thinning of the Markov chain with every 200th draw stored.

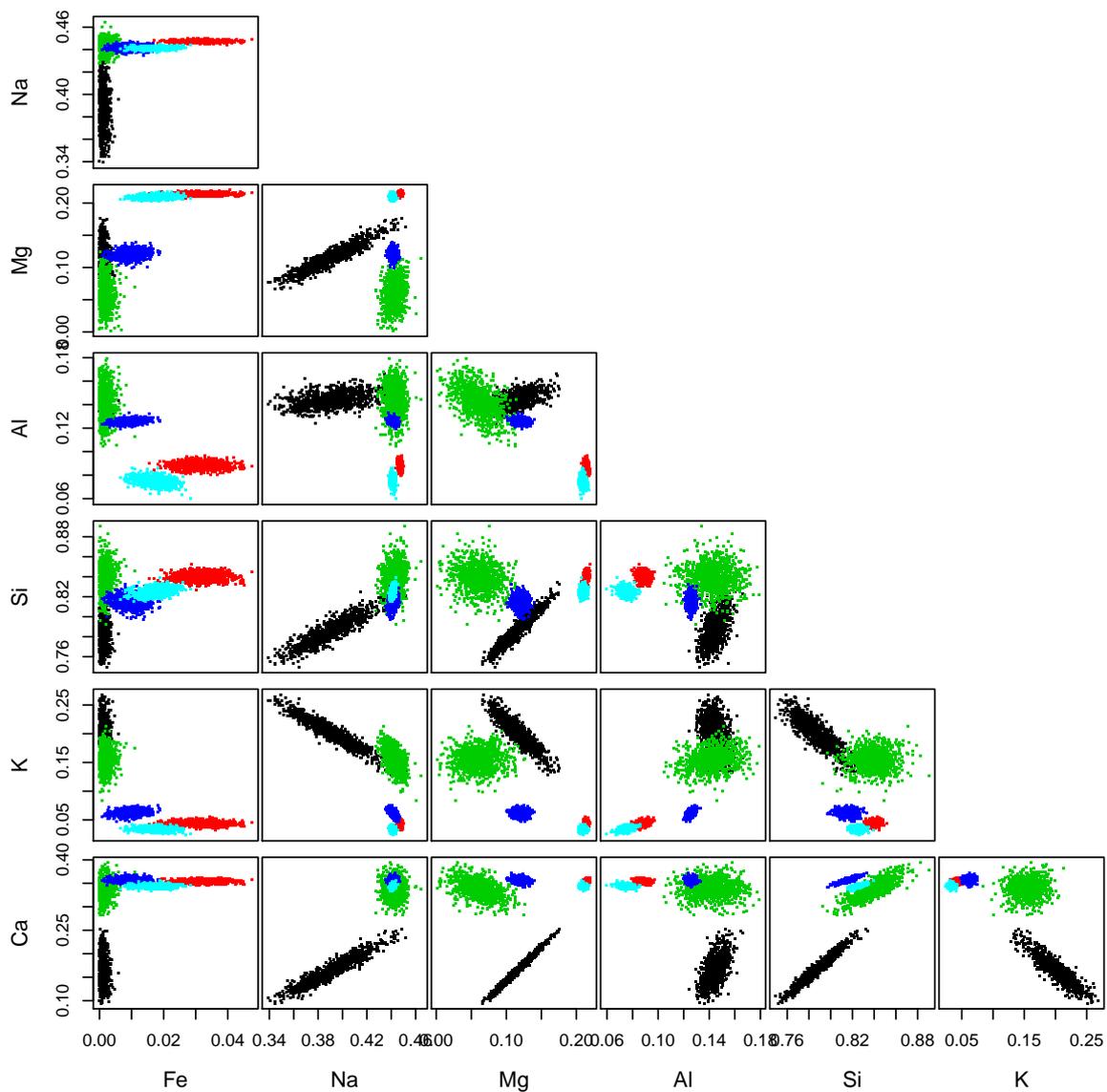


Figure 3.2: Scatterplots of draws from θ_t from the Bayesian hierarchical model with all compositional zeros unaltered. The different coloured points correspond to the five use type categories: bulb, car window, headlamp, container and building window.

The standard deviations from the variance-covariance matrices corresponding to each of the random effects are shown in Table 3.2. As mentioned earlier the variability at item level was allowed to change by use type with the introduction of different Ω_t^{-1} for each use type. This was mainly due to the increased variability seen between bulbs, i.e. $t = 1$, which is shown in the first row of Table 3.2. Comparing the fragment level variability, Ψ^{-1} , with the variability at measurement level, Λ^{-1} , little difference is observed between them. The variability at fragment level is slightly greater than that at measurement level, with the variability between items much greater than that found within items, as would be expected.

Table 3.2: Standard deviations (multiplied by 10) from covariance matrices Ω_t^{-1} , Ψ^{-1} and Λ^{-1} from the Bayesian hierarchical model with unaltered compositional zeros. For Ω_t^{-1} , $t = 1, \dots, 5$ correspond to use types: bulb, car window, headlamp, container and building window.

	Na	Mg	Al	Si	K	Ca	Fe
Ω_1^{-1}	0.96	0.95	0.28	0.65	1.18	1.35	0.07
Ω_2^{-1}	0.08	0.18	0.27	0.31	0.34	0.26	0.45
Ω_3^{-1}	0.20	0.87	0.49	0.48	0.68	0.73	0.09
Ω_4^{-1}	0.15	0.52	0.16	0.41	0.39	0.34	0.24
Ω_5^{-1}	0.11	0.25	0.32	0.29	0.32	0.25	0.34
Ψ^{-1}	0.06	0.04	0.08	0.30	0.08	0.24	0.03
Λ^{-1}	0.04	0.04	0.04	0.15	0.05	0.12	0.03

The next section goes on to detail the two different approaches incorporated into the Bayesian hierarchical model to handle the compositional zeros present in the glass database, as well as display the posterior samples ob-

tained from each approach.

3.2 Modelling compositional zeros

The compositional zeros present in the glass database can be considered as either rounded zeros or essential zeros. As zeros in compositional data are most often treated as rounded zeros, i.e. traces of a component below levels of detection, the first approach to modelling the zeros detailed below treats them as such. The second approach considers the zeros as being essential zeros, i.e. as being truly absent from an item's composition, and splits the data up according to the presence or absence of the elements iron and potassium, as described in Section 2.3.

3.2.1 Data augmentation approach

The easiest approach to handling zeros considered as rounded zeros is to add a constant value which is below the limits of detection to all zero values (Neocleous *et al.*, 2011). However, as this introduces some artificial correlation as noted by Palarea-Albaladejo *et al.* (2007), the approach taken here is to augment all of the zero values during the MCMC procedure so that the same value is not added to all the zeros. This is done by updating each of the zeros using a univariate truncated normal distribution, where the mean and variance of the distribution come from the current values of ξ , \mathbf{b} and \mathbf{c} generated from the MCMC. That is to say that for each iteration of the MCMC sampler all zero values present in the glass database are updated with non-

zero values below the limits of detection. From Bayesian hierarchical model (3.1) it was assumed that

$$\mathbf{z}_{tijk} = \boldsymbol{\theta}_t + \mathbf{b}_{ti} + \mathbf{c}_{tij} + \boldsymbol{\epsilon}_{tijk},$$

$$\mathbf{b}_{ti} \stackrel{\text{iid}}{\sim} N_p(\mathbf{0}, \Omega_t^{-1}), \quad \mathbf{c}_{tij} \stackrel{\text{iid}}{\sim} N_p(\mathbf{0}, \Psi^{-1}), \quad \boldsymbol{\epsilon}_{tijk} \stackrel{\text{iid}}{\sim} N_p(\mathbf{0}, \Lambda^{-1}).$$

With the subscripts t , i , j and k suppressed for ease of notation, let the l -th element of \mathbf{z} be denoted by \mathbf{z}_l . The square root ratio transformation of $\mathbf{z}_l = \sqrt{\frac{\mathbf{w}_l}{\mathbf{w}_D}}$, where the common divisor \mathbf{w}_D is the element oxygen. Then, $\mathbf{w}_l = 0$ would suggest that the weight percentage for the l -th element is below the detection limit of the measuring equipment. As SEM-EDX analysis returns the weight percentages to two decimal places it will return zero values for elements where the weight percentage is below 0.005, that is, when $\mathbf{w}_l < 0.005$. This implies that the non-zero values generated from the normal distribution should be truncated over the range $R = \left(0, \sqrt{\frac{0.005}{w_D}}\right)$. Then, when $\mathbf{w}_l = 0$, the l -th element of \mathbf{z}_l is assumed to come from the univariate normal distribution

$$\mathbf{z}_l \sim N(\boldsymbol{\mu}_l, \Lambda_l^{-1})I(\mathbf{z}_l \in R).$$

Conditioning on the remaining elements of \mathbf{z} , denoted by \mathbf{z}_{-l} , with the subscript $-l$ denoting all elements minus the l -th element, the univariate normal distribution of the l -th element, given that it is a rounded zero is

$$\mathbf{z}_l \mid \mathbf{z}_{-l} \sim N(\boldsymbol{\mu}_l + \mathbf{C}_l^T \Lambda_{-l}(\mathbf{z}_{-l} - \boldsymbol{\mu}_{-l}), \Lambda_l^{-1} - \mathbf{C}_l^T \Lambda_{-l} \mathbf{C}_l)I(\mathbf{z}_l \in R), \quad (3.13)$$

where $\boldsymbol{\mu}_l = \boldsymbol{\theta}_l + \mathbf{b}_l + \mathbf{c}_l$; $\Lambda_l^{-1} = \text{Var}(\mathbf{z}_l)$ and $\mathbf{C}_l = \text{Cov}(\mathbf{z}_l, \mathbf{z}_{-l})$. This approach treats each rounded zero \mathbf{z}_l as an additional parameter which is partially observed given it is assumed below limits of detection, and samples from its

full conditional distribution restricted to the set R . This allows for different non-zero values to be generated for all zeros in the glass database, thus reducing the artificial correlation induced by only adding the same constant value to all zeros.

3.2.2 Posterior samples from data augmentation

The posterior samples shown here were obtained from the Bayesian hierarchical model using the data augmentation approach above and the MCMC procedure detailed in Section 3.1.1 for the square root ratios. Due to augmenting the zeros present in the database at each iteration of the MCMC sampler, the time taken to obtain the simulated model results was longer than running the model without taking the presence of zeros into consideration, and was approximately 20 hours. As with the earlier posterior samples from the Bayesian hierarchical model with unaltered zeros, a burn-in period of 10,000 was used, and thinning of the Markov chain with every 200th draw kept and the rest discarded. The acceptance rate for Metropolis-Hastings move M-H 2 was 43%, and the acceptance rate for Metropolis-Hastings move M-H 3 was 39%. Time series plots of the draws obtained for the sampled fixed effect θ_t are shown in Figure 3.3 and appear to show improved convergence, most so for bulbs (θ_1), over the posterior samples obtained from the Bayesian hierarchical model leaving the compositional zeros alone. This is also reflected in the improved effective sample sizes for bulbs and headlamps seen in Table 3.3. It is also seen in Figure 3.3 from the time series plots of iron for bulbs (θ_1) and headlamps (θ_3) with the points no longer clustered close to the zero boundary. This is also displayed in Figure 3.4 which shows

scatterplots of the draws of θ_t from using data augmentation. From the first column of plots in Figure 3.4 for iron the point mass is no longer clustered close to zero for the use types bulb and headlamp.

Table 3.3: Effective sample size from using data augmentation for the square root ratios for the mean vector θ_t . For θ_t , $t = 1, \dots, 5$ correspond to use types: bulb, car window, headlamp, container and building window.

	Na	Mg	Al	Si	K	Ca	Fe
θ_1	77.0	68.8	258.1	99.9	72.7	89.0	1000.0
θ_2	1000.0	1000.0	1000.0	1000.0	859.2	1000.0	1000.0
θ_3	722.56	886.9	723.7	735.3	300.2	564.6	1000.0
θ_4	857.5	1000.0	905.8	1000.0	836.2	1000.0	906.5
θ_5	1174.0	1000.0	836.1	1000.0	879.0	887.4	1000.0

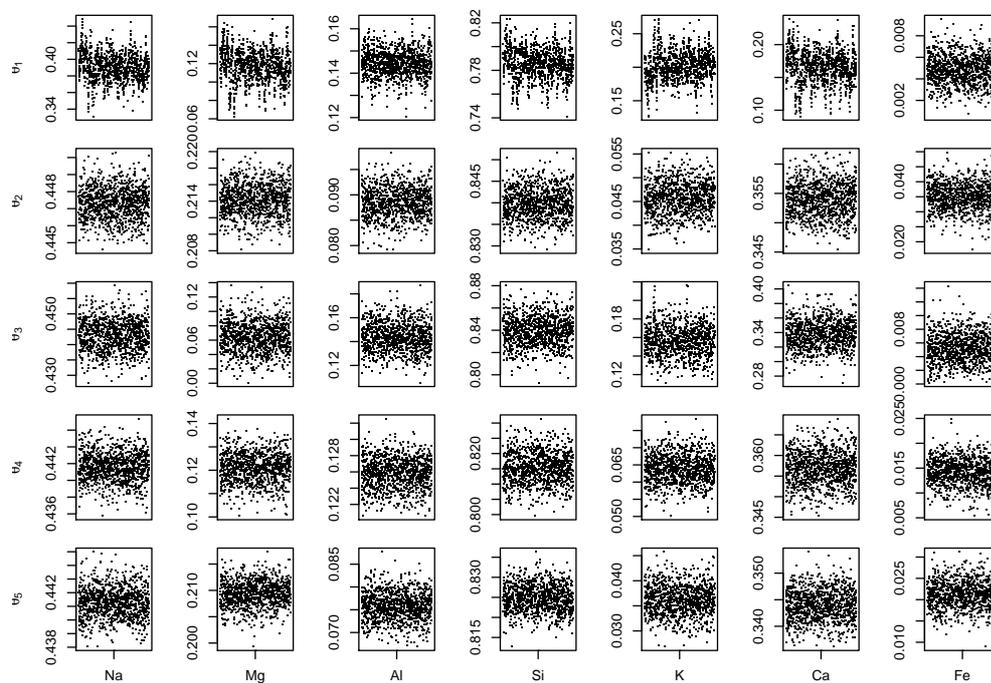


Figure 3.3: Trace plots of the mean θ_t using data augmentation for the square root ratios. A burn-in period of 10,000 was used, and thinning of the Markov chain with every 200th draw stored.

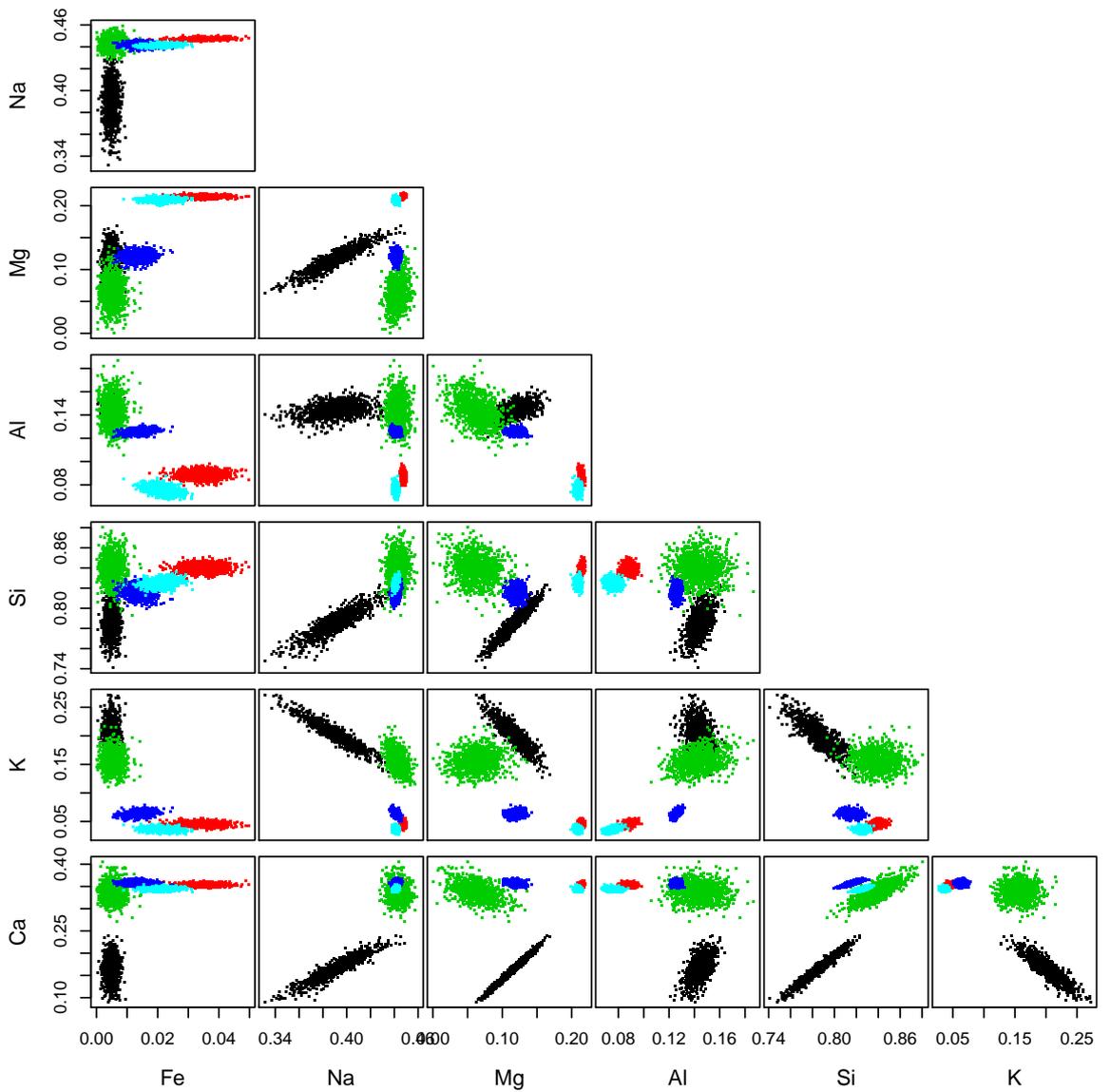


Figure 3.4: Scatterplots of draws from θ_t using data augmentation for the square root ratios. The different coloured points correspond to the five use type categories: bulb, car window, headlamp, container and building window.

Table 3.4 displays the standard deviations from the variance-covariance matrices corresponding to each of the random effects obtained from using data augmentation. The standard deviations obtained are very similar, with some slightly smaller, than those obtained from the Bayesian hierarchical model with the compositional zeros left alone; see Table 3.2.

Table 3.4: Standard deviations (multiplied by 10) from covariance matrices Ω_t^{-1} , Ψ^{-1} and Λ^{-1} from using data augmentation for the square root ratios. For Ω_t^{-1} , $t = 1, \dots, 5$ correspond to use types: bulb, car window, headlamp, container and building window.

	Na	Mg	Al	Si	K	Ca	Fe
Ω_1^{-1}	0.95	0.92	0.28	0.64	1.17	1.32	0.07
Ω_2^{-1}	0.08	0.18	0.27	0.31	0.32	0.26	0.43
Ω_3^{-1}	0.20	0.85	0.47	0.48	0.67	0.74	0.09
Ω_4^{-1}	0.15	0.51	0.16	0.41	0.37	0.34	0.22
Ω_5^{-1}	0.11	0.25	0.31	0.29	0.29	0.25	0.32
Ψ^{-1}	0.06	0.04	0.08	0.30	0.08	0.24	0.03
Λ^{-1}	0.04	0.04	0.04	0.15	0.05	0.12	0.04

The next approach now treats zeros as being essential, meaning that an element is absent from an item's composition if it is zero. More specifically it looks at whether the elements iron and potassium are present or absent from an item's composition.

3.2.3 Composite model approach

This approach makes use of the splitting up of the data according to the presence or absence of the elements iron and potassium that was described as part of the exploratory analysis of the data in Section 2.3. Based on the exploratory analysis in Section 2.3 this approach should satisfy the assumptions of normality made by model (3.1) more so than the previous data augmentation approach. As the data augmentation approach of Section 3.2.1 updates any zero values with non-zero values in the range $R = \left(0, \sqrt{\frac{0.005}{w_D}}\right)$, it does not necessarily solve the issue of having a large mass of zero values influencing the distribution of the data. This is due to the values being added to all of the zero values still being extremely small, with the zero values not being perturbed by much due to the detection limit restriction. In this alternative approach the Bayesian hierarchical model (3.1) is specified separately for all items of each of the four configurations $\mathcal{C}_{\mathbf{z}} = m$ in Table 2.3, conditional on their known use types $\mathcal{T}_{\mathbf{z}}$. This section demonstrates how these four separate models relating to each of the four configurations can be collectively pulled together to form a single model.

First denote the complete reference database as $D = \{\mathbf{z}_{ti}, i = 1, \dots, I_t, t = 1, \dots, T\}$, where the number of items of each use type t , denoted by I_t are under the control of the experimenter collecting the data. Then, if D represents the entire database, let $D_m = \{\mathbf{z} \in D : \mathcal{C}_{\mathbf{z}} = m\}$ be the subset of D that contains all glass items with elemental configuration m . Therefore, the distribution for a given glass item of $\mathbf{z} \in D$ is given by the hierarchical

model (3.1) as

$$p(\mathbf{z}|\mathcal{T}_{\mathbf{z}} = t, \mathcal{C}_{\mathbf{z}} = m, \xi) = p(\mathbf{z}|\mathcal{T}_{\mathbf{z}} = t, \mathcal{C}_{\mathbf{z}} = m, \xi_m), \quad (3.14)$$

where ξ_m denotes the collection of parameters associated with items with elemental configuration m , and $\xi = \{\xi_m\}_{m=1}^M$ is the collection of parameters across all elemental configurations, i.e. over all $M = 4$ configurations. Hence, if the item \mathbf{z} has configuration $\mathcal{C}_{\mathbf{z}} = m$ then the only parameters of interest are those associated with the m -th configuration, i.e. ξ_m . To be more specific, the distribution in (3.14) is given by formulae (3.2) and (3.3), with ξ replaced by ξ_m . The elemental configurations are treated as being independent, but it may also be possible to link them in such a way that allows for the different elemental configurations with some similarities between them to borrow strength from one another. Future work involving an additional level to the model hierarchy may be able to achieve this by including probabilistic models for the configurations.

The probabilities associated with an item of use type t having the m -th elemental configuration are then obtained. Let $\varphi_t = (\varphi_{t1}, \dots, \varphi_{tM})$ be an unknown vector containing the configuration probabilities of an item \mathbf{z} of use type t . Then the probability, φ_{tm} , of the item \mathbf{z} of use type t being of elemental configuration m is given as

$$\varphi_{tm} = p(\mathcal{C}_{\mathbf{z}} = m|\mathcal{T}_{\mathbf{z}} = t, \varphi, \xi) = p(\mathcal{C}_{\mathbf{z}} = m|\mathcal{T}_{\mathbf{z}} = t, \varphi_t). \quad (3.15)$$

The elemental configuration probabilities $\varphi = \{\varphi_t\}_{t=1}^T$ are assumed to be a priori independent of the model parameters ξ from hierarchical model (3.1), with independent Dirichlet prior distributions:

$$\varphi_t|\xi \sim Dir(\alpha_{t1}, \dots, \alpha_{tM}), \quad t = 1, \dots, T, \quad (3.16)$$

where the α parameters are constants which indicate any prior knowledge about the elemental configurations for each use type.

The distribution of a single glass item $\mathbf{z} \in D$, given its use type $\mathcal{T}_{\mathbf{z}}$ and the parameters ξ and φ , is

$$\begin{aligned} p(\mathbf{z}|\mathcal{T}_{\mathbf{z}} = t, \xi, \varphi) &= \sum_{r=1}^M p(\mathcal{C}_{\mathbf{z}} = r|\mathcal{T}_{\mathbf{z}} = t, \xi, \varphi) p(\mathbf{z}|\mathcal{T}_{\mathbf{z}} = t, \mathcal{C}_{\mathbf{z}} = r, \xi, \varphi) \\ &= \sum_{r=1}^M p(\mathcal{C}_{\mathbf{z}} = r|\mathcal{T}_{\mathbf{z}} = t, \varphi_t) p(\mathbf{z}|\mathcal{T}_{\mathbf{z}} = t, \mathcal{C}_{\mathbf{z}} = r, \xi_r) \quad (3.17) \\ &= \sum_{r=1}^M \varphi_{tr} p(\mathbf{z}|\mathcal{T}_{\mathbf{z}} = t, \mathcal{C}_{\mathbf{z}} = r, \xi_r). \end{aligned}$$

The distribution of the reference data D , given ξ and φ , and the use types of all glass items, is

$$p(D|\xi, \varphi) = \prod_{t=1}^T \prod_{i=1}^{I_t} \left\{ \sum_{r=1}^M \varphi_{tr} p(\mathbf{z}_{ti}|\mathcal{T}_{\mathbf{z}_{ti}} = t, \mathcal{C}_{\mathbf{z}_{ti}} = r, \xi_r) \right\}. \quad (3.18)$$

Since the observed data D is fixed the elemental configurations $\mathcal{C}_{\mathbf{z}_{ti}}$ for all items are known, which implies that the \sum_r in (3.18) contains only a term corresponding to the observed configuration of each glass item \mathbf{z}_{ti} . This is a special case of a mixture model, but where upon inspection of the elemental configuration of an item with transformed elemental composition \mathbf{z} , the mixture component to which \mathbf{z} is associated with is immediately known. This was also recognised by [Stewart and Field \(2011\)](#), see their formula (3.1). The distribution of the reference database D in (3.18) can then be written as

$$\begin{aligned} p(D|\xi, \varphi) &= \prod_{t=1}^T \prod_{i=1}^{I_t} \varphi_{tm} p(\mathbf{z}_{ti}|\mathcal{T}_{\mathbf{z}_{ti}} = t, \mathcal{C}_{\mathbf{z}_{ti}} = m, \xi_m) \\ &= \left\{ \prod_{t=1}^T \prod_{m=1}^M \varphi_{tm}^{N_{tm}} \right\} \cdot \left\{ \prod_{m=1}^M \prod_{t=1}^T \prod_{i \in E_{tm}} p(\mathbf{z}_{ti}|\mathcal{T}_{\mathbf{z}_{ti}} = t, \mathcal{C}_{\mathbf{z}_{ti}} = m, \xi_m) \right\}, \quad (3.19) \end{aligned}$$

where $E_{tm} = \{i : T_{\mathbf{z}_{ii}} = t, \mathcal{C}_{\mathbf{z}_{ii}} = m\}$ is the collection of all items in D which are of use type t that have elemental configuration m . The counts from Table 2.3 that refer to the number of items in the database of use type t and configuration m are then denoted by $N_{tm} = \#E_{tm}$.

As ξ and φ are assumed a priori independent, the above factorisation of $p(D|\xi, \varphi)$ in (3.19) also implies that ξ and φ are also a posteriori independent. This leads to the configuration probabilities φ_t also having independent Dirichlet posterior distributions:

$$\varphi_t|\xi, D \sim Dir(\alpha_{t1} + N_{t1}, \dots, \alpha_{tM} + N_{tM}), \quad t = 1, \dots, T. \quad (3.20)$$

This is obtained by combining the first term on the right-hand side of (3.19) with the prior distribution on the φ_t 's in (3.16).

Since ξ and φ are a posteriori independent a sample from their joint distribution can be obtained from two independent steps: (i) samples of φ can be obtained from the independent Dirichlet posterior distributions in (3.20) and (ii) samples of the model parameters ξ_m , for each elemental configuration m , can be obtained from the hierarchical model (3.1) using the MCMC procedure detailed in Section 3.1.1.

The next section reports posterior draws from hierarchical model (3.1) for all items with elemental configuration $m = 2$, i.e. iron absent and potassium present, as that configuration has the most items associated with it in the database, including items from each of the five use types. The posterior draws obtained from items with elemental configurations $m = 1, 3, 4$ can be found in Appendix C.

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

The posterior samples shown are those from items with configuration $m = 2$ ($\overline{\text{Fe}}$, K) from Table 2.3 for the square root ratios. A burn-in period of 10,000 was used, and thinning of the Markov chain where every 200th draw from the chain was kept and the rest discarded. As the posterior samples displayed here were obtained from a subset of the glass database, i.e. all glass items with iron absent and potassium present, they should not be directly compared with the posterior samples shown in sections 3.1.2 and 3.2.2, which were obtained using the entire database. However, due to the composite model approach reducing the data dimension in some configurations and also not having to update the compositional zeros with non-zero values below limits of detection, less time is needed to obtain the results, with approximately 10 hours required across all four elemental configurations. Since each of the four subsets of the data are independent, the posterior samples could be obtained even quicker by running all four configurations in parallel, with the configuration associated with the most glass items, configuration $m = 2$, taking approximately 6 hours. The acceptance rate for Metropolis-Hastings move M-H 2 was 31%, and the acceptance rate for Metropolis-Hastings move M-H 3 was 54%. Time series plots of the posterior draws for the fixed effect θ_t associated with the composite model for items with elemental configuration $m = 2$ are shown in Figure 3.5. The time series plots indicate good convergence for use types car window, headlamp, container and building window, and for bulbs less so with the same reason mentioned above in Section 3.1.2 mostly accountable for this. Table 3.5 displays the effective sample sizes for

configuration $m = 2$ and are, for the most part, greater than those obtained from using data augmentation. Figure 3.6 shows scatterplots for the mean θ_t and again shows clear separation between the use type categories. Also, since all glass items with iron absent are not present in elemental configuration $m = 2$ the point mass at zero seen in earlier plots is no longer an issue.

Table 3.5: Effective sample size from the composite model for items with configuration $m = 2$ (Fe, K) for the square root ratios for the mean vector θ_t . For θ_t , $t = 1, \dots, 5$ correspond to use types: bulb, car window, headlamp, container and building window.

	Na	Mg	Al	Si	K	Ca
θ_1	73.0	66.8	293.5	70.9	74.4	99.5
θ_2	1000.0	1000.0	1057.9	1000.0	910.2	1000.0
θ_3	1000.0	1000.0	1000.0	1000.0	1000.0	1000.0
θ_4	1000.0	1000.0	1000.0	1000.0	1000.0	1173.74
θ_5	1000.0	1146.5	1064.4	1000.0	1000.0	1000.0

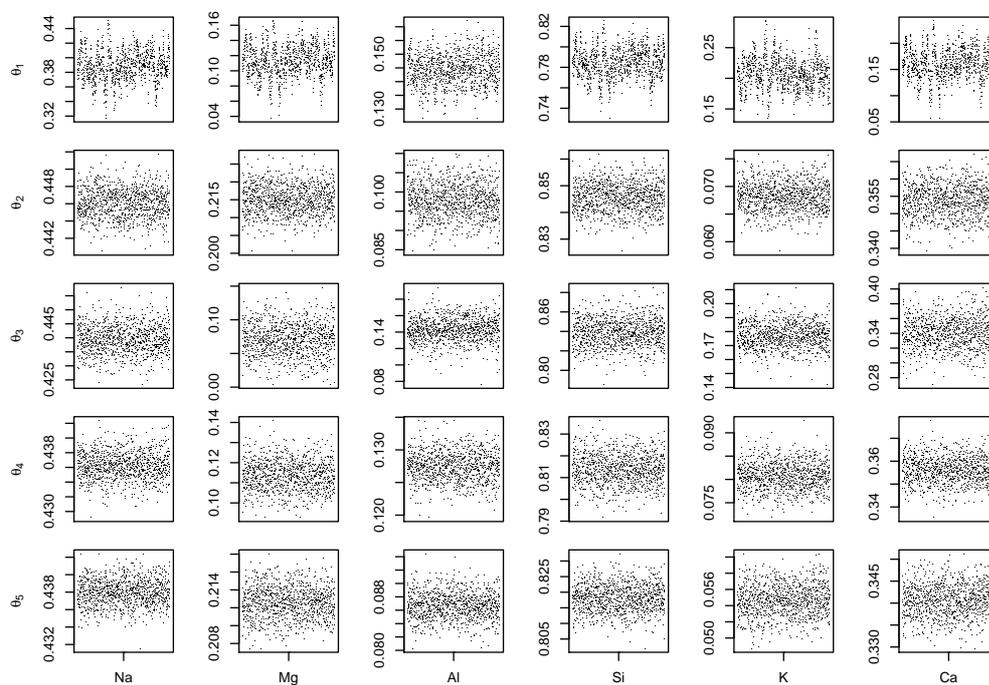


Figure 3.5: Trace plots of the mean θ_t for configuration $m = 2$ ($\overline{\text{Fe}}$, K) for the square root ratios. A burn-in period of 10,000 was used, and thinning of the Markov chain with every 200^{th} draw stored.

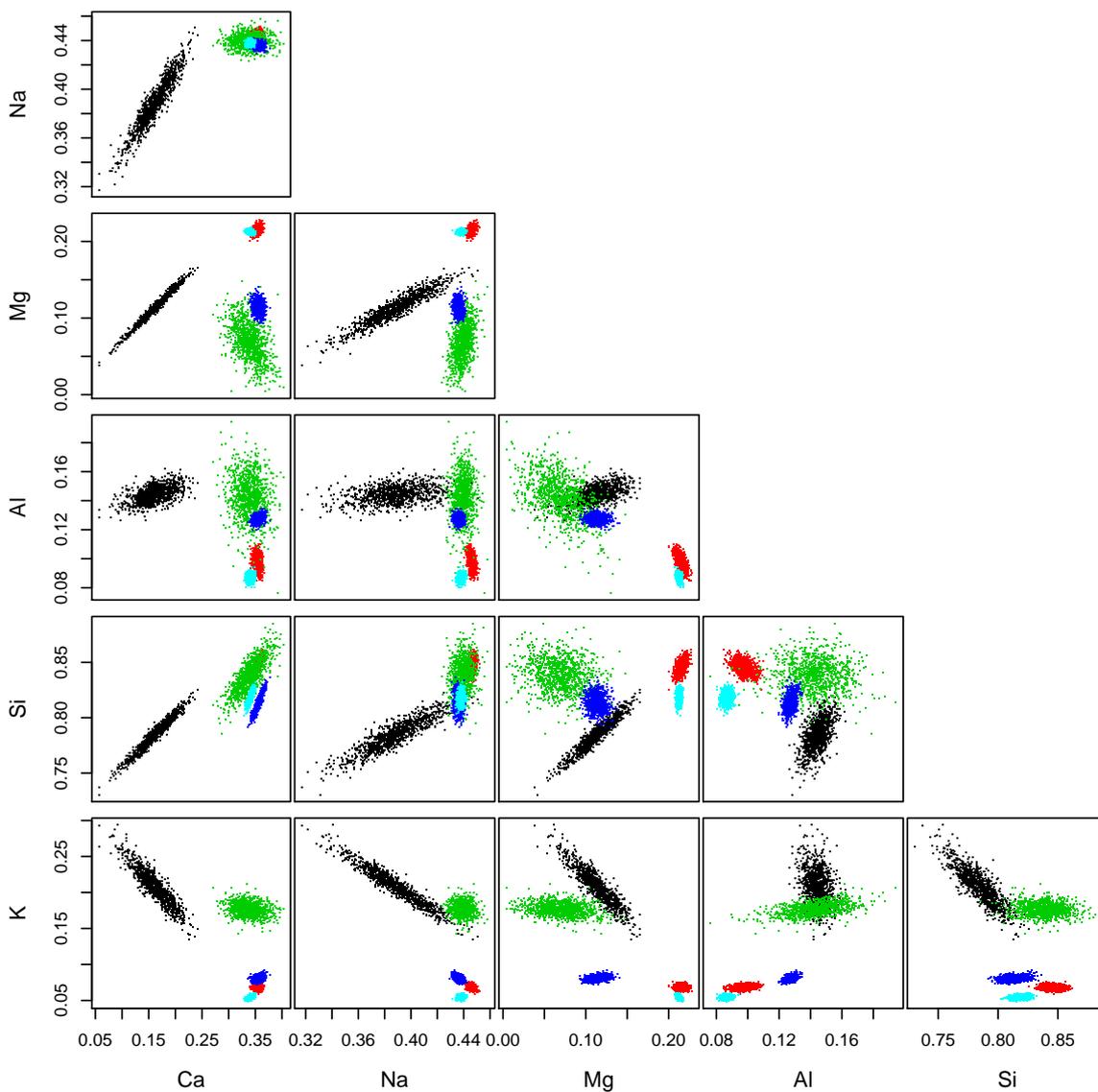


Figure 3.6: Scatterplots of draws from θ_t for items with configuration $m = 2$, i.e. iron absent and potassium present, for the square root ratios. The different coloured points correspond to the five use type categories: bulb, car window, headlamp, container and building window.

The standard deviations from the variance-covariance matrices corresponding to each of the random effects for the composite model for items with elemental configuration $m = 2$ are shown in Table 3.6. Comparing the fragment level variability, Ψ^{-1} , with the variability at measurement level, Λ^{-1} , little difference is observed between them. The variability at fragment level is slightly greater than that at measurement level. The variability between items is much greater than that found within items for the use types bulb and headlamp, but less so for the other three use types. All of the results shown here for the other three elemental configurations can be found in Appendix C.

Table 3.6: Standard deviations (multiplied by 10) from covariance matrices Ω_t^{-1} , Ψ^{-1} and Λ^{-1} obtained from the composite model for items with configuration $m = 2$ ($\overline{\text{Fe}}, \text{K}$) for the square root ratios. For Ω_t^{-1} , $t = 1, \dots, 5$ 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

The next section will look at posterior samples obtained using the log-ratios instead of the square root of the compositional ratios, and will include some necessary changes to how the Bayesian hierarchical model is implemented.

3.3 Log-ratio transformation

A few changes have to be considered before applying the logarithm to the compositional ratios. First, since it is not possible to obtain the logarithm of zero, all compositional zeros present are replaced by constants that are below the detection limits of the measuring equipment. Since SEM-EDX returns the percentage weights of the elements rounded to two decimal places, any traces of an element with a concentration below 0.005 will be returned as zero. The easiest thing to do then is to add the constant 0.005 to all zero values present in the glass database before applying the logarithm to the ratios, where again oxygen is chosen to be the common divisor. Second, since constants have been added to the zeros present, they are taken as being rounded zeros and not essential zeros. This means that the constants can be updated as part of the MCMC process using the data augmentation approach, which should reduce any artificial correlation that may be induced from adding the same constant to all zeros. It is also possible to use the composite model approach but it would mean considering some zeros as being essential and others as being rounded. For example, the four configurations are due to observing the presence and absence of iron and potassium and so they would be considered essential. However, in order to apply the logarithm to the four separate subsets of the data obtained from the four configurations any zeros present from any other elements would have to be replaced by constants. This would mean they would be considered as rounded zeros, and could either be replaced by constants or updated using data augmentation. A combination of both of the compositional zero approaches can be done, but from a philosophical point of view it may not make sense to treat the

compositional zeros differently; however for the sake of comparison both the composite model and data augmentation approaches will be used for the log-ratios.

Here let \mathbf{z}_{tijk} denote the log-ratios, with the subscripts t, i, j and k denoting the k -th measurement from the j -th fragment from the i -th item of use type t . Each \mathbf{z}_{tijk} is then assumed, as in (3.1), to be

$$\mathbf{z}_{tijk} = \boldsymbol{\theta}_t + \mathbf{b}_{ti} + \mathbf{c}_{tij} + \boldsymbol{\epsilon}_{tijk}, \quad (3.21)$$

$$\mathbf{b}_{ti} \stackrel{\text{iid}}{\sim} N_p(\mathbf{0}, \Omega_t^{-1}), \quad \mathbf{c}_{tij} \stackrel{\text{iid}}{\sim} N_p(\mathbf{0}, \Psi^{-1}), \quad \boldsymbol{\epsilon}_{tijk} \stackrel{\text{iid}}{\sim} N_p(\mathbf{0}, \Lambda^{-1}).$$

The distribution of a glass item \mathbf{z} is then given as in (3.2) and (3.3) with prior distributions on the covariance matrices associated with the random effects as in (3.5). However, for the square root transformed ratios the prior placed on the fixed mean effect $\boldsymbol{\theta}_t$ was a multivariate normal distribution restricted to the positive orthant, as shown in (3.4). As the log-ratios are negative this restriction on $\boldsymbol{\theta}_t$ is removed with the prior on the fixed effects $\boldsymbol{\theta}_t$ now assumed to be independent multivariate normal distributions without any restrictions:

$$\boldsymbol{\theta}_t \stackrel{\text{iid}}{\sim} N_p(\mathbf{0}, \Phi^{-1}), \quad t = 1, \dots, T. \quad (3.22)$$

This means that Metropolis-Hastings move 1 (M-H 1) used for the square root transformation, which was accepted with probability 1 or 0 depending on whether the drawn candidate was in the positive orthant or not, is simply replaced by a regular Gibbs sampling move, where draws of $\boldsymbol{\theta}_t$ are made from its full conditional distribution. The posterior draws shown in Section 3.3.1 for the log-ratios use Gibbs sampling and Metropolis-Hastings move 3 (M-H 3) detailed earlier, which was the joint move on both $\boldsymbol{\theta}_t$ and \mathbf{b}_{ti} . The only change to this move for the log-ratios is in the δ_{ll} values used for the interval

widths that govern the magnitude of the change from current to candidate values, and were changed to reflect the influence of the logarithmic transformation. Also, for the implementation of the data augmentation approach to updating the compositional zeros, the only change is in the range R in which the log-ratios are truncated over, with $R = \left(-\infty, \log \frac{0.005}{w_D} \right)$. For the composite model any compositional zeros for the chemical elements not being observed as present or absent are assumed present with the constant value of 0.005 added to all corresponding zeros.

3.3.1 Log-ratio data augmentation posterior samples

The posterior samples obtained for the log-ratios used the same burn-in period of 10,000 draws, and also the same thinning with every 200th draw stored as the previous posterior samples for the square root transformation. The acceptance rate for Metropolis-Hastings move M-H 3 was 48%. Figure 3.7 displays the time series plots obtained from using the log-ratios, and clearly shows poor convergence for some chemical elements across all use types. The most obvious issues can be seen in the time series plots for iron across each use type. This is due to, as was seen in Figure 2.2, the strong influence the logarithmic transformation has over the large proportion of zeros present for iron. Even when iron is present the influence of the logarithmic transformation remains strong, and is due to all concentrations of iron, when present, being less than 1%. Very poor convergence can be seen for the elements magnesium and calcium for bulbs (θ_1). This is due to nine of the 26 bulbs in the database containing no magnesium or calcium, and so the stronger influence of the logarithmic transformation over these small concentrations is

reflected by the convergence issues. The same problem exists for magnesium in headlamps (θ_3) due to eight of the 16 headlamps containing no magnesium. The poor convergence issues for bulbs and headlamps, and for some of the elements across all use types are reflected in the very small effective sample sizes seen in Table 3.7. Figure 3.8 displays scatterplots of the draws of θ_t obtained using the log-ratios, while Table 3.8 contains the standard deviations obtained from the random effects covariance matrices.

Table 3.7: Effective sample size from the Bayesian hierarchical model using data augmentation and log-ratios for the mean vector θ_t . For θ_t , $t = 1, \dots, 5$ correspond to use types: bulb, car window, headlamp, container and building window.

	Na	Mg	Al	Si	K	Ca	Fe
θ_1	23.8	4.1	36.5	20.2	65.0	2.6	2.7
θ_2	1015.8	879.2	877.4	910.4	108.0	934.1	87.5
θ_3	132.1	16.3	67.4	75.7	51.4	34.4	5.7
θ_4	405.9	285.1	333.4	549.4	250.7	824.7	3.3
θ_5	1037.8	1000.0	1047.0	1000.0	6.2	1000.0	23.2

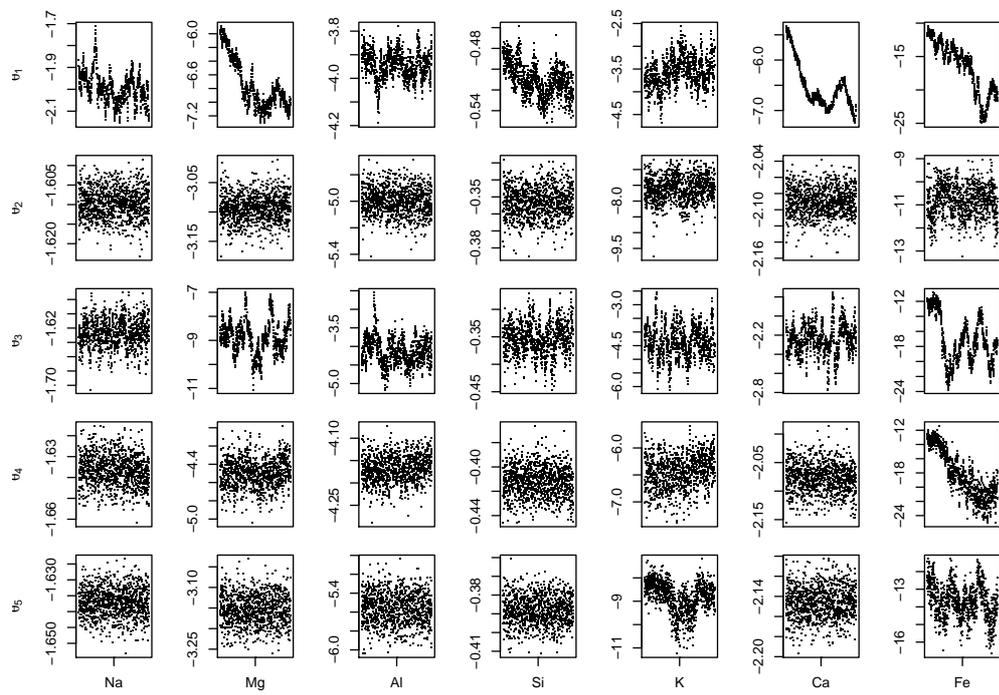


Figure 3.7: Trace plots of the mean θ_t from the Bayesian hierarchical model using data augmentation and the log-ratio transformation. A burn-in period of 10,000 was used, and thinning of the Markov chain with every 200th draw stored.

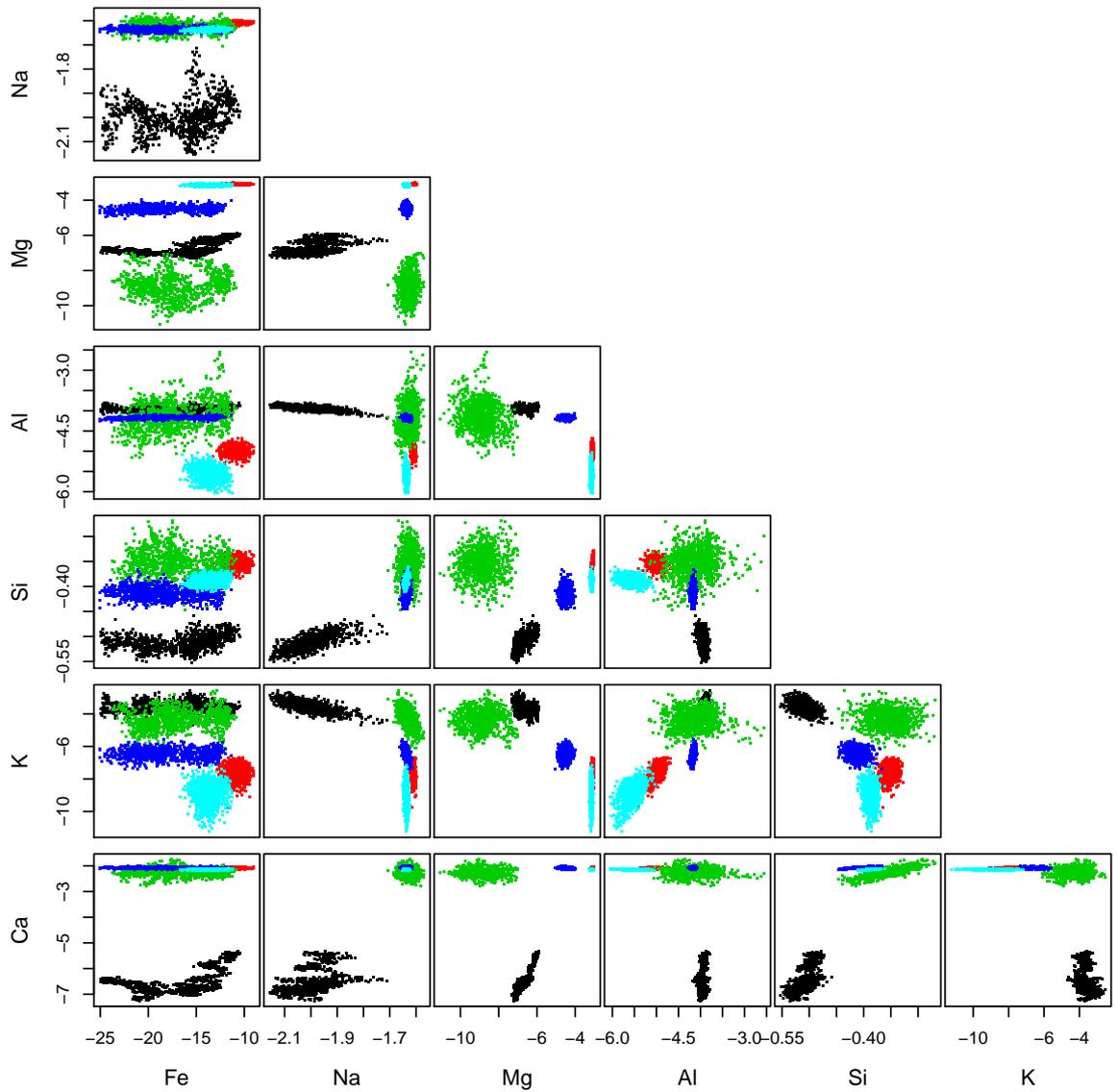


Figure 3.8: Scatterplots of draws from θ_t from the Bayesian hierarchical model using data augmentation and the log-ratio transformation. The different coloured points correspond to the five use type categories: bulb, car window, headlamp, container and building window.

Table 3.8: Standard deviations (multiplied by 10) from covariance matrices Ω_t^{-1} , Ψ^{-1} and Λ^{-1} from the Bayesian hierarchical model using data augmentation and log-ratios. For Ω_t^{-1} , $t = 1, \dots, 5$ correspond to use types: bulb, car window, headlamp, container and building window.

	Na	Mg	Al	Si	K	Ca	Fe
Ω_1^{-1}	4.95	38.56	3.71	1.65	18.54	44.11	26.36
Ω_2^{-1}	0.32	2.10	9.62	0.74	34.16	1.45	51.97
Ω_3^{-1}	0.79	48.54	15.48	1.15	24.00	6.52	19.73
Ω_4^{-1}	0.65	13.45	2.54	1.02	26.73	2.10	116.40
Ω_5^{-1}	0.44	3.64	16.81	0.68	45.60	1.43	58.12
Ψ^{-1}	0.31	0.46	1.26	0.74	2.20	1.40	1.48
Λ^{-1}	0.19	1.18	1.31	0.37	2.06	0.68	3.08

The next section looks at the posterior samples obtained from the composite model using the log-ratio transformation for glass items with elemental configuration $m = 2$.

3.3.2 Log-ratio composite model posterior samples

The posterior samples shown here are for items with configuration $m = 2$ ($\overline{\text{Fe}}$, K) using the log-ratio transformation. The results for the other three configurations can be found in Appendix D. As with previous draws a burn-in period of 10,000 was used, and thinning of the Markov chain where every 200th draw was stored and the rest discarded. The time taken to obtain the draws across all four elemental configurations was approximately 10 hours. The composite model approach used the same sampling scheme as that mentioned earlier for the log-ratio transformation. Figure 3.9 shows the time series plots obtained for the fixed mean effect θ_t for all glass items with configuration $m = 2$ when using the log-ratio transformation. Figure 3.9 clearly displays issues with convergence for the use types bulb (θ_1) and headlamp (θ_3). The posterior draws obtained for the other three configurations do not suffer from convergence as poor as displayed in Figure 3.9, however they are not without fault, as seen in Appendix D. The issues displayed in Figure 3.9 for bulbs and headlamps may well be down to configuration $m = 2$ containing a mixture of both bulbs and headlamps that do and do not contain concentrations of magnesium. This is seen from Table 2.2 containing the original 10 configurations of items in the database. This may account for the poor convergence associated with magnesium for these two use types, with the spread of the distribution of magnesium when using the log-ratio transformation making it difficult for the sampler to converge to a happy medium for bulbs and headlamps with and without magnesium present. The same issue is seen for calcium in bulbs with there being a split between bulbs that do and do not contain calcium in configuration $m = 2$. These changes

in the levels of concentration of magnesium and calcium in bulbs may also be the reason why the remaining elements also have convergence problems but to a lesser extent. These convergence issues are not problematic for the square root of the ratios as seen in Section 3.2.4 as the presence of zero or small concentrations of elements such as magnesium and calcium has a much greater influence on the distribution of the data for the log-ratios, as was shown in Figure 2.2. This would mean that the spread or distribution of the data for items of use types bulb and headlamp that do and do not contain magnesium, for example, is much greater than that for the square root ratios. Table 3.9 contains the effective sample sizes obtained from items with configuration $m = 2$ using the log-ratio transformation. Figure 3.10 displays the scatterplots corresponding to the posterior samples from Figure 3.9, while Table 3.10 contains the standard deviations obtained from the random effects covariance matrices. It should be noted that implementing the data augmentation approach to update the zero concentrations unaffected by the configuration process does not help improve the issues with convergence.

Table 3.9: Effective sample size from the composite model for items with configuration $m = 2$ ($\overline{\text{Fe}}$, K) using log-ratios for the mean vector θ_t . For θ_t , $t = 1, \dots, 5$ correspond to use types: bulb, car window, headlamp, container and building window.

	Na	Mg	Al	Si	K	Ca
θ_1	17.8	5.0	50.4	7.2	33.9	5.2
θ_2	799.5	347.6	500.9	651.3	1105.3	798.2
θ_3	235.2	29.2	53.2	164.8	142.3	29.6
θ_4	1000.0	716.2	869.3	845.2	838.1	800.9
θ_5	1000.0	785.4	815.5	1000.0	844.3	1000.0

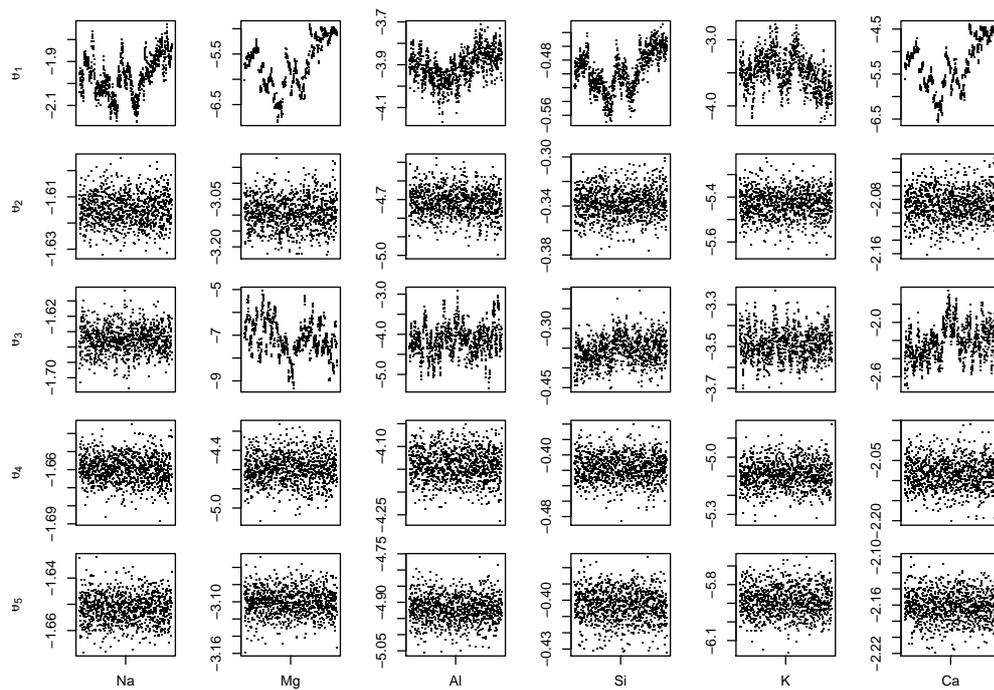


Figure 3.9: Trace plots of the mean θ_t from the composite model using the log-ratio transformation for configuration $m = 2$ ($\overline{\text{Fe}}$, K). A burn-in period of 10,000 was used, and thinning of the Markov chain with every 200th draw stored.

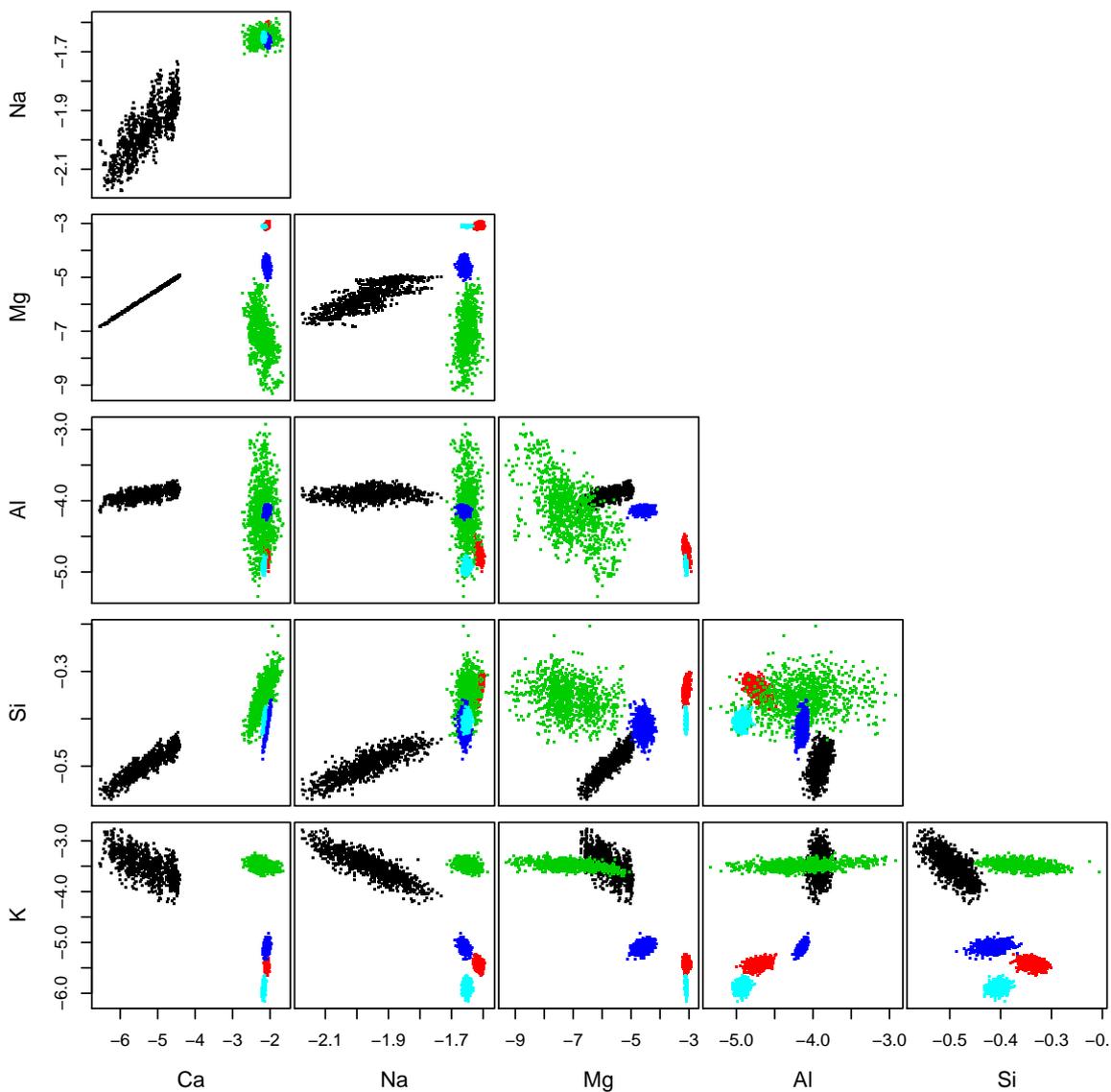


Figure 3.10: Scatterplots of draws from θ_t from the composite model using the log-ratio transformation for items with configuration $m = 2$, i.e. iron absent and potassium present. The different coloured points correspond to the five use type categories: bulb, car window, headlamp, container and building window.

Table 3.10: Standard deviations (multiplied by 10) from covariance matrices Ω_t^{-1} , Ψ^{-1} and Λ^{-1} from the composite model using the log-ratio transformation for configuration $m = 2$ ($\overline{\text{Fe}}$, K). For Ω_t^{-1} , $t = 1, \dots, 5$ correspond to use types: bulb, car window, headlamp, container and building window.

	Na	Mg	Al	Si	K	Ca
Ω_1^{-1}	4.96	29.53	3.75	1.68	14.09	32.99
Ω_2^{-1}	0.31	3.09	4.73	0.64	3.67	1.22
Ω_3^{-1}	0.68	29.64	15.46	1.22	2.45	7.03
Ω_4^{-1}	0.48	11.02	2.18	1.11	4.46	2.38
Ω_5^{-1}	0.36	1.07	3.07	0.56	5.38	1.05
Ψ^{-1}	0.37	0.62	1.37	0.78	2.54	1.54
Λ^{-1}	0.19	0.72	0.87	0.36	2.07	0.66

Various different Metropolis-Hastings moves not detailed in this thesis were also implemented specifically for the log-ratio transformation in order to see if the convergence problems could be alleviated, but little improvement over the posterior draws shown was obtained. This issue, along with the problem of having to manipulate any compositional zeros in order to apply any logarithmic transformation, is why the square root of the compositional zeros was chosen as the appropriate transformation for the glass database. The next section will include model diagnostics in the form of checking for convergence, and also include results obtained from simulated datasets using the acquired posterior draws, for the square root ratio transformation.

3.4 Model diagnostics

To formally check whether the MCMC samples have converged, the Gelman and Rubin diagnostic \hat{R} (Gelman *et al.*, 2004, Chapter 11, pages 296-297) for monitoring MCMC convergence can be obtained. Using their notation, \hat{R} is defined as

$$\hat{R} = \sqrt{\frac{\hat{\text{var}}^+(\psi|y)}{W}}. \quad (3.23)$$

The model parameter of interest, ψ , contains draws simulated from the model for multiple chains with different initial starting values, i.e. ψ_{ij} where $i = 1, \dots, n$ is the number of simulated draws from the model, and $j = 1, \dots, c$ is the number of parallel chains used to test for convergence. From (3.23) the marginal posterior variance, denoted by $\hat{\text{var}}^+(\psi|y)$, can be estimated by

$$\hat{\text{var}}^+(\psi|y) = \frac{n-1}{n}W + \frac{1}{n}B, \quad (3.24)$$

where B denotes the between-chain variability, and W the within-chain variability, with both given as follows:

$$B = \frac{n}{c-1} \sum_{j=1}^c (\bar{\psi}_{\cdot j} - \bar{\psi}_{\cdot\cdot})^2, \quad W = \frac{1}{c} \sum_{j=1}^c s_j^2, \quad (3.25)$$

where $\bar{\psi}_{\cdot j} = \frac{1}{n} \sum_{i=1}^n \psi_{ij}$, $\bar{\psi}_{\cdot\cdot} = \frac{1}{c} \sum_{j=1}^c \bar{\psi}_{\cdot j}$ and $s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (\psi_{ij} - \bar{\psi}_{\cdot j})^2$. Formula (3.23) provides an estimate of the scale reduction in the distribution of ψ for when the number of simulations $n \rightarrow \infty$. A value of \hat{R} close to 1 suggests that the chains have converged to their target distributions, with values below 1.1 considered acceptable. A value greater than 1.1 suggests that the number of simulation draws n may need to be increased to reach convergence, with perhaps a larger burn-in period also required.

3.4.1 Results of model diagnostics

For both the data augmentation and composite model approaches, draws were simulated from the Bayesian hierarchical model for $c = 3$ parallel chains, with each chain beginning from different initial starting values for the model parameters. The number of simulated draws for each was $n = 1000$ with each chain having a burn-in period of 10,000 and thinning where every 200th draw was kept and the rest discarded. For both model approaches all \hat{R} values obtained for each of the parameters were close to 1, with all of the values less than 1.1, indicating convergence of the MCMC chains. To further inspect the convergence of the models and the reliability of the posterior draws obtained, data is simulated from the Bayesian hierarchical model using both the data augmentation and composite model approaches.

3.4.2 Simulating data from the models

To simulate data from the Bayesian hierarchical model, posterior means of each of the parameters were first computed from MCMC draws from the model. Datasets were then generated from the model with the posterior means from the parameters assumed to be the ‘true’ parameter values from the model. Posterior samples are then obtained from the model using the newly simulated datasets in order to check how well they can recover these ‘true’ parameter values.

Simulated data using data augmentation

Twenty datasets were simulated, each containing the same total number of glass items as the glass database. As the number of items of each use type in the glass database do not reflect the prevalence of such use types in a real world setting, the number of items of each use type in the simulated datasets have been evenly distributed, with 64 glass items simulated for each use type. The datasets were simulated from the Bayesian hierarchical model using formulae (3.2) and (3.3), which were given as:

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

with Σ_t given by

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

The known values of the parameters in $\xi = \{\boldsymbol{\theta}, \Omega, \Psi, \Lambda\}$ that were used to simulate the datasets are from the posterior means of the draws obtained from the Bayesian hierarchical model, with data augmentation, applied to the real glass database. The simulated square root ratios \mathbf{z} are then used to obtain the simulated percentage weights of all of the glass items. Also, as the data augmentation approach was used, there are no compositional zeros found in any of the simulated datasets. The Bayesian hierarchical model was then re-estimated on each of the 20 datasets to assess how well it recovers the ‘true’ values used in the above formulae.

Figure 3.11 displays boxplots of the posterior mean draws of $\boldsymbol{\theta}_t$ from the 20 simulated datasets. The red dots show the true values of $\boldsymbol{\theta}_t$ used to simulate the data, and for the most part these estimates have been recaptured well

by the Bayesian hierarchical model. However, for the chemical element iron, which was absent from 79% of the measurements in the glass database, the model appears to overestimate the true value of θ_t used for that element, and recovers a value that is approximately twice that used to obtain the simulated datasets. A similar issue is seen for magnesium in headlamps, where again the model has overestimated the true value. Ten of the 16 headlamps in the glass database did not contain any magnesium. This may be due to there being less information for these elements, particularly iron, leading to the overestimates by the model. Figures 3.12-3.16 contain boxplots of the posterior mean draws for the between-item random effect covariance matrices Ω_t^{-1} , and as with θ_t , recover the true values well for the elements with relatively high concentrations that are common to each use type. Similarly, in Figure 3.17 for the within-item covariance matrix Ψ^{-1} , and Figure 3.18 for the measurement error covariance matrix Λ^{-1} , issues lie with underestimating the relationship between iron and the other chemical elements.

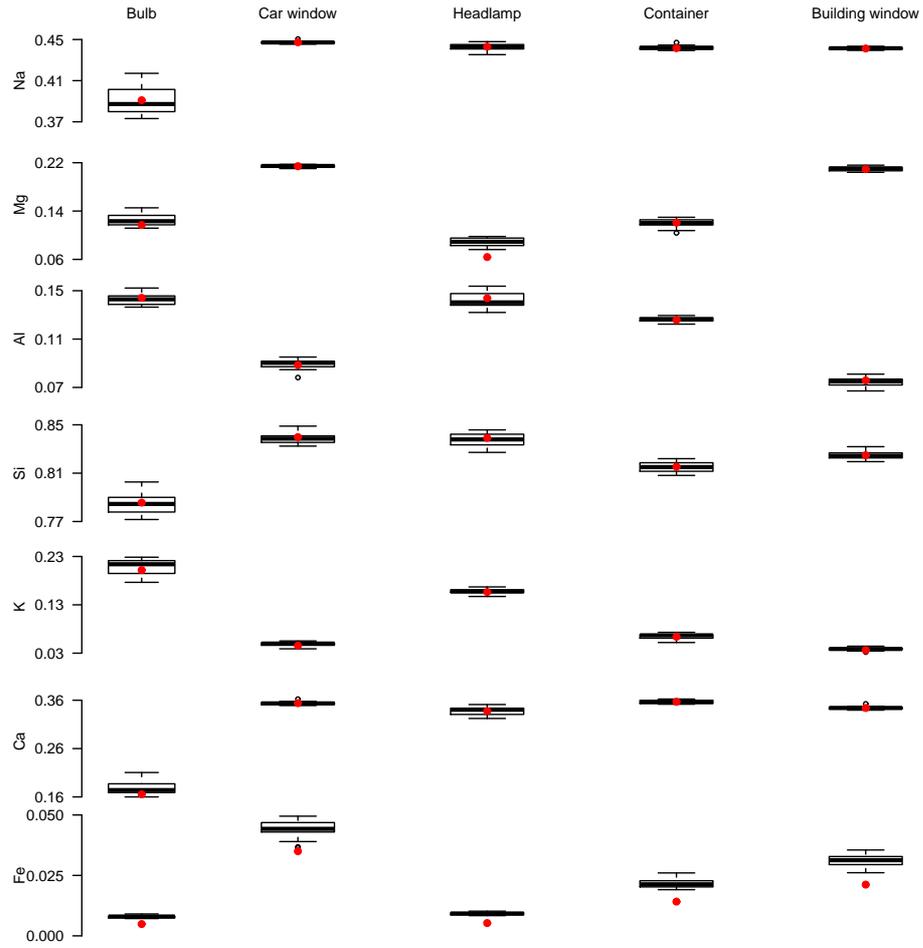


Figure 3.11: Boxplots containing the mean posterior draws of θ_t from 20 simulated datasets from data augmentation for the square root ratios. The red dots indicate the true values of θ_t used to simulate the datasets.

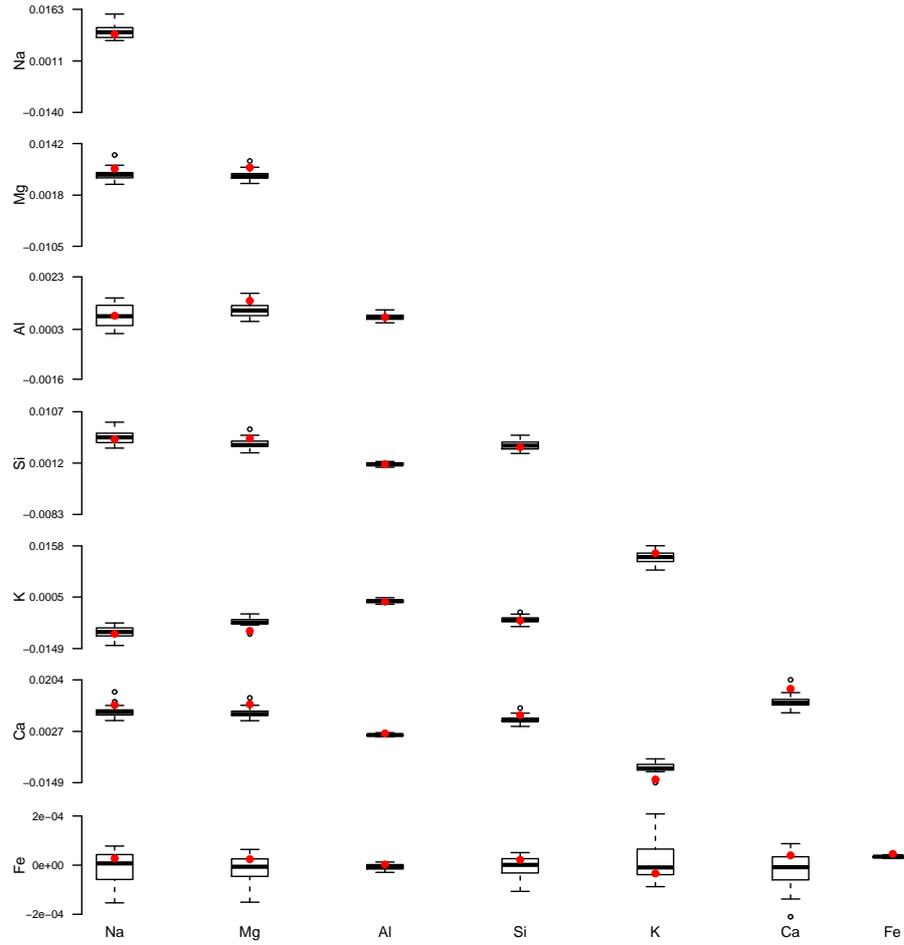


Figure 3.12: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 1$ (bulb) from 20 simulated datasets from data augmentation for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

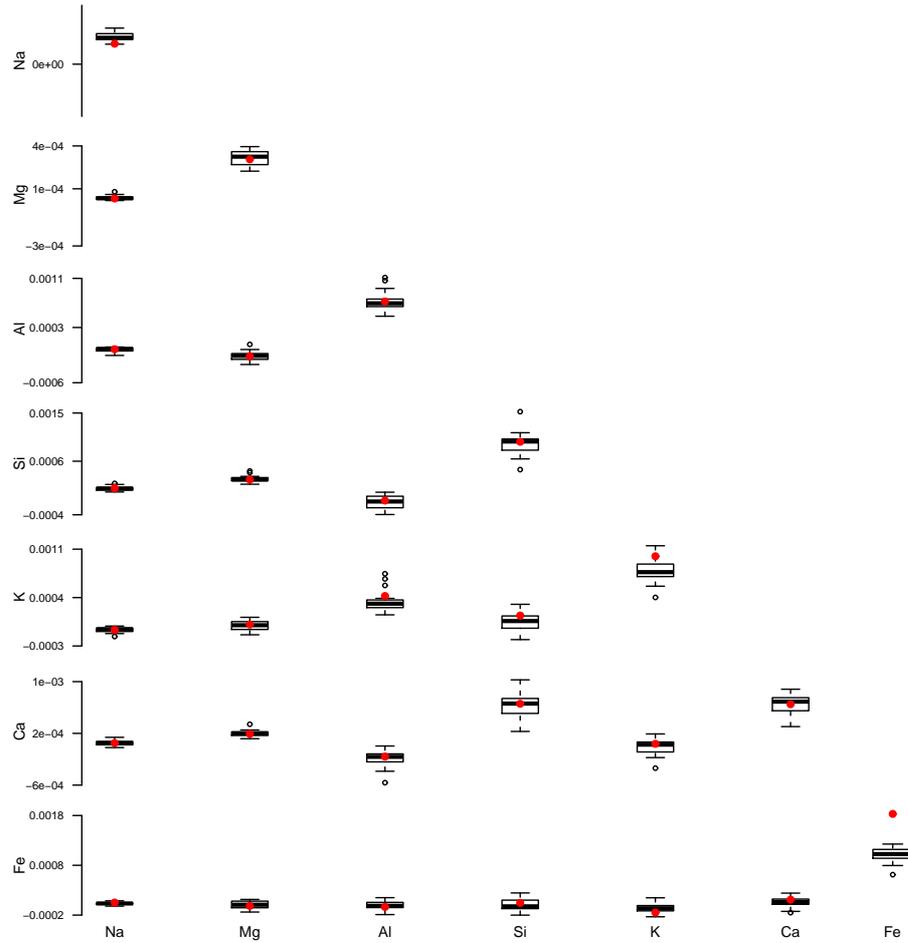


Figure 3.13: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 2$ (car window) from 20 simulated datasets from data augmentation for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

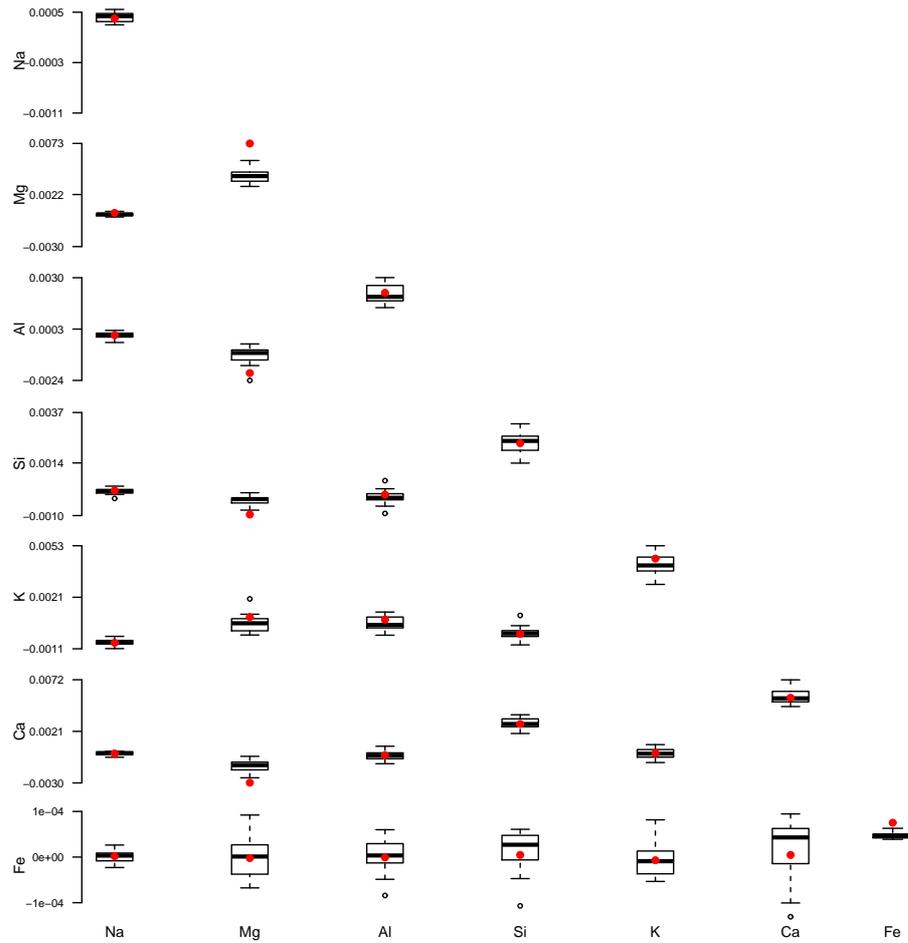


Figure 3.14: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 3$ (headlamp) from 20 simulated datasets from data augmentation for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

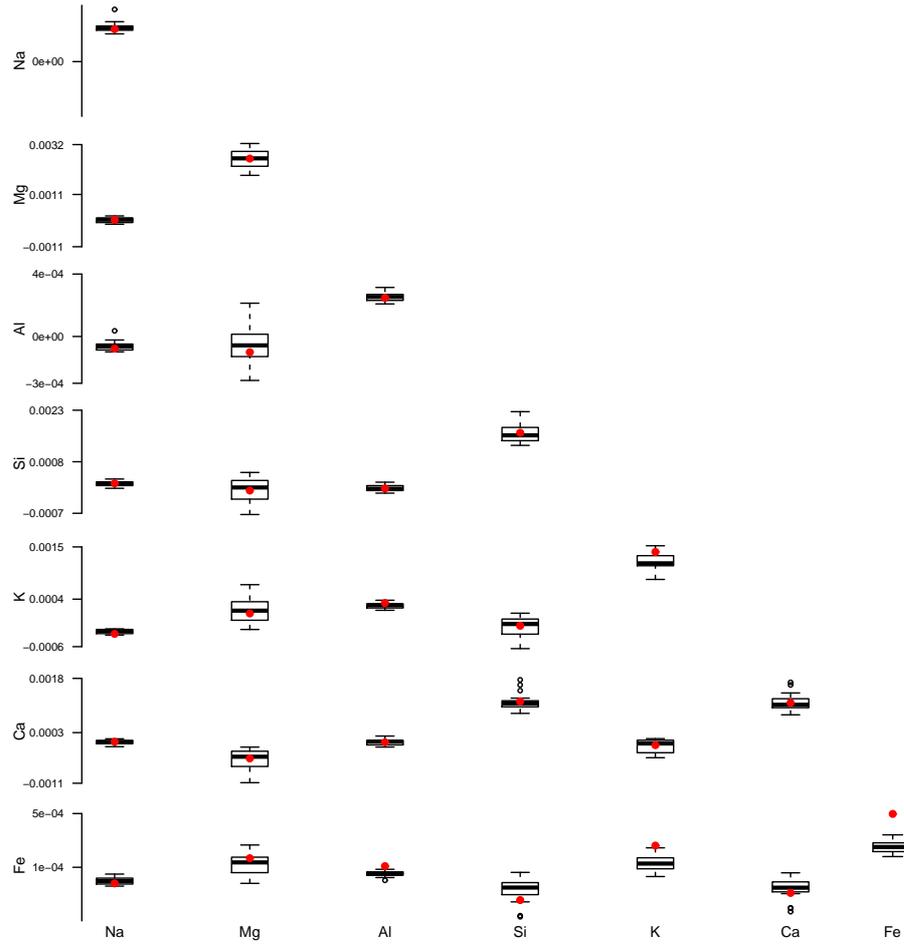


Figure 3.15: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 4$ (container) from 20 simulated datasets from data augmentation for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

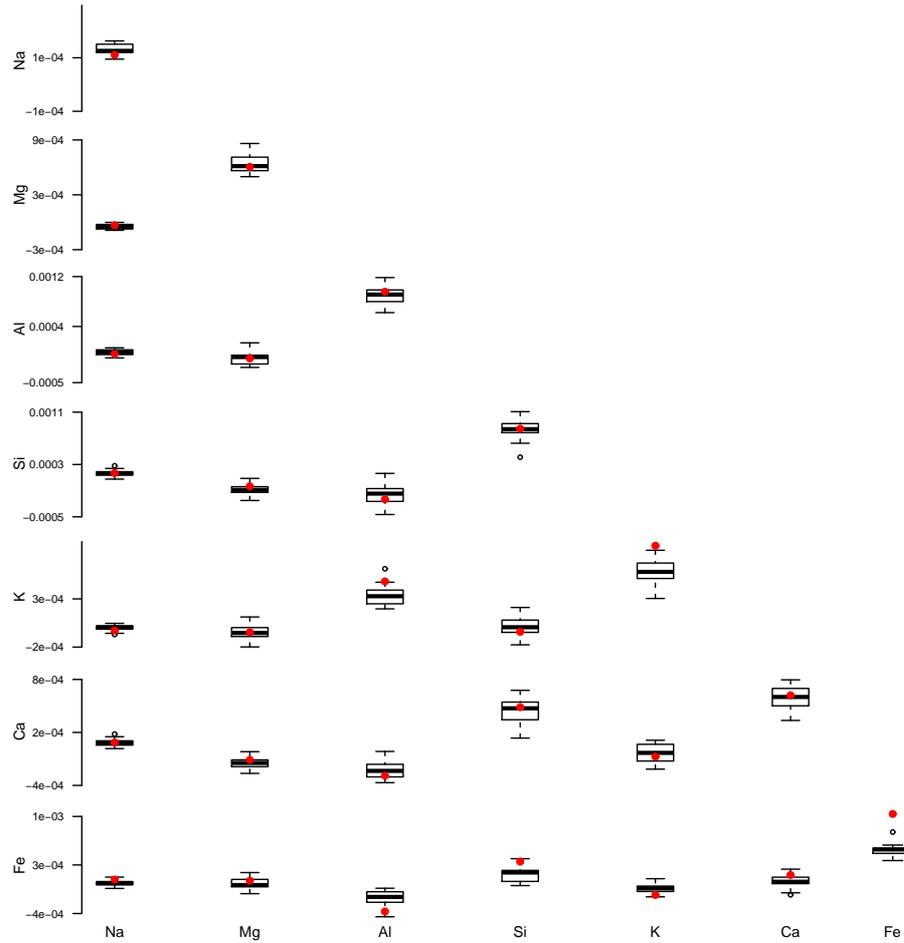


Figure 3.16: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 5$ (building window) from 20 simulated datasets from data augmentation for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

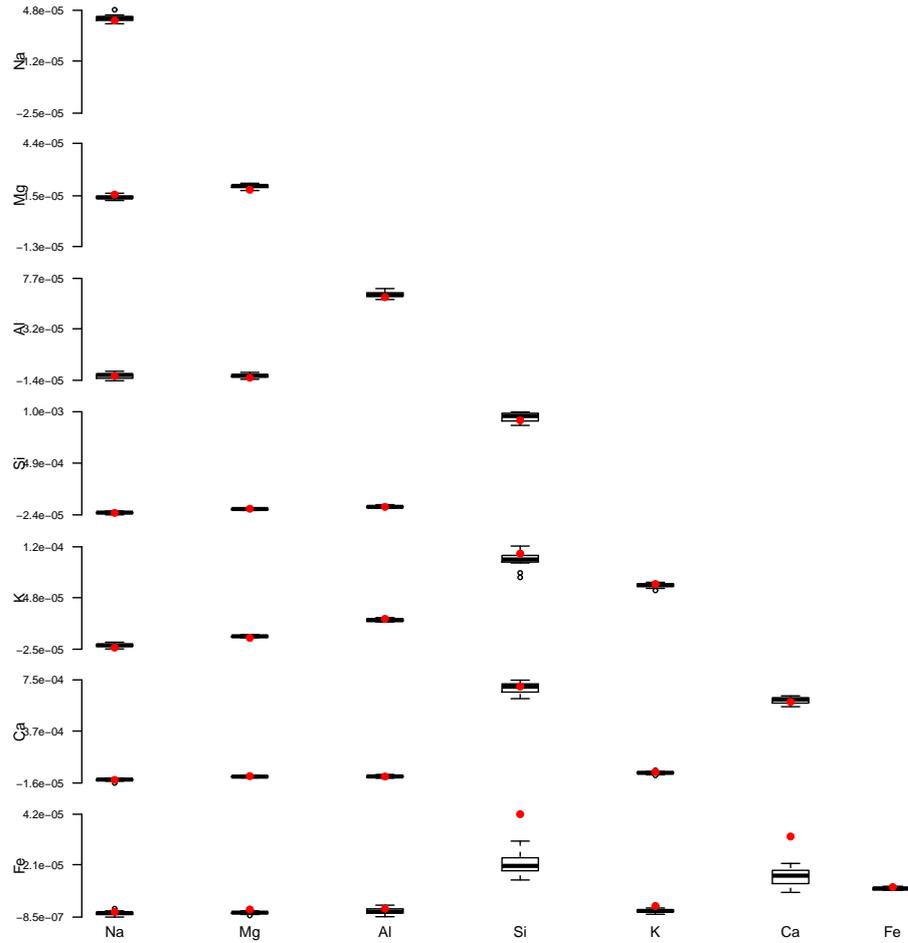


Figure 3.17: Boxplots containing the mean posterior draws of Ψ^{-1} from 20 simulated datasets from data augmentation for the square root ratios. The red dots indicate the true values of Ψ^{-1} used to simulate the datasets.

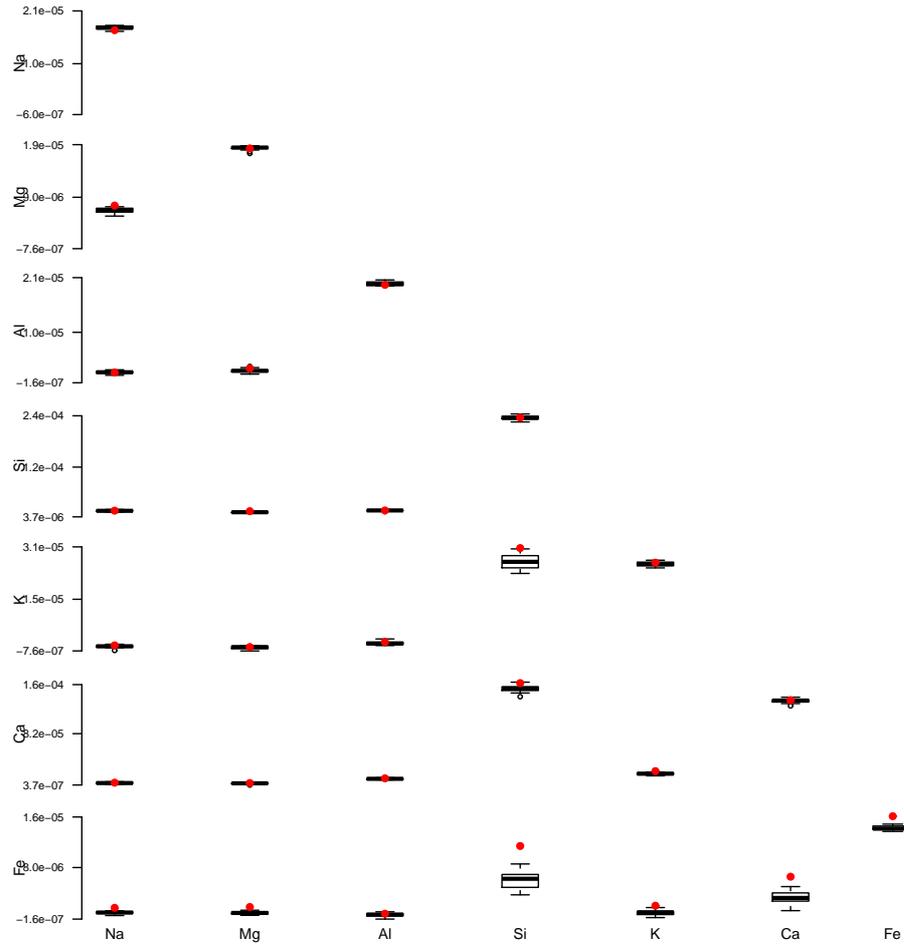


Figure 3.18: Boxplots containing the mean posterior draws of Λ^{-1} from 20 simulated datasets from data augmentation for the square root ratios. The red dots indicate the true values of Λ^{-1} used to simulate the datasets.

Simulated data using the composite model

Twenty datasets, each containing 64 items from each of the five use types, for a total of 320 glass items per dataset, were simulated using the composite model approach. As the composite model takes the elemental configuration of a glass item into account, the configuration of each simulated glass item has to be determined. Each glass item is assigned to one of the four elemental configurations using a multinomial $Mu(1, \varphi_t)$ distribution. The probabilities, φ_t come from formula (3.20):

$$\varphi_t | \xi, D \sim Dir(\alpha_{t1} + N_{t1}, \dots, \alpha_{tM} + N_{tM}), \quad t = 1, \dots, T.$$

The φ_t 's are generated from a Dirchlet distribution with parameter values $\alpha_{tm} = 0.1$, and the N_{tm} 's given by the number of items of use type t and configuration m in the glass database, as shown in Table 2.3. Once the configuration of a glass item has been allocated, it is simulated from the Bayesian hierarchical model as follows:

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

with Σ_t given by

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

Values for estimates of the parameters associated with the m -th configuration, $\xi_m = \{\boldsymbol{\theta}, \Omega, \Psi, \Lambda\}$, were obtained from the mean posterior draws from the Bayesian hierarchical model using the composite model approach applied to the glass database.

The results from simulating data shown here are those for glass items with elemental configuration $m = 2$ ($\overline{\text{Fe}}, K$). The results for glass items with ele-

mental configurations $m = 1, 3, 4$ can be found in Appendix E. Figure 3.19 displays boxplots obtained from the posterior mean draws of θ_t from the 20 simulated datasets for glass items with elemental configuration $m = 2$. As the composite model examines the presence and absence of the elements iron and potassium the potential issue of overestimating iron that was seen for the data augmentation approach is removed, with the model able to recover the true values of θ_t used to simulate the datasets. However, the same issue of overestimating magnesium for headlamps that was observed for data augmentation still persists. This may be due to 10 headlamps in the database not containing magnesium, while six do contain magnesium, while for all simulated items magnesium is present, with only iron and potassium deemed present or absent. The simulated results for the between-item random effect covariance matrices for items with configuration $m = 2$ can be seen in figures 3.20-3.24. For the most part, the model is able to recover the true estimates used for the Ω_t 's, with a couple of exceptions. Due to nine out of the 26 bulbs in the glass database not containing any magnesium, the covariances, as seen in Figure 3.20, are underestimated, with one overestimated, by the model. A similar problem is seen in Figure 3.22 for magnesium in headlamps. With the odd exception, the true estimates used to obtain the simulated data have been recovered at the within-item and measurement error levels, as shown in Figures 3.25 and 3.26, for covariance matrices Ψ^{-1} and Λ^{-1} , respectively. The composite model appears to be able to recover the true estimates better than the data augmentation approach, but a potential problem can arise depending on whether the glass items are, or are not, simulated for a particular configuration for specific use types. This is due to draws being made from prior distributions if there is no data available for a use type or configuration.

This can be seen in Appendix E.

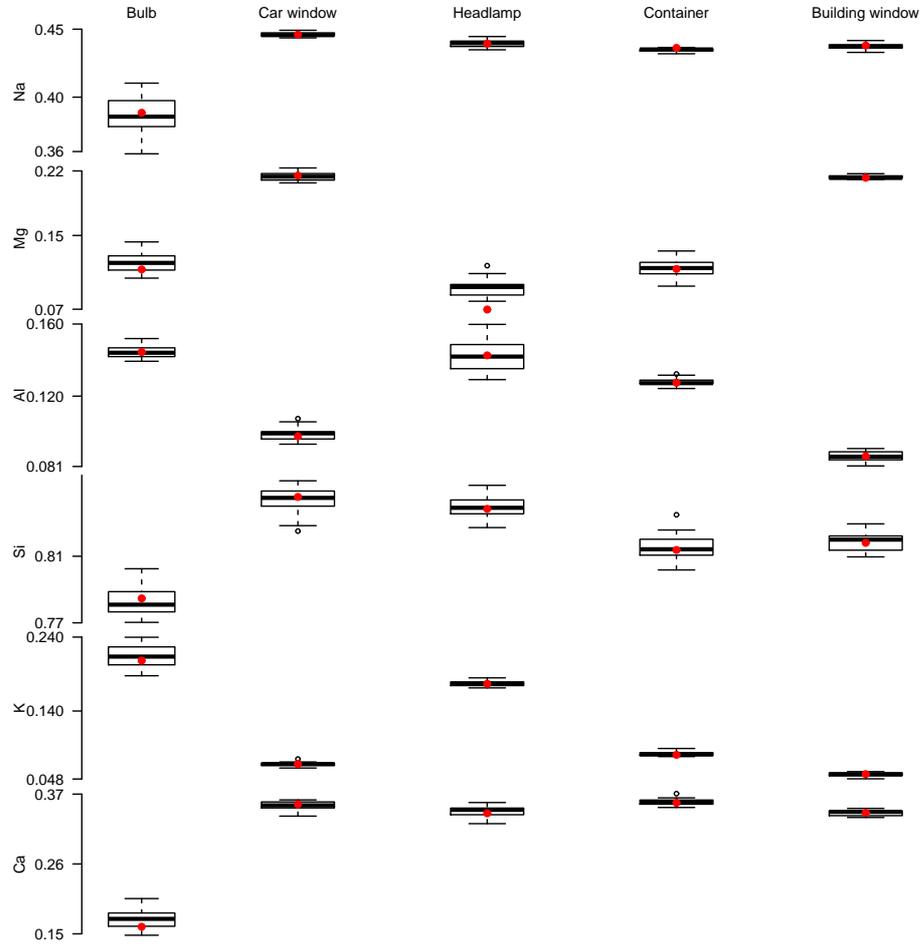


Figure 3.19: Boxplots containing the mean posterior draws of θ_t from 20 simulated datasets for configuration $m = 2$ ($\overline{\text{Fe}}, \text{K}$) for the square root ratios. The red dots indicate the true values of θ_t used to simulate the datasets.

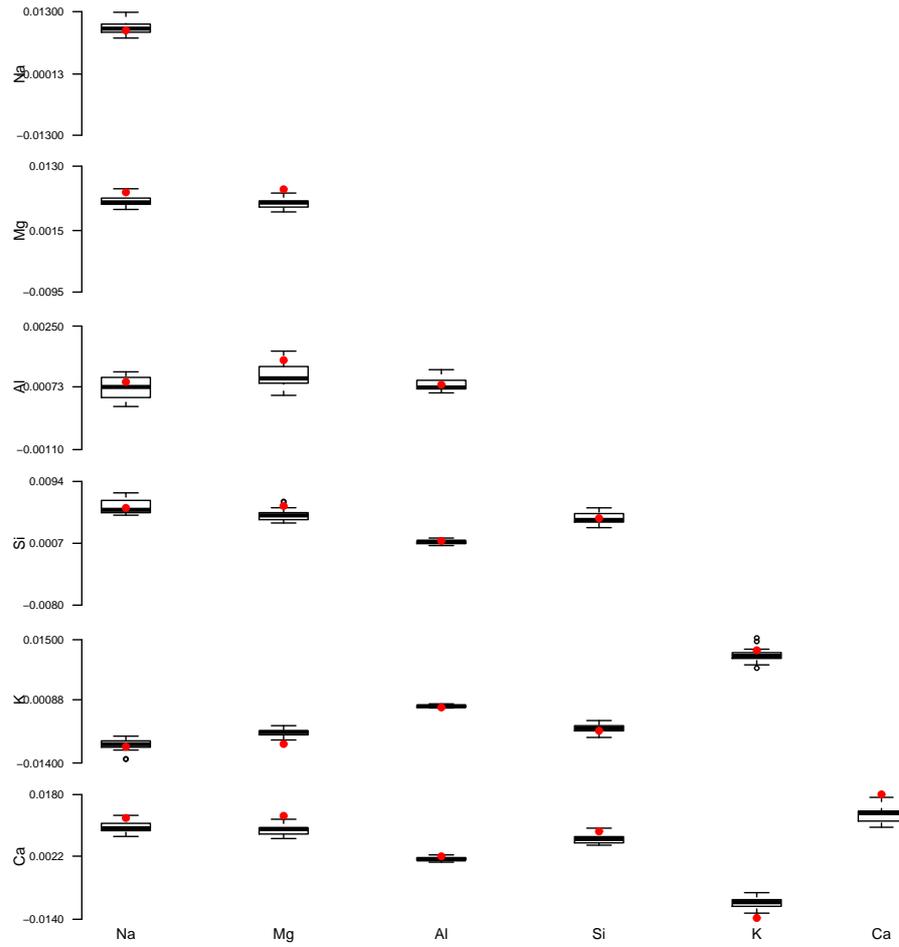


Figure 3.20: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 1$ (bulb) from 20 simulated datasets for configuration $m = 2$ ($\overline{\text{Fe}}$, K) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

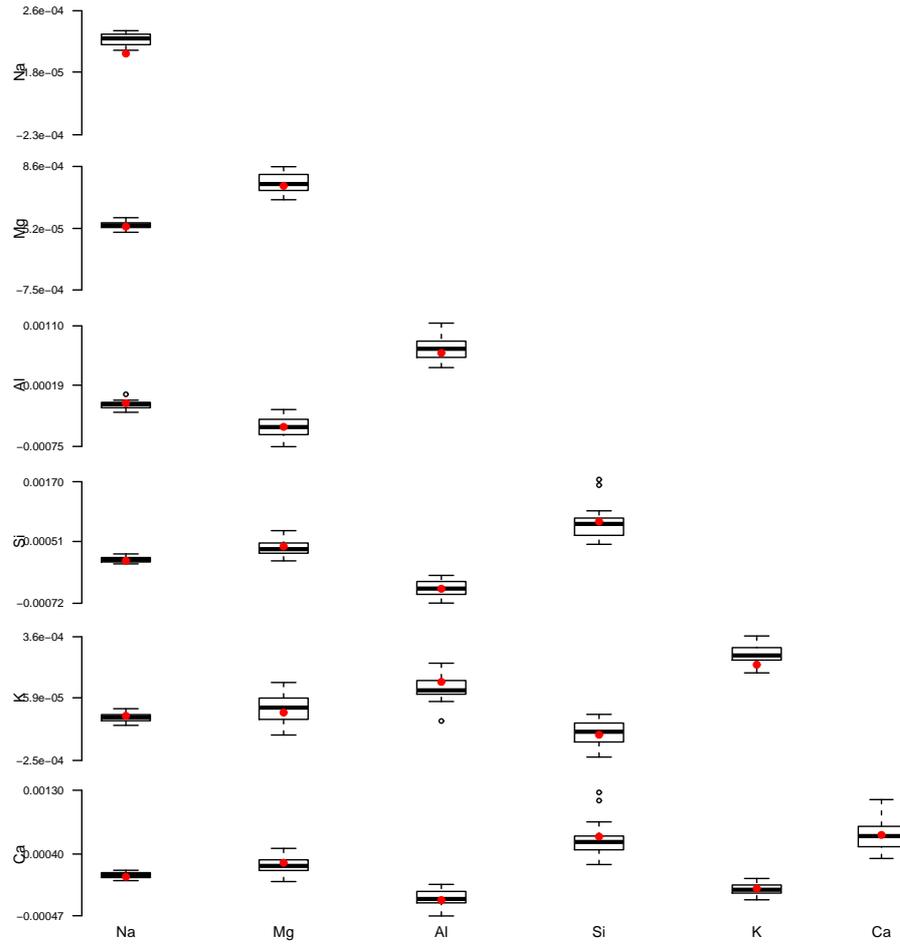


Figure 3.21: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 2$ (car window) from 20 simulated datasets for configuration $m = 2$ ($\overline{\text{Fe}}$, K) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

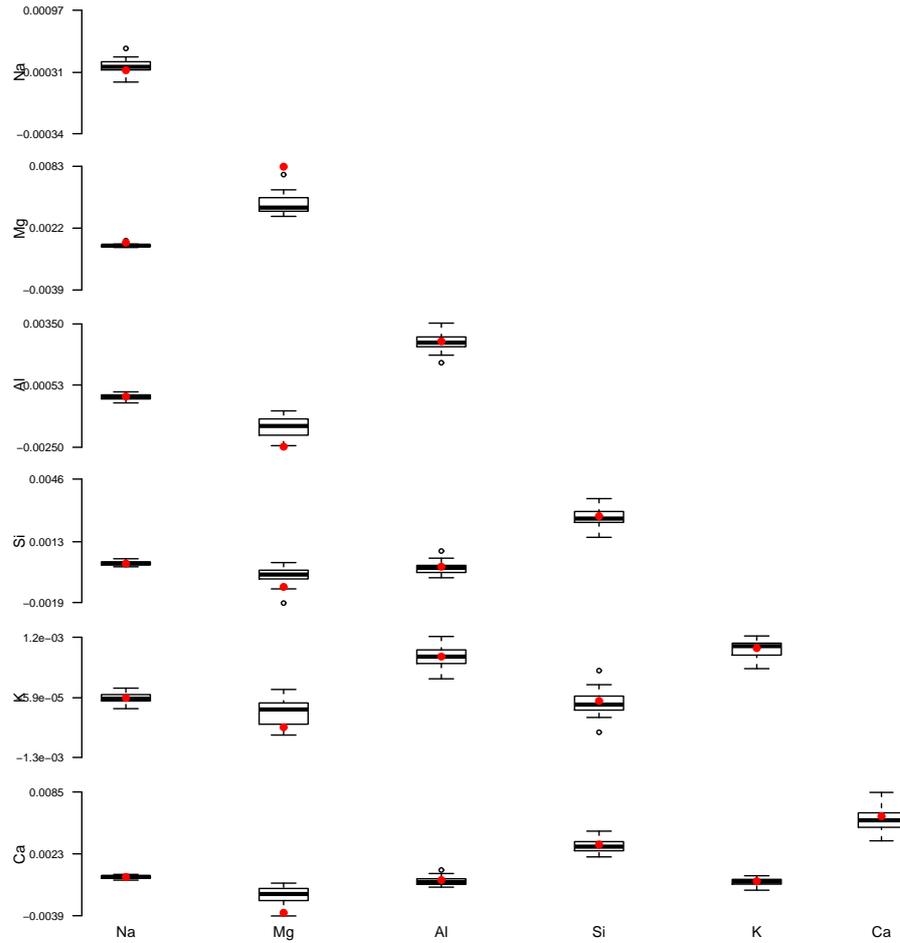


Figure 3.22: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 3$ (headlamp) from 20 simulated datasets for configuration $m = 2$ ($\overline{\text{Fe}}, \text{K}$) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

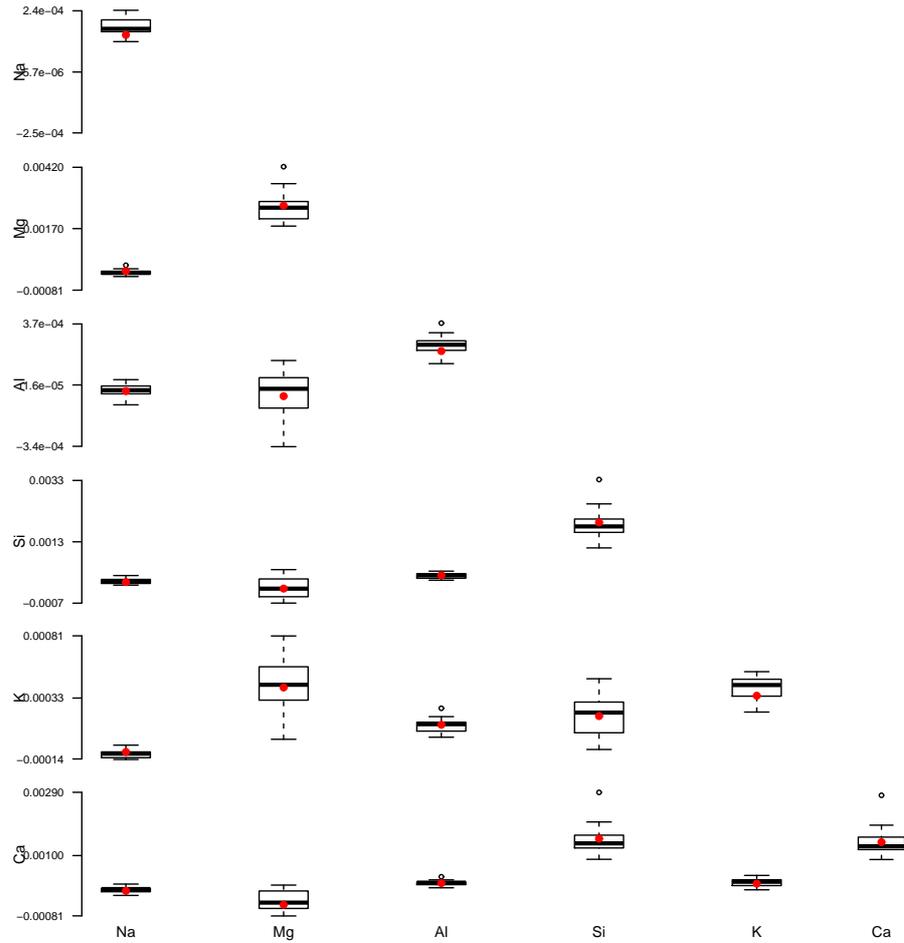


Figure 3.23: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 4$ (container) from 20 simulated datasets for configuration $m = 2$ ($\overline{\text{Fe}}, \text{K}$) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

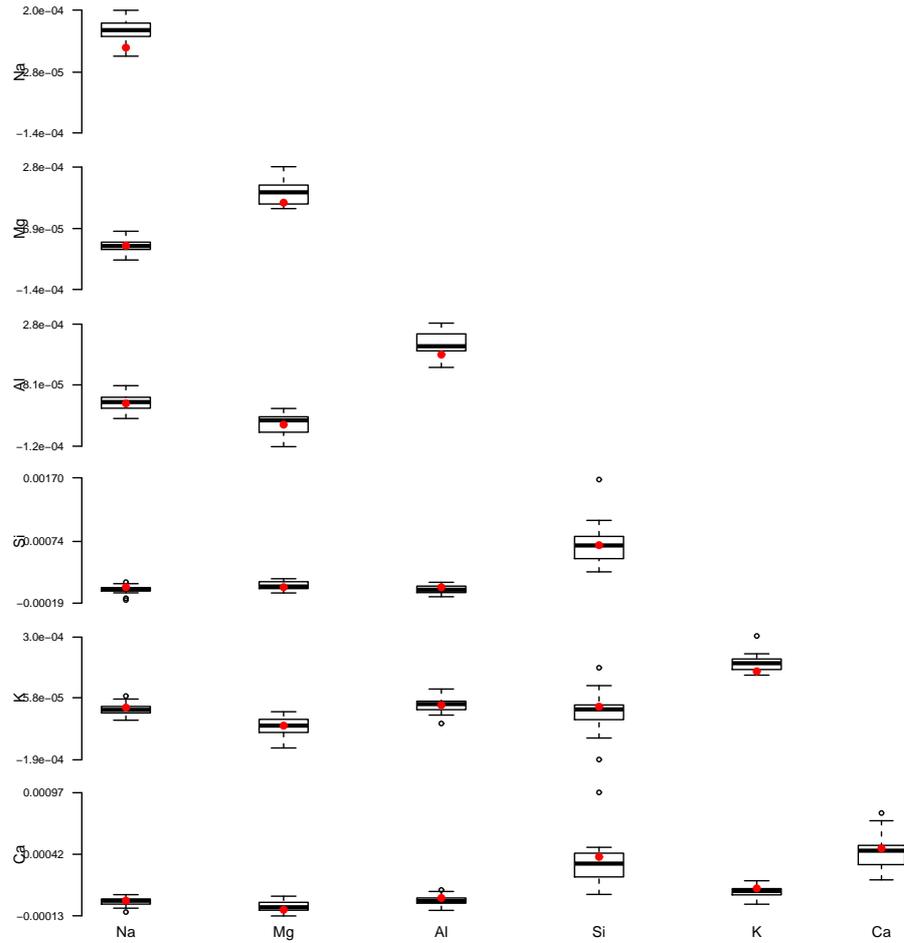


Figure 3.24: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 5$ (building window) from 20 simulated datasets for configuration $m = 2$ ($\overline{\text{Fe}}$, K) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

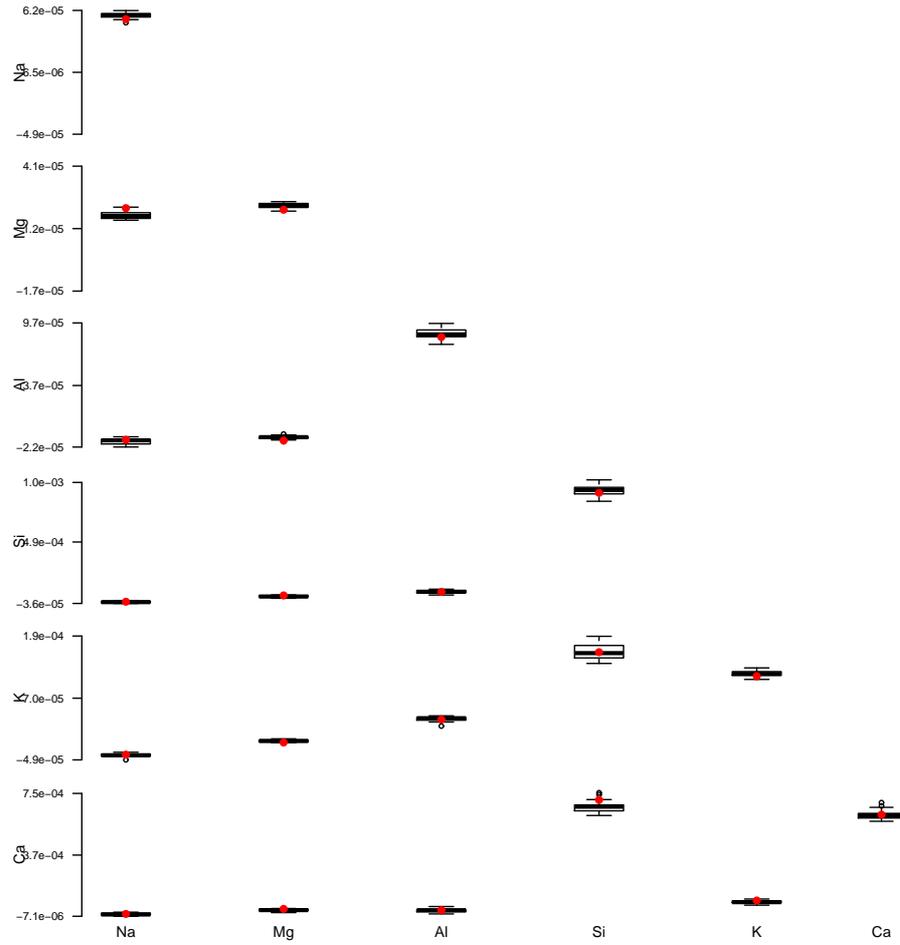


Figure 3.25: Boxplots containing the mean posterior draws of Ψ^{-1} from 20 simulated datasets for configuration $m = 2$ ($\bar{\text{Fe}}, \text{K}$) for the square root ratios. The red dots indicate the true values of Ψ^{-1} used to simulate the datasets.

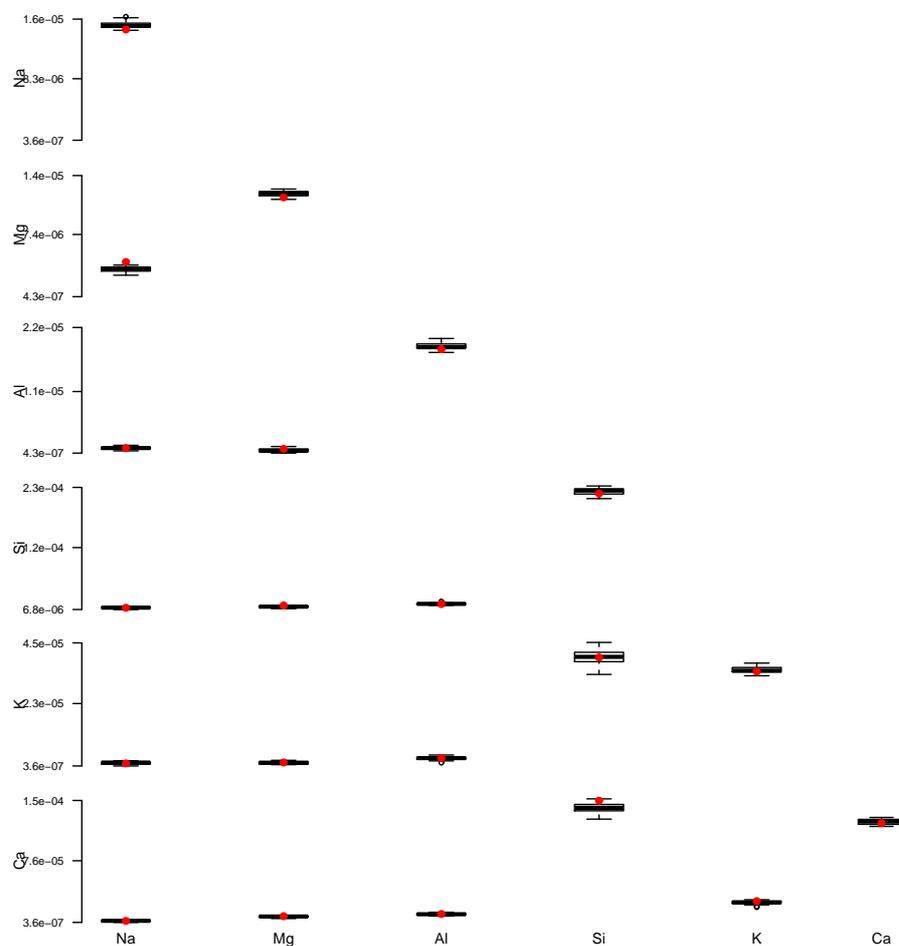


Figure 3.26: Boxplots containing the mean posterior draws of Λ^{-1} from 20 simulated datasets for configuration $m = 2$ ($\overline{\text{Fe}}, \text{K}$) for the square root ratios. The red dots indicate the true values of Λ^{-1} used to simulate the datasets.

The next chapter will use the models and MCMC draws obtained from them to classify glass items into different use type categories.

Chapter 4

Glass classification

This chapter will use hierarchical model (3.1) and the two methods of handling zeros detailed in Chapter 3 to classify glass items into use type categories. Denote by \mathbf{y} the transformed elemental composition of a newly observed glass item, with its unknown use type denoted by $\mathcal{T}_{\mathbf{y}}$, it is of interest to be able to correctly classify the new item to its use type. This can be done using the posterior distribution, $p(\mathcal{T}_{\mathbf{y}}|\mathbf{y}, D)$, of the use type $\mathcal{T}_{\mathbf{y}}$ of the newly observed item \mathbf{y} , given the reference database D and the newly obtained measurements from \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,

$$\begin{aligned} p(\mathcal{T}_{\mathbf{y}} = t|\mathbf{y}, D) &= p(\mathcal{T}_{\mathbf{y}} = t|\mathbf{y}, \mathcal{C}_{\mathbf{y}} = m, D) \\ &= \frac{p(\mathcal{T}_{\mathbf{y}} = t|\mathcal{C}_{\mathbf{y}} = m, D) p(\mathbf{y}|\mathcal{T}_{\mathbf{y}} = t, \mathcal{C}_{\mathbf{y}} = m, D)}{\sum_{s=1}^T p(\mathcal{T}_{\mathbf{y}} = s|\mathcal{C}_{\mathbf{y}} = m, D) p(\mathbf{y}|\mathcal{T}_{\mathbf{y}} = s, \mathcal{C}_{\mathbf{y}} = m, D)} \\ &\propto p(\mathcal{T}_{\mathbf{y}} = t|\mathcal{C}_{\mathbf{y}} = m, D) p(\mathbf{y}|\mathcal{T}_{\mathbf{y}} = t, \mathcal{C}_{\mathbf{y}} = m, D). \end{aligned} \quad (4.1)$$

Next, expressions for the two terms on the right-hand side of (4.1) are derived. Beginning with the first quantity and using again Bayes' Theorem, one has

$$\begin{aligned} p(\mathcal{T}_{\mathbf{y}} = t | \mathcal{C}_{\mathbf{y}} = m, D) &= \frac{p(\mathcal{T}_{\mathbf{y}} = t | D) p(\mathcal{C}_{\mathbf{y}} = m | \mathcal{T}_{\mathbf{y}} = t, D)}{\sum_{s=1}^T p(\mathcal{T}_{\mathbf{y}} = s | D) p(\mathcal{C}_{\mathbf{y}} = m | \mathcal{T}_{\mathbf{y}} = s, D)} \\ &\propto p(\mathcal{T}_{\mathbf{y}} = t | D) p(\mathcal{C}_{\mathbf{y}} = m | \mathcal{T}_{\mathbf{y}} = t, D). \end{aligned} \quad (4.2)$$

On its own, the reference database D is not informative about the use type $\mathcal{T}_{\mathbf{y}}$ of a newly observed glass item \mathbf{y} , since the number of glass items of each use type in the reference data do not reflect the prevalence of such use types as forensic samples in a real world setting. Therefore, $p(\mathcal{T}_{\mathbf{y}} = t | D) = p(\mathcal{T}_{\mathbf{y}} = t)$ and (4.2) becomes

$$p(\mathcal{T}_{\mathbf{y}} = t | \mathcal{C}_{\mathbf{y}} = m, D) \propto p(\mathcal{T}_{\mathbf{y}} = t) p(\mathcal{C}_{\mathbf{y}} = m | \mathcal{T}_{\mathbf{y}} = t, D). \quad (4.3)$$

As mentioned above, the prior distribution $p(\mathcal{T}_{\mathbf{y}} = t)$ associated with each use type should be indicative of real life forensic samples, and so this information should be reflected in the prior knowledge. If no prior information about the prevalence of use types is known then the prior probabilities $p(\mathcal{T}_{\mathbf{y}} = t)$ can be set to be equal to one another, i.e. a discrete uniform distribution where $p(\mathcal{T}_{\mathbf{y}} = t) = 1/T, t = 1, \dots, T$. The second term on the right-hand side of (4.3) can be computed as follows:

$$\begin{aligned} p(\mathcal{C}_{\mathbf{y}} = m | \mathcal{T}_{\mathbf{y}} = t, D) &= \int p(\mathcal{C}_{\mathbf{y}} = m | \mathcal{T}_{\mathbf{y}} = t, \varphi_t, D) p(\varphi_t | \mathcal{T}_{\mathbf{y}} = t, D) d\varphi_t \\ &= \int p(\mathcal{C}_{\mathbf{y}} = m | \mathcal{T}_{\mathbf{y}} = t, \varphi_t) p(\varphi_t | D) d\varphi_t \\ &= \int \varphi_{tm} p(\varphi_t | D) d\varphi_t \\ &= E_{\varphi_t | D}[\varphi_{tm}] \\ &= \frac{\alpha_{tm} + N_{tm}}{\sum_{r=1}^M (\alpha_{tr} + N_{tr})}, \end{aligned} \quad (4.4)$$

where $\varphi_{tm} = p(\mathcal{C}_{\mathbf{y}} = m | \mathcal{T}_{\mathbf{y}} = t, \varphi_t)$ as defined in (3.15), and by the posterior Dirichlet distribution of φ_t in (3.20). Substituting (4.4) into (4.3) yields the posterior distribution of the use type $\mathcal{T}_{\mathbf{y}}$ conditional only on the reference database D and the elemental configuration $\mathcal{C}_{\mathbf{y}}$, but without conditioning on the actual new glass item with transformed elemental composition \mathbf{y} :

$$p(\mathcal{T}_{\mathbf{y}} = t | \mathcal{C}_{\mathbf{y}} = m, D) \propto p(\mathcal{T}_{\mathbf{y}} = t) \frac{\alpha_{tm} + N_{tm}}{\sum_{r=1}^M (\alpha_{tr} + N_{tr})}. \quad (4.5)$$

The values obtained for (4.5) with $p(\mathcal{T}_{\mathbf{y}} = t) = 1/T$, $\alpha_{tm} = 0.1$ for all t and m , and the N_{tm} 's given in Table 2.3, are given in Table 4.1. The relative frequencies of use types in the reference data being unaccounted for in (4.5) for reasons mentioned earlier is reflected in Table 4.1. For example, in Table 4.1 the probability of the use type headlamp given configuration 2 is 0.256, but if the relative frequency of headlamps in the reference data was used then this would be significantly reduced. Also, the choice of α in (4.5) does not seem to matter much, with the classification results in Section 4.1.1 the same for values of α between 0.1 and 0.5.

The second term in (4.1) is the posterior predictive distribution of the newly observed item with transformed elemental composition \mathbf{y} given its use type and elemental configuration as well as the reference data, and is written as

$$\begin{aligned} p(\mathbf{y} | \mathcal{T}_{\mathbf{y}} = t, \mathcal{C}_{\mathbf{y}} = m, D) &= \\ &= \int p(\mathbf{y} | \mathcal{T}_{\mathbf{y}} = t, \mathcal{C}_{\mathbf{y}} = m, \xi_m, D) p(\xi_m | \mathcal{T}_{\mathbf{y}} = t, \mathcal{C}_{\mathbf{y}} = m, D) d\xi_m \\ &= \int p(\mathbf{y} | \mathcal{T}_{\mathbf{y}} = t, \mathcal{C}_{\mathbf{y}} = m, \xi_m) p(\xi_m | D_m) d\xi_m \\ &= E_{\xi_m | D_m} [p(\mathbf{y} | \mathcal{T}_{\mathbf{y}} = t, \mathcal{C}_{\mathbf{y}} = m, \xi_m)], \end{aligned} \quad (4.6)$$

where $E_{\xi_m | D_m}$ denotes the expectation with respect to the posterior distribution of ξ_m . Hierarchical model (3.1) gives the density $p(\mathbf{y} | \mathcal{T}_{\mathbf{y}} = t, \mathcal{C}_{\mathbf{y}} = m, \xi_m)$

in (4.6) for elemental configuration m . That is, for a newly observed glass item \mathbf{y} of use type t and elemental configuration m consisting of \tilde{J} fragments each with \tilde{K} repeated measurements, then the distribution of $\mathbf{y}|\mathcal{T}_{\mathbf{y}} = t, \xi_m$ is as given by formulae (3.2) and (3.3), with J replaced by \tilde{J} and K by \tilde{K} :

$$\mathbf{y}|\mathcal{T}_{\mathbf{y}} = t, \xi_m \sim N_{\tilde{J}\tilde{K}p}(\mathbf{1}_{\tilde{J}\tilde{K}} \otimes \boldsymbol{\theta}_t, \Sigma_t), \quad (4.7)$$

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

Plugging (4.5) and (4.6) into (4.1) gives the use type probability for a newly observed glass item \mathbf{y} :

$$p(\mathcal{T}_{\mathbf{y}} = t|\mathbf{y}, D) \propto p(\mathcal{T}_{\mathbf{y}} = t) \frac{\alpha_{tm} + N_{tm}}{\sum_{r=1}^M (\alpha_{tr} + N_{tr})} E_{\xi_m|D_m} [p(\mathbf{y}|\mathcal{T}_{\mathbf{y}} = t, \mathcal{C}_{\mathbf{y}} = m, \xi_m)]. \quad (4.8)$$

The expectation term in (4.8) is estimated by taking the average of the densities $p(\mathbf{y}|\mathcal{T}_{\mathbf{y}} = t, \mathcal{C}_{\mathbf{y}} = m, \xi_m)$ with the ξ_m 's given by the MCMC sample, as detailed in Section 3.1.1.

Table 4.1: 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

The use type probability of a newly observed item given in (4.8) is for when the elemental configuration of an item is considered as in the composite model

approach of Section 3.2.3. For the data augmentation approach of Section 3.2.1 the use type probability of a future item \mathbf{y} is given as

$$p(\mathcal{T}_{\mathbf{y}} = t | \mathbf{y}, D) \propto p(\mathcal{T}_{\mathbf{y}} = t) p(\mathbf{y} | \mathcal{T}_{\mathbf{y}} = t, D). \quad (4.9)$$

The posterior predictive distribution $p(\mathbf{y} | \mathcal{T}_{\mathbf{y}} = t, D)$ conditional on only the use type and reference data is

$$\begin{aligned} p(\mathbf{y} | \mathcal{T}_{\mathbf{y}} = t, D) &= \int p(\mathbf{y} | \mathcal{T}_{\mathbf{y}} = t, \xi, D) p(\xi | \mathcal{T}_{\mathbf{y}} = t, D) d\xi \\ &= \int p(\mathbf{y} | \mathcal{T}_{\mathbf{y}} = t, \xi) p(\xi | D) d\xi \\ &= E_{\xi | D}[p(\mathbf{y} | \mathcal{T}_{\mathbf{y}} = t, \xi)]. \end{aligned} \quad (4.10)$$

This leads to the use type probability of an item \mathbf{y} under the data augmentation approach being

$$p(\mathcal{T}_{\mathbf{y}} = t | \mathbf{y}, D) \propto p(\mathcal{T}_{\mathbf{y}} = t) E_{\xi | D}[p(\mathbf{y} | \mathcal{T}_{\mathbf{y}} = t, \xi)], \quad (4.11)$$

where the distribution of $\mathbf{y} | \mathcal{T}_{\mathbf{y}} = t, \xi$ is as in (4.7) with ξ replacing ξ_m .

4.1 Classification simulation study

In order to assess how well the models perform at classifying glass items into use type categories, a simulation study was carried out. Each glass item was classified into one of five use type categories (bulb, car window, headlamp, container or building window). The probabilities associated with classifying a set of fragments from the same item were estimated using expression (4.8) for the composite model approach, and estimated using (4.11) for the data augmentation approach. Each glass item is classified as the use type with

the highest probability. The classification results for each of the model approaches will also be compared with results obtained from one of the leading classification methods - support vector machines (SVM) (Cortes and Vapnik, 1995). Using the *svm* function as part of the *e1071* package in R (R Development Core Team, 2011), two separate results from using SVM were obtained: SVM.sub and SVM.full. SVM.sub splits the data into four different subsets by observing the presence or absence of the elements iron and potassium, as seen in Table 2.3, in the same way as the composite model approach. All glass items within each subset are then classified before bringing all items back together to obtain the overall classification results from the database. SVM.full is more akin to the data augmentation approach in that the elemental configurations were not taken into account, but instead of updating any zero values present as is the case with data augmentation, SVM.full uses the square root ratios with all zeros unchanged for each glass item in the database. The results for the SVM approaches use the vector means of each glass item, i.e. the mean across fragments and repeated measurements. The *tune.svm* function was used to tune the hyperparameters, with the polynomial kernel giving the best classification results for the two SVM approaches.

Five-fold cross-validation is used as part of the simulation study, where the reference data D was randomly split into five parts, each containing 64 glass items. One of the five parts is then selected to be the test dataset consisting of unobserved glass items \mathbf{y} . The remaining four parts are then collated to form the training dataset consisting of reference glass items \mathbf{z} , from which the model parameters are estimated. This process is repeated five times with

each of the five parts of D being the test dataset in order to classify all 320 glass items in the reference data.

4.1.1 Classification results

Tables 4.2 and 4.3 report the results of classifying all 320 glass items in the reference data into five use type categories for the composite model and data augmentation approaches, respectively. The respective results for the SVM approaches SVM.sub and SVM.full can be found in Tables 4.4 and 4.5. Of the four approaches the composite model is able to correctly classify the most glass items with a misclassification rate of 20.6%, with the data augmentation approach having the second best outcome with a misclassification rate of 21.9%. SVM.sub and SVM.full report misclassification rates of 22.8% and 22.5%, respectively. The misclassification rates for the composite model and data augmentation approaches reflect good performance in the classification of bulbs, headlamps and containers, and poorer performance when classifying car and building windows. In contrast, both SVM approaches seem to be better at classifying building windows, but fall short for the other four use types. For each approach, misclassification of a window type is most often to the other window type. This is due to both types of windows having very similar elemental compositions, thus making it very difficult to distinguish between them based on their elemental composition alone. Zadora (2009) 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 4.2: Composite model: classification of each glass item into one of five use type categories.

Classification	Glass type					Total
	bulb	car window	headlamp	container	building window	
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

Table 4.3: Data augmentation: classification of each glass item into one of five use type categories.

Classification	Glass type					Total
	bulb	car window	headlamp	container	building window	
bulb	26	0	1	1	0	28
car window	0	75	0	2	33	110
headlamp	0	1	13	2	1	17
container	0	1	1	71	6	79
building window	0	17	1	3	65	86
Total	26 (100.0%)	94 (79.8%)	16 (81.3%)	79 (89.9%)	105 (61.9%)	320

Table 4.4: SVM.sub: classification of each glass item into one of five use type categories.

Classification	Glass type					Total
	bulb	car window	headlamp	container	building window	
bulb	23	0	1	0	0	24
car window	0	69	1	7	26	103
headlamp	0	0	10	1	0	11
container	2	3	4	69	3	81
building window	1	22	0	2	76	101
Total	26 (88.5%)	94 (73.4%)	16 (62.5%)	79 (87.3%)	105 (72.4%)	320

Table 4.5: SVM.full: classification of each glass item into one of five use type categories.

Classification	Glass type					Total
	bulb	car window	headlamp	container	building window	
bulb	23	0	1	0	1	25
car window	0	61	1	3	22	87
headlamp	2	0	12	0	0	14
container	1	2	1	71	1	76
building window	0	31	1	5	81	118
Total	26 (88.5%)	94 (64.9%)	16 (75.0%)	79 (89.9%)	105 (77.1%)	320

Tables 4.6 and 4.7 display the expected number of items, according to the predictive distributions (4.8) and (4.11), for the composite model and data augmentation approaches. The expected counts were obtained by taking the probability associated with each use type for each item and summing them together for all items of that use type. For example, for two bulbs each given 0.9 probability of being bulbs and 0.1 probability of being headlamps, the expected counts obtained would then be 1.8 bulbs and 0.2 headlamps. With the exception of bulbs, the expected number of glass items correctly classified are all closer to the actual number of items of each use type in the database for the composite model. This is due to only one of the bulbs in the reference data having elemental configuration $m = 4$; see the second bulb in Figure 4.1. The expected counts for both SVM approaches are much smaller, indicating greater uncertainty in the classification of glass items using this approach; see Tables 4.8 and 4.9.

Table 4.6: Composite model: expected counts of each glass item into one of five use type categories.

Classification	Glass type					Total
	bulb	car window	headlamp	container	building window	
bulb	24.39	0.02	1.24	0.40	0.53	26.58
car window	0.51	63.93	0.00	3.96	29.13	97.53
headlamp	0.62	1.01	14.39	1.29	1.06	18.37
container	0.47	4.07	0.37	69.80	6.07	80.78
building window	0.02	24.97	0.00	3.55	68.21	96.75
Total	26	94	16	79	105	320

Table 4.7: Data augmentation: expected counts of each glass item into one of five use type categories.

Classification	Glass type					Total
	bulb	car window	headlamp	container	building window	
bulb	25.05	0.00	1.33	0.49	0.33	27.20
car window	0.00	63.21	0.13	3.26	40.07	106.67
headlamp	0.93	0.61	12.09	2.74	1.72	18.09
container	0.02	3.44	1.55	68.05	6.02	79.08
building window	0.00	26.74	0.89	4.46	56.86	88.95
Total	26	94	16	79	105	320

Table 4.8: SVM.sub: expected counts of each glass item into one of five use type categories.

Classification	Glass type					Total
	bulb	car window	headlamp	container	building window	
bulb	17.77	1.36	1.52	1.89	1.98	24.52
car window	1.68	42.04	1.81	16.90	35.55	97.98
headlamp	2.99	2.05	5.98	3.72	2.97	17.71
container	1.64	12.01	5.60	45.51	11.83	76.59
building window	1.92	36.54	1.09	10.98	52.67	103.20
Total	26	94	16	79	105	320

Table 4.9: SVM.full: expected counts of each glass item into one of five use type categories.

Classification	Glass type					Total
	bulb	car window	headlamp	container	building window	
bulb	19.31	1.03	0.89	1.16	1.28	23.67
car window	1.26	40.66	2.75	13.38	37.08	95.13
headlamp	2.42	2.12	7.02	2.92	2.66	17.14
container	1.64	8.99	4.48	49.28	11.17	75.56
building window	1.38	41.20	0.86	12.26	52.80	108.50
Total	26	94	16	79	105	320

Tables 4.10 and 4.11 display the expected probabilities for the composite model and data augmentation approaches, respectively. The expected probabilities were obtained by taking an average of the five probabilities associated with the five use types across all items of the same use type. They give an indication of how likely a glass item of use type bulb, say, is on average classified as a bulb. For example, for two bulbs, if one had probabilities 0.85 and 0.15 of being a bulb or headlamp, and the other had probabilities 0.90 and 0.10 for the same use types; the expected probabilities from those two bulbs associated with use types bulb and headlamp would be 0.875 and 0.125, respectively. Again, with the exception of bulbs, the composite model outperforms the data augmentation approach, with both greatly outperforming both SVM approaches; see Tables 4.12 and 4.13. Figures 4.1 and 4.2 display the two largest probabilities associated with each individual glass item for the composite model and data augmentation approaches, respectively. Both approaches show good separation between the two largest probabilities for glass items of use types bulb, headlamp and container, with more uncertainty surrounding the classification of the two window types. The performance of both approaches is generally similar, with the composite model slightly outperforming the data augmentation approach; however, the biggest difference between the two approaches can be seen from the uncertainty surrounding the classification of building windows, with better performance seen in favour of the composite model. Figures 4.3 and 4.4 display the results for the SVM approaches with both figures emphasising the much greater levels of uncertainty associated with these approaches, especially when it comes to classifying car and building windows.

Table 4.10: Composite model: expected probability of each glass item into one of five use type categories. The values in parentheses are the standard deviations.

Classification	Glass type					Total
	bulb	car window	headlamp	container	building window	
bulb	0.94 (0.21)	0.00 (0.00)	0.08 (0.25)	0.01 (0.04)	0.01 (0.05)	1.04
car window	0.02 (0.10)	0.68 (0.29)	0.00 (0.00)	0.05 (0.19)	0.28 (0.33)	1.03
headlamp	0.02 (0.09)	0.01 (0.10)	0.90 (0.26)	0.02 (0.10)	0.01 (0.08)	0.96
container	0.02 (0.09)	0.04 (0.13)	0.02 (0.09)	0.88 (0.27)	0.06 (0.18)	1.02
building window	0.00 (0.00)	0.27 (0.27)	0.00 (0.00)	0.04 (0.15)	0.65 (0.37)	0.96
Total	1.00	1.00	1.00	1.00	1.01	5.01

Table 4.11: Data augmentation: expected probability of each glass item into one of five use type categories. The values in parentheses are the standard deviations.

Classification	Glass type					Total
	bulb	car window	headlamp	container	building window	
bulb	0.96 (0.09)	0.00 (0.00)	0.08 (0.23)	0.01 (0.06)	0.00 (0.02)	1.05
car window	0.00 (0.00)	0.67 (0.26)	0.01 (0.02)	0.04 (0.13)	0.38 (0.26)	1.10
headlamp	0.04 (0.09)	0.01 (0.06)	0.76 (0.33)	0.03 (0.09)	0.02 (0.10)	0.86
container	0.00 (0.00)	0.04 (0.08)	0.10 (0.19)	0.86 (0.24)	0.06 (0.16)	1.06
building window	0.00 (0.00)	0.28 (0.23)	0.06 (0.20)	0.06 (0.13)	0.54 (0.28)	0.94
Total	1.00	1.00	1.01	1.00	1.00	5.01

Table 4.12: SVM.sub: expected probability of each glass item into one of five use type categories. The values in parentheses are the standard deviations.

Classification	Glass type					Total
	bulb	car window	headlamp	container	building window	
bulb	0.68 (0.29)	0.01 (0.01)	0.10 (0.09)	0.02 (0.02)	0.02 (0.01)	0.83
car window	0.06 (0.09)	0.45 (0.18)	0.11 (0.14)	0.21 (0.13)	0.34 (0.17)	1.17
headlamp	0.11 (0.09)	0.02 (0.02)	0.37 (0.33)	0.05 (0.04)	0.03 (0.02)	0.58
container	0.06 (0.12)	0.13 (0.12)	0.35 (0.29)	0.58 (0.21)	0.11 (0.13)	1.23
building window	0.07 (0.13)	0.39 (0.17)	0.07 (0.09)	0.14 (0.11)	0.50 (0.17)	1.17
Total	0.98	1.00	1.00	1.00	1.00	4.98

Table 4.13: SVM.full: expected probability of each glass item into one of five use type categories. The values in parentheses are the standard deviations.

Classification	Glass type					Total
	bulb	car window	headlamp	container	building window	
bulb	0.74 (0.27)	0.01 (0.00)	0.06 (0.09)	0.01 (0.00)	0.01 (0.02)	0.83
car window	0.05 (0.07)	0.43 (0.14)	0.17 (0.28)	0.17 (0.11)	0.35 (0.14)	1.17
headlamp	0.09 (0.12)	0.02 (0.01)	0.44 (0.36)	0.04 (0.02)	0.03 (0.01)	0.62
container	0.06 (0.10)	0.10 (0.12)	0.28 (0.26)	0.62 (0.22)	0.11 (0.08)	1.17
building window	0.05 (0.10)	0.44 (0.13)	0.05 (0.10)	0.16 (0.16)	0.50 (0.15)	1.20
Total	0.99	1.00	1.00	1.00	1.00	4.99

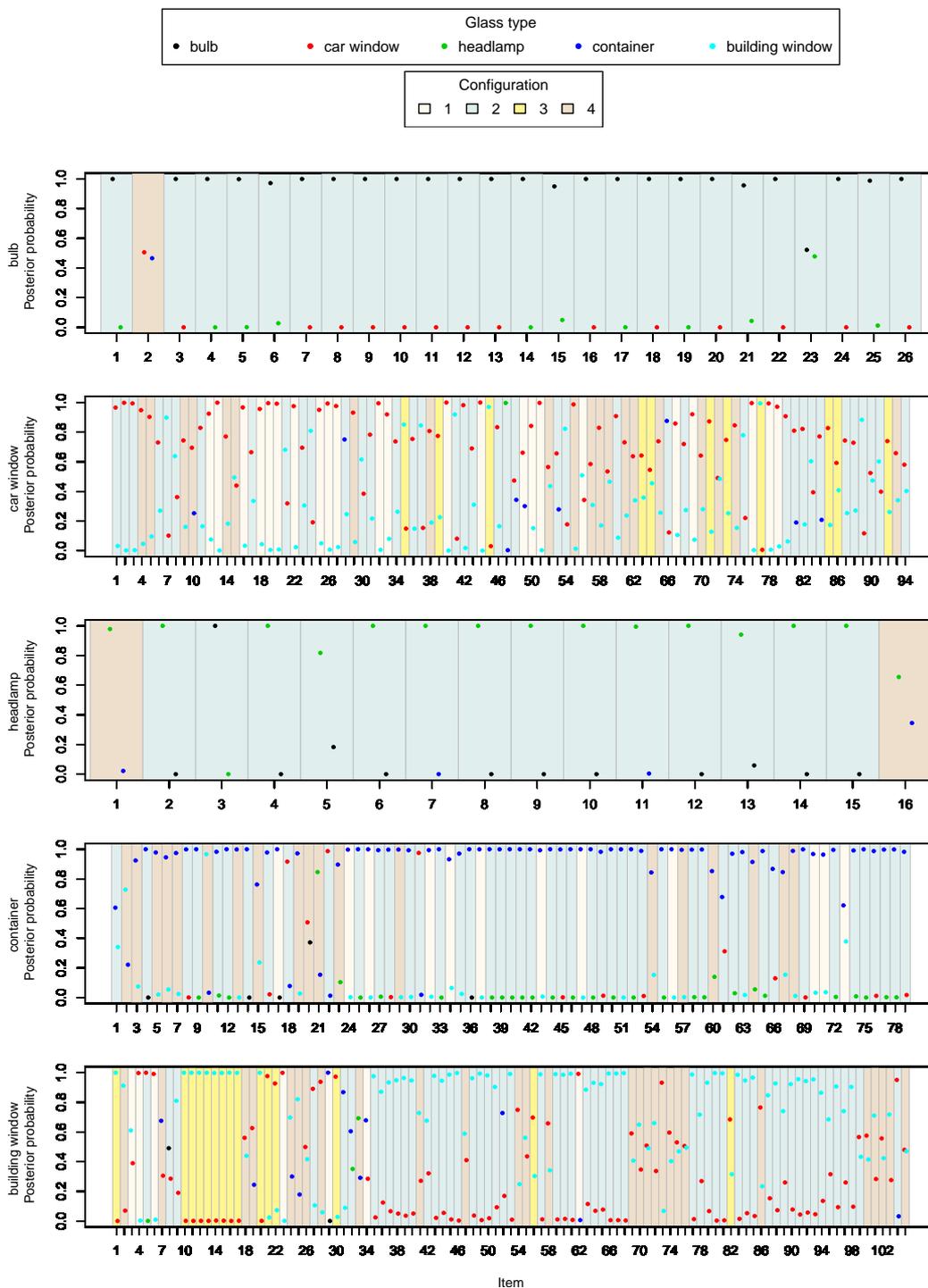


Figure 4.1: Composite model: Posterior probabilities for the classification of each glass item. The two largest posterior probabilities are displayed for each item.

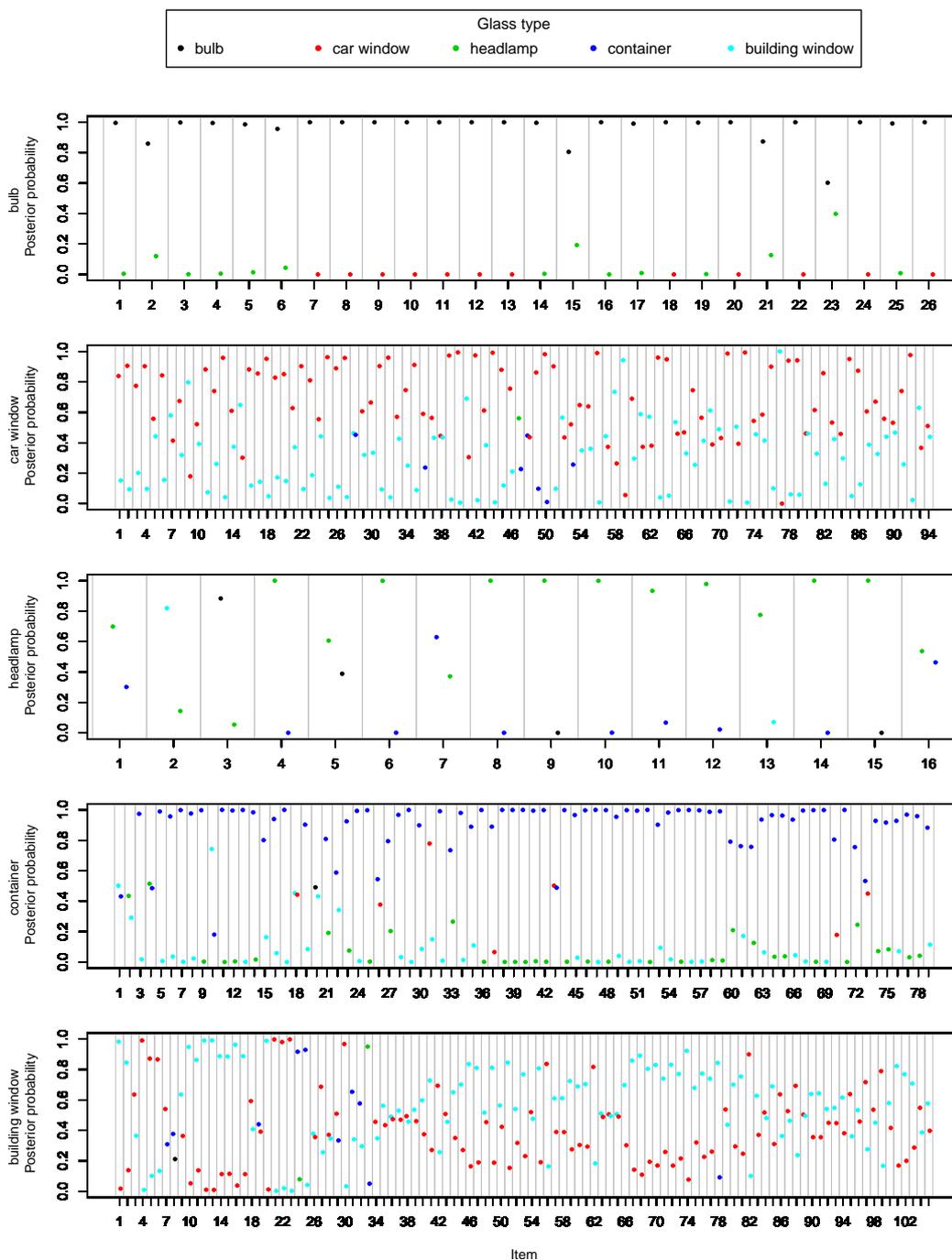


Figure 4.2: Data augmentation: Posterior probabilities for the classification of each glass item. The two largest posterior probabilities are displayed for each item.

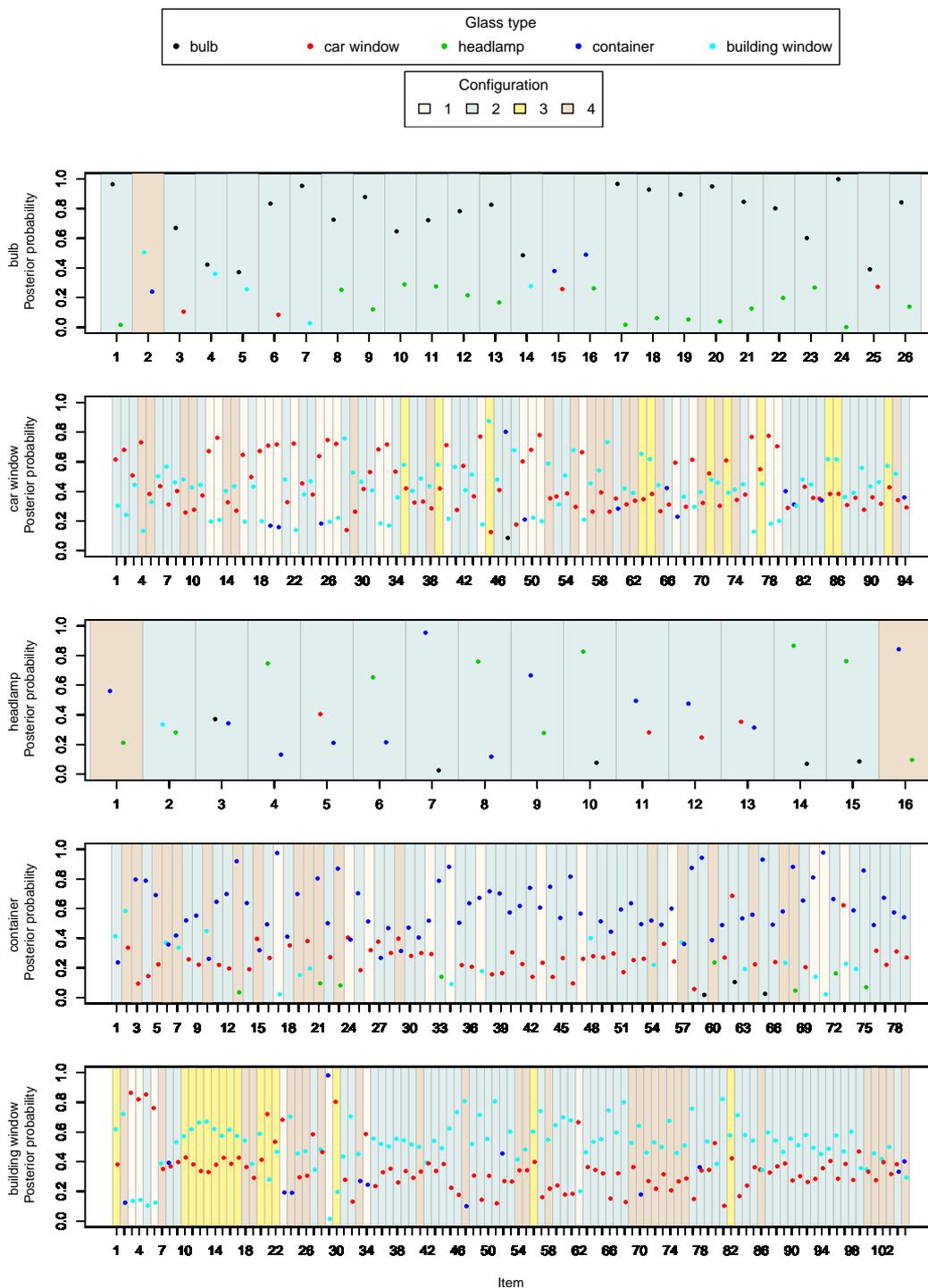


Figure 4.3: SVM.sub: Posterior probabilities for the classification of each glass item. The two largest posterior probabilities are displayed for each item.

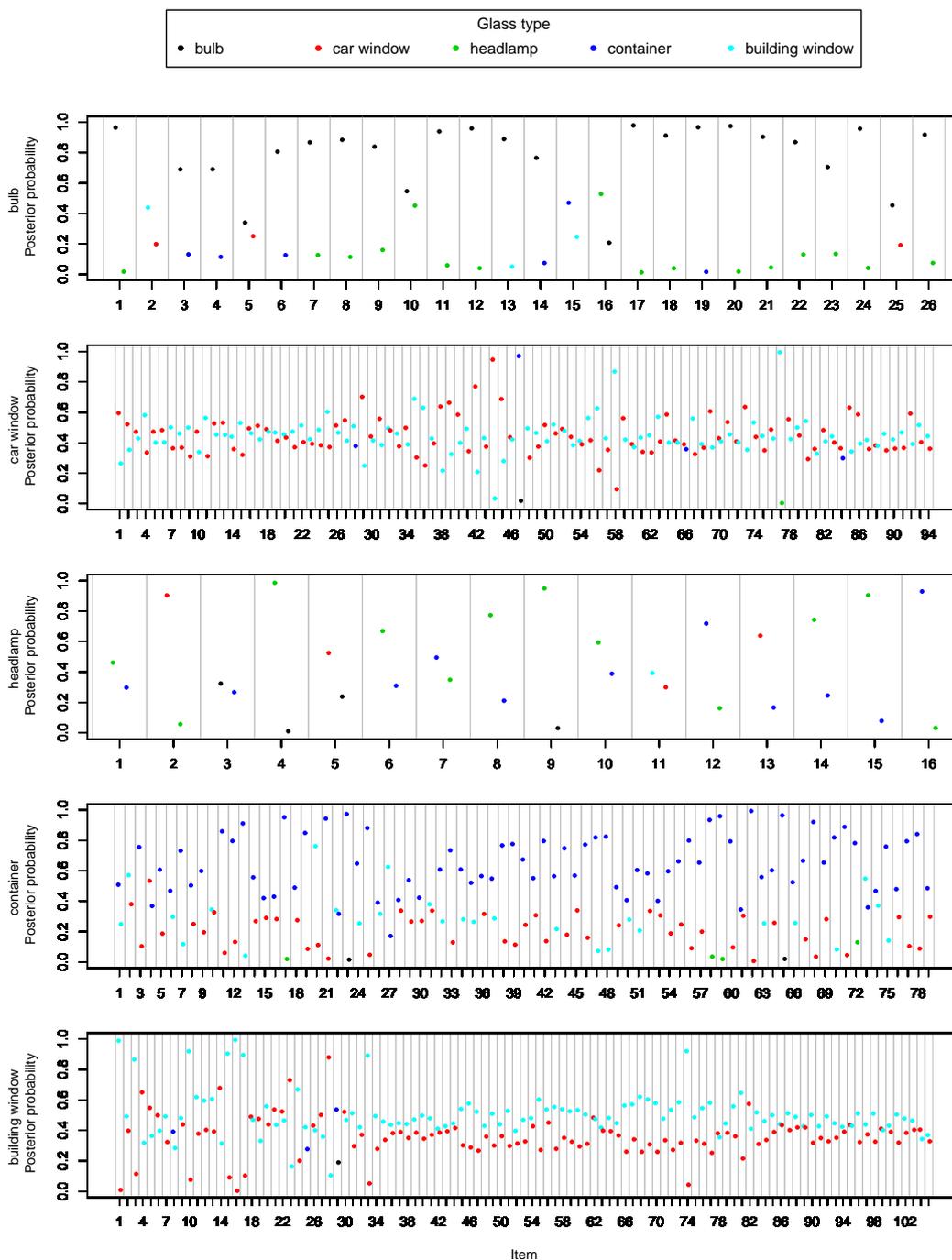


Figure 4.4: SVM.full: Posterior probabilities for the classification of each glass item. The two largest posterior probabilities are displayed for each item.

4.2 Classification performance measures

To measure the classification performance of the two different models and SVM approaches, a number of performance measures can be calculated. Each approach will be compared using four different performance measures. Three of the measures are detailed in [Agresti \(1990\)](#), whose notation is adopted. The classification measures τ and U in sections [4.2.1](#) and [4.2.2](#) are measures of association and examine the proportional reduction in prediction error. In this case both measures would be examining how much the predicted use types deviate from the true use types. The classification measure κ in Section [4.2.3](#) is a measure of the strength of agreement, e.g. between the predicted and true use types. Both measures of association and the measure of agreement are examined as it is possible to have strong association without having strong agreement. The final classification measure in Section [4.2.4](#) is the Brier score and is the only measure of the four to use the actual prediction probabilities obtained from the classification procedure, and measures the strength of those predictions. For a given contingency table, such as [Table 4.2](#), let π_{ij} denote the probability that a newly observed glass item of true use type j is classified as being of use type i ; $\pi_{+j} = \sum_i \pi_{ij}$ and $\pi_{i+} = \sum_j \pi_{ij}$ be the marginal probabilities obtained from summing the probabilities across the rows and columns, respectively. Then, the performance measures are given as follows:

4.2.1 Goodman and Kruskal's tau

Goodman and Kruskal's tau (Agresti, 1990, p. 24) is a measure of the reduction in the expected conditional variability in comparison to the marginal variability. Goodman and Kruskal's tau is

$$\tau = \frac{\sum_i \sum_j \pi_{ij}^2 / \pi_{i+} - \sum_j \pi_{+j}^2}{1 - \sum_j \pi_{+j}^2}, \quad (4.12)$$

and ranges between 0 and 1. A value of $\tau = 1$ implies that the expected conditional variation is equal to 0, which occurs when the first term in the numerator of (4.12) is equal to 1; see (Agresti, 1990, p. 24) for details. This happens when the joint distribution probabilities on the main diagonal of the contingency table are equal to the corresponding marginal probabilities, i.e. $\pi_{ii} = \pi_{i+} = \pi_{+i}$ for all $i = j$, thus indicating perfect classification performance. A value of $\tau = 0$ is equivalent to the rows and columns being independent, indicating poor performance.

4.2.2 Theil's U

Theil's U (Agresti, 1990, p. 25) provides an alternative measure of variation to that of Goodman and Kruskal's tau and is given as

$$U = -\frac{\sum_i \sum_j \pi_{ij} \log(\pi_{ij} / \pi_{i+} \pi_{+j})}{\sum_j \pi_{+j} \log \pi_{+j}},$$

where much like the values of τ , values of U close to 1 indicate good performance, and values close to 0 indicate poor performance.

4.2.3 Cohen's kappa

Cohen's kappa ([Agresti, 1990](#), p. 366) is a measure of agreement that takes into account any agreement that can occur by chance. Cohen's kappa is given as

$$\kappa = \frac{\sum \pi_{ii} - \sum \pi_{i+}\pi_{+i}}{1 - \sum \pi_{i+}\pi_{+i}} = \frac{\Pi_o - \Pi_e}{1 - \Pi_e},$$

where $\Pi_o = \sum \pi_{ii}$ is the probability of agreement and $\Pi_e = \sum \pi_{i+}\pi_{+i}$ is the probability of agreement by chance. A value of $\kappa = 0$ is equivalent to that of agreement by chance, whereas perfect agreement would have a value of $\kappa = 1$.

4.2.4 Brier score

The Brier score ([Brier, 1950](#)) is the average of the squared differences between the probabilities associated with the classification of glass items into use types and those glass items' true use types, which are denoted by binary indicators. For example, a glass item classified into two use types, $t = 1, 2$, with $t = 1$ the true use type, may have classification probabilities 0.75 and 0.25 for $t = 1$ and $t = 2$, respectively. The true use type would then be indicated by 1 for $t = 1$ and 0 for $t = 2$; taking the squared differences would then result in a Brier score for this glass item of 0.125. For multiple glass items the Brier score is given by

$$BS = \frac{1}{I} \sum_{i=1}^I \sum_{t=1}^T (\pi_{ti} - o_{ti})^2,$$

where $I = \sum_t I_t$ is the total number of items being classified; π_{ti} is the probability that the i -th item is of use type t ; and o_{ti} indicates the actual

use type of the i -th item, i.e. $o_{ti} = 1$ if item i is of use type t , and $o_{ti} = 0$ if the i -th item is not of use type t . The Brier score ranges from 0 to 1 with a value of $BS = 0$ indicating perfect predictions. An assessment of some classification methods can be found in [Hand \(2012\)](#).

4.2.5 Classification performance results

Table 4.14 reports the results from the four performance measures, as well as the misclassification rate of each approach. The composite model performs best for each measure except for that of Theil's U where the data augmentation approach is slightly better. Table 4.14 also includes the classification performance for the Bayesian hierarchical model without the implementation of data augmentation. This approach does not perform as well as the composite model and data augmentation approaches, however it outperforms SVM for each of the performance measures minus the misclassification rate of SVM.full.

From the results of classification in Section 4.1.1 and classification performance in Section 4.2.5 the composite model outperforms the data augmentation approach, which in turn, outperforms both the SVM approaches. The classification of glass items of use types bulb, headlamp and container displayed good performance, but due to having very similar elemental compositions, the classification of car and building windows resulted in poorer performance. This leads to what appears to be a large misclassification rate of approximately 20%. However, although the windows are tricky to correctly classify, the composite model appears to give classification results that

are the best among all methods applied to this glass database.

Table 4.14: Classification performance measures for each classification approach.

Approach	Classification performance				
	τ	U	κ	BS	% mis.
Composite model	0.560	0.591	0.721	0.319	20.6%
Data augmentation	0.551	0.608	0.706	0.330	21.9%
BHM without DA	0.538	0.588	0.693	0.338	22.8%
SVM.sub	0.515	0.534	0.688	0.447	22.8%
SVM.full	0.536	0.565	0.693	0.431	22.5%

The next chapter will use the hierarchical model to compute the evidential value of glass fragments under two competing propositions about their source.

Chapter 5

Evidence evaluation

The hierarchical model given in Section 3.1 will now be used to obtain the evidential value of glass fragments under two complementary propositions. As a brief reminder from Chapter 1, the evidential value, V , of the evidence E is defined as

$$V = \frac{\Pr(E|H_p, I)}{\Pr(E|H_d, I)},$$

i.e. the value for H_p and against H_d , on evidence E , where I denotes additional background information related to the case in question.

In terms of glass fragments as forensic evidence, the evidence E is obtained from two separate sources. Let \mathbf{x} denote measurements obtained from a sample of glass fragments that were collected from the crime scene. Here \mathbf{x} is referred to as the control sample or source evidence upon which similar evidence obtained from a suspect is compared. Let \mathbf{y} denote measurements obtained from glass fragments obtained from a suspect. Here \mathbf{y} is referred to as the recovered sample or receptor object. As the source of the frag-

ments obtained from a suspect is less clear, or uncertain, it is assumed that all fragments obtained from a suspect's person come from the same glass item, which in some cases may not be true. Both sets of measurements on fragments \mathbf{x} and \mathbf{y} are what forms the evidence $E = (\mathbf{x}, \mathbf{y})$ under inspection. The prosecution proposition H_p would be that the fragments \mathbf{y} obtained from the suspect come from the same glass item as the control fragments \mathbf{x} . The defence proposition H_d would be that the fragments obtained from the suspect originated from some source not found at the crime scene, i.e. \mathbf{y} and \mathbf{x} are from two different glass items. The value V of the evidence can then be written as

$$V = \frac{p(\mathbf{x}, \mathbf{y} | H_p, I)}{p(\mathbf{x}, \mathbf{y} | H_d, I)}. \quad (5.1)$$

Typically (5.1) is often referred to as the likelihood ratio in forensics literature. Here, with a Bayesian approach being used to find V , the densities in (5.1) are found by integrating out the model parameters to obtain the marginal probabilities of the evidence \mathbf{x} and \mathbf{y} under both propositions, thus strictly speaking, (5.1) is a Bayes factor. For details on Bayes factors see [Kass and Raftery \(1995\)](#).

The value, V , of the evidence in (5.1) will now be derived with the elemental configurations of \mathbf{x} and \mathbf{y} in mind. Upon inspection of both the fragments of \mathbf{x} and \mathbf{y} , if their elemental configurations do not match, i.e. $\mathcal{C}_{\mathbf{x}} \neq \mathcal{C}_{\mathbf{y}}$, then it is assumed that the fragments that form \mathbf{x} and \mathbf{y} come from two different glass items. This would give a value of $V = 0$. It should be noted that, although this assumption might appear restrictive, in practice it only affects a small proportion of glass items in the database. From the database only two of the 320 glass items have at least one fragment with a different elemental

configuration than the others. From this point onwards the derivation of V assumes that the elemental configurations of \mathbf{x} and \mathbf{y} are equal, i.e. $\mathcal{C}_{\mathbf{x}} = \mathcal{C}_{\mathbf{y}} = \mathcal{C} = m$. The use types of both sets of fragments also have to be accounted for. Since the set of fragments \mathbf{x} were collected from the crime scene the source of their origin is known, therefore their use type $\mathcal{T}_{\mathbf{x}} = t$ is known. Under the prosecution proposition H_p the use types of both sets of fragments are assumed equal, i.e. $\mathcal{T}_{\mathbf{y}} = \mathcal{T}_{\mathbf{x}} = t$. However, under H_d there is uncertainty over the origin of the fragments \mathbf{y} obtained from the suspect, therefore their use type is also uncertain. Replacing the conditioning on I in (5.1) with the known use type $\mathcal{T}_{\mathbf{x}}$ of \mathbf{x} , the known elemental configuration \mathcal{C} , and the reference data set D used to assess the evidence under the two competing propositions, the value V of the evidence is

$$\begin{aligned} V &= \frac{p(\mathbf{x}, \mathbf{y} | \mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, D, H_p)}{p(\mathbf{x}, \mathbf{y} | \mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, D, H_d)} \\ &= \frac{\int p(\mathbf{x}, \mathbf{y} | \mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m, D, H_p) p(\xi_m | \mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, D, H_p) d\xi_m}{\int p(\mathbf{x}, \mathbf{y} | \mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m, D, H_d) p(\xi_m | \mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, D, H_d) d\xi_m}. \end{aligned} \quad (5.2)$$

For the numerator in (5.2) the first term in the integrand can be written as

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

This is so as the prosecution proposition H_p implies that the use type of the recovered fragments \mathbf{y} is the same as that of the control fragments \mathbf{x} , i.e. $\mathcal{T}_{\mathbf{y}} = \mathcal{T}_{\mathbf{x}} = t$. The notation $\mathcal{T}_{(\mathbf{x}, \mathbf{y})}$ is used to denote that under H_p both sets

of fragments are believed to be from the same glass item, and therefore to the same use type. The numerator of (5.2) can then be written as

$$\int p(\mathbf{x}, \mathbf{y} | \mathcal{T}_{(\mathbf{x}, \mathbf{y})} = t, \mathcal{C} = m, \xi_m) p(\xi_m | D_m) d\xi_m = E_{\xi_m | D_m} [p(\mathbf{x}, \mathbf{y} | \mathcal{T}_{(\mathbf{x}, \mathbf{y})} = t, \mathcal{C} = m, \xi_m)]. \quad (5.3)$$

Under H_d both sets of fragments \mathbf{x} and \mathbf{y} are believed to come from different glass items and are therefore considered independent. The first term in the integrand in the denominator of (5.2) can then be written as

$$\begin{aligned} p(\mathbf{x}, \mathbf{y} | \mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m, D, H_d) &= \\ p(\mathbf{x} | \mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m, D, H_d) p(\mathbf{y} | \mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m, D, H_d) &= \\ = p(\mathbf{x} | \mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m) p(\mathbf{y} | \mathcal{C} = m, \xi_m, D). \end{aligned} \quad (5.4)$$

Under H_d the use type $\mathcal{T}_{\mathbf{y}}$ of the recovered fragments \mathbf{y} is uncertain and so the second term in (5.4) 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 (4.2). The denominator of (5.2) then becomes

$$\begin{aligned} &\int p(\mathbf{x} | \mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m) \\ &\quad \left[\sum_{s=1}^T p(\mathbf{y} | \mathcal{T}_{\mathbf{y}} = s, \mathcal{C} = m, \xi_m) p(\mathcal{T}_{\mathbf{y}} = s | \mathcal{C} = m, D) \right] p(\xi_m | D_m) d\xi_m = \\ &= \sum_{s=1}^T p(\mathcal{T}_{\mathbf{y}} = s | \mathcal{C} = m, D) \\ &\quad E_{\xi_m | D_m} [p(\mathbf{x} | \mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m) p(\mathbf{y} | \mathcal{T}_{\mathbf{y}} = s, \mathcal{C} = m, \xi_m)]. \end{aligned} \quad (5.5)$$

Substituting (5.3) and (5.5) into (5.2) gives the value of the evidence V as follows:

$$V = \frac{E_{\xi_m|D_m} [p(\mathbf{x}, \mathbf{y} | \mathcal{T}_{(\mathbf{x}, \mathbf{y})} = t, \mathcal{C} = m, \xi_m)]}{\sum_{s=1}^T p(\mathcal{T}_{\mathbf{y}} = s | \mathcal{C} = m, D) E_{\xi_m|D_m} [p(\mathbf{x} | \mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m) p(\mathbf{y} | \mathcal{T}_{\mathbf{y}} = s, \mathcal{C} = m, \xi_m)]}, \quad (5.6)$$

where $p(\mathcal{T}_{\mathbf{y}} = s | \mathcal{C} = m, D)$ is given in (4.2). The density in the numerator of (5.6) is 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 across both the recovered and control fragments, \tilde{K} is the number of repeated measurements taken on each fragment, and the covariance matrix Σ_t is given in (4.7). The densities $p(\mathbf{x} | \mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m)$ for \mathbf{x} and $p(\mathbf{y} | \mathcal{T}_{\mathbf{y}} = s, \mathcal{C} = m, \xi_m)$ for \mathbf{y} in the denominator are $N_{\tilde{J}_{\mathbf{x}}\tilde{K}p}(\mathbf{1}_{\tilde{J}_{\mathbf{x}}\tilde{K}} \otimes \boldsymbol{\theta}_t, \Sigma_{t_{\mathbf{x}}})$ and $N_{\tilde{J}_{\mathbf{y}}\tilde{K}p}(\mathbf{1}_{\tilde{J}_{\mathbf{y}}\tilde{K}} \otimes \boldsymbol{\theta}_t, \Sigma_{t_{\mathbf{y}}})$ distributions, respectively, where $\Sigma_{t_{\mathbf{x}}}$ and $\Sigma_{t_{\mathbf{y}}}$ are given by formula (4.7), with \tilde{J} replaced by $\tilde{J}_{\mathbf{x}}$ and $\tilde{J}_{\mathbf{y}}$.

If the elemental configuration of the glass fragments is not taken into consideration, as is the case with the data augmentation approach, then the value of the evidence V is given as

$$V = \frac{E_{\xi|D} [p(\mathbf{x}, \mathbf{y} | \mathcal{T}_{(\mathbf{x}, \mathbf{y})} = t, \xi)]}{\sum_{s=1}^T p(\mathcal{T}_{\mathbf{y}} = s) E_{\xi|D} [p(\mathbf{x} | \mathcal{T}_{\mathbf{x}} = t, \xi) p(\mathbf{y} | \mathcal{T}_{\mathbf{y}} = s, \xi)]}. \quad (5.7)$$

See Appendix F for details. The next section will now look at different performance measures used to validate the evidential values V obtained with this method.

5.1 Evidence evaluation performance measures

The ability to obtain an evidential value V from any type of forensic evidence is not enough for a method to be incorporated into real-life casework. The most important reason for this is due to the method producing values of V in strong favour of the wrong proposition, which is referred to as *strongly misleading evidence* (Royall, 2000). This can lead to the wrong decisions being made in cases, with the worst outcome seeing an innocent person being convicted of a crime they did not commit. In order to interpret the value of V obtained and to ensure methods produce valid values that do not lead to strongly misleading evidence, validation procedures are implemented that analyse the values of V obtained from simulation studies where the ‘true’ outcomes are known.

5.1.1 Measuring performance

To measure the performance of a method, simulation studies can be undertaken to determine whether the method produces values associated with strongly misleading evidence. To do this, values of V are obtained from the method using what is known as a *validation* database consisting of forensic data that may be used as a source of evidence in real-life casework. Performance can then be assessed in terms of the percentage of false positive (FP) and false negative (FN) answers produced from the simulation study. A false positive occurs when two pieces of evidence are from different sources, but

if $V > v$ for some critical value v , then they are evaluated as coming from the same source. A false negative occurs when two pieces of evidence from the same source are evaluated as originating from different sources, i.e. when $V \leq v$. Typically FP and FN errors are produced for a critical value of $v = 1$, however the critical value v for whether a FP or FN answer is obtained may be varied to compare and contrast error rates. To obtain these error rates, each object in the validation database has a *ground-truth* label corresponding to its origin. These labels allow for the error rates to be determined when objects are compared as part of simulation studies assessing the performance of a method at computing the evidential value V . For example, if the evidence for a particular case involved measurements from glass fragments, then the method would have to have been validated using a validation database consisting of glass fragments of the same type of measurements. In this case the value of V would be obtained under two competing propositions: the prosecution proposition H_p and the defence proposition H_d . The ground-truth labels associated with fragments being compared would then indicate which of the propositions is true, thus signalling whether the method has produced an error resulting in misleading evidence.

The V values obtained from a simulation can then be examined to check the percentage of errors produced and also to see whether the method produces values of V associated with strongly misleading evidence. The easiest way to look at the values of V associated with when the prosecution H_p is true and for when H_d is true would be to display them using histograms. For example, if the evidence under investigation is on glass fragments as mentioned above then the prosecution proposition H_p being true would mean that the control

and recovered fragments come from the same glass item, while H_d being true would mean that they come from different items. Figure 5.1 shows the V values produced from an artificial simulation study. The histogram labelled H_p true displays the V values obtained for when H_p is true, thus producing a FN error if $\log_{10}(V) \leq \log_{10}(v)$; the histogram labelled H_d true shows the V values obtained for when H_d is true, thus producing a FP error when $\log_{10}(V) > \log_{10}(v)$. By changing the critical value v the percentage of FP and FN errors will change, with increasing values of v resulting in less FP errors but more FN errors, and vice versa for decreasing values of v . For example, for a threshold of $\log_{10}(v) = 0$ the percentage of false positive and false negative errors in Figure 5.1 are 11% and 9%, respectively. Any value in Figure 5.1 that produces an error means that the method has produced a value of V that has resulted in *misleading evidence*. Strongly misleading evidence in favour of the wrong proposition is represented by large values of the absolute value of $\log_{10}(V)$. However, when providing support for the correct proposition these absolute values are expected to be large. The ideal situation would result in H_p true values of V being as large as possible, and H_d true values of V being as small as possible. What is referred to as the *discriminating power* of a method can also be seen from the histograms, with the ideal outcome of large V values for when H_p is true and small values for when H_d is true resulting in there being little overlap between the two histograms, thus indicating that the method is good at discriminating between both propositions. Histograms like those in Figure 5.1 are a simple way of visualising the performance of a method, however there are also better ways of visually checking the discriminating power.

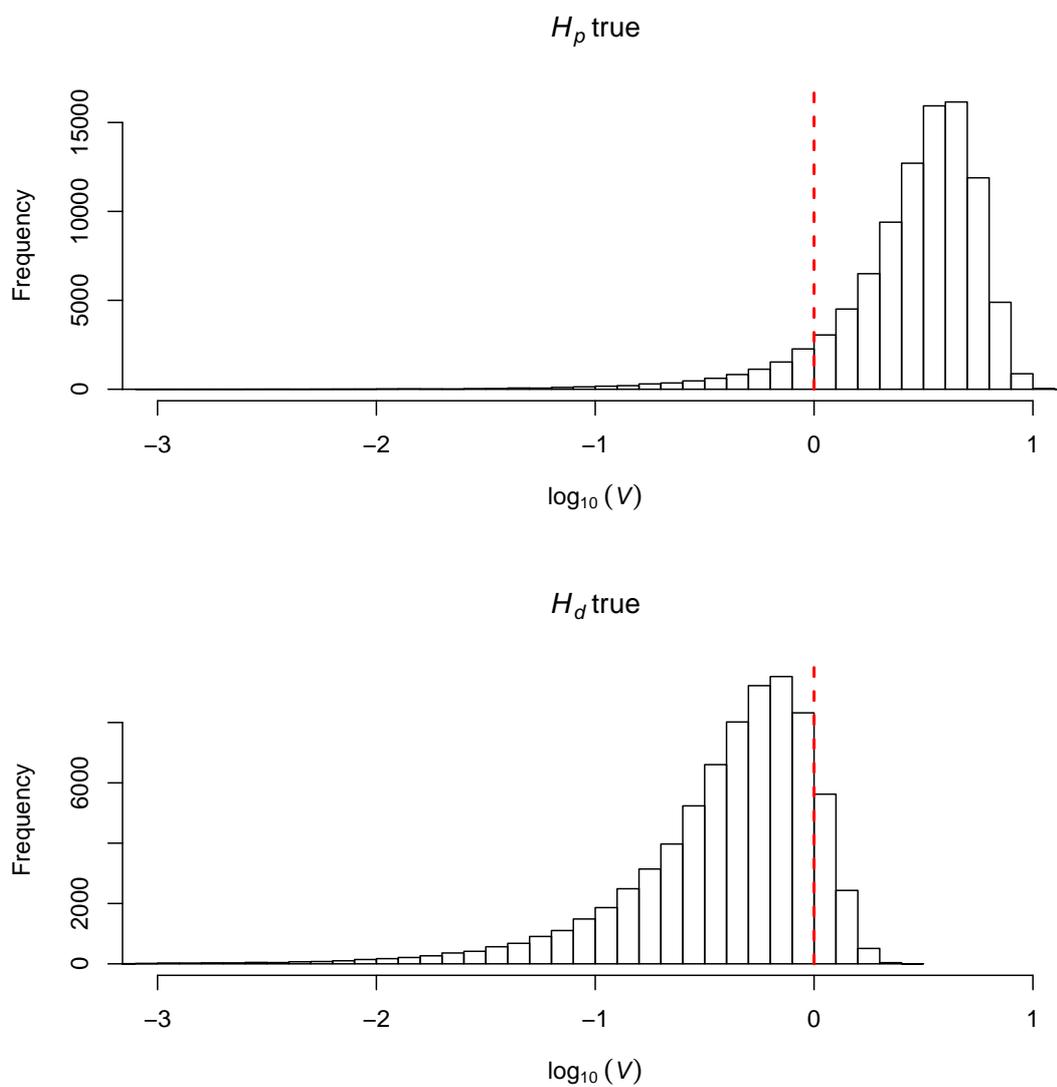


Figure 5.1: Histograms of $\log_{10}(V)$ values for when H_p is true and for when H_d is true. The dashed lines correspond to the value of $\log_{10}(V)$ for the critical value $v = 1$.

Discriminating power

The discriminating power is a measure of how well a method can differentiate between the prosecution proposition H_p and the defence proposition H_d . The discriminating power of a method is linked to the false positive and false negative error rates the method produces for varying critical values v , with $v = 1$ the most typical critical value used as it equates to the probabilities of the evidence E under the two propositions being equal, i.e. $\Pr(E|H_p, I) = \Pr(E|H_d, I)$. The discriminating power of a given set of V values is defined, see [Zadora *et al.* \(2014\)](#), by:

1. The discriminating power of two separate simulated sets of V values V_1 and V_2 is equal if for all critical values v in V_1 , there is a critical value v' in V_2 such that the FP and FN error rates are equal; and vice versa for all critical values in V_2 .
2. A set of V values V_1 is said to have better discriminating power than a different set V_2 for a specific critical value v if for that critical value in V_1 there is a critical value v' in V_2 that produces higher FP and FN error rates.

A better visual representation of the discriminating power of a method than histograms is Tippett plots, which display on the x-axis the $\log_{10}(V)$ of all of the V values obtained from a simulation against the proportion of cases where the method produces misleading evidence for when H_p and H_d are true, respectively. They show the percentage of FP and FN errors produced for the default critical value of $v = 1$, i.e. $\log_{10}(v) = 0$. [Figure 5.2](#) shows

the Tippett plots produced from the same set of simulated V values as those from the histograms in Figure 5.1.

The Tippett plots display two different curves: a curve for when H_p is true and one for when H_d is true. The curves show the proportion of times satisfying $V > v$ plotted against $\log_{10}(v)$ for when H_p and H_d are true, respectively. Under H_d the required outcome is $\log_{10}(V) \leq v$, and so the curve for when H_d is true shows the FP percentage error rate for a given critical value v . For example, Figure 5.2 shows a FP error rate of 11% for the critical value $v = 1$, i.e. $\log_{10}(v) = 0$. Under H_p the desired outcome is $\log_{10}(V) > v$, and so the FN error rate is obtained by subtracting from 100% the proportion at $\log_{10}(V) > v$ given by the curve for when H_p is true. For example, the FN error rate given in Figure 5.2 for the critical value $v = 1$ is $100\% - 91\% = 9\%$.

Misleading evidence can also be observed from Tippett plots. For when the proposition H_p is true if the proportion of $\log_{10}(V) > 0$ is smaller than 100% then these values indicate misleading evidence, i.e. values of V which are in favour of the wrong proposition, H_d , being true. Any $\log_{10}(V)$ values that are much smaller than zero then indicate strongly misleading evidence in support of the wrong proposition. When H_d is true the proportion of values greater than zero indicate misleading evidence in favour of H_p , with values much greater than zero providing strong support for the wrong proposition. The discriminating power can also be observed from Tippett plots by looking at how well the curves are vertically separated, with larger separation between the curves illustrating greater discriminating power.

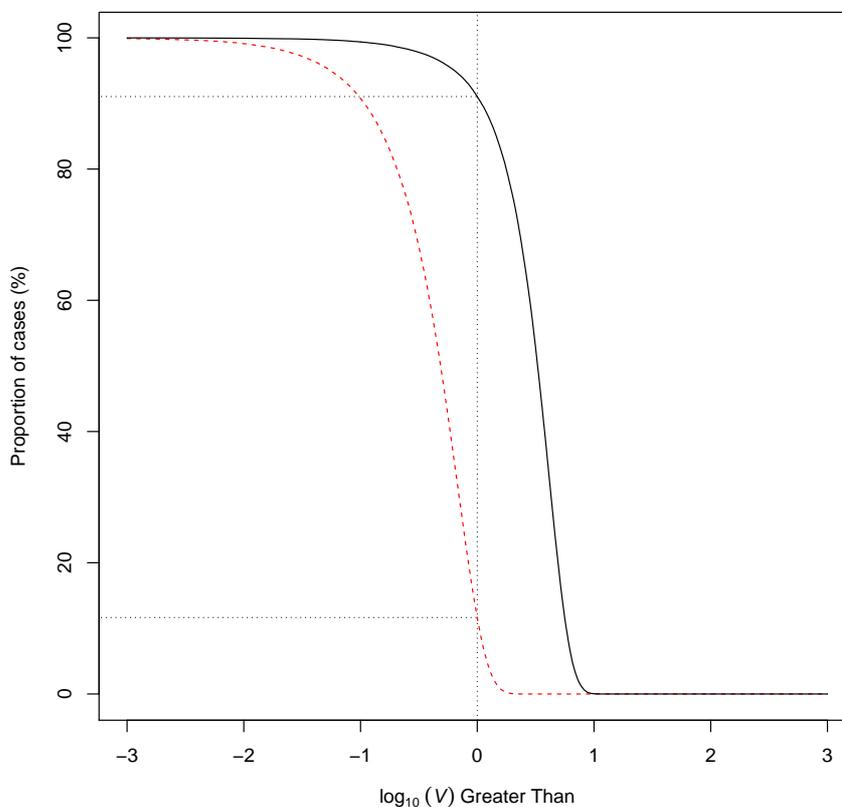


Figure 5.2: Tippet plots for a simulated set of V values. The solid curve is for when the proposition H_p is true, while the dashed curve is for when H_d is true.

As the discriminating power of a method is strongly linked to the FP and FN error rates an illustration of the discriminating power can be shown in a *detection error trade-off* (DET) plot, like the one shown in Figure 5.3. The DET curve is linked to the Tippet plots in such a way that it is essentially a plot of all of the FP and FN error rates observed from Tippet plots for varying critical values v . Figure 5.3 displays the DET curve from the same set of simulated V values with a circle indicating the same FP and FN rates previously mentioned, i.e. 11% and 9%, respectively. The closer the curve

gets to the bottom left-hand corner of the plot the stronger the discriminating power of the V values as this demonstrates small FP and FN error rates. A measure that can be observed from the DET curve arises when both the FP and FN error rates are equal. This gives what is known as the *equal error rate* (EER) (Brümmer and du Preez, 2006) and occurs for a specific critical value v , which may not be when $v = 1$.

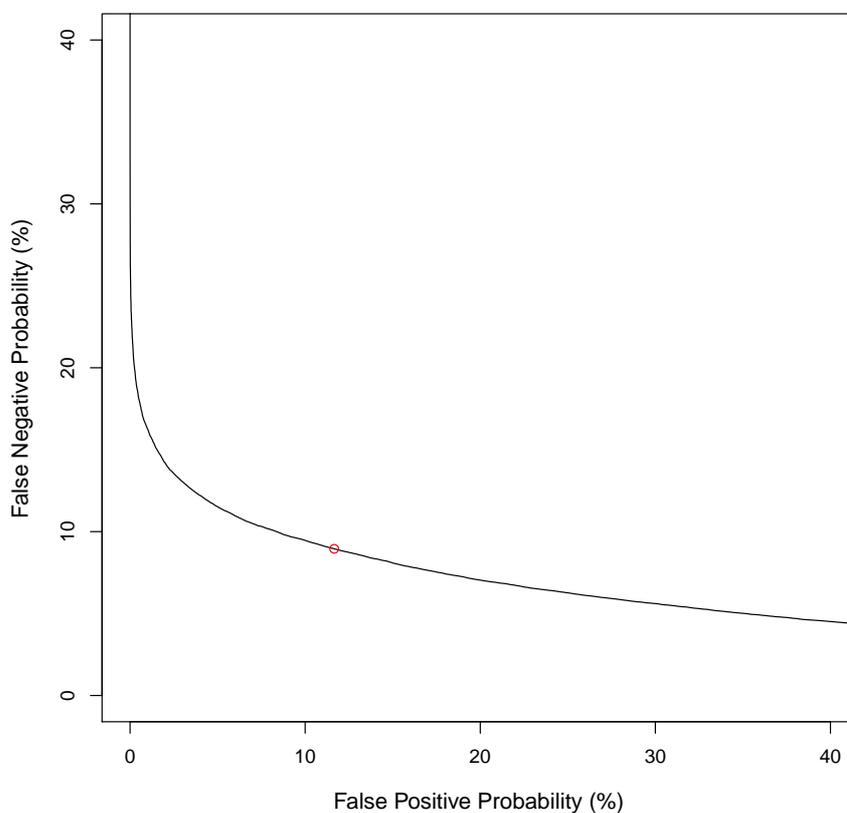


Figure 5.3: DET curve with a circle indicating the FP and FN rates for a critical value of $v = 1$.

Although the discriminating power of a set of V values is helpful in determining whether a method is capable of differentiating between H_p and H_d ,

by itself it can also be misleading about the actual size of the V values. This is shown in Figure 5.4 which displays Tippett plots and DET curves for two separate sets of simulated V values. From the definition of discriminating power and from the interpretation of both Tippett plots and DET curves the second set of V values displayed in the bottom row of Figure 5.4 have better discriminating power than the first set of values shown in the top row of Figure 5.4. This is observed from the DET curve being closer to the bottom left-hand corner of the plot, and also from the greater separation between the curves displayed in the Tippett plots. However, if the actual V values themselves are examined it can be seen from the Tippett plots that the second set of V values does not provide as much support in favour of the proposition H_p being true with a larger proportion of the V values in support of the wrong proposition H_d , and hence the possibility of strongly misleading evidence in support of H_d is much more likely to occur for this set of V values. Therefore, in order for a method to be used in real-life casework it is not enough to look at the discriminating power alone, but to look at it in conjunction with other measures of accuracy.

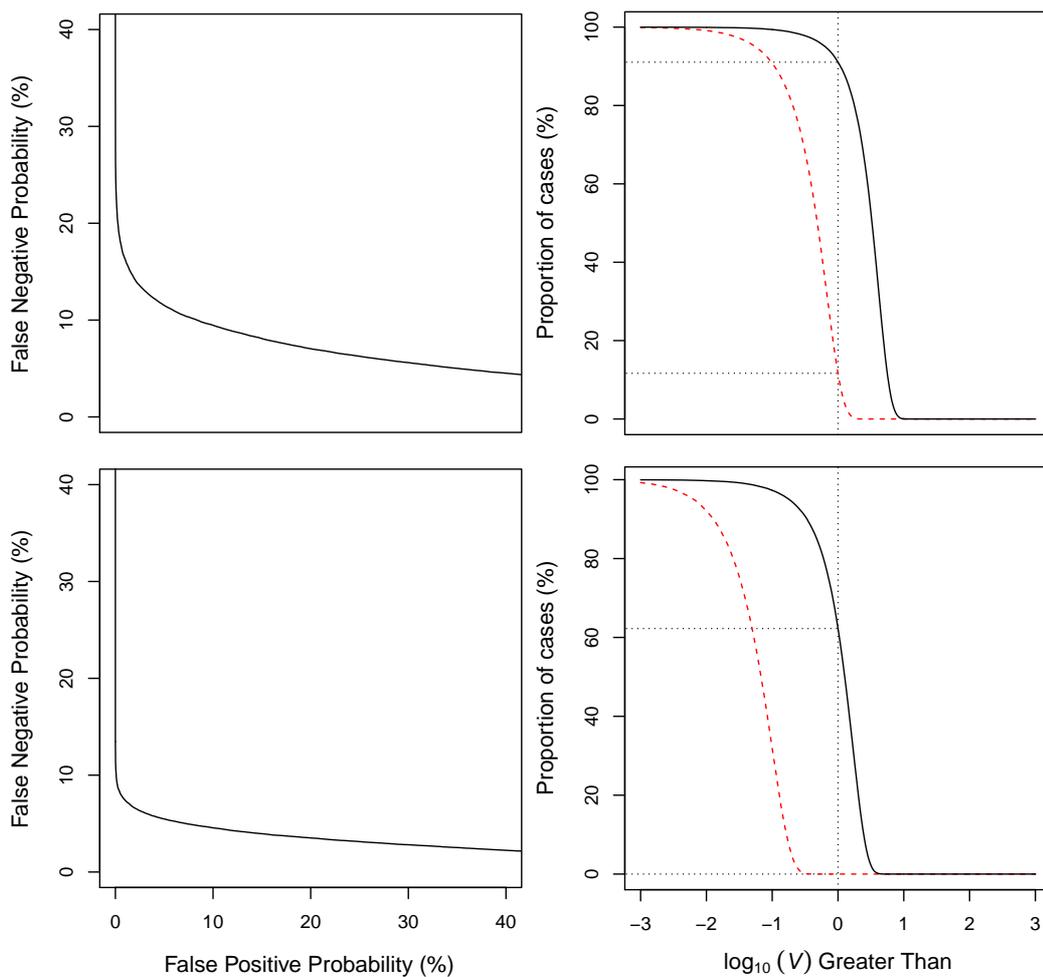


Figure 5.4: DET curves (left) and Tippett plots (right) displayed for two different sets of simulated V values. In the Tippett plots the solid curves are for when H_p is true, while the dashed curves are for when H_d is true.

Accuracy: discriminating power and calibration

To further inspect whether a method provides suitable V values that could be used in real-life casework, the accuracy of those values needs to be examined. The accuracy of a set of V values is a combination of how good the discriminating power of the V values is, along with how well calibrated they are. *Calibration* is a measure of the reliability of a set of V values obtained from a method. It relates the probability associated with an outcome with how often such an outcome occurs. If a method associates P probability to an outcome and that is equal to p , the proportion of times in which that outcome occurs, then the method has good calibration. For a method to be well calibrated the probabilities have to be between $p + \epsilon$ and $p - \epsilon$, where ϵ provides the error allowed for the method to be well calibrated. A well calibrated set of V values could have either weak or strong discriminating power. For example, if the probabilities P are not close to 0 or 1 then the discriminating power will be weak, but if the probabilities are close to 0 or 1 then the discriminating power will be strong. For more details on the calibration of V values see [Ramos and Gonzalez-Rodriguez \(2013\)](#).

To assess the accuracy of the probabilities P obtained from a method, *strictly proper scoring rules* (SPSR) ([Gneiting and Raftery, 2007](#)) can be used, which are functions of the probability assigned to an unknown outcome, as well as the actual value of that outcome. For example, the logarithmic SPSR of an event with two possible outcomes, such as propositions H_p and H_d , is defined as

$$\begin{aligned} & -\log_2(P) && \text{if } H_p \text{ is true;} \\ & -\log_2(1 - P) && \text{if } H_d \text{ is true,} \end{aligned} \tag{5.8}$$

where P is the probability associated with H_p being true. The SPSR assigns penalties linked to how much weight is given to the probabilities P when H_p is true. For example, if H_p was true and the weight of the probability P assigned to it was small, then a high penalty would be assigned to P by the SPSR, and conversely a low penalty would be assigned to a large value of P , which can be seen from Figure 5.5.

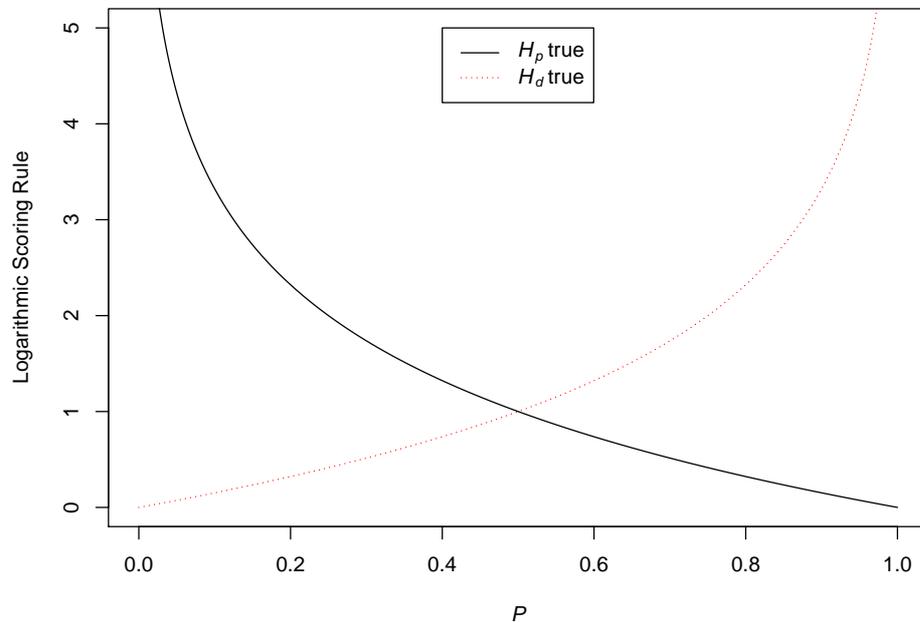


Figure 5.5: Plot of SPSR (5.8) for when H_p is true (solid line) and for when H_d is true (dashed line).

The SPSR in (5.8) is a measure of a single instance of H_p being true, but in order to assess the overall performance of say, a simulated set of V values, an overall average of all of the SPSR scores across the values of V would need

to be obtained. This is given by

$$L = -\frac{1}{N_p} \sum_{i:H_p \text{ is true}} \log_2(P_i) - \frac{1}{N_d} \sum_{j:H_d \text{ is true}} \log_2(1 - P_j), \quad (5.9)$$

where the number of times H_p is true is denoted by N_p and the probabilities by P_i ; and N_d and P_j the corresponding number of times and probabilities for when H_d is true. The overall SPSR score given by (5.9) can be interpreted as a global measure of the accuracy of the method under examination. It is similar to the Brier score (Brier, 1950) in that it is a measure of how much the probabilities deviate from the result of the true outcome. For example, if H_p is true it would have a ground-truth label indicating such, and if the probability assigned to it was $P = 1$, then this would result in a perfect match leading to no penalty. However, if H_p is true and the probability designated to it was $P = 0$, this would result in the maximum penalty as the method has assigned zero probability to an outcome which happened to occur.

Incorporating prior probabilities

The probabilities P used in obtaining SPSR scores are posterior probabilities that have been assigned to different outcomes by a method or experimenter. This means that in order to obtain the SPSR scores in a forensic setting for a set of simulated V values, the prior probabilities for H_p and H_d need to be taken into consideration. This is due to the prior probabilities being incorporated into the posterior odds as follows:

$$\begin{aligned} \frac{\Pr(H_p|E, I)}{\Pr(H_d|E, I)} &= \frac{\Pr(E|H_p, I)}{\Pr(E|H_d, I)} \times \frac{\Pr(H_p|I)}{\Pr(H_d|I)} \\ &= V \times O(H_p|I), \end{aligned} \quad (5.10)$$

where $O(H_p|I) = \frac{\Pr(H_p|I)}{\Pr(H_d|I)}$ is the prior odds in favour of the prosecution proposition H_p . A measure of the accuracy is then obtained across different prior odds values as the forensic examiner is only interested in acquiring the evidential value of the evidence V and has no input into the prior probabilities set for H_p and H_d . However, for a method to be suitable for real-life casework its accuracy is measured over a wide range of possible prior probabilities. This is done using a measure known as *empirical cross-entropy* (ECE), which is the same as the SPSR in (5.9), but with the prior probabilities incorporated to obtain a weighted average. This is given, see Brümmer (2010), as

$$\begin{aligned} ECE = & -\frac{\Pr(H_p|I)}{N_p} \sum_{i:H_p \text{ is true}} \log_2 \Pr(H_p|E_i, I) \\ & -\frac{\Pr(H_d|I)}{N_d} \sum_{j:H_d \text{ is true}} \log_2 \Pr(H_d|E_j, I), \end{aligned} \quad (5.11)$$

where E_i denotes the evidence used to obtain the N_p simulated V values for when H_p is true, and E_j the evidence used to obtain the N_d simulated V values for when H_d is true. As a validation set contains the evidential values V of the evidence, the ECE in (5.11) can be expressed in terms of V and the prior odds $O(H_p|I)$:

$$\begin{aligned} ECE = & \frac{\Pr(H_p|I)}{N_p} \sum_{i:H_p \text{ is true}} \log_2 \left(1 + \frac{1}{V_i \times O(H_p|I)} \right) \\ & + \frac{\Pr(H_d|I)}{N_d} \sum_{j:H_d \text{ is true}} \log_2(1 + V_j \times O(H_p|I)). \end{aligned} \quad (5.12)$$

Figure 5.6 displays an example of an ECE plot for a set of simulated V values where the ECE is calculated across a wide range of prior odds values, as seen on the x-axis. The solid curve displays the ECE from (5.12) and illustrates the accuracy of a set of simulated V values. The dashed curve represents perfect calibration of the validation set of V values, hence showing

the curve for a set of optimised V values (Brümmer, 2010). The dotted curve is referred to as the neutral reference and shows the performance that would be obtained from a set of V values that are all equal to one. This is the opposite of the calibrated curve in that it displays what would be considered the poorest performance, thus the solid curve should be lower than it for all prior odds values. The closer the solid and dashed curves are to one another the better the calibration of the V values. The *comparison* package in R (R Development Core Team, 2011) by David Lucy includes functions that can produce ECE plots.

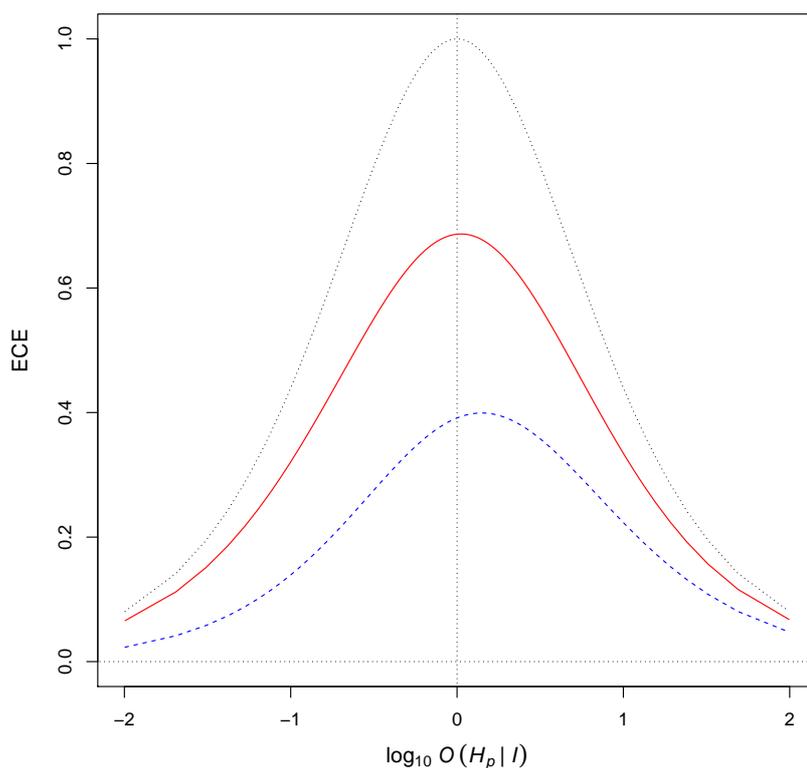


Figure 5.6: ECE plot of a set of simulated V values. The solid curve represents the accuracy; the dashed curve the calibrated accuracy; and the dotted curve the neutral reference.

5.2 Evidence evaluation simulation study

The performance of the hierarchical model using both the composite model and data augmentation approaches in evaluating glass fragments as evidence was assessed by performing a simulation study. As with glass classification, five-fold cross-validation was used to obtain a simulated set of V values and to estimate the percentage of false positive and false negative answers. The five test sets used were randomly allocated 64 glass items each. The percentage of false negative answers was obtained by randomly choosing two fragments from each glass item to be the source evidence, \mathbf{x} , which are then compared with the remaining two fragments from the same glass item as the recovered evidence, \mathbf{y} . This gives 64 comparisons per test set for a total of $N_p = 320$ same-source comparisons. The percentage of false positive answers was obtained by taking all 12 measurements from each glass item in a test set to be the source evidence, \mathbf{x} , and all 12 measurements from another glass item in the test set as the recovered evidence, \mathbf{y} . This gives a total of $\binom{64}{2} = 2016$ item pairs per test set, thus giving a total of $N_d = 5 \times \binom{64}{2} = 10,080$ different-source comparisons. As many more different-source comparisons were made, the uncertainty surrounding the false negative rates is greater than that for the false positive rates. The value of the evidence V is computed using formulae (5.6) and (5.7) for the composite model and data augmentation approaches, respectively.

Due to the way the data is separated into different subsets in the composite model through identifying the presence/absence of the chemical elements iron and potassium, two sets of fragments are assumed to have not originated

from the same glass item if their elemental configurations do not match, i.e. $\mathcal{C}_x \neq \mathcal{C}_y$, returning $V = 0$. However, setting $V = 0$ can be seen as providing too much support in favour of the defence proposition H_d . It also causes problems in computing the ECE, as seen from (5.12). In order to restrict potentially overwhelming support in favour of one proposition over the other, a threshold, v_T , can be placed upon the values of V obtained from the simulation study. The bound is given, see Royall (2000), by

$$\Pr(V \geq v_T | H_d) \leq \frac{1}{v_T}. \quad (5.13)$$

This bounds the probability of wrong support in favour of the prosecution proposition H_p being true, when in fact the defence proposition H_d is true, by the reciprocal of the threshold v_T . Thresholds have been used to aid the reporting of the evidential value V of evidence by using different scales. The quantitative value of V was interpreted on a verbal scale introduced by Evett *et al.* (2000) and is displayed in Table 5.1; see also Chapter 3, page 107 of Aitken and Lucy (2004). Table 5.2 shows the scale introduced by the Swedish National Laboratory of Forensic Science (SKL) which assigns positive or negative values to the value of V based upon the range of values in which it lies; see Nordgaard *et al.* (2012) for details.

Table 5.1: Verbal interpretation of the value of the evidence V in support of H_p against H_d (Evett *et al.*, 2000).

1	$< V \leq$	10	Limited evidence to support
10	$< V \leq$	100	Moderate evidence to support
100	$< V \leq$	1000	Moderately strong evidence to support
1000	$< V \leq$	10000	Strong evidence to support
10000	$< V$		Very strong evidence to support

Table 5.2: The Swedish National Laboratory of Forensic Science (SKL) scale for interpreting the evidential value V (Nordgaard *et al.*, 2012).

+4:	$10^6 \leq V$
+3:	$6000 \leq V < 10^6$
+2:	$100 \leq V < 6000$
+1:	$6 \leq V < 100$
0:	$1/6 < V < 6$
-1:	$1/100 < V \leq 1/6$
-2:	$1/6000 < V \leq 1/100$
-3:	$1/10^6 < V \leq 1/6000$
-4:	$V \leq 1/10^6$

5.2.1 Evidence evaluation simulation results

Histograms of the simulated sets of V values from the composite model and data augmentation approaches can be seen in figures 5.7 and 5.8, respectively. The false negative and false positive error rates from each are given in Table 5.3 for $v = 1$. These results are an improvement on previous publications using a similar glass database; see Neocleous *et al.* (2011). Table 5.3 also includes the error rates obtained from the Bayesian hierarchical model without data augmentation being implemented, with slightly higher false negative and false positive rates being obtained.

The two types of errors obtained have different levels of seriousness associated with them in terms of forensics. More emphasis is placed on reducing the false positive error rate as it is believed to be worse to wrongly imprison an innocent person than it is to let someone who is guilty go free. Therefore, if two sets of fragments are incorrectly evaluated as having originated from the same glass item, then this one piece of evidence may contribute to the

Table 5.3: Percentage of false negative (FN) and false positive (FP) answers produced from the simulation study for $v = 1$.

Approach	Error rate	
	FN	FP
Composite model	4.4%	1.4%
Data augmentation	4.4%	1.5%
BHM without DA	5.0%	1.6%

conviction of a suspect who is innocent. By varying the critical value v corresponding to the error rates, a different set of false negative and false positive percentage error rates can be obtained, and are displayed in receiver operating characteristic (ROC) curves in figures 5.9 and 5.10. In the ROC curves the true positive (TP) rate (TP rate = $1 - \text{FN rate}$) is plotted against the false positive rate for varying values of v . An ideal outcome produces an area under the curve (AUC) value of 1, with the ROC curve very steep with zero false positive rate and a true positive rate of 1. Here the AUC is 0.99 for both approaches with the curves very steep in the region of false positive rates close to zero. To improve upon the false positive rates of 1.4% and 1.5% obtained for a critical value of $v = 1$ would come at the cost of an increased false negative rate, with false positive rates smaller than 1% resulting in false negative rates of around 10%. An equal error rate (EER) of 2.8% was obtained for the composite model for a critical value of $v = 0.007$, while the same EER was found for a critical value of $v = 0.001$ for the Bayesian hierarchical model using data augmentation. With importance placed on reducing the amount of false positive errors produced, both model approaches are able to maintain false positive rates smaller than 3% for critical values

much smaller than 1.

By placing the threshold v_T from (5.13) upon the reported V values we can restrict potentially wrong support in favour of the wrong proposition. Figures 5.11 and 5.12 display Tippett plots, and figures 5.13 and 5.14 DET curves for four different thresholds v_T placed upon the simulated V values. These thresholds had no affect on the FP and FN error rates obtained in Table 5.3, but from the Tippett plots it can be seen that lowering the threshold v_T reduces the amount of favour given to either proposition, and thus reduces wrong support whenever misleading evidence is produced. From the Tippett plots it appears that the V values associated with the data augmentation approach have slightly greater discriminating power with larger values seen in support of the prosecution proposition H_p , and similar values for when H_d is true. However, as will be seen from the ECE plots in figures 5.15 and 5.16 they are not as well calibrated as the values obtained from the composite model. From the DET curves, for a given FN error rate, the composite model has lower FP error rate than the data augmentation approach for FP errors between 0% and 2%, and also a lower FN error rate for FP errors greater than 4%. Again, as was the case for the ROC curves in figures 5.9 and 5.10, it can also be seen from the DET curves that reducing the FP error rate through increasing the critical value v , comes at the cost of increasing the FN error rate.

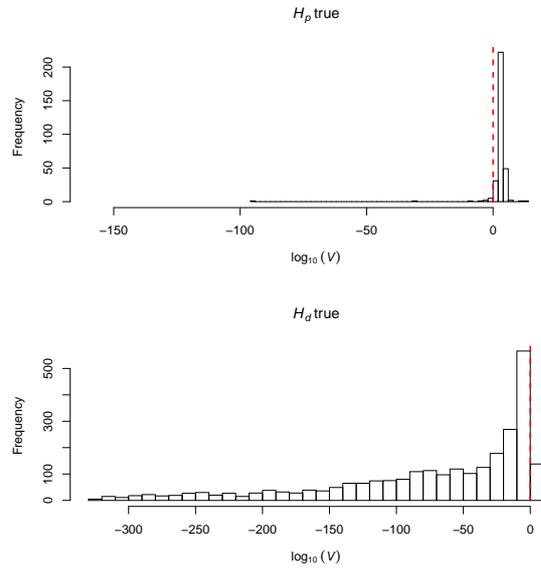


Figure 5.7: Histograms of $\log_{10}(V)$ values for when H_p is true and for when H_d is true obtained from the composite model. The dashed lines correspond to the value of $\log_{10}(V)$ for the critical value $v = 1$.

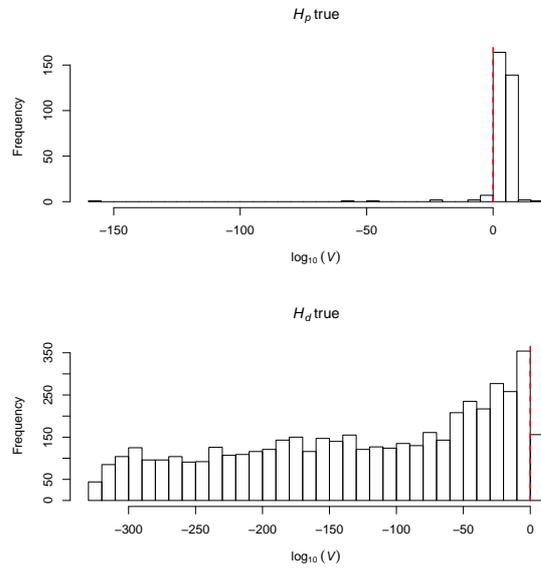


Figure 5.8: Histograms of $\log_{10}(V)$ values for when H_p is true and for when H_d is true obtained from data augmentation. The dashed lines correspond to the value of $\log_{10}(V)$ for the critical value $v = 1$.

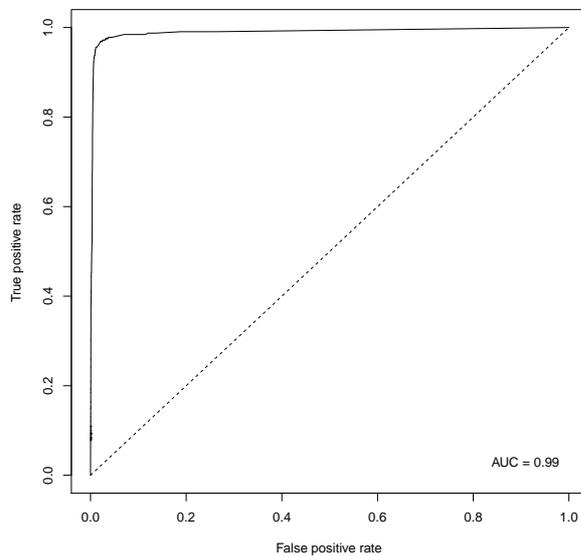


Figure 5.9: ROC curve for different values of v obtained from the composite model.

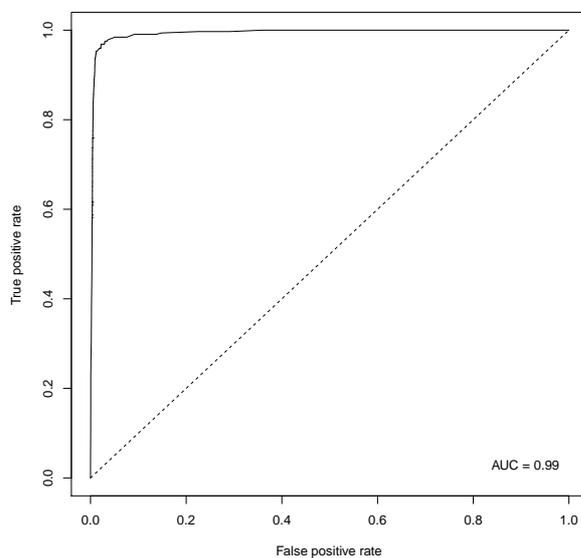


Figure 5.10: ROC curve for different values of v obtained from data augmentation.

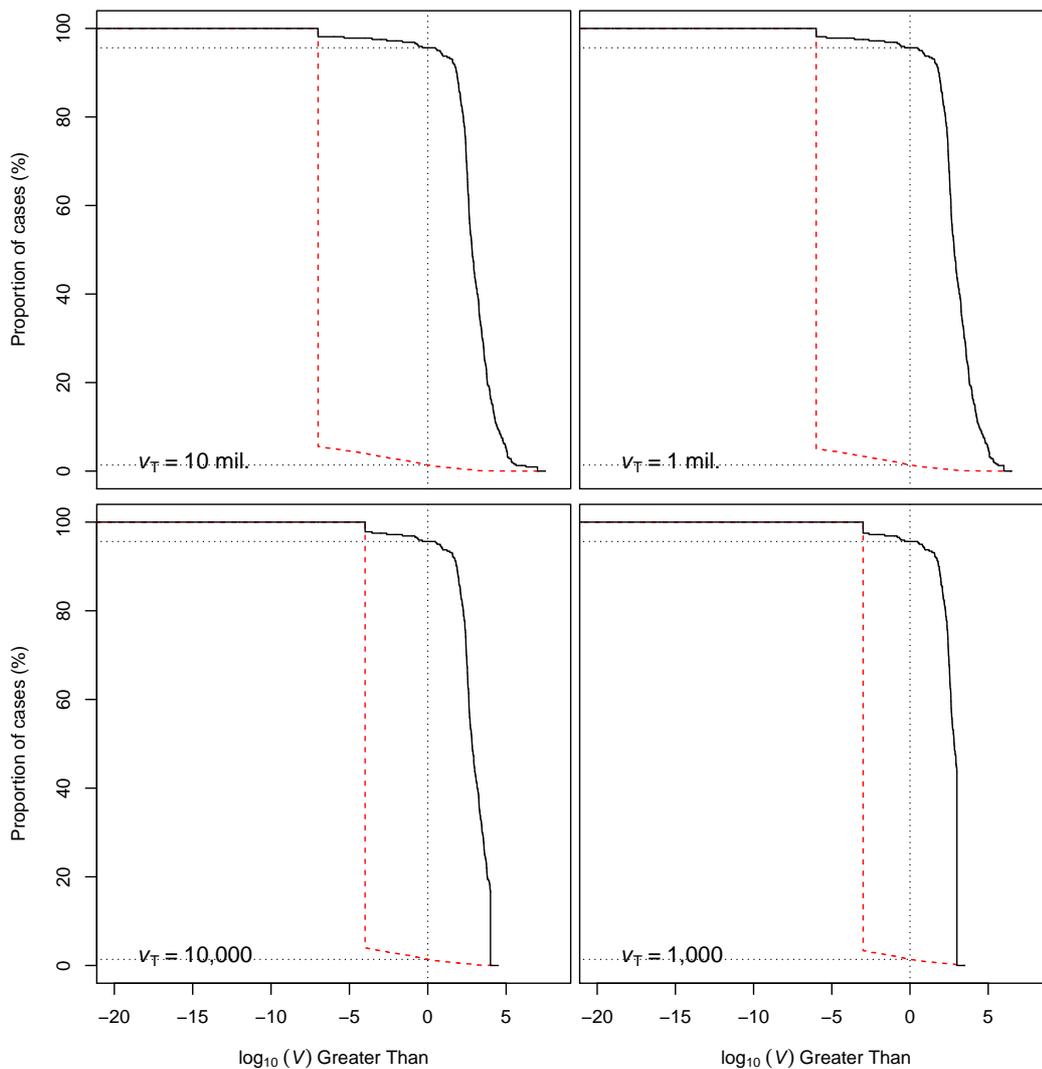


Figure 5.11: Tippett plots obtained from the simulation study with different thresholds v_T on V from the composite model. The solid curves are for when H_p is true, while the dashed curves are for when H_d is true.

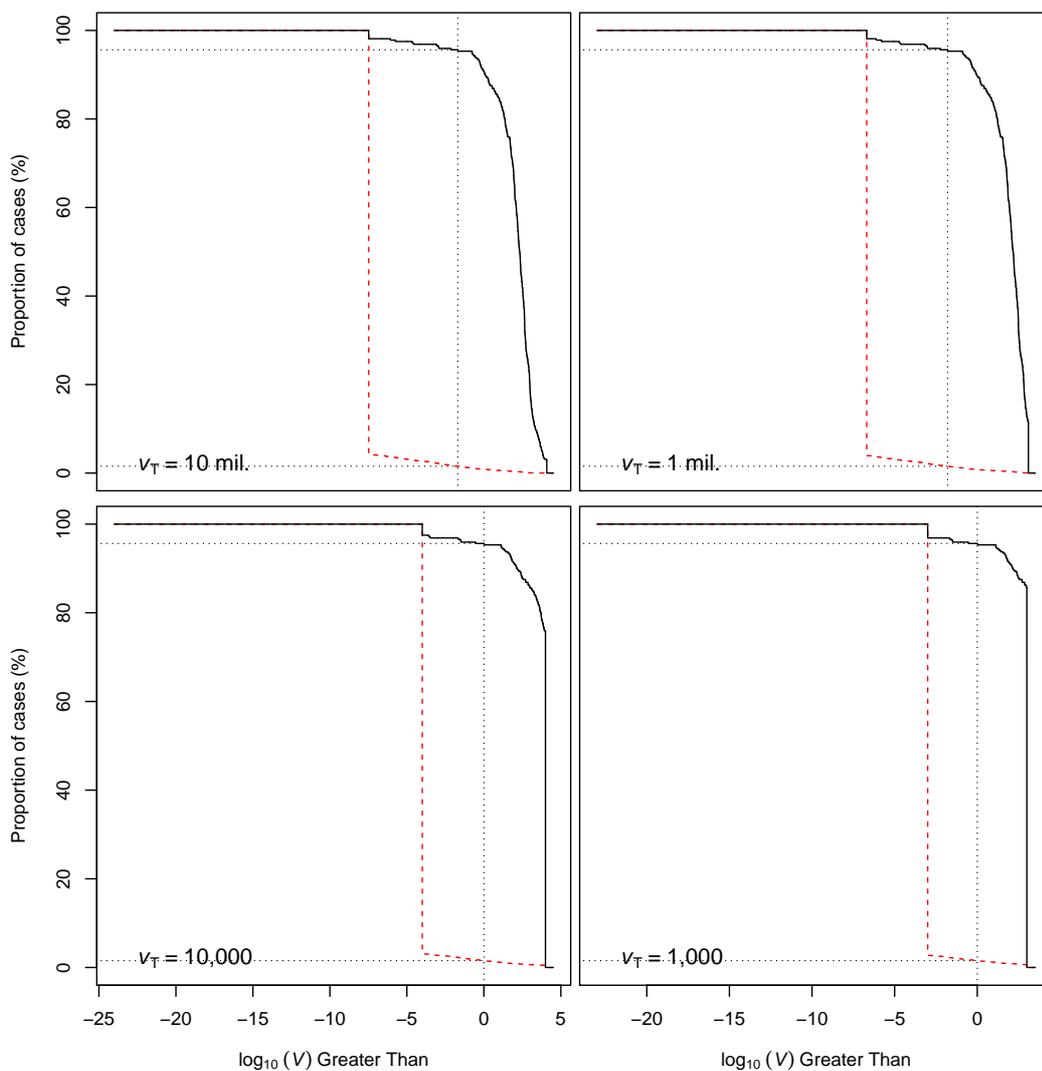


Figure 5.12: Tippett plots obtained from the simulation study with different thresholds v_T on V from data augmentation. The solid curves are for when H_p is true, while the dashed curves are for when H_d is true.

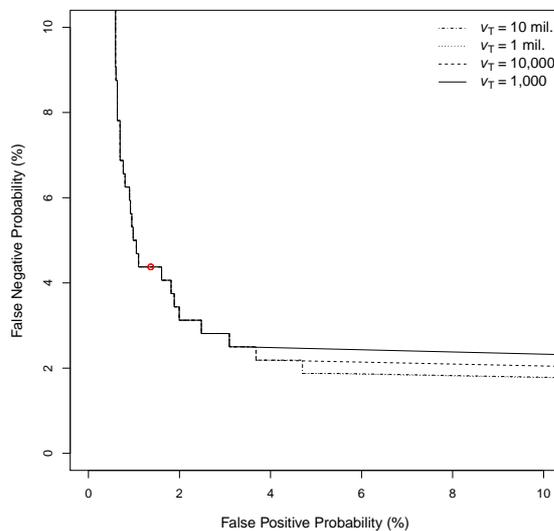


Figure 5.13: DET curves from the simulation study with different thresholds v_T on V from the composite model. The circles indicate the FP and FN rates for a critical value of $v = 1$.

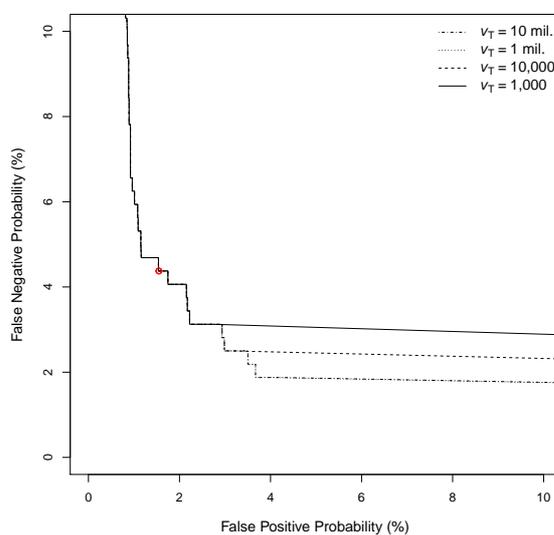


Figure 5.14: DET curves from the simulation study with different thresholds v_T on V from data augmentation. The circles indicate the FP and FN rates for a critical value of $v = 1$.

Figures 5.15 and 5.16 are ECE plots obtained from the simulated V values from the composite model and data augmentation approaches, respectively, for four different thresholds v_T on V . The calibration of the values obtained from each approach are very good for prior odds $O(H_p|I) \leq 1$ but deteriorate for increasing prior odds values in favour of the prosecution proposition H_p , i.e. as $O(H_p|I)$ tends towards 100 ($\log_{10} O(H_p|I) = 2$). The composite model appears to have better calibrated values than those of the data augmentation approach with there being less separation between the solid and dashed curves, especially for low prior odds values. The deterioration in the ECE is improved by reducing the threshold v_T placed upon the V values. However, for large prior odds values in favour of H_p both models do not perform as well as the neutral reference curve, with the composite model showing better improvement than the data augmentation approach. This may lead to potential reservations about the models, but for prior odds values not largely in favour of the prosecution proposition H_p , it performs much better than the neutral reference. Performance for large prior odds values is less questionable for a threshold value of $v_T = 1000$, thus this would be the recommended threshold value for this type of evidence. A threshold value of $v_T = 1000$ would correspond to this type of evidence giving, at most, moderately strong evidence in support of H_p (see Table 5.1), while using larger thresholds may potentially produce strong support in favour of the wrong proposition.

The next chapter will provide a walkthrough of an online application of the model used to classify glass fragments and to obtain their evidential value.

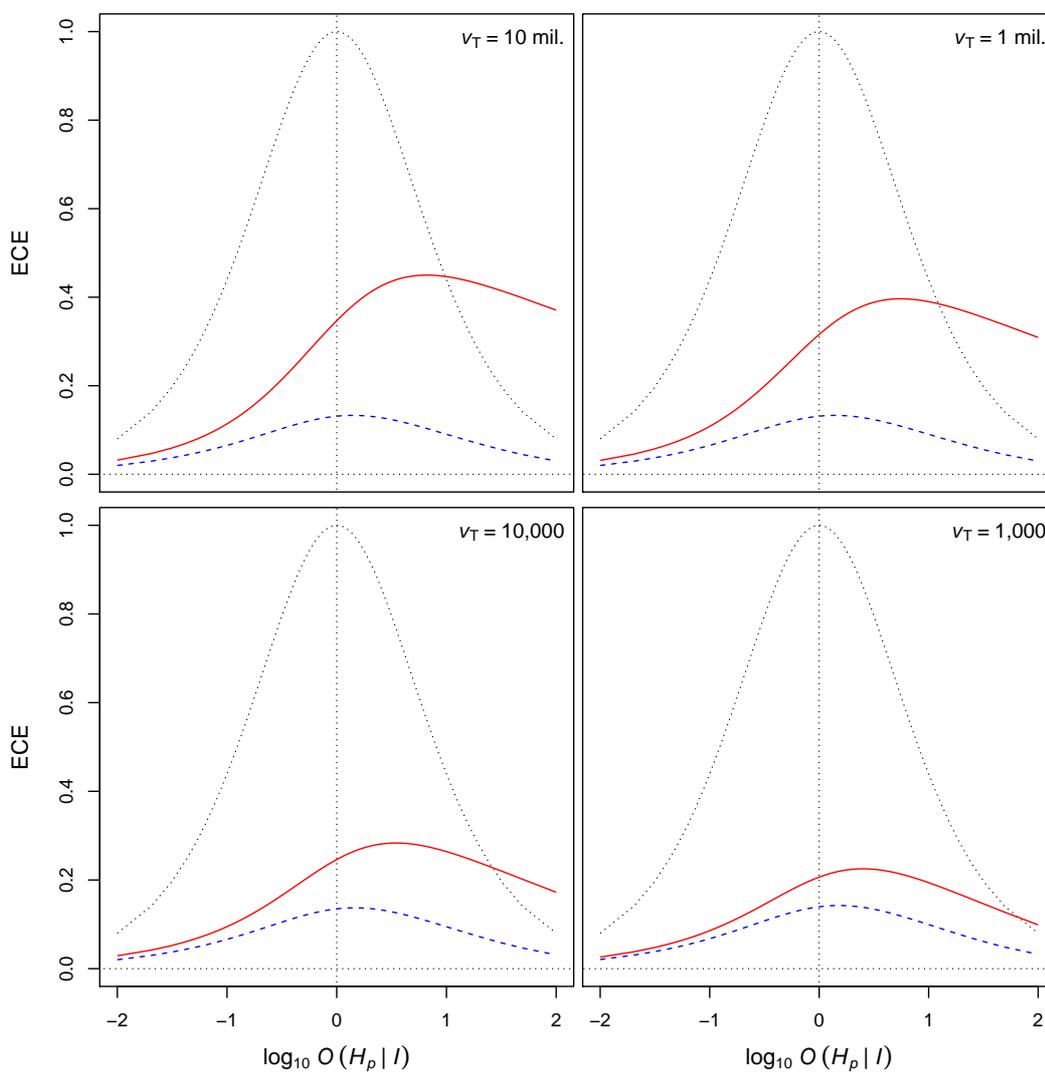


Figure 5.15: ECE plots of the simulation study with different thresholds v_T on V from the composite model. The solid curves represent the accuracy; the dashed curves the calibrated accuracy; and the dotted curves the neutral reference.

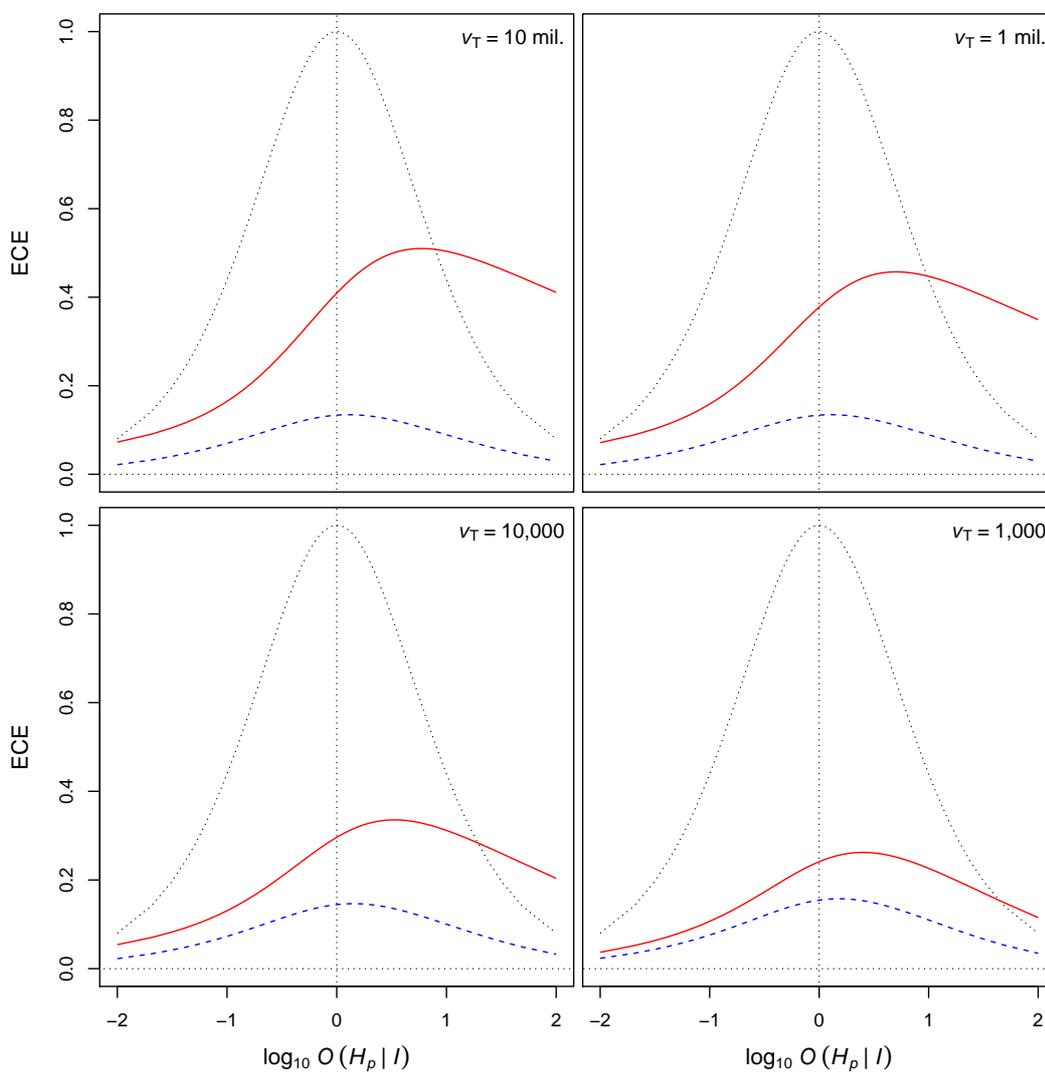


Figure 5.16: ECE plots of the simulation study with different thresholds v_T on V from data augmentation. The solid curves represent the accuracy; the dashed curves the calibrated accuracy; and the dotted curves the neutral reference.

Chapter 6

Web application

This short chapter contains a description of an online application of the classification and evidence evaluation procedures of Chapters 4 and 5. The application was created using the *shiny* package as part of the statistical programming language R (R Development Core Team, 2011), containing in the background, the R code written to obtain the classification and evidence evaluation results. The web application is accessible from the following web address: <http://gnapier.shinyapps.io/GlassClassificationAndEvaluation/>.

To make use of the application a user only requires data on the elemental composition of glass fragments, as described in Chapter 2, contained in a *.txt* or *.csv* file. Each file containing the data should contain nine columns: the first column indicates which fragment the measurements belong to, as well as the number of replicate measurements associated with that fragment. The remaining eight columns contain the measurements for the chemical elements oxygen (O), sodium (Na), magnesium (Mg), aluminium (Al), silicon (Si),

potassium (K), calcium (Ca) and iron (Fe). Table 6.1 gives an example *.txt* or *.csv* file containing measurements from a glass item with two fragments, with each fragment having two replicate measurements. The default web application page also instructs the user on how to appropriately load the data as seen in Figure 6.1. The web application can perform two separate tasks given measurements on the elemental content of glass fragments: (i) classify glass fragments by use type and (ii) compute the evidential value of two sets of fragments. First the classification part of the application will be detailed.

Table 6.1: Example of how a file containing data to be analysed should be loaded into the application.

fragment	O	Na	Mg	Al	Si	K	Ca	Fe
1	46.5	11.2	2.2	1.7	32.3	2.5	3.3	0.3
1	46.7	11.2	2.1	1.5	32.7	2.4	3.2	0.2
2	46.6	11.3	2.4	1.4	32.5	2.3	3.1	0.4
2	46.8	11.1	2.3	1.3	32.8	2.2	3.3	0.2

Classification and evaluation of glass fragments

This application is for the classification and evaluation of glass fragments as evidence, where the measurements have been obtained by SEM-EDX analysis. The procedures can be selected using the tabs below, where upon selection a description of the procedure will appear.

Any file uploaded for analysis should contain nine columns. The first column should contain numbers indicating the fragments, which are repeated to indicate the number of repeated measurements on each fragment. The percentage weights (wt%) of the chemical elements are stored in columns 2-9 as follows: oxygen (O), sodium (Na), magnesium (Mg), aluminium (Al), silicon (Si), potassium (K), calcium (Ca) and iron (Fe). An example data file is displayed in the panel below. Note that the column names do not need to match those of the example file, but please ensure that the order of the columns is the same. Once a file has been uploaded the example data file below will be replaced by the apps perception of the loaded data file. Note that for parity with the data used to produce the app, any data uploaded is taken to two decimal places.

Classification

Evidence evaluation

The classification procedure classifies glass fragments into one of five use type categories: bulb, car window, headlamp, container and building window. Glass fragments are classified into these use type categories by obtaining posterior probabilities corresponding to each use type, with the fragments being classified as the use type with the largest posterior probability.

Enter a .txt/.csv file containing measurements from the glass fragments that are to be classified:

Browse...

No file selected.

Example data file

```
fragment 0 Na Mg Al Si K Ca Fe
1 47.5 10.2 2.0 1.5 33.5 1.0 3.3 0.0
1 47.7 10.2 1.8 1.5 33.7 0.8 3.1 0.2
2 47.6 10.1 2.1 1.3 33.4 1.2 2.0 0.3
2 47.5 10.0 2.2 1.2 33.8 1.0 2.3 0.0
```

Glass classification result

NULL

Figure 6.1: Snapshot of the default application web page.

6.1 Classification

As the composite model outperformed the data augmentation approach in the classification simulation study, as seen in Section 4.1.1 of Chapter 4, the web application employs the composite model approach of Section 3.2.3 instead of the data augmentation approach. This means that the formula used in order to compute the classification probabilities of the glass data measurements, \mathbf{y} , loaded into the application is formula (4.8) from Chapter 4 which was found to be

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

In order to compute the classification probabilities, the application already uses the MCMC draws of ξ_m from the composite model approach that were obtained from the reference database D , which means that the expectation term in the formula above can be computed. The use type priors are assumed uniform with $p(\mathcal{T}_{\mathbf{y}} = t) = 1/T$ for all t , the α values are all equal to 0.1, with values for the second expression on the right-hand side of the formula above obtained from the reference database and were given in Table 4.1.

To perform the classification of glass fragments using the web application, the user first has to select the classification task by clicking on the corresponding tab (default). The application can only classify glass fragments into one of the five use types contained in the reference database analysed in this thesis: bulb, car window, headlamp, container and building window. Figure 6.1 shows where the data file containing the glass fragments to be classified should be uploaded into the application. Once a file has been uploaded into the application it will take a few seconds to obtain the classification results.

The classification results consist of the five posterior probabilities linking the glass fragments with the five use type categories mentioned above, including which use type the fragments have been classified as. The application classifies glass fragments to the use type with the largest posterior probability. Figure 6.2 shows an example of the classification results output. The uploaded data file is also displayed alongside the classification results so that the user can easily check it was uploaded correctly. It should be noted that if a user wants to classify more than one set of fragments at a time, then they will need to upload separate files containing the corresponding measurements.

Classification and evaluation of glass fragments

This application is for the classification and evaluation of glass fragments as evidence, where the measurements have been obtained by SEM-EDX analysis. The procedures can be selected using the tabs below, where upon selection a description of the procedure will appear.

Any file uploaded for analysis should contain nine columns. The first column should contain numbers indicating the fragments, which are repeated to indicate the number of repeated measurements on each fragment. The percentage weights (wt%) of the chemical elements are stored in columns 2-9 as follows: oxygen (O), sodium (Na), magnesium (Mg), aluminium (Al), silicon (Si), potassium (K), calcium (Ca) and iron (Fe). An example data file is displayed in the panel below. Note that the column names do not need to match those of the example file, but please ensure that the order of the columns is the same. Once a file has been uploaded the example data file below will be replaced by the apps perception of the loaded data file. Note that for parity with the data used to produce the app, any data uploaded is taken to two decimal places.

Classification
Evidence evaluation

The classification procedure classifies glass fragments into one of five use type categories: bulb, car window, headlamp, container and building window. Glass fragments are classified into these use type categories by obtaining posterior probabilities corresponding to each use type, with the fragments being classified as the use type with the largest posterior probability.

Enter a .txt/.csv file containing measurements from the glass fragments that are to be classified:

Loaded data file

fragment	0	Na	Mg	Al	Si	K	Ca	Fe
1	47.71	10.41	1.62	1.62	34.00	0.94	3.70	0
2	47.73	10.27	1.61	1.57	34.23	0.91	3.68	0
3	47.69	10.51	1.68	1.58	33.88	0.98	3.68	0
4	48.96	11.41	1.65	1.50	32.05	0.82	3.61	0
5	48.79	11.32	1.72	1.54	32.20	0.85	3.58	0
6	48.44	11.29	1.66	1.57	32.60	0.85	3.59	0
7	46.12	10.45	1.65	1.55	35.05	0.96	4.22	0
8	45.51	10.32	1.54	1.55	35.64	1.00	4.44	0
9	44.76	9.79	1.51	1.54	36.68	1.15	4.57	0
10	48.39	10.85	1.66	1.54	32.95	0.90	3.71	0
11	47.86	11.12	1.67	1.53	33.20	0.89	3.73	0
12	48.46	10.99	1.67	1.53	32.85	0.84	3.66	0

Glass classification result

Classification posterior probabilities:

Bulb	0.975	0.000	0.025	0.000	0.000	0.000
Car window	0.000	0.000	0.000	0.000	0.000	0.000
Headlamp	0.000	0.000	0.000	0.000	0.000	0.000
Container	0.000	0.000	0.000	0.000	0.000	0.000
Building window	0.000	0.000	0.000	0.000	0.000	0.000

[1] "Fragments classified as use type: Bulb"

Figure 6.2: Snapshot of the application displaying classification results consisting of the five use type posterior probabilities and the use type the fragments have been classified as.

6.2 Evidence evaluation

As with the classification procedure, the evidence evaluation procedure of the application uses the composite model approach to evaluating glass fragments given in Chapter 5. That is to say that the evidence $E = (\mathbf{x}, \mathbf{y})$ is evaluated under two competing propositions: the prosecution proposition H_p and the defence proposition H_d . Here H_p would be that both sets of fragments \mathbf{x} and \mathbf{y} come from the same item; and H_d would be that \mathbf{x} and \mathbf{y} come from different items. The set of fragments \mathbf{x} corresponds to the control fragments obtained from a crime scene, with \mathbf{y} the fragments obtained from a suspect. The evidential value of E in (5.1) was then

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

and upon evaluation using the composite model approach was shown to be (5.6):

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

As with classification, the necessary MCMC posterior draws of ξ_m are already stored in the application. To obtain the evidential value of glass fragments, the user must first select the evidence evaluation task by clicking on the corresponding tab as shown in Figure 6.3. The application will then change and display a brief description of the procedure and show where the two files containing the control and recovered measurements should be uploaded. As the control measurements, \mathbf{x} , were obtained from the crime scene, their use type is known and so the application provides a drop-down menu for the user

to select the use type of the control measurements. Again, as with classification there are five use types for the user to choose from as the reference data used to obtain the posterior draws consisted of these five use types. There is no option to select the use type of the recovered measurements as they were obtained from the suspect and so there is uncertainty surrounding their origin, which is why the denominator in the value of V above is a weighted average across use types.

When the two files containing the measurements from the control and recovered samples have been uploaded into the application, it will compute and return the evidential value, V , of the evidence. It also returns the verbal scale of [Evetts *et al.* \(2000\)](#) that was shown in [Table 5.1](#), and the scale used by the Swedish National Laboratory of Forensic Science (SKL) ([Nordgaard *et al.*, 2012](#)) as reference points for the value obtained. The application will also notify the user if the elemental configurations of the control and recovered fragments do not match, thus under the assumption that the control and recovered fragments are therefore from different glass items, will return the evidential value $V = 0$, as shown in [Figure 6.5](#).

Classification and evaluation of glass fragments

This application is for the classification and evaluation of glass fragments as evidence, where the measurements have been obtained by SEM-EDX analysis. The procedures can be selected using the tabs below, where upon selection a description of the procedure will appear.

Any file uploaded for analysis should contain nine columns. The first column should contain numbers indicating the fragments, which are repeated to indicate the number of repeated measurements on each fragment. The percentage weights (wt%) of the chemical elements are stored in columns 2-9 as follows: oxygen (O), sodium (Na), magnesium (Mg), aluminium (Al), silicon (Si), potassium (K), calcium (Ca) and iron (Fe). An example data file is displayed in the panel below. Note that the column names do not need to match those of the example file, but please ensure that the order of the columns is the same. Once a file has been uploaded the example data file below will be replaced by the apps perception of the loaded data file. Note that for parity with the data used to produce the app, any data uploaded is taken to two decimal places.

Classification
Evidence evaluation

The evidence evaluation procedure evaluates the value, V , of two sets of glass fragments. The first set of glass fragments are obtained from the crime scene (control sample) and the second set of glass fragments are found on the suspect (recovered sample). The glass fragments obtained from the suspect are used as evidence in support (or against) the proposition that the suspect was at the scene of the crime. Values of $V > 1$ provide support for the prosecution, while $V < 1$ provides support for the defence.

As the glass fragments found at the crime scene are of known origin, their use type is also known. Select the use type of the glass fragments found at the crime scene:

Bulb
▼

Enter a .txt/.csv file containing measurements found at crime scene:

Browse...
No file selected.

Since the origin of the glass fragments found on the suspect is questionable, the use type of these fragments is uncertain, with the calculation of V taking this into consideration.

Enter a .txt/.csv file containing measurements found on the suspect:

Browse...
No file selected.

Example data file

Fragment	O	Na	Mg	Al	Si	K	Ca	Fe
1	47.5	10.2	2.0	1.5	33.5	1.0	3.3	0.0
1	47.7	10.2	1.8	1.5	33.7	0.8	3.1	0.2
2	47.6	10.1	2.1	1.3	33.4	1.2	2.0	0.3
2	47.5	10.0	2.2	1.2	33.8	1.0	2.3	0.0

Evidence evaluation result

NULL

Figure 6.3: Snapshot of the evidence evaluation application screen showing were the two files containing the control and recovered fragments should be uploaded.

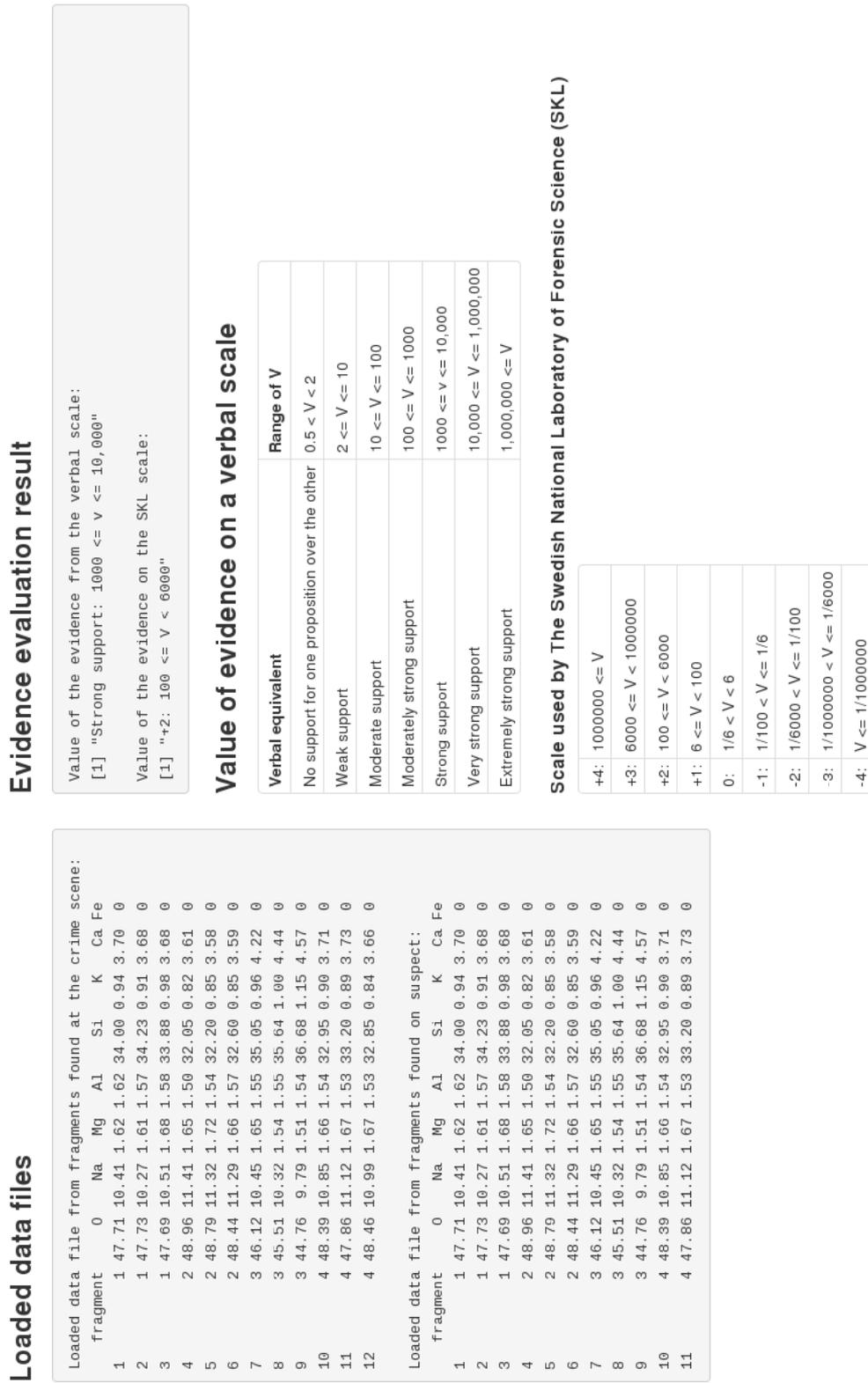


Figure 6.4: Snapshot of the evidence evaluation results screen displaying the evidence evaluation result obtained along with two reference scales for V .

Loaded data files

Loaded data file from fragments found at the crime scene:

Fragment	0	Na	Mg	Al	Si	K	Ca	Fe
1	1	47.71	10.41	1.62	1.62	34.00	0.94	3.70
2	1	47.73	10.27	1.61	1.57	34.23	0.91	3.68
3	1	47.69	10.51	1.68	1.58	33.88	0.98	3.68
4	2	48.96	11.41	1.65	1.50	32.05	0.82	3.61
5	2	48.79	11.32	1.72	1.54	32.20	0.85	3.58
6	2	48.44	11.29	1.66	1.57	32.60	0.85	3.59
7	3	46.12	10.45	1.65	1.55	35.05	0.96	4.22
8	3	45.51	10.32	1.54	1.55	35.64	1.00	4.44
9	3	44.76	9.79	1.51	1.54	36.68	1.15	4.57
10	4	48.39	10.85	1.66	1.54	32.95	0.90	3.71
11	4	47.86	11.12	1.67	1.53	33.20	0.89	3.73
12	4	48.46	10.99	1.67	1.53	32.85	0.84	3.66

Loaded data file from fragments found on suspect:

Fragment	0	Na	Mg	Al	Si	K	Ca	Fe
1	1	50.80	9.67	2.27	0.30	31.64	0	5.32
2	1	51.49	9.81	2.26	0.30	31.01	0	5.13
3	1	52.85	10.12	2.32	0.25	29.80	0	4.65
4	2	49.62	9.49	2.21	0.25	32.78	0	5.65
5	2	49.62	9.50	2.27	0.21	32.79	0	5.61
6	2	49.55	9.50	2.28	0.23	32.76	0	5.68
7	3	52.98	10.00	2.29	0.24	29.81	0	4.68
8	3	52.55	9.79	2.25	0.22	30.25	0	4.94
9	3	52.71	9.82	2.25	0.31	30.07	0	4.84
10	4	50.88	9.65	2.15	0.28	31.81	0	5.23
11	4	50.67	9.68	2.26	0.24	31.85	0	5.30
12	4	49.73	9.50	2.15	0.21	32.84	0	5.57

Evidence evaluation result

Elemental configurations of both sets of fragments do not match.
 Elemental configuration of fragments obtained at crime scene:2
 Elemental configuration of fragments obtained from suspect: 4
 Elemental configurations
 1: Fe and K present
 2: Fe absent, K present
 3: Fe present, K absent
 4: Fe and K absent
 The two sets of fragments are assumed to come from different objects. Note that this assumption may not always hold as it is possible for two sets of fragments from the same object to have non-matching elemental configurations, although this occurs very rarely.
 V = 0.001

Value of evidence on a verbal scale

Verbal equivalent	Range of V
No support for one proposition over the other	0.5 < V < 2
Weak support	2 <= V <= 10
Moderate support	10 <= V <= 100
Moderately strong support	100 <= V <= 1000
Strong support	1000 <= V <= 10,000
Very strong support	10,000 <= V <= 1,000,000
Extremely strong support	1,000,000 <= V

Figure 6.5: Snapshot of the evidence evaluation results screen when the configurations of the control and recovered fragments do not match.

Chapter 7

Discussion and conclusions

Two different modelling approaches were presented in this thesis to deal with a large point mass at zero for various components of the elemental compositions of glass fragments. The main objectives were in classifying newly observed glass fragments into use type categories, and to compute the evidential value of glass fragments relating to two competing propositions about their source.

The data analysed was obtained from SEM-EDX analysis ([Zadora, 2009](#)) which returns the percentage weights (wt%) - to two decimal places - of the eight main elements comprising the composition of a glass fragment. Due to the compositional nature of the data, constraints on the data such as each component being non-negative and the sum of all components summing to 100% needed to be taken into consideration, with [Chapter 2](#) providing a review of methods used to analyse such data. Typically logarithmic transformations are used to handle the constraints on compositional data with

the most commonly used transformation being that introduced by [Aitchison \(1982\)](#); the additive log-ratio (alr). Other transformations applied to compositional data include the centred log-ratio (clr) and multiplicative log-ratio (mlr) ([Aitchison, 1982](#)); and more recently the isometric log-ratio (ilr) has been adopted by researchers of compositional data. However, all of these transformations mentioned involve logarithmic terms meaning that they cannot be directly applied to compositional datasets containing zero components, which are a frequent occurrence in such data. In this thesis the square root of the ratio of one component part to the remaining ones was applied to the compositions of glass fragments and was found to improve variance stability and normality in the data over the alr transformation, as seen in [Section 2.3](#). Another important reason why square root transformed compositions were used instead of a logarithmic transformation was due to the abundance of compositional zeros found in the glass database. These compositional zeros would have to be altered in some way in order to apply a logarithmic transformation to them. Also, the presence of many compositional zeros, alongside very small concentrations, can have a much greater influence on the distribution of the data on a log scale. Regardless of which transformation is applied to the database, the presence of many zeros was still a major factor to be taken into consideration when modelling the data.

[Chapter 3](#) introduced the statistical models presented for modelling the glass database containing a large concentration of zeros. Previous models applied to similar datasets have been from a frequentist perspective and incorporated two levels of variation; see [Aitken and Lucy \(2004\)](#) and [Neocleous *et al.* \(2011\)](#). In this thesis a Bayesian hierarchical model was introduced

that takes into account three levels of variation: between-item, within-item and measurement error. In addition to the extra level of variability, the model also takes into account the use type of the glass items to aid in the classification procedure. Two different ways of handling the compositional zeros are then incorporated into the Bayesian hierarchical model. First, the simplest way to handle compositional zeros would be to add a small constant value to them treating them as rounded zeros, i.e. non-zero traces below limits of detection. More sophisticated methods of replacing zeros with non-zero values have been proposed as seen in Section 2.2. The first approach implemented into the Bayesian hierarchical model treats the zeros as such and is referred to in Section 3.2.1 as the data augmentation approach. This approach essentially updates all of the compositional zeros in the glass database with non-zero values that are below the limits of detection of the measuring equipment. It does this as part of the MCMC sampler by using the current estimates of the model parameters to update the zeros from a truncated normal distribution. Updating the zeros with different non-zero values reduces the artificial correlation that would otherwise be introduced; see Palarea-Albaladejo *et al.* (2007). The results obtained when using data augmentation in both the classification and evidence evaluation tasks are also improved over leaving the compositional zeros unaltered; see tables 4.14 and 5.3. However, due to the detection limit of the measuring equipment being very small (0.005) updating the zeros using data augmentation does not necessarily alleviate the problems pertaining to a large point mass at zero. This was shown in Section 3.3.1 when applying the alr transformation to the glass database, with the zeros replaced by small concentrations still having a strong influence on the distribution of the data. These problems

can be alleviated by the second approach to modelling compositional zeros and is referred to as the composite model approach.

The composite model - see Section 3.2.3 - treats the compositional zeros as being essential zeros, i.e. concentrations considered truly zero. The composite model procedure consists of (i) partitioning the glass database into distinct subsets characterised by the same pattern of the presence or absence of chemical elements in the composition of a glass item and (ii) fitting a separate Bayesian hierarchical model to the square root transformed compositions corresponding to glass items comprising each of the distinct subsets of the database. The glass database consisted of 10 distinct presence/absence configurations as was shown in Table 2.2; however as very few glass items were associated with some of the configurations the choice to only observe the presence or absence of the chemical elements iron (Fe) and potassium (K) was made. Table 2.3 provides the results of only observing the presence/absence of these two elements with the original ten configurations being reduced to four. The elements iron and potassium were chosen as they are responsible for 87.9% of zeros in the glass database as shown in Table 2.1. Splitting the database up into different presence/absence configurations helps to reduce the influence the zeros have on the distribution of the data. Improvements in both the classification and evidence evaluation tasks can also be found; see tables 4.14 and 5.3. The composite model approach is not without its faults however. The choice of only observing the presence/absence of two chemical elements may be seen as being ad-hoc, but due to there being so few items associated with some of the configurations a compromise had to be made. Ideally, given enough data the composite model would use all possible

elemental configurations, and in turn be more likely to produce even better results. Also, since some of the configurations that were combined contain a mixture of glass items of certain use types (bulbs and headlamps) where some items do and others do not contain an element with the presence/absence being unobserved, problems still persist when using the alr transformation as seen in Section 3.3.2. The Bayesian hierarchical model also assumes normality, which although improved by modelling the presence/absence of elements, may be questionable. Future work involving the use of mixture models would perhaps improve upon the satisfactory results obtained by the Bayesian hierarchical model.

The main uses of the Bayesian hierarchical model are to correctly classify glass fragments into use type categories and to evaluate the strength of glass fragments as evidence under two competing propositions. First, the composite model approach outperformed the data augmentation approach in classifying glass items into use type categories (misclassification rates: 20.6% vs 21.9%). Both approaches to handling the zeros in the glass database outperformed the Bayesian hierarchical model with the compositional zeros left untouched (22.8% misclassification rate); and also outperformed the support vector machines classification method; see Table 4.14. Classification results for glass items of use types bulb, headlamp and container were very good, with the relatively high overall misclassification rate due to difficulties in distinguishing between car and building windows. This is due to both window types being manufactured in a similar manner and having very similar elemental compositions. However, whenever a window is misclassified it is most often misclassified as the other window type. Improvements in the classifi-

cation rates of car and building windows can be obtained when, in addition to using the elemental compositions, the refractive index before and after annealing is also used (Zadora, 2009). Although a misclassification rate of 20.6% seems high, the composite model performs very well in the classification task given the data available.

In the evidence evaluation task the composite model again outperformed the data augmentation approach, which in turn was an improvement over the Bayesian hierarchical model with no treatment given to the compositional zeros (see Table 5.3). The false positive and false negative rates obtained using cross-validation for the composite model were 1.4% and 4.4% respectively, which are improvements over previous publications using similar glass databases; see Neocleous *et al.* (2011). One possible point of concern when using the composite model is in the assumption that if the elemental configuration of the control fragments obtained from the crime scene differs from the elemental configuration of the fragments obtained from a suspect, then both sets of fragments are assumed to have originated from different sources, thus setting the value of the evidence $V = 0$. This assumption was made to ensure that the dimensions of the densities in the numerator and denominator of the Bayes factor used to compute V were the same. However, it is possible for two sets of fragments from the same source to have different elemental configurations, but for the glass database this was very rare. Table 2.4 highlighted that only eight out of the 320 glass items have a chemical element with its 12 measurements not all positive or all zero. Of those eight glass items only two have different within-item configurations due to an element being present in one fragment but absent from the other three

fragments, which is less than 1% of all items in the glass database. When performance was assessed as part of the evidence evaluation study in Section 5.2.1, a threshold value of $v_T = 1000$ was found to be appropriate for this type of evidence. This was found after calibration of the simulated V values, with larger threshold values potentially producing support in favour of the wrong proposition for large prior odds values in favour of H_p (see Figure 5.15). This restricts the support for this type of evidence to being, at most, moderately strong in favour of H_p against H_d (see Table 5.1). However, this also helps to prevent potentially strong support being given in favour of the wrong proposition.

The composite model described in this thesis and also the classification and evidence evaluation procedures associated with it have been implemented into an online web application. The web application can be found at the following web address: <http://gnapier.shinyapps.io/GlassClassificationAndEvaluation/>. The application is easily accessible and usable from any device with a web browser with the user only needing data on the elemental compositions of glass fragments contained in a *.txt* or *.csv* file. If a user wants to classify glass fragments then the application will return the probabilities associated with those fragments being one of five use types. The application can only classify fragments into five different use type categories as the glass database used to obtain the necessary results from the model consisted of these five use types. Given two sets of measurements the application can also return the evidential value of those fragments.

This thesis presented a comprehensive statistical modelling approach for elemental composition data relating to glass fragments, with the purpose of clas-

sifying glass fragments into use type categories, and evaluating its strength as forensic evidence. For this type of glass measurements the models proposed produced results which outperform existing methods in both the classification and evidence evaluation tasks. Assessment of the performance in the evidence evaluation task led to the recommendation to restrict the strength of the evidence reported for this type of evidence. This should help reduce potentially wrong support being provided to the wrong proposition. These contributions are easily accessible from a web application. The web application is quick and easy to use by forensic scientists who have to deal with such data in real-life casework.

Appendix A

Elemental configuration scatterplots

Figures [A.1](#), [A.2](#) and [A.3](#) are scatterplots containing the item means for all items in the reference database that have elemental configuration $m = 1, 3, 4$, respectively.

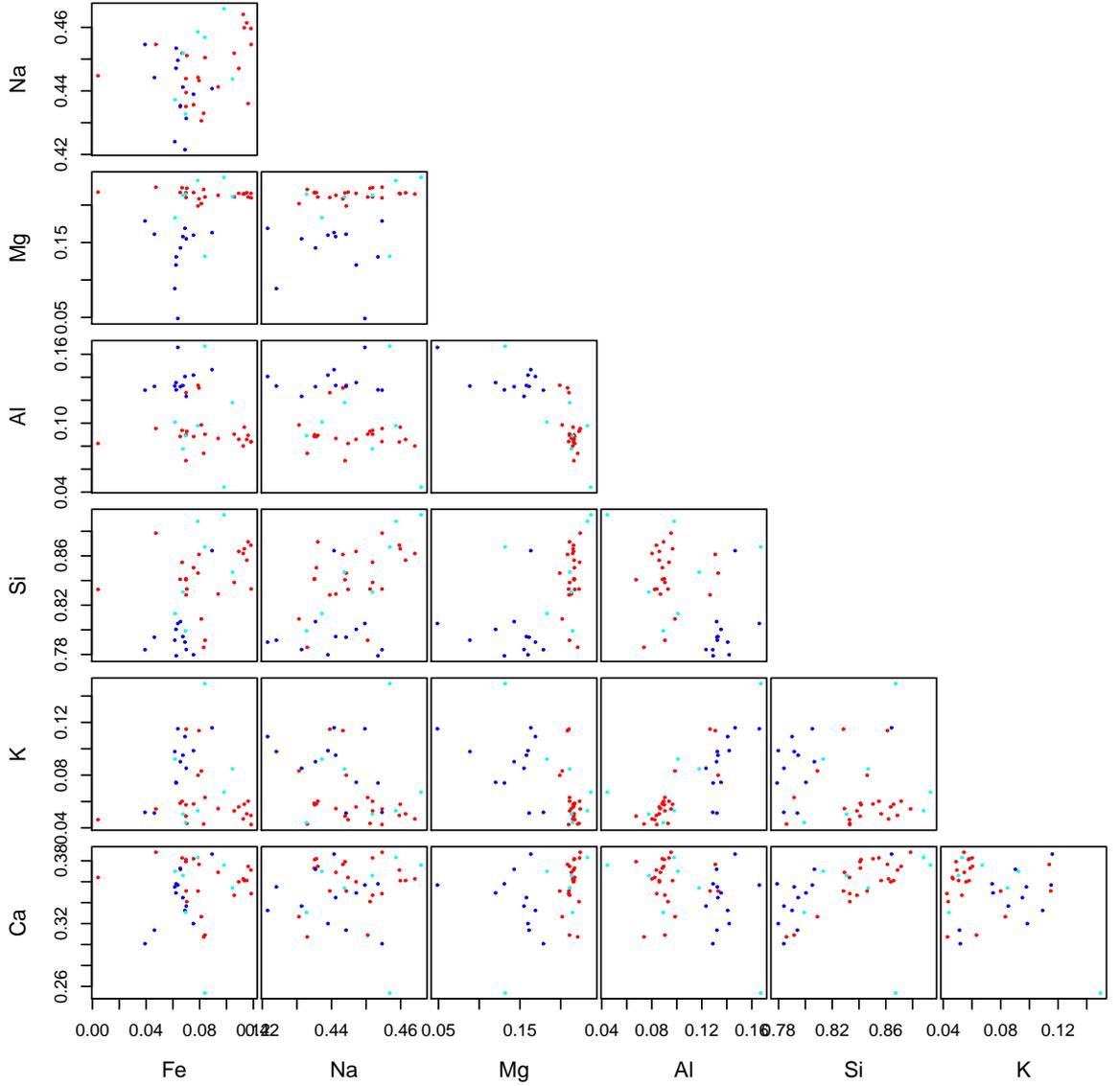


Figure A.1: Scatterplots of the square root transformed ratios of the item means for items with configuration 1 from Table 2.3. The different coloured points correspond to the use type categories: **car window**, **container** and **building window**.

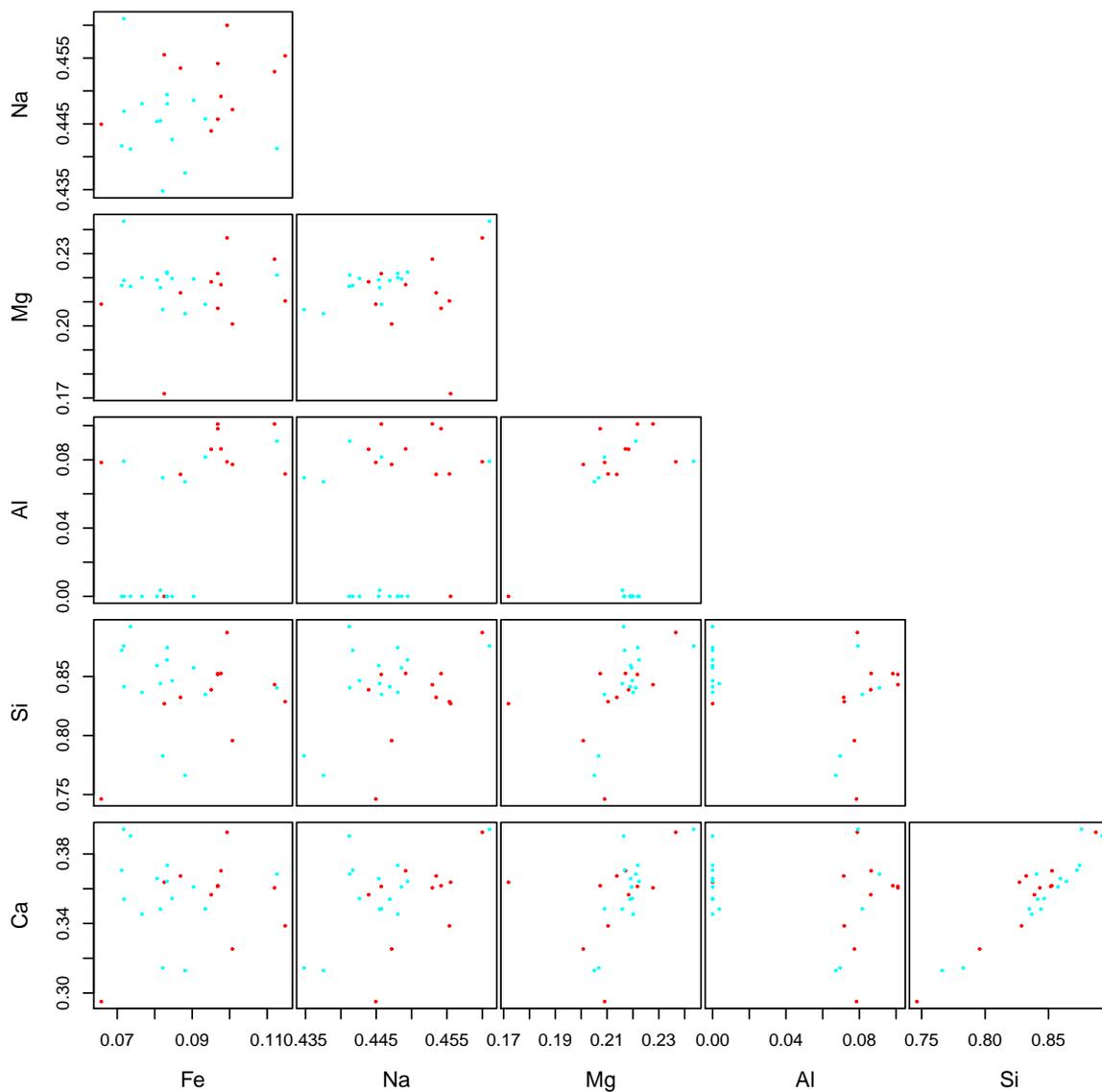


Figure A.2: Scatterplots of the square root transformed ratios of the item means for items with configuration 3 from Table 2.3. The different coloured points correspond to the use type categories: **car window** and **building window**.

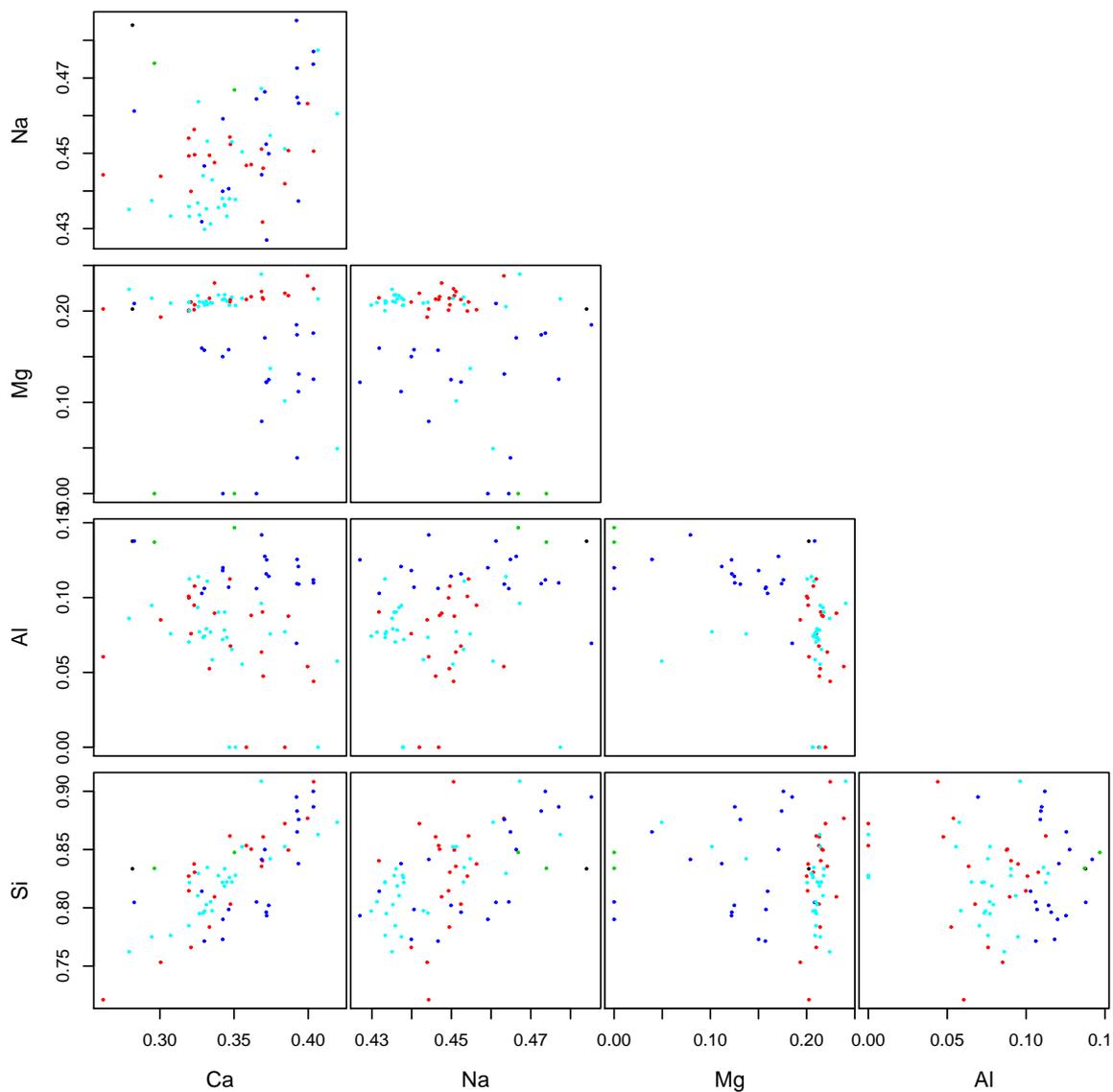


Figure A.3: Scatterplots of the square root transformed ratios of the item means for items with configuration 4 from Table 2.3. The different coloured points correspond to the five use type categories: bulb, car window, headlamp, container and building window.

Appendix B

MCMC

B.1 Full conditional distributions

All of the full conditional distributions of each of the unknown quantities in the model are reported below. The use of “ $|\dots$ ” means “conditionally on all the other variables”.

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

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

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

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

- $\mathbf{c}_{tij} | \dots \sim N_p(\tilde{\boldsymbol{\psi}}_{tij}, \tilde{\Psi}^{-1}),$

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

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

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

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

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

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

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

B.2 M-H moves interval widths

The interval widths for Metropolis-Hastings moves M-H 2 and M-H 3 from Section 3.1.1 are given in Table B.1 below.

Table B.1: The δ_{tl} 's used for the interval widths in Metropolis-Hastings moves M-H 2 and M-H 3.

δ_{tl}	l						
t	1:Na	2: Mg	3:Al	4:Si	5:K	6:Ca	7:Fe
1:bulb	0.0100	0.0080	0.0030	0.0060	0.0100	0.0120	0.0020
2:car window	0.0016	0.0040	0.0060	0.0060	0.0030	0.0060	0.0060
3:headlamp	0.0060	0.0300	0.0160	0.0160	0.0100	0.0240	0.0100
4:container	0.0020	0.0080	0.0030	0.0080	0.0030	0.0060	0.0040
5:building window	0.0016	0.0020	0.0020	0.0050	0.0024	0.0060	0.0040

Appendix C

Posterior samples from the composite model

The posterior samples shown here are for the square root transformed ratios. They were obtained from the composite model for elemental configurations $m = 1, 3, 4$, and as the posterior samples for configuration $m = 2$ ($\overline{\text{Fe}}, K$) shown in Section 3.2.4, where obtained from MCMC with a burn-in period of 10,000, and thinning of the Markov chain where every 200th draw was stored and the rest discarded.

C.1 Posterior samples for elemental configuration $m = 1$ (Fe, K)

The Metropolis-Hastings acceptance rate for configuration $m = 1$ (Fe, K) for M-H 3 was 76%. M-H 2 is not performed for configuration $m = 1$ as there are no bulbs, i.e. θ_1 , that have the elemental configuration $m = 1$. When no data is associated with a use type for a given configuration draws are made from the prior distribution of θ_t . This is why only the draws of θ_t for $t = 2, 4, 5$ are displayed for configuration $m = 1$, with $t = 1, 3$ being use types bulb and headlamp, with both having no items in the database with this elemental configuration, as shown in Table 2.3. From Figure C.2 it is clear to see why it is difficult to distinguish between the two window use types with the cluster of draws for each heavily overlapping one another. Table C.1 contains the effective sample sizes for θ_t for items with configuration $m = 1$. The values for the standard deviations for the variance-covariance matrices shown in Table C.2 show increased values for some elements when compared to those for items of elemental configuration $m = 2$ in Table 3.6, but this is most likely due to there being fewer items associated with configuration $m = 1$ for these use types, as seen in Table 2.3.

APPENDIX C. POSTERIOR SAMPLES FROM COMPOSITE MODEL183

Table C.1: Effective sample size from the composite model for items with configuration $m = 1$ (Fe, K) for the square root ratios for the mean vector θ_t . For θ_t , $t = 2, 4, 5$ correspond to use types: car window, container and building window.

	Na	Mg	Al	Si	K	Ca	Fe
θ_2	864.8	1000.0	857.6	1000.0	813.2	1000.0	1000.0
θ_4	1000.0	1000.0	1000.0	1000.0	693.5	1000.0	797.3
θ_5	507.0	110.6	105.8	812.2	120.0	146.7	1153.1

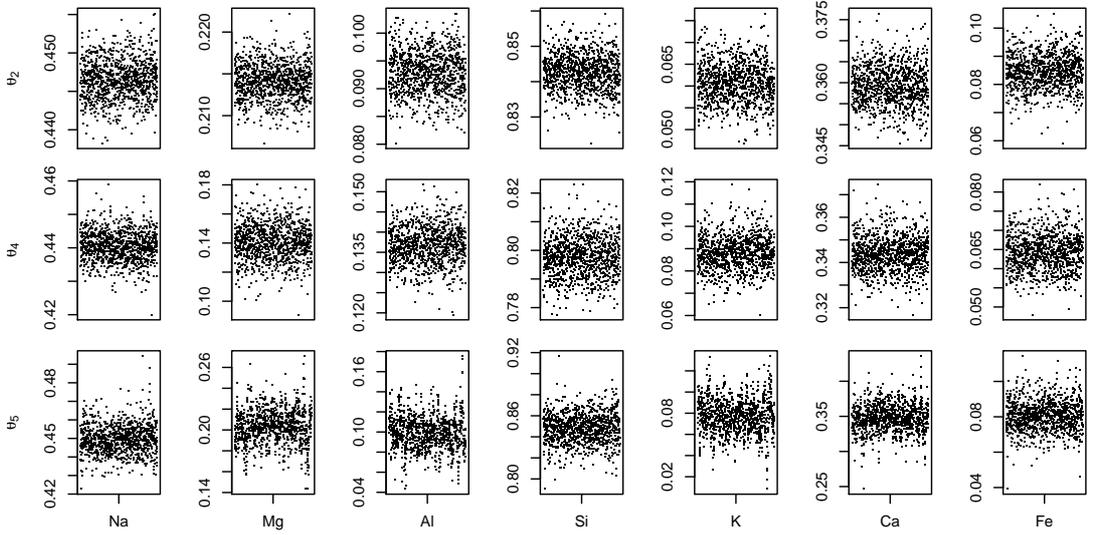


Figure C.1: Trace plots of the mean θ_t for configuration $m = 1$ (Fe, K) for the square root ratios. A burn-in period of 10,000 was used, and thinning of the Markov chain with every 200th draw stored.

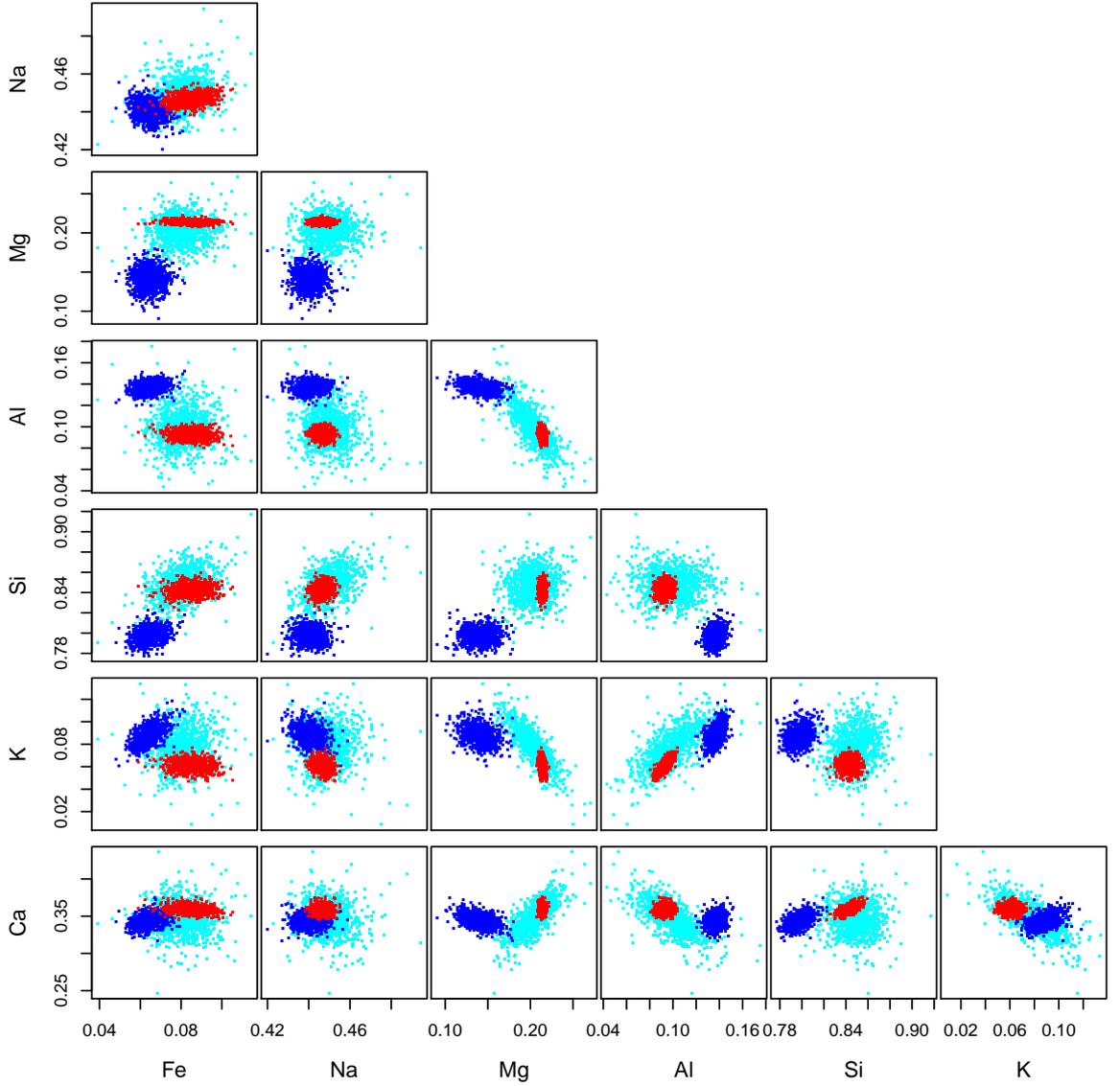


Figure C.2: Scatterplots of draws from θ_t for items with configuration $m = 1$, i.e. Fe and K present, for the square root ratios. The different coloured points correspond to the use type categories: **car window**, **container** and **building window**.

Table C.2: Standard deviations (multiplied by 10) from covariance matrices Ω_t^{-1} , Ψ^{-1} and Λ^{-1} for configuration $m = 1$ (Fe, K). For Ω_t^{-1} , $t = 2, 4, 5$ correspond to use types: car window, container and building window.

	Na	Mg	Al	Si	K	Ca	Fe
Ω_2^{-1}	0.12	0.09	0.18	0.17	0.21	0.17	0.28
Ω_4^{-1}	0.15	0.41	0.15	0.18	0.24	0.23	0.16
Ω_5^{-1}	0.19	0.42	0.45	0.35	0.43	0.50	0.22
Ψ^{-1}	0.05	0.05	0.05	0.31	0.07	0.24	0.08
Λ^{-1}	0.05	0.06	0.04	0.16	0.05	0.14	0.08

C.2 Posterior samples for elemental configuration $m = 3$ (Fe, $\bar{\mathbf{K}}$)

The Metropolis-Hastings acceptance rate for configuration $m = 3$ (Fe, $\bar{\mathbf{K}}$) was 58% for M-H 3. Again, as with configuration $m = 1$ M-H 2 is not performed due to there being no bulbs associated with this configuration. The only two use types associated with configuration $m = 3$ are car window and building window as seen from the time series plots for θ_2 and θ_5 in Figure C.3. From Figure C.4 it is easier to see more of a separation between the window use types for this configuration than for items of those use types with configuration $m = 1$. Table C.3 contains the effective sample sizes for items with configuration $m = 3$, while Table C.4 contains the standard deviations from the variance-covariance matrices.

Table C.3: Effective sample size from the composite model for items with configuration $m = 3$ ($\text{Fe}, \bar{\text{K}}$) for the square root ratios for the mean vector θ_t . For θ_t , $t = 2, 5$ correspond to use types: car window and building window.

	Na	Mg	Al	Si	Ca	Fe
θ_2	751.7	1003.5	844.2	875.9	769.6	1051.5
θ_5	896.7	1026.6	249.8	535.8	759.2	737.2

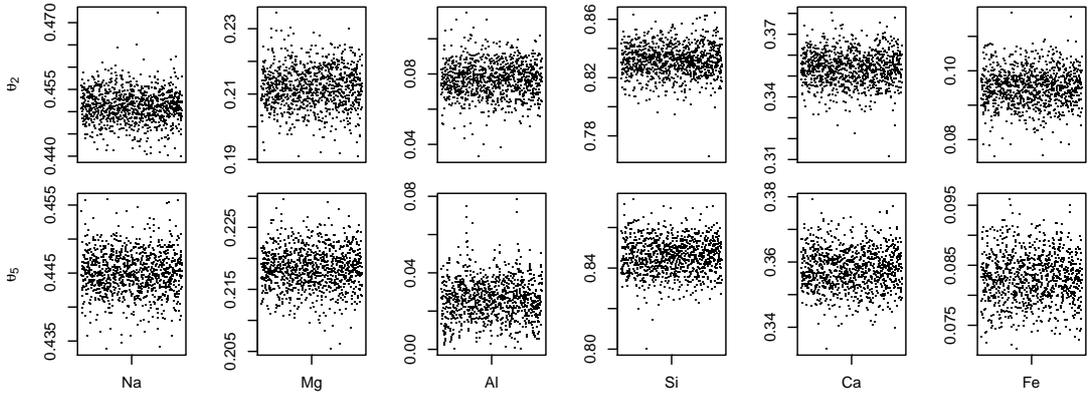


Figure C.3: Trace plots of the mean θ_t for configuration $m = 3$ ($\text{Fe}, \bar{\text{K}}$) for the square root ratios. A burn-in period of 10,000 was used, and thinning of the Markov chain with every 200th draw stored.

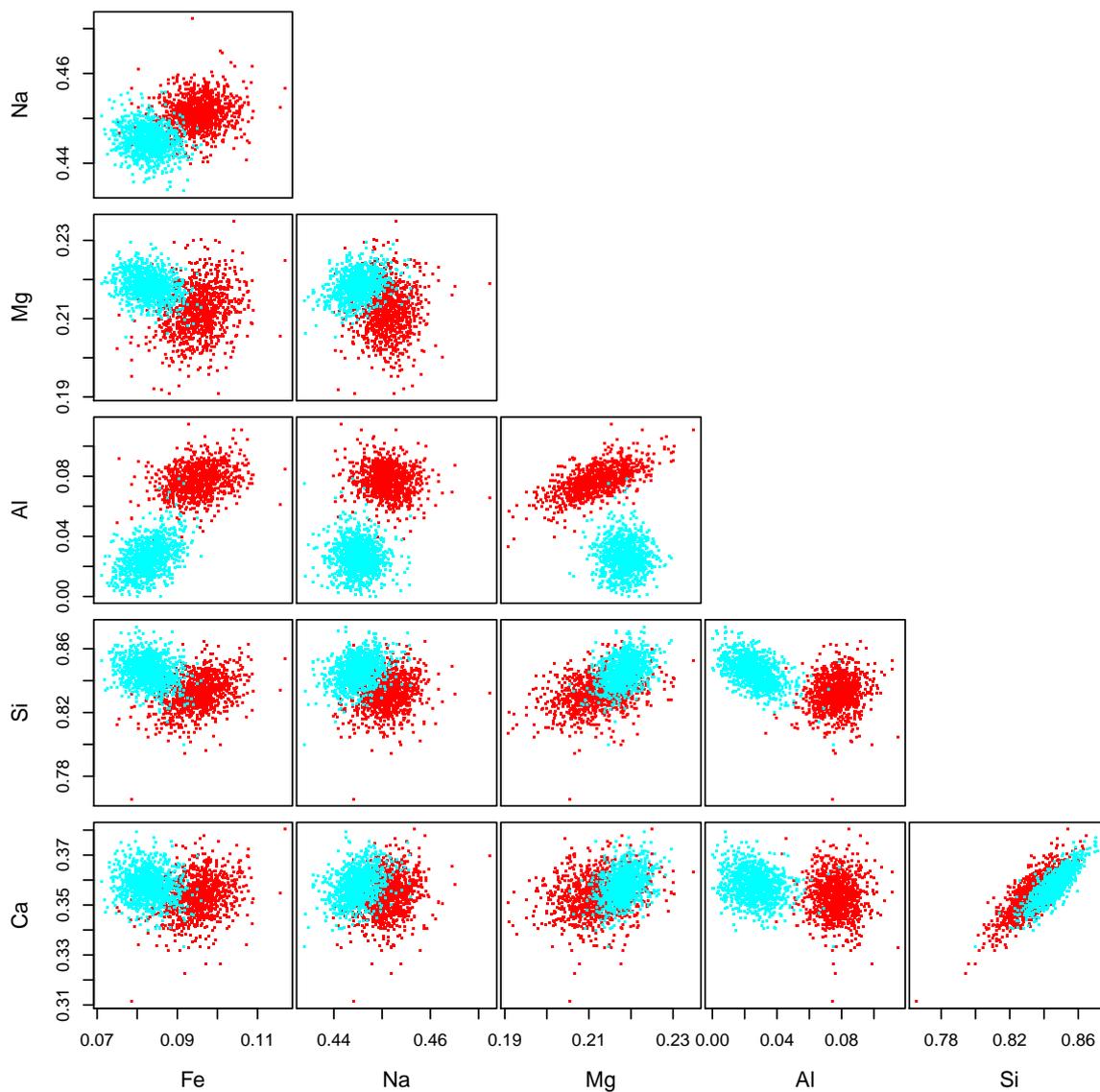


Figure C.4: Scatterplots of draws from θ_t for items with configuration $m = 3$, i.e. Fe present and K absent, for the square root ratios. The different coloured points correspond to the use type categories: **car window** and **building window**.

Table C.4: Standard deviations (multiplied by 10) from covariance matrices Ω_t^{-1} , Ψ^{-1} and Λ^{-1} for configuration $m = 3$ ($\overline{\text{Fe}}, \overline{\text{K}}$). For Ω_t^{-1} , $t = 2, 5$ correspond to use types: car window and building window.

	Na	Mg	Al	Si	Ca	Fe
Ω_2^{-1}	0.12	0.20	0.31	0.30	0.23	0.17
Ω_5^{-1}	0.11	0.12	0.40	0.28	0.20	0.14
Ψ^{-1}	0.06	0.05	0.05	0.32	0.24	0.09
Λ^{-1}	0.06	0.04	0.04	0.16	0.11	0.07

C.3 Posterior samples for elemental configuration $m = 4$ ($\overline{\text{Fe}}, \overline{\text{K}}$)

The Metropolis-Hastings acceptance rates from configuration $m = 4$ ($\overline{\text{Fe}}, \overline{\text{K}}$) were 85% for M-H 2 and 95% for M-H 3. Due to a lack of data associated with the use types bulb and headlamp for configuration $m = 4$, the draws from the model in Figure C.5 for those use types do not appear to have converged as well as the other use types. There is only one bulb and two headlamps in the glass database that have elemental configuration $m = 4$ as shown in Table 2.3. Also, from Figure C.5 the values drawn for Mg for the headlamps, θ_3 , are all very close to zero. This is due to these two headlamps associated with configuration $m = 4$ containing no Mg as shown in Figure A.3 and also from Table 2.2 before the original ten elemental configurations were reduced to four. Figure C.6 displays scatterplots of the draws of θ_t for items with configuration $m = 4$. Table C.5 contains the effective sample sizes

for items with configuration $m = 4$. The lack of data associated with the use types bulb and headlamp also results in increased standard deviations for those use types as shown in Table C.6.

Table C.5: Effective sample size from the composite model for items with configuration $m = 4$ ($\overline{\text{Fe}}$, $\overline{\text{K}}$) for the square root ratios for the mean vector $\boldsymbol{\theta}_t$. For $\boldsymbol{\theta}_t$, $t = 1, \dots, 5$ correspond to use types: bulb, car window, headlamp, container and building window.

	Na	Mg	Al	Si	Ca
$\boldsymbol{\theta}_1$	184.7	127.8	106.9	208.8	304.8
$\boldsymbol{\theta}_2$	1000.0	1000.0	1000.0	1000.0	1000.0
$\boldsymbol{\theta}_3$	246.2	743.2	768.3	601.8	653.9
$\boldsymbol{\theta}_4$	1000.0	1000.0	914.3	856.4	717.8
$\boldsymbol{\theta}_5$	842.0	452.0	709.9	843.3	673.0

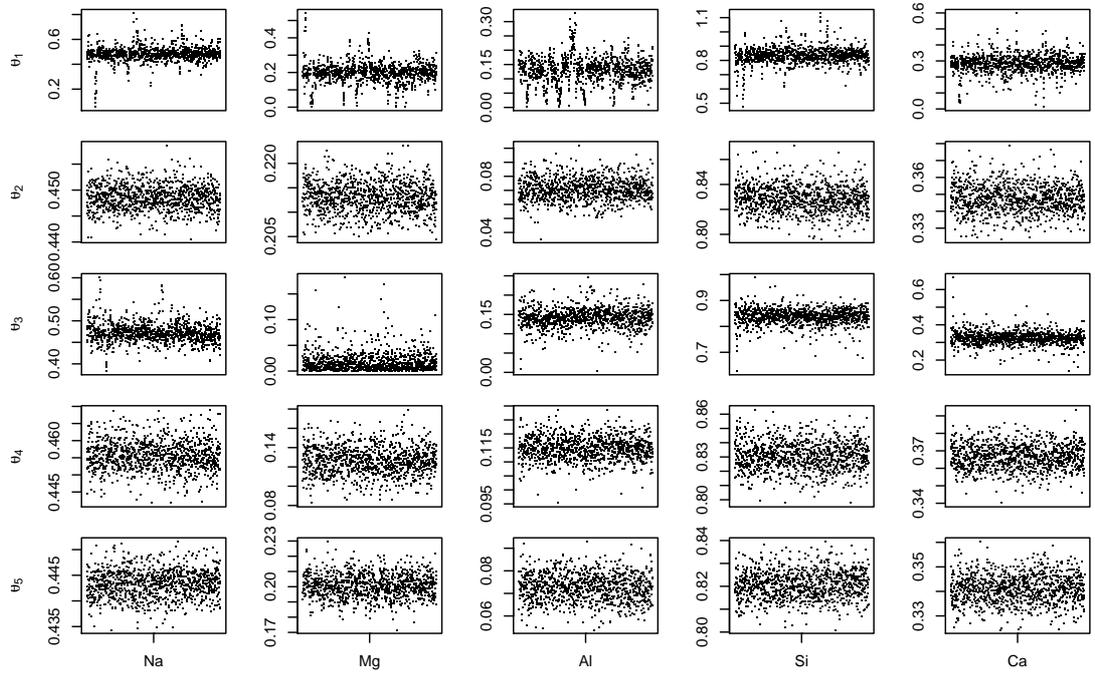


Figure C.5: Trace plots of the mean θ_t for configuration $m = 4$ ($\overline{\text{Fe}}, \overline{\text{K}}$) for the square root ratios. A burn-in period of 10,000 was used, and thinning of the Markov chain with every 200th draw stored.

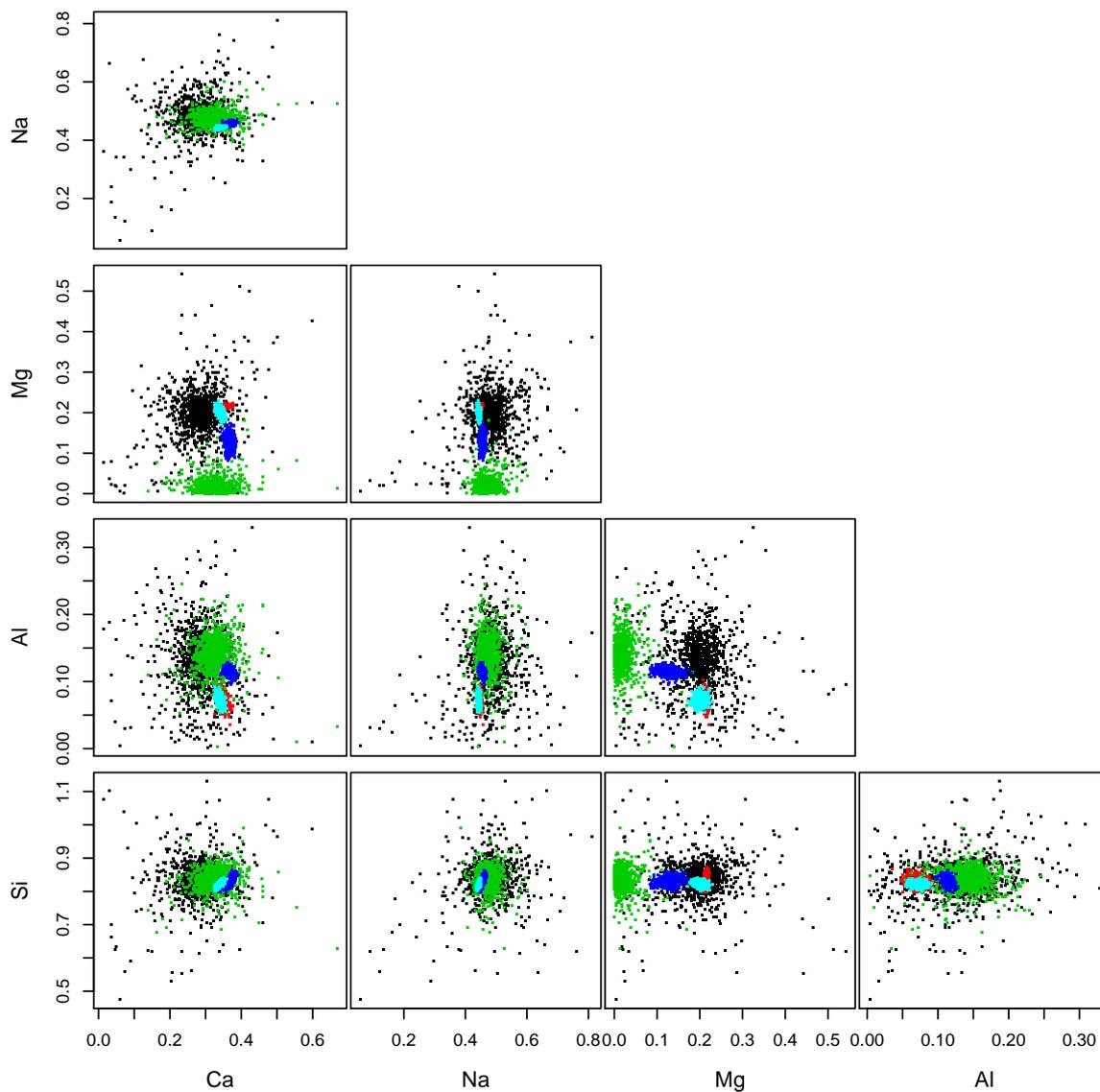


Figure C.6: Scatterplots of draws from θ_t for items with configuration $m = 4$, i.e. Fe and K absent, for the square root ratios. The different coloured points correspond to the five use type categories: bulb, car window, headlamp, container and building window.

Table C.6: Standard deviations (multiplied by 10) from covariance matrices Ω_t^{-1} , Ψ^{-1} and Λ^{-1} for configuration $m = 4$ ($\overline{\text{Fe}}, \overline{\text{K}}$). For Ω_t^{-1} , $t = 1, \dots, 5$ correspond to use types: bulb, car window, headlamp, container and building window.

	Na	Mg	Al	Si	Ca
Ω_1^{-1}	1.34	1.28	1.23	1.49	1.30
Ω_2^{-1}	0.10	0.13	0.33	0.42	0.33
Ω_3^{-1}	0.42	0.44	0.46	0.48	0.65
Ω_4^{-1}	0.18	0.61	0.17	0.39	0.29
Ω_5^{-1}	0.14	0.40	0.31	0.27	0.28
Ψ^{-1}	0.06	0.05	0.08	0.31	0.25
Λ^{-1}	0.04	0.03	0.05	0.16	0.12

Appendix D

Posterior samples from the composite model using log-ratios

The posterior samples shown here were obtained from the composite model for elemental configurations $m = 1, 3, 4$ using the log-ratio transformation, and as the posterior samples for configuration $m = 2$ ($\overline{\text{Fe}}, \text{K}$) shown in Section [3.3.2](#), where obtained from MCMC with a burn-in period of 10,000, and thinning of the Markov chain where every 200th draw was stored and the rest discarded.

D.1 Log-ratio posterior samples for configuration $m = 1$ (Fe, K)

The time series plots for the log-ratio posterior draws of items with configuration $m = 1$ (Fe, K) are displayed in Figure D.1. As mentioned in C.1 the draws shown are only for the use types car window, container and building window, i.e. θ_2 , θ_4 and θ_5 , as there are no bulbs or headlamps associated with this configuration. The time series plots, although not perfect, do not show any of the convergence problems that were obvious for items of configuration $m = 2$. Figure D.2 displays scatterplots of the posterior draws, while Table D.1 contains the effective sample sizes for θ_t for items with configuration $m = 1$ using the log-ratio transformation. Table D.2 contains standard deviations obtained from the random effects covariance matrices.

Table D.1: Effective sample size from the composite model for items with configuration $m = 1$ (Fe, K) using log-ratios for the mean vector θ_t . For θ_t , $t = 2, 4, 5$ correspond to use types: car window, container and building window.

	Na	Mg	Al	Si	K	Ca	Fe
θ_2	670.9	673.5	298.7	1000.0	303.1	856.5	603.1
θ_4	518.8	277.6	360.8	696.5	578.9	740.9	584.7
θ_5	165.8	49.5	70.7	133.5	77.8	48.6	130.4

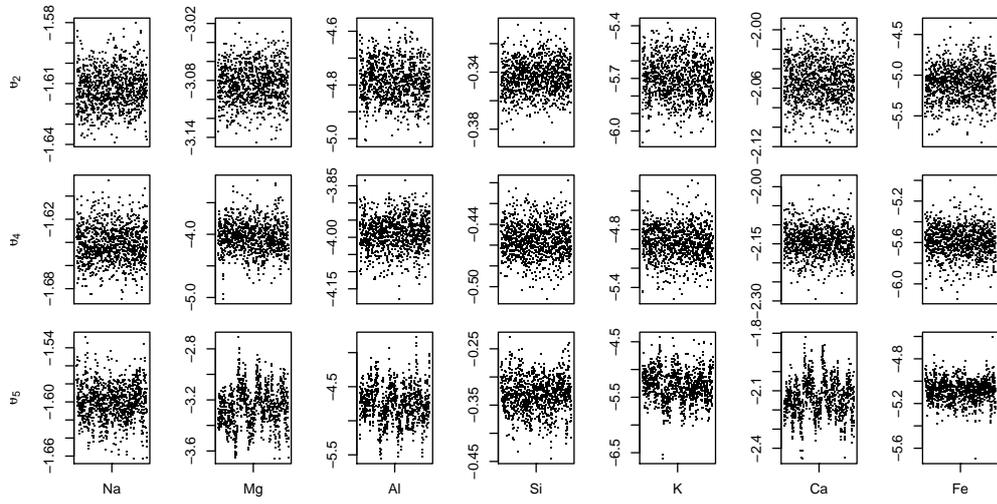


Figure D.1: Trace plots of the mean θ_t from the composite model using the log-ratio transformation for configuration $m = 1$ (Fe, K). A burn-in period of 10,000 was used, and thinning of the Markov chain with every 200th draw stored.

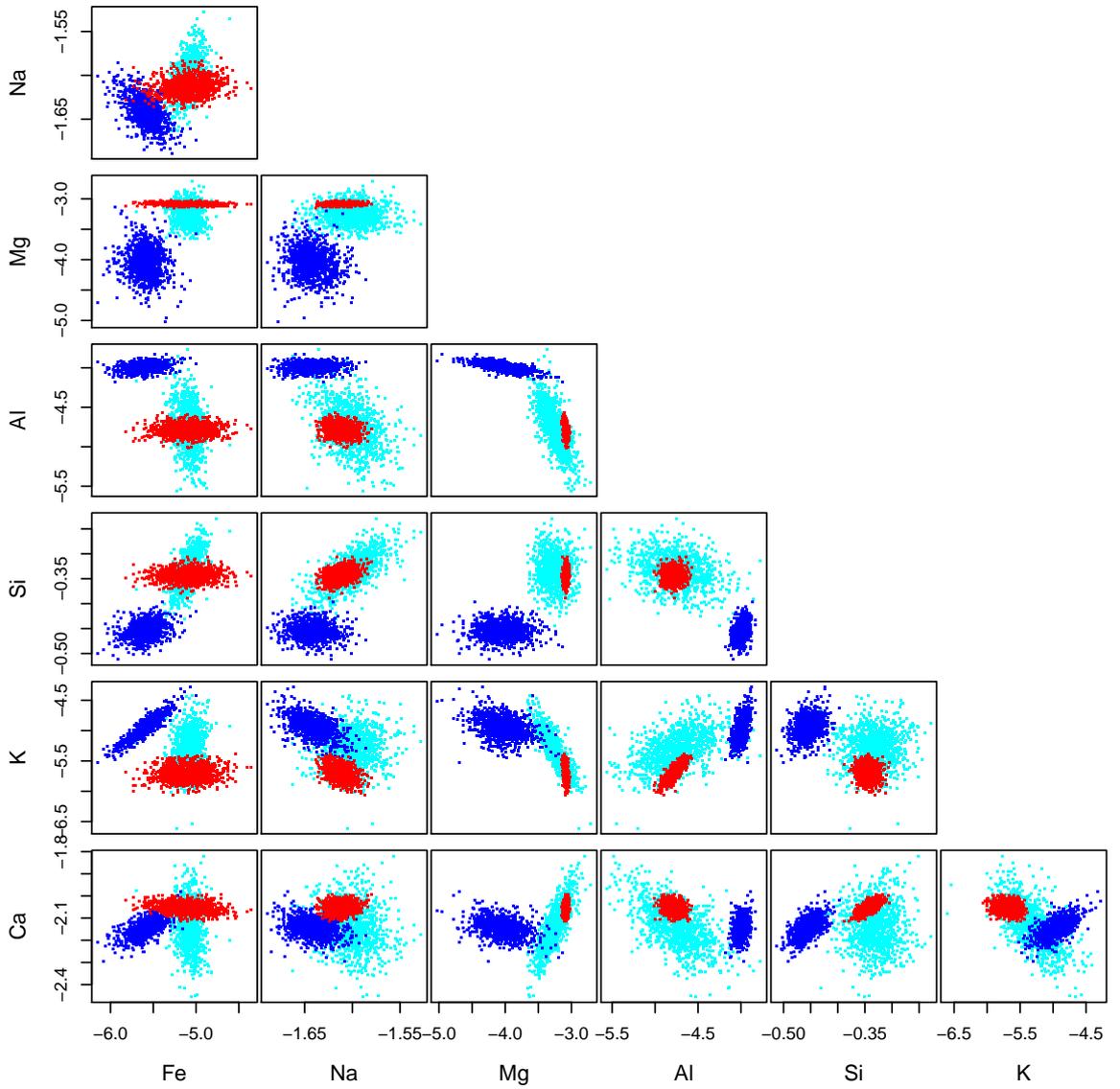


Figure D.2: Scatterplots of draws from θ_t from the composite model using the log-ratio transformation for items with configuration $m = 1$, i.e. Fe and K present. The different coloured points correspond to the five use type categories: bulb, car window, headlamp, container and building window.

Table D.2: Standard deviations (multiplied by 10) from covariance matrices Ω_t^{-1} , Ψ^{-1} and Λ^{-1} from the composite model using the log-ratio transformation for configuration $m = 1$ (Fe, K). For Ω_t^{-1} , $t = 2, 4, 5$ correspond to use types: car window, container and building window.

	Na	Mg	Al	Si	K	Ca	Fe
Ω_2^{-1}	0.44	0.54	3.31	0.34	5.49	0.70	9.73
Ω_4^{-1}	0.51	7.86	1.53	0.35	5.67	1.10	4.70
Ω_5^{-1}	0.55	5.04	8.56	0.77	8.87	3.00	2.25
Ψ^{-1}	0.18	0.32	0.72	0.84	1.47	1.44	1.73
Λ^{-1}	0.20	2.50	0.87	0.39	2.03	0.79	3.88

D.2 Log-ratio posterior samples for configuration $m = 3$ (Fe, $\bar{\mathbf{K}}$)

The time series plots for the log-ratio posterior draws of items with configuration $m = 3$ (Fe, $\bar{\mathbf{K}}$) are shown in Figure D.3. The results shown are those for car windows (θ_2) and building windows (θ_5) as those are the only two use types with items associated with elemental configuration $m = 3$. Similarly to the posterior samples for configuration $m = 1$, those for $m = 3$ do not contain the convergence problems seen for items of configuration $m = 2$. There are possible issues with the element Al, and as with the problems for configuration $m = 2$, in this instance this could be due to the mixture of car and building windows in configuration $m = 3$ that do and do not contain concentrations of Al. Figure D.4 shows the scatterplots from the posterior

draws of Figure D.3, with the effective sample sizes found in Table D.3. The standard deviations obtained from the random effects covariance matrices are found in Table D.4.

Table D.3: Effective sample size from the composite model for items with configuration $m = 3$ ($\text{Fe}, \bar{\text{K}}$) using log-ratios for the mean vector θ_t . For θ_t , $t = 2, 5$ correspond to use types: car window and building window.

	Na	Mg	Al	Si	Ca	Fe
θ_2	538.9	92.9	102.6	889.1	794.6	685.6
θ_5	751.7	653.8	71.1	161.7	224.2	329.2

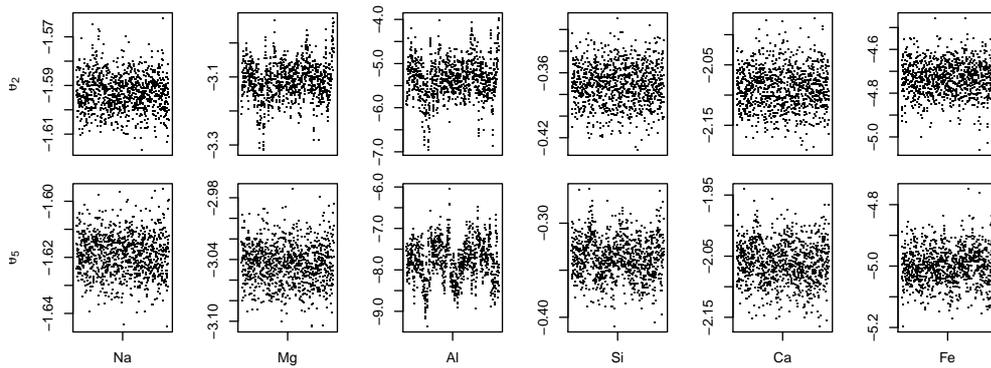


Figure D.3: Trace plots of the mean θ_t from the composite model using the log-ratio transformation for configuration $m = 3$ ($\text{Fe}, \bar{\text{K}}$). A burn-in period of 10,000 was used, and thinning of the Markov chain with every 200th draw stored.

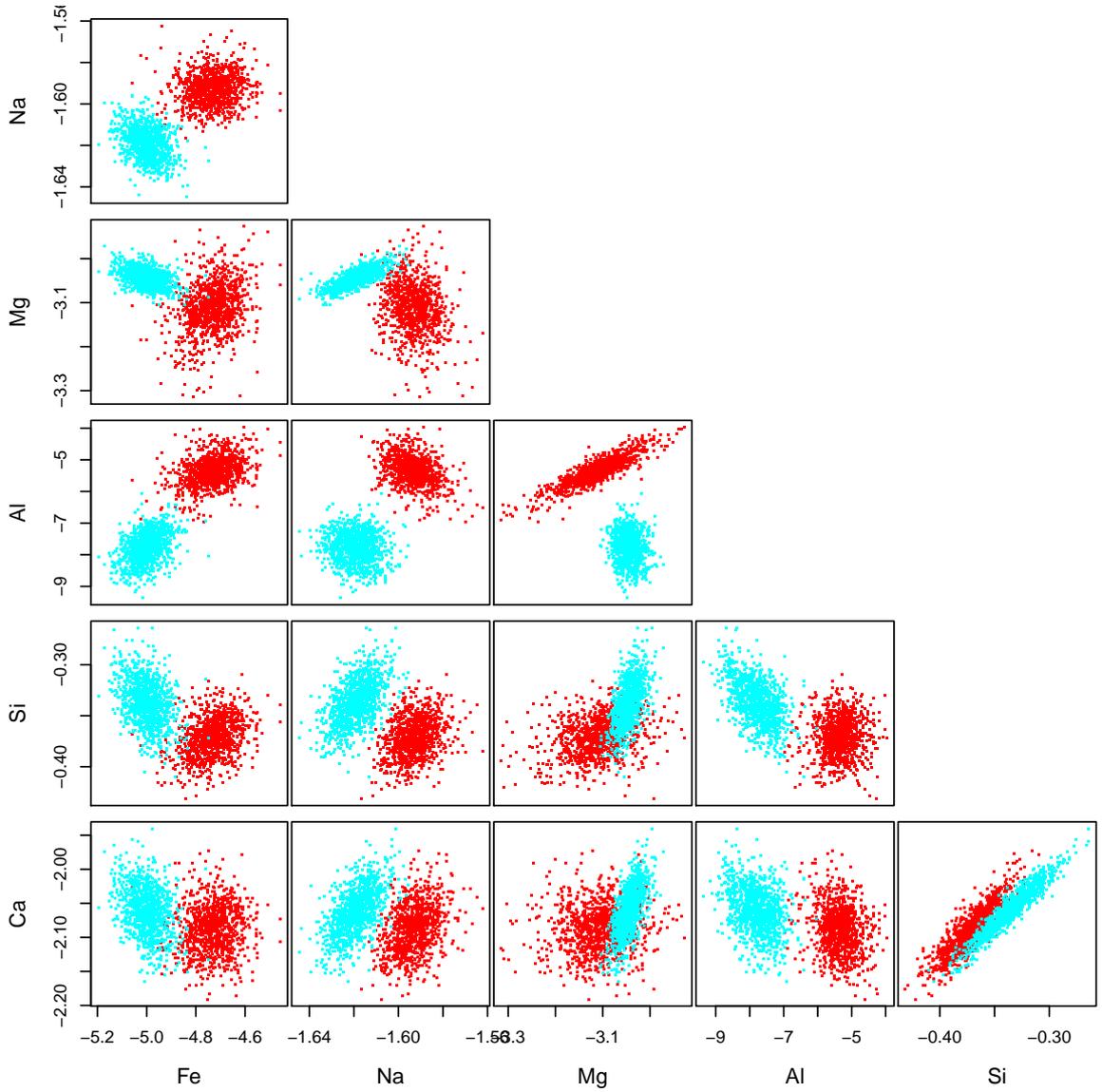


Figure D.4: Scatterplots of draws from θ_t from the composite model using the log-ratio transformation for items with configuration $m = 3$, i.e. Fe present and K absent. The different coloured points correspond to the five use type categories: bulb, car window, headlamp, container and building window.

Table D.4: Standard deviations (multiplied by 10) from covariance matrices Ω_t^{-1} , Ψ^{-1} and Λ^{-1} from the composite model using the log-ratio transformation for configuration $m = 3$ ($\overline{\text{Fe}}, \overline{\text{K}}$). For Ω_t^{-1} , $t = 2, 5$ correspond to use types: car window and building window.

	Na	Mg	Al	Si	Ca	Fe
Ω_2^{-1}	0.21	1.70	13.74	0.38	0.73	2.33
Ω_5^{-1}	0.26	0.76	20.21	0.72	1.07	2.08
Ψ^{-1}	0.21	0.29	0.54	0.90	1.54	1.94
Λ^{-1}	0.24	0.31	1.84	0.37	0.57	1.53

D.3 Log-ratio posterior samples for configuration $m = 4$ ($\overline{\text{Fe}}, \overline{\text{K}}$)

Figure D.5 displays the time series plots for the log-ratio posterior draws of items with elemental configuration $m = 4$ ($\overline{\text{Fe}}, \overline{\text{K}}$). Figure D.5 has posterior plots for all five use types with convergence for use types car window (θ_2), container (θ_4) and building window (θ_5) displaying no clear issues with convergence, however there are potential problems for bulbs (θ_1) and headlamps (θ_3). This is due to there being only one bulb and two headlamps, respectively, that have elemental configuration $m = 4$, with improvements in convergence most likely obtained for an increased number of glass items for each use type. Scatterplots of the posterior draws for glass items with elemental configuration $m = 4$ are shown in Figure D.6, and the effective samples sizes found in Table D.5. The standard deviations obtained from

the random effects covariance matrices found in Table D.6.

Table D.5: Effective sample size from the composite model for items with configuration $m = 4$ ($\overline{\text{Fe}}, \overline{\text{K}}$) using log-ratios for the mean vector $\boldsymbol{\theta}_t$. For $\boldsymbol{\theta}_t$, $t = 1, \dots, 5$ correspond to use types: bulb, car window, headlamp, container and building window.

	Na	Mg	Al	Si	Ca
$\boldsymbol{\theta}_1$	101.8	303.6	357.4	110.3	235.4
$\boldsymbol{\theta}_2$	1000.0	931.4	530.1	887.5	764.4
$\boldsymbol{\theta}_3$	121.8	1386.0	357.0	67.5	100.5
$\boldsymbol{\theta}_4$	840.4	844.7	825.8	1043.2	1000.0
$\boldsymbol{\theta}_5$	909.2	594.8	1000.0	861.1	733.6

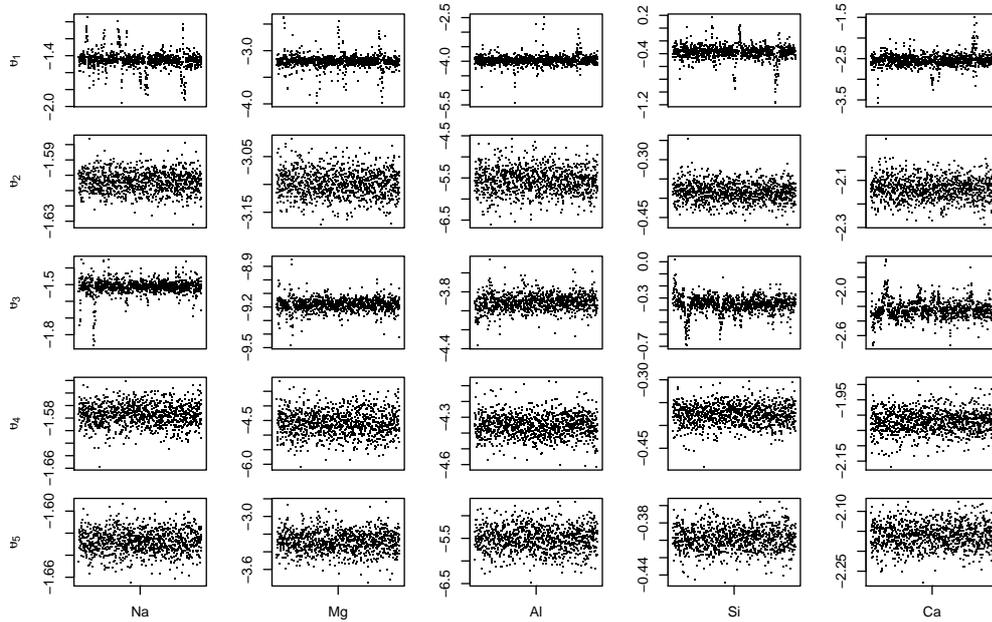


Figure D.5: Trace plots of the mean $\boldsymbol{\theta}_t$ from the composite model using the log-ratio transformation for configuration $m = 4$ ($\overline{\text{Fe}}, \overline{\text{K}}$). A burn-in period of 10,000 was used, and thinning of the Markov chain with every 200th draw stored.

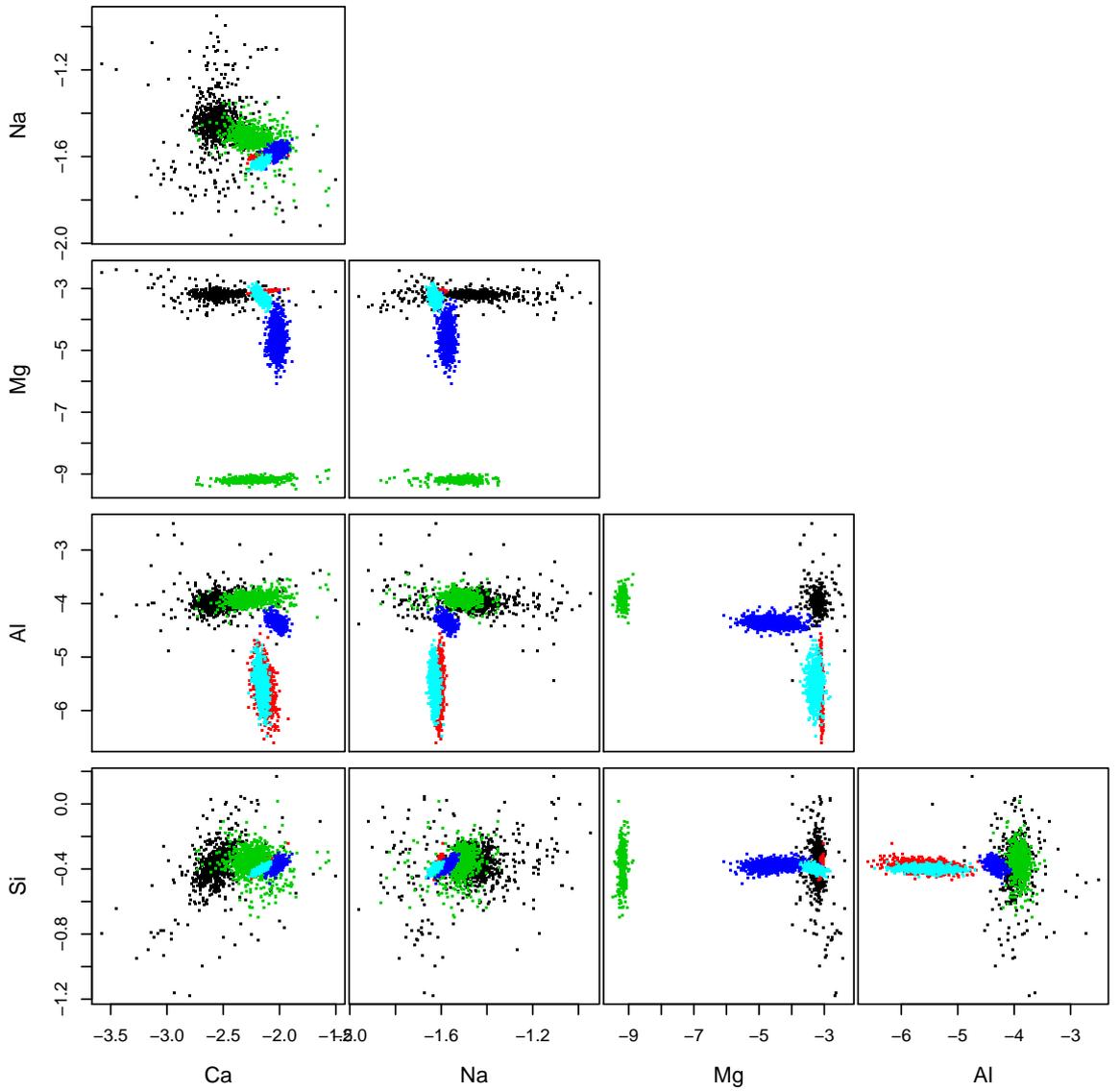


Figure D.6: Scatterplots of draws from θ_t from the composite model using the log-ratio transformation for items with configuration $m = 4$, i.e. Fe and K absent. The different coloured points correspond to the five use type categories: bulb, car window, headlamp, container and building window.

Table D.6: Standard deviations (multiplied by 10) from covariance matrices Ω_t^{-1} , Ψ^{-1} and Λ^{-1} from the composite model using the log-ratio transformation for configuration $m = 4$ ($\overline{\text{Fe}}, \overline{\text{K}}$). For Ω_t^{-1} , $t = 1, \dots, 5$ correspond to use types: bulb, car window, headlamp, container and building window.

	Na	Mg	Al	Si	Ca
Ω_1^{-1}	3.66	2.92	4.24	2.11	2.46
Ω_2^{-1}	0.27	0.99	13.93	0.95	1.85
Ω_3^{-1}	0.67	0.66	1.02	1.09	1.96
Ω_4^{-1}	0.72	18.29	2.87	0.83	1.29
Ω_5^{-1}	0.55	6.41	13.72	0.58	1.44
Ψ^{-1}	0.25	0.49	1.19	0.84	1.52
Λ^{-1}	0.19	0.76	2.00	0.39	0.66

Appendix E

Simulated data from the composite model

The results from simulating datasets from the Bayesian hierarchical model using the composite model approach for glass items with elemental configurations $m = 1, 3, 4$ are given below. Some of the figures below will show the odd extreme outlier and produce what looks like very wrong estimates. This is due to the model making draws from the prior distributions whenever there is no data associated with a glass use type, or an elemental configuration. For example, in Figure [E.1](#), for the use type building window there is one extreme outlier, which is due to draws being made from the prior distribution for one of the twenty simulated datasets. This is due to one of the twenty datasets for building windows having no building windows with elemental configuration 1. The issue is not noticeable in Figure [E.1](#) for the use types bulb and headlamp due to all but one or two of the simulated datasets containing

no bulbs or headlamps with elemental configuration 1, with the draws for these use types being generated from the prior distributions. The number of simulated datasets containing at least one glass item for each elemental configuration across use types is shown in Table E.1.

Table E.1: Number (out of 20) of simulated datasets containing at least one item of use type t with configuration m .

Glass type	Configuration m			
	1: Fe, K	2: $\overline{\text{Fe}}$, K	3: Fe, $\overline{\text{K}}$	4: $\overline{\text{Fe}}$, $\overline{\text{K}}$
bulb	2	20	5	15
car window	20	20	20	20
headlamp	1	20	3	20
container	20	20	2	20
building window	19	20	20	20

Simulated data from glass items with configuration $m = 1$ (Fe, K)

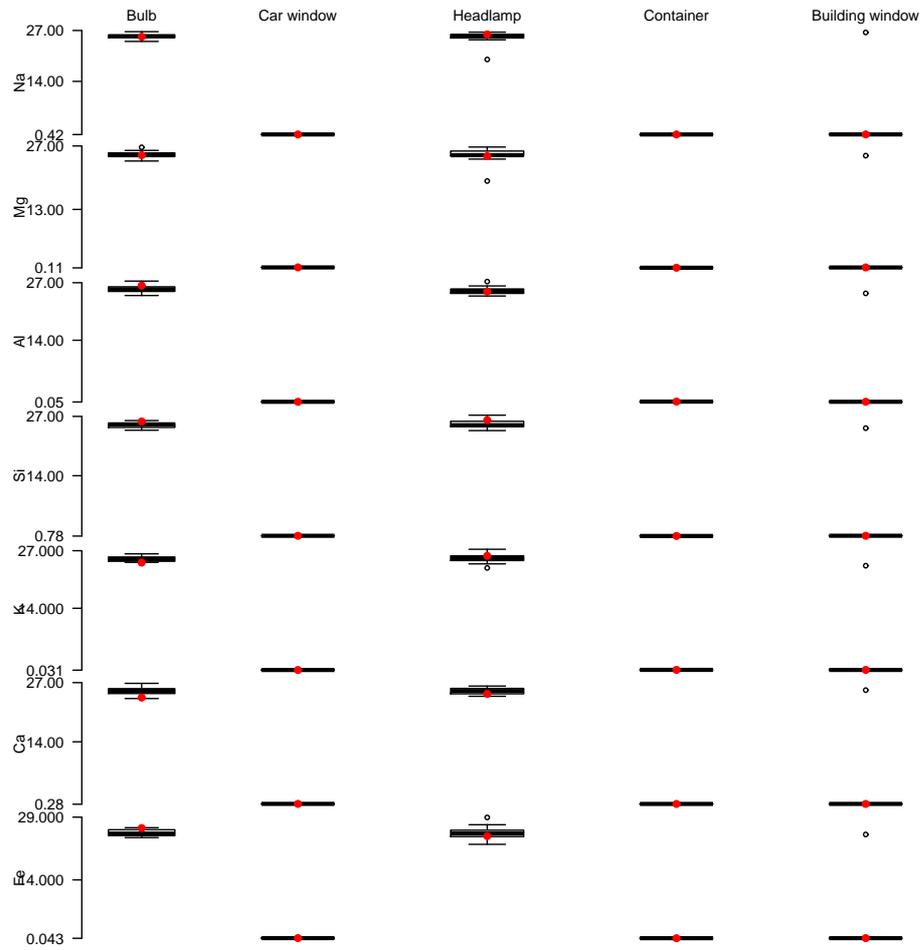


Figure E.1: Boxplots containing the mean posterior draws of θ_t from 20 simulated datasets for configuration $m = 1$ (Fe, K) for the square root ratios. The red dots indicate the true values of θ_t used to simulate the datasets.

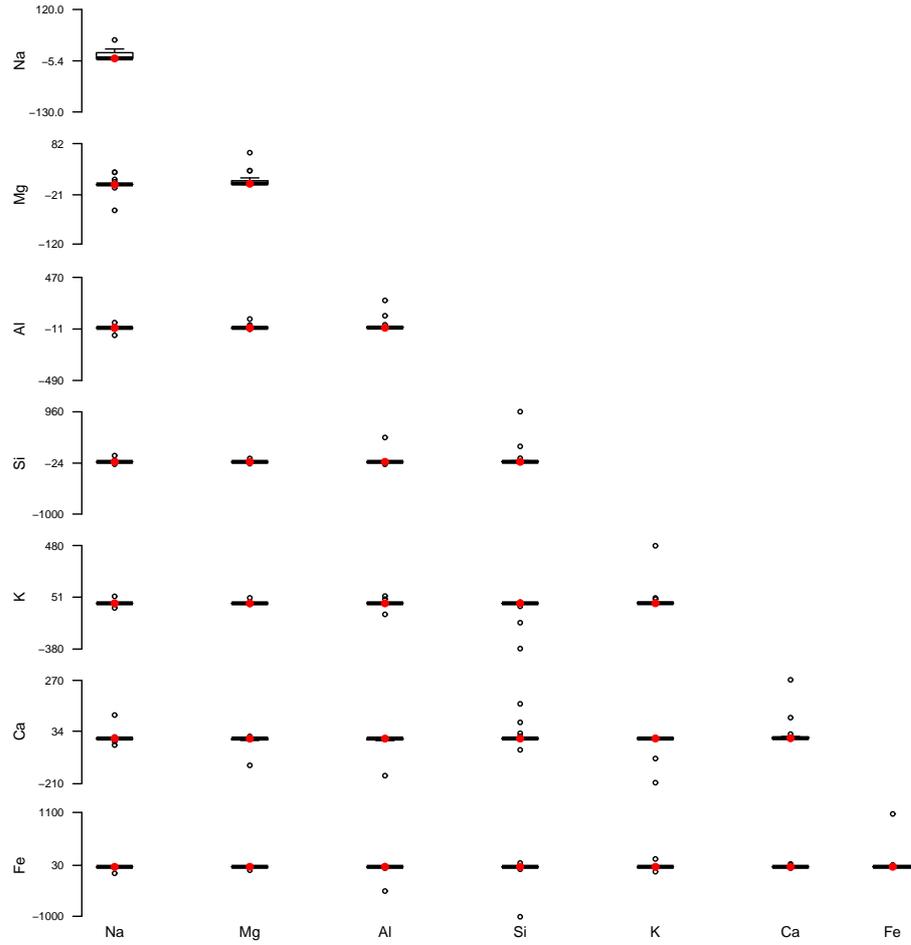


Figure E.2: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 1$ (bulb) from 20 simulated datasets for configuration $m = 1$ (Fe, K) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

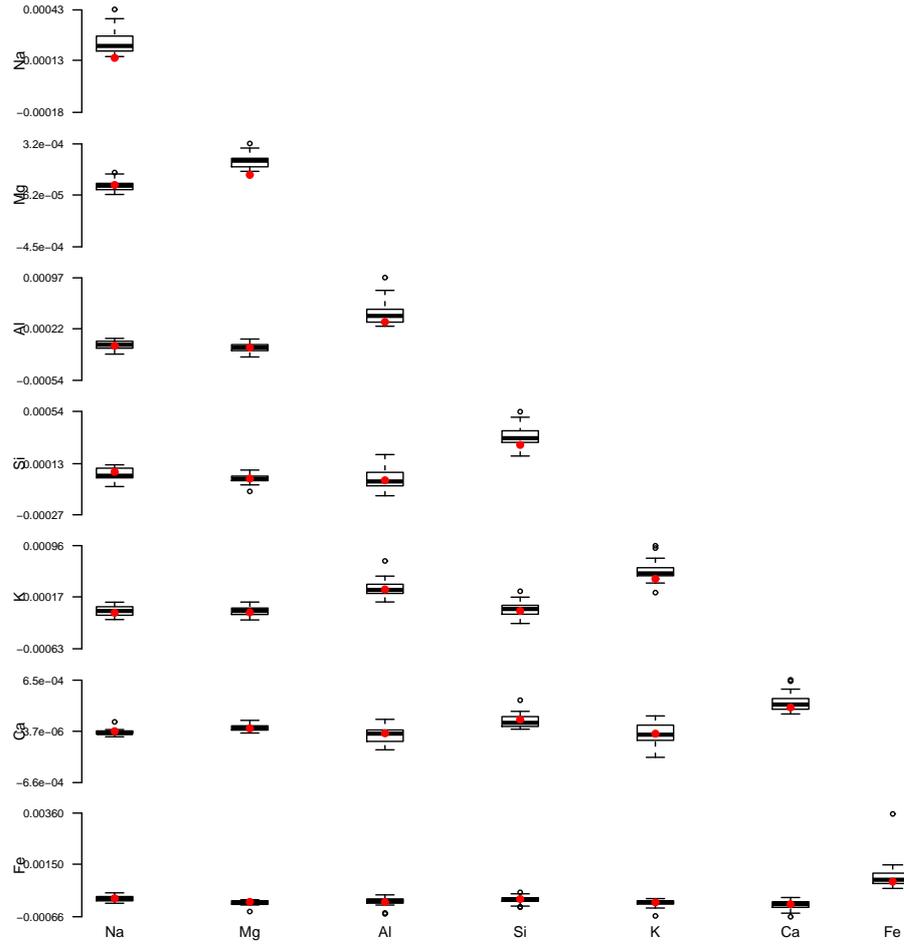


Figure E.3: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 2$ (car window) from 20 simulated datasets for configuration $m = 1$ (Fe, K) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

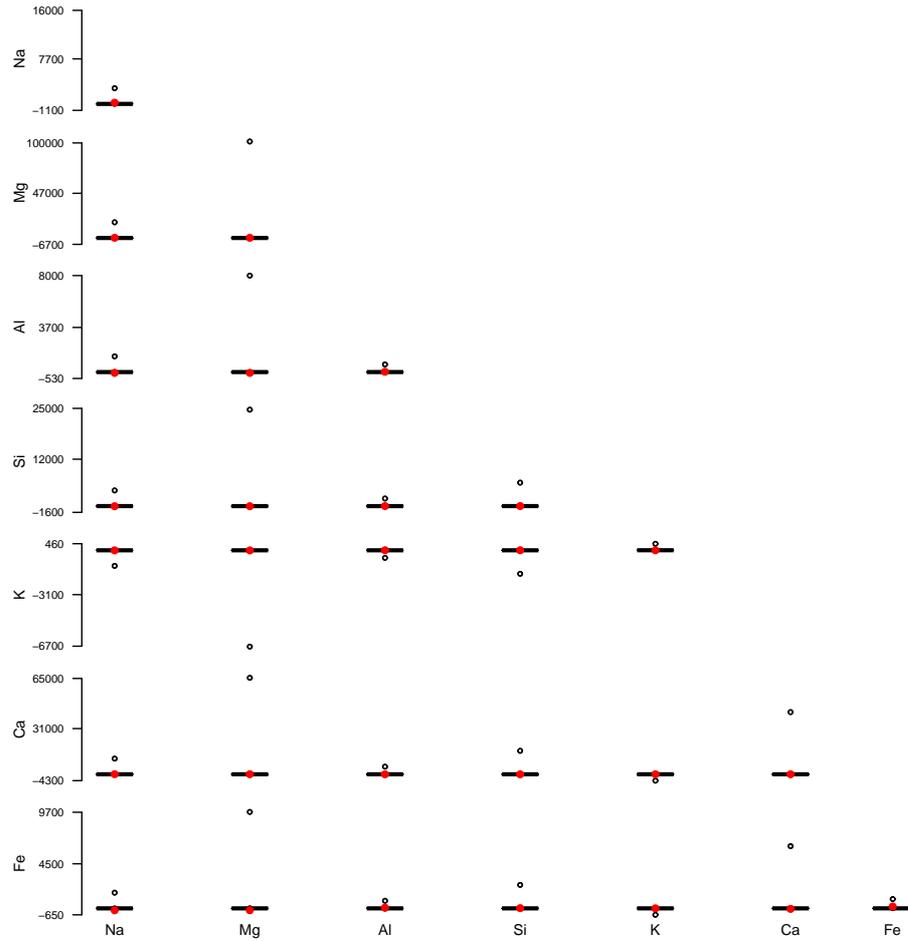


Figure E.4: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 3$ (headlamp) from 20 simulated datasets for configuration $m = 1$ (Fe, K) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

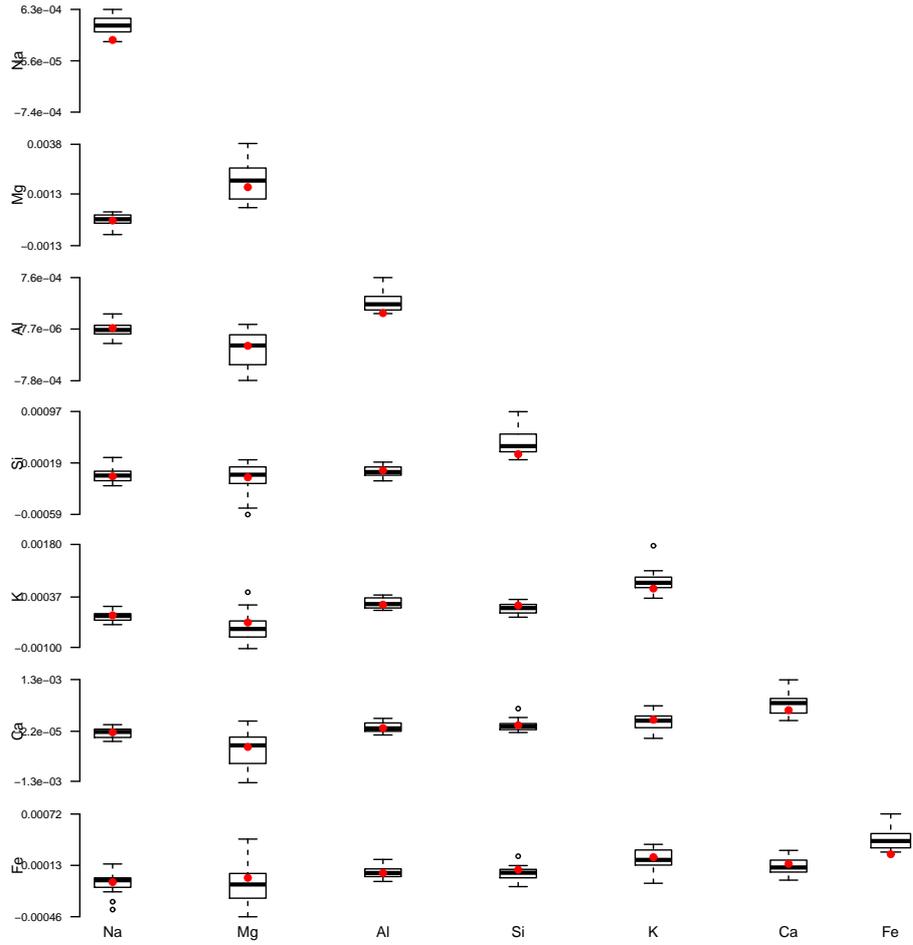


Figure E.5: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 4$ (container) from 20 simulated datasets for configuration $m = 1$ (Fe, K) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

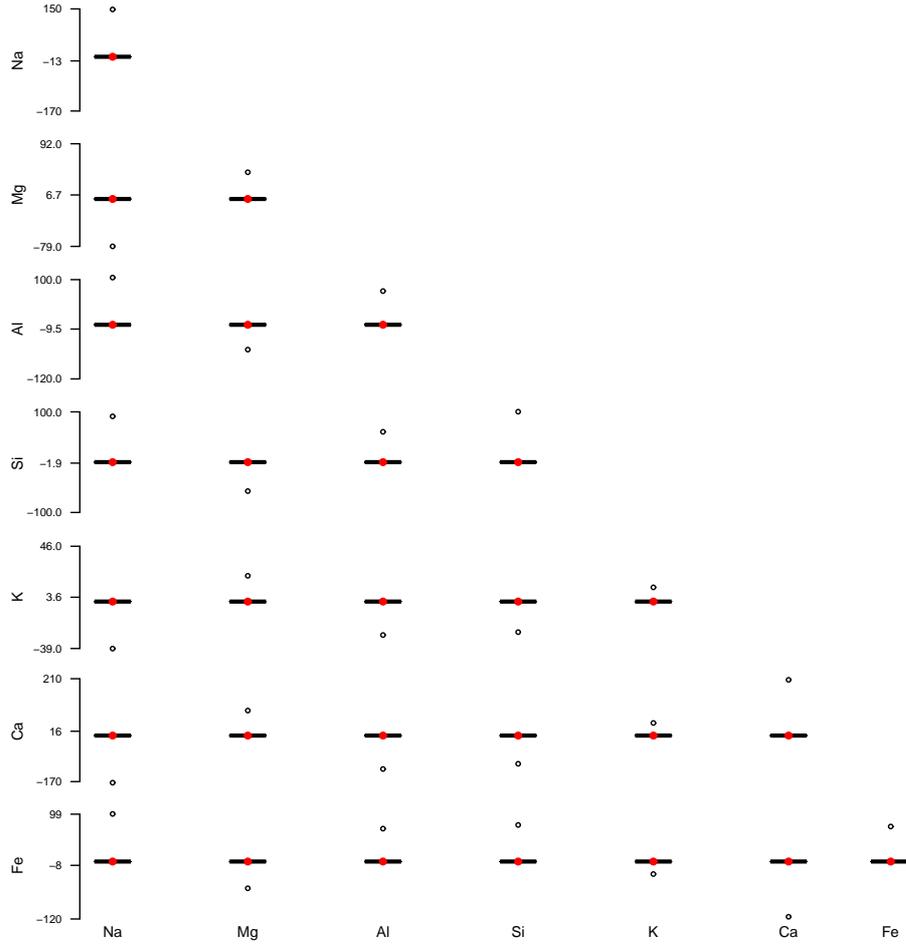


Figure E.6: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 5$ (building window) from 20 simulated datasets for configuration $m = 1$ (Fe, K) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

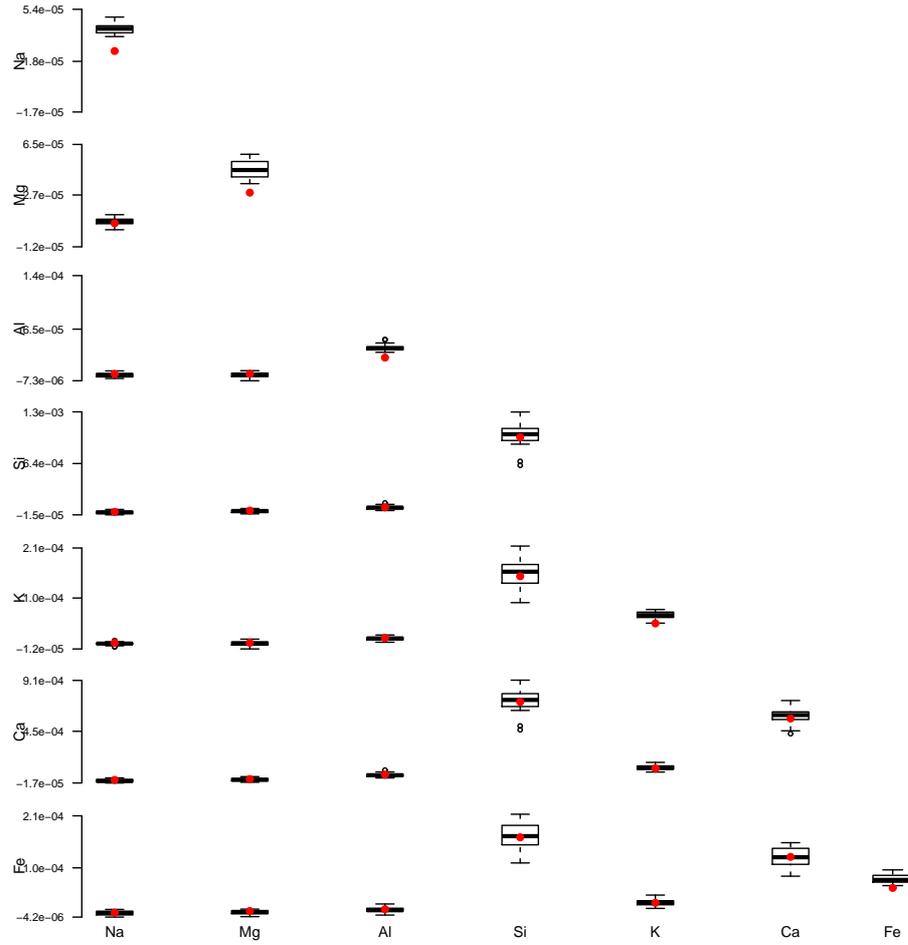


Figure E.7: Boxplots containing the mean posterior draws of Ψ^{-1} from 20 simulated datasets for configuration $m = 1$ (Fe, K) for the square root ratios. The red dots indicate the true values of Ψ^{-1} used to simulate the datasets.

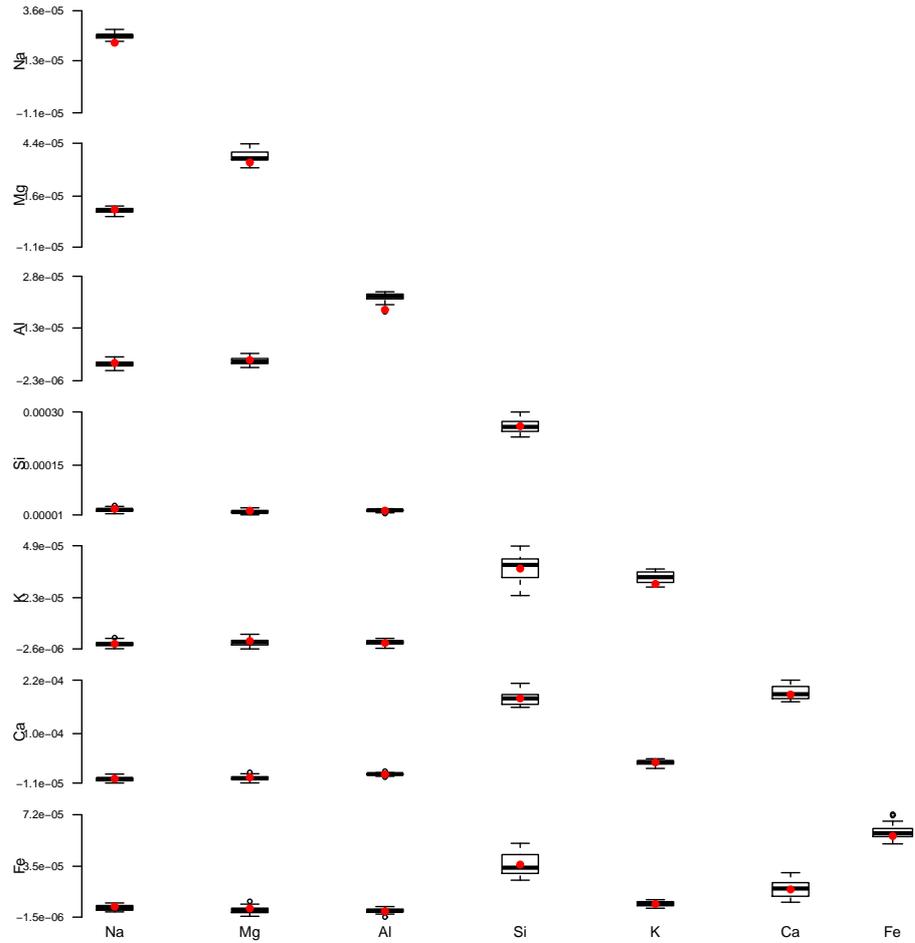


Figure E.8: Boxplots containing the mean posterior draws of Λ^{-1} from 20 simulated datasets for configuration $m = 1$ (Fe, K) for the square root ratios. The red dots indicate the true values of Λ^{-1} used to simulate the datasets.

Simulated data from glass items with configuration $m = 3$ ($\text{Fe}, \bar{\text{K}}$)

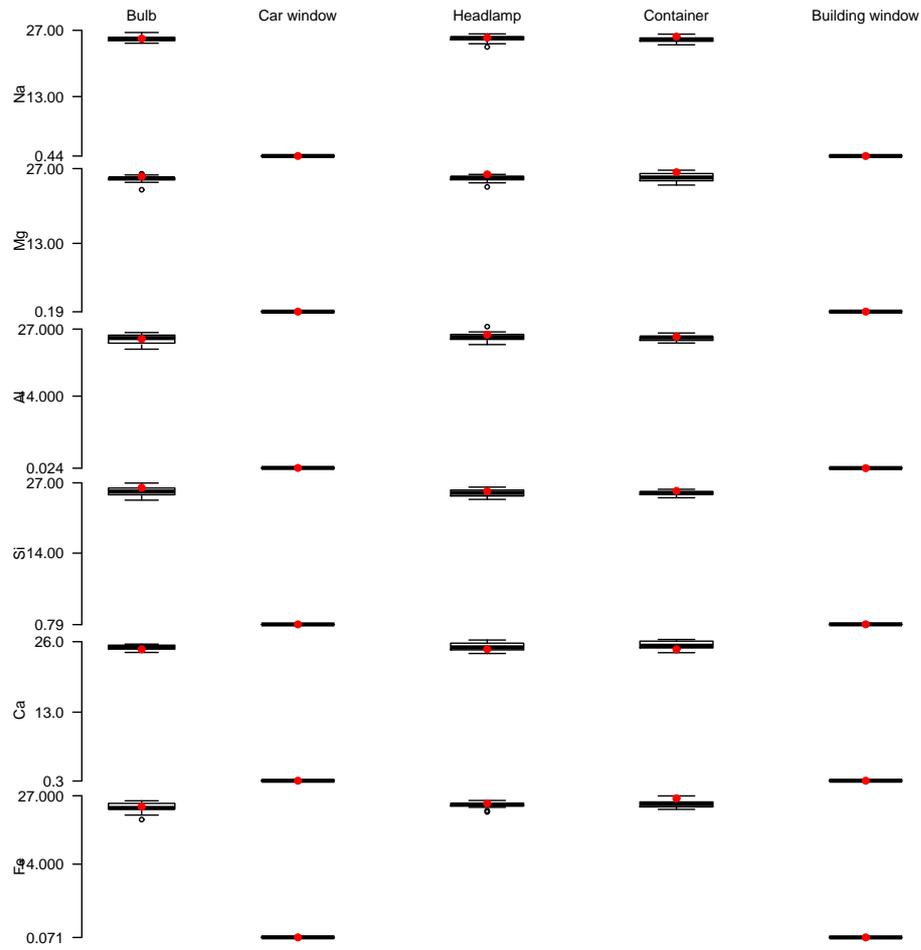


Figure E.9: Boxplots containing the mean posterior draws of θ_t from 20 simulated datasets for configuration $m = 3$ ($\text{Fe}, \bar{\text{K}}$) for the square root ratios. The red dots indicate the true values of θ_t used to simulate the datasets.

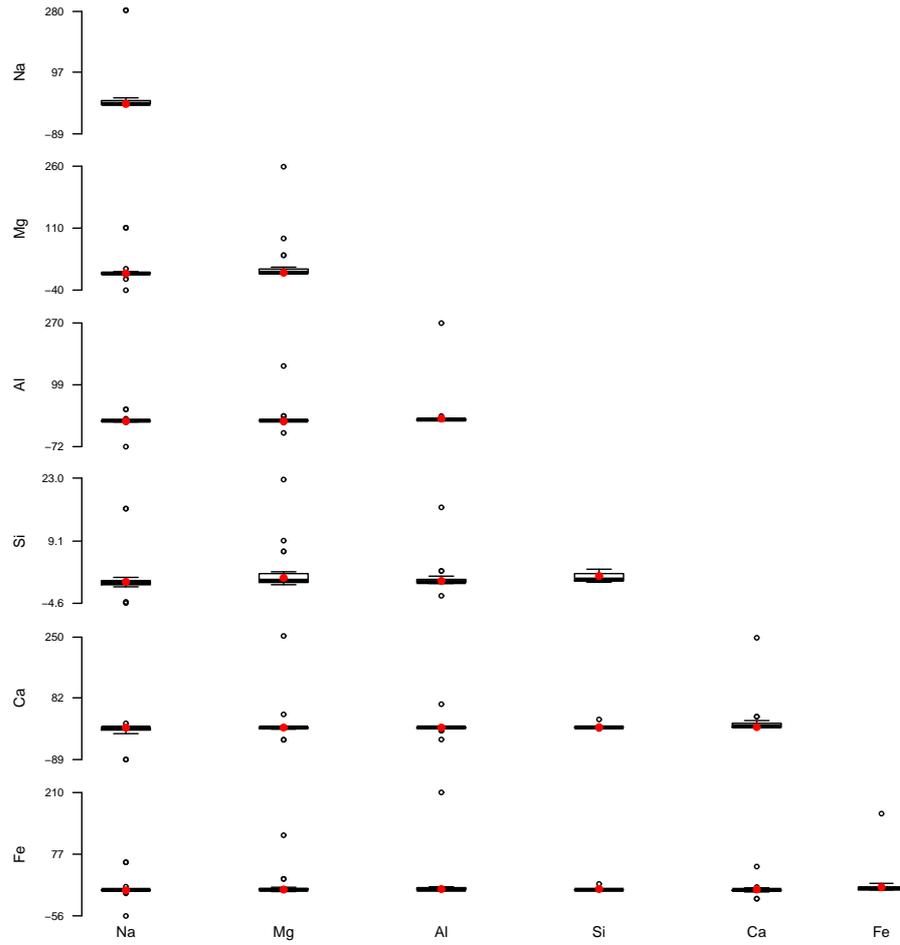


Figure E.10: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 1$ (bulb) from 20 simulated datasets for configuration $m = 3$ ($\text{Fe}, \bar{\mathbf{K}}$) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

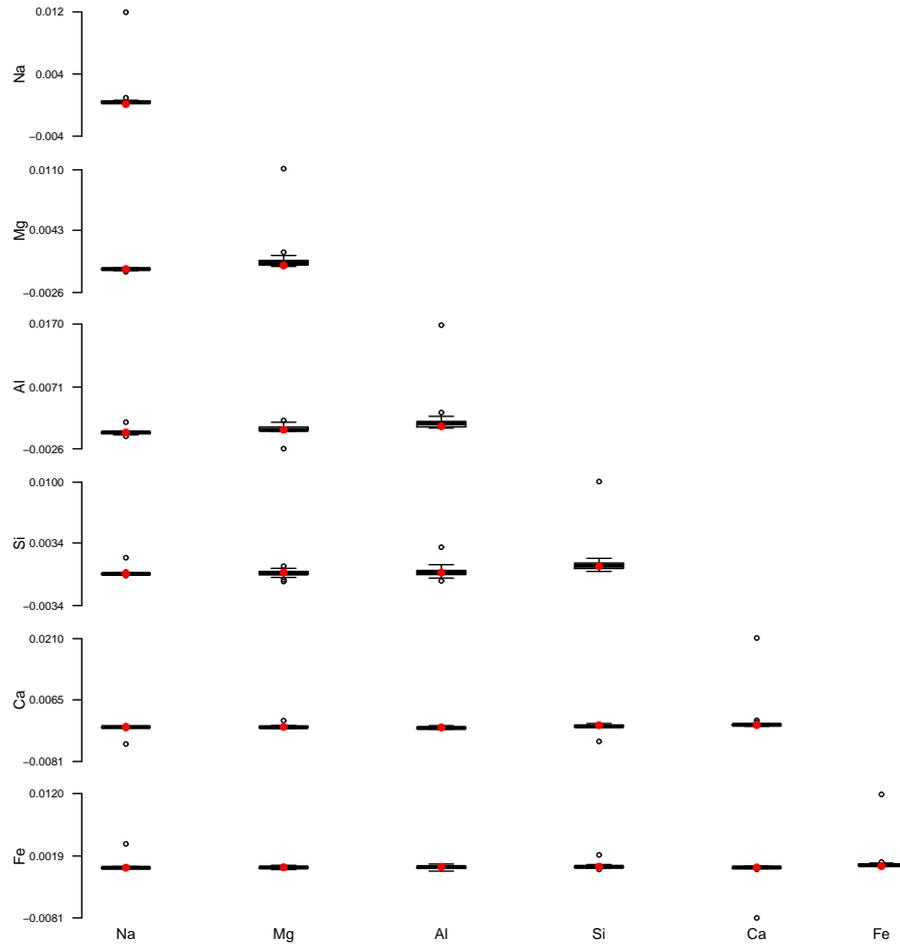


Figure E.11: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 2$ (car window) from 20 simulated datasets for configuration $m = 3$ (Fe, \bar{K}) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

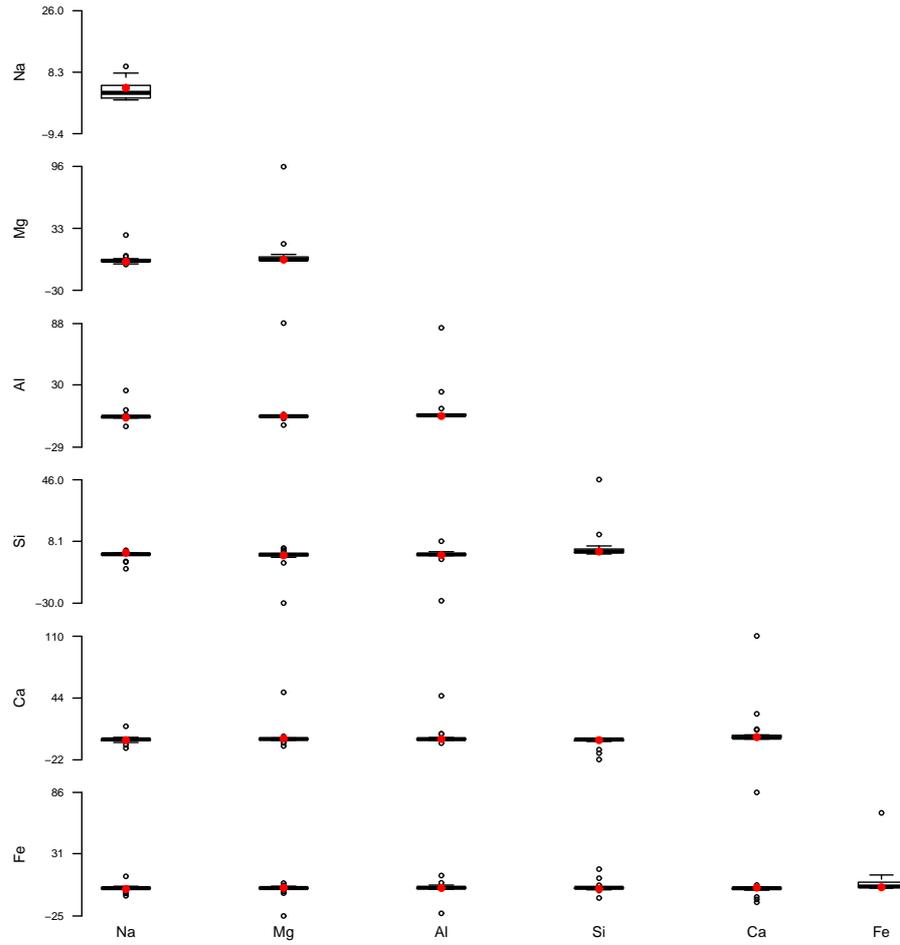


Figure E.12: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 3$ (headlamp) from 20 simulated datasets for configuration $m = 3$ (Fe, \bar{K}) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

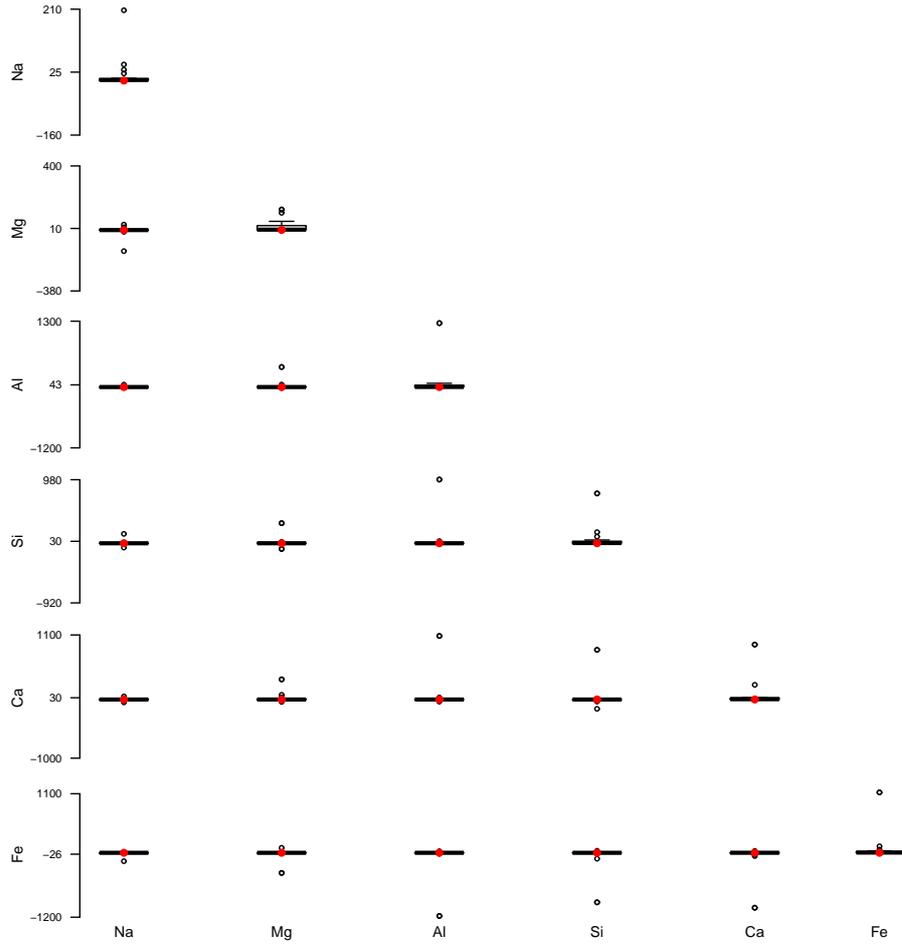


Figure E.13: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 4$ (container) from 20 simulated datasets for configuration $m = 3$ (Fe, \bar{K}) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

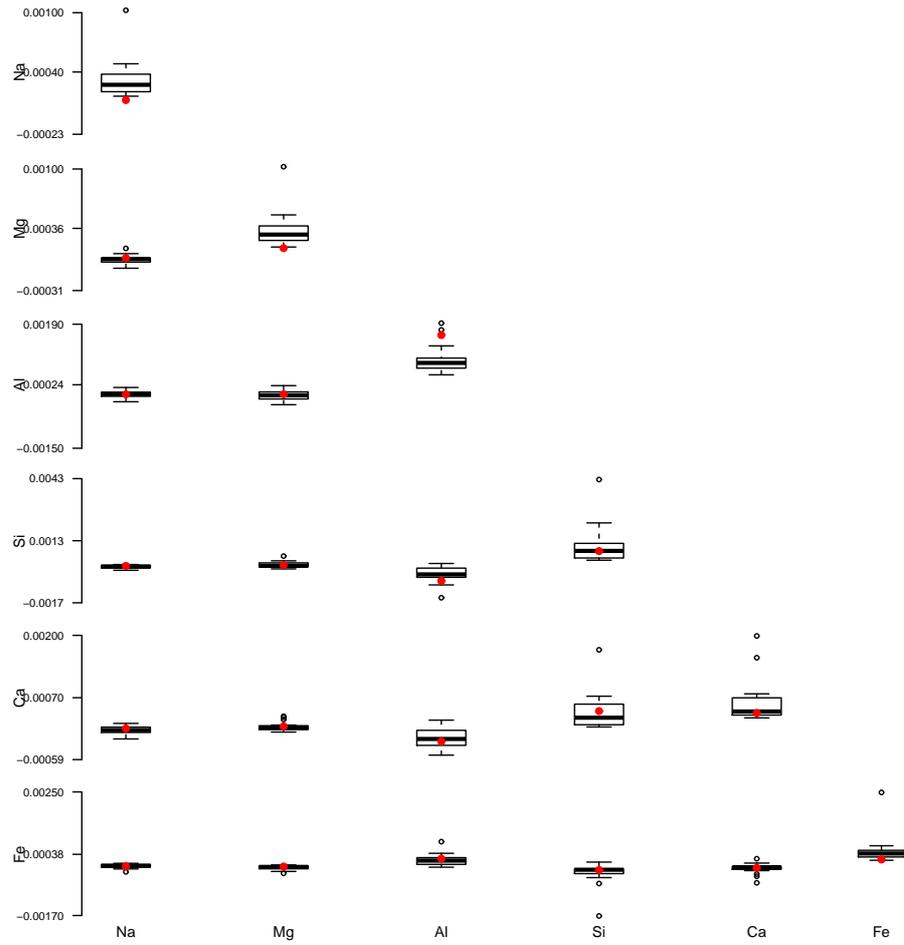


Figure E.14: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 5$ (building window) from 20 simulated datasets for configuration $m = 3$ ($\text{Fe}, \bar{\text{K}}$) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

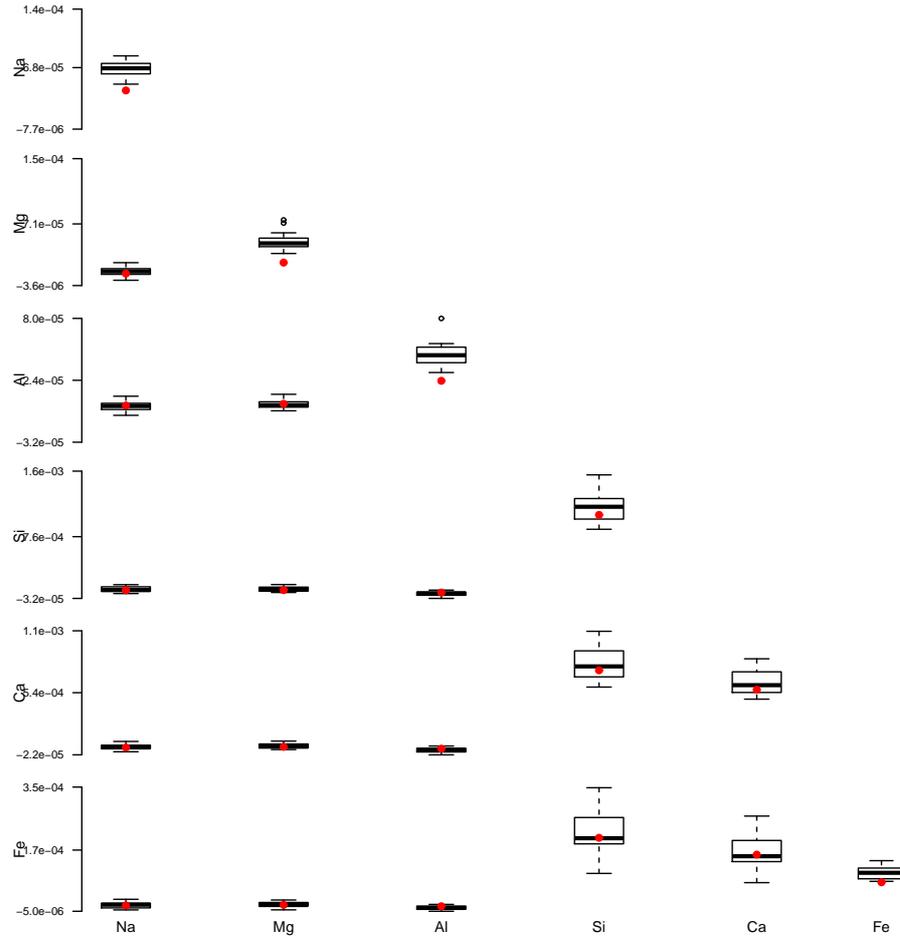


Figure E.15: Boxplots containing the mean posterior draws of Ψ^{-1} from 20 simulated datasets for configuration $m = 3$ ($\text{Fe}, \bar{\mathbf{K}}$) for the square root ratios. The red dots indicate the true values of Ψ^{-1} used to simulate the datasets.

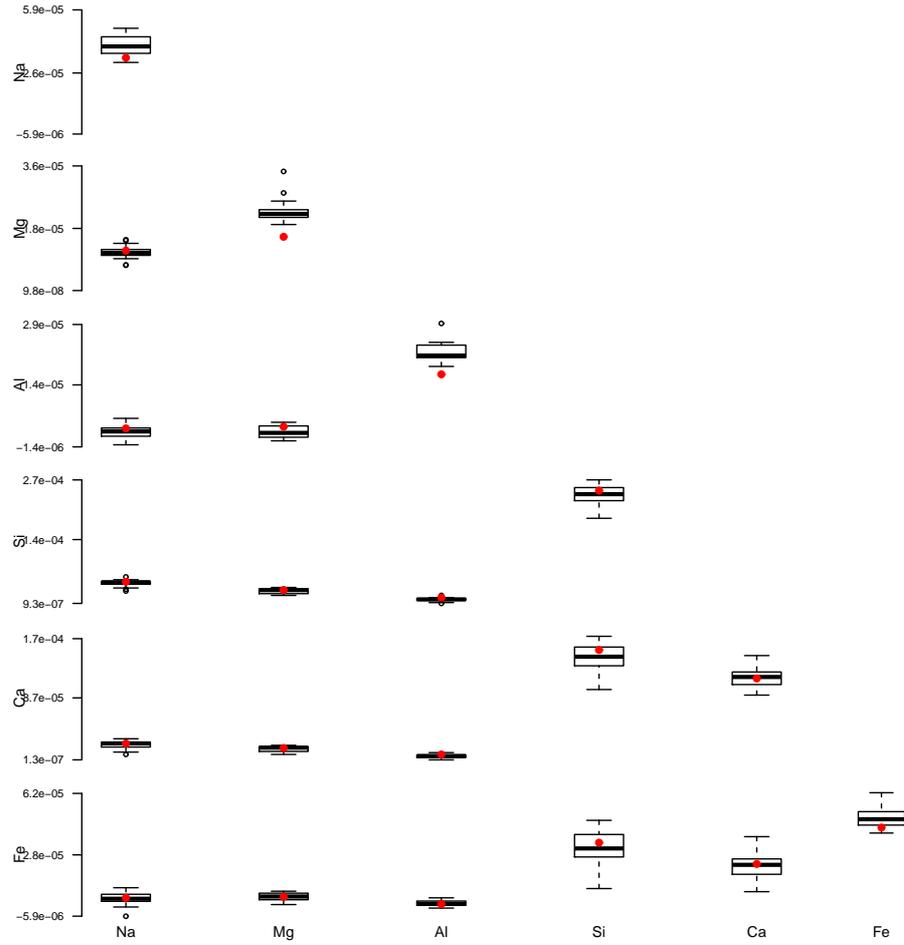


Figure E.16: Boxplots containing the mean posterior draws of Λ^{-1} from 20 simulated datasets for configuration $m = 3$ ($\text{Fe}, \bar{\mathbf{K}}$) for the square root ratios. The red dots indicate the true values of Λ^{-1} used to simulate the datasets.

Simulated data from glass items with configuration $m = 4$ ($\overline{\text{Fe}}, \overline{\text{K}}$)

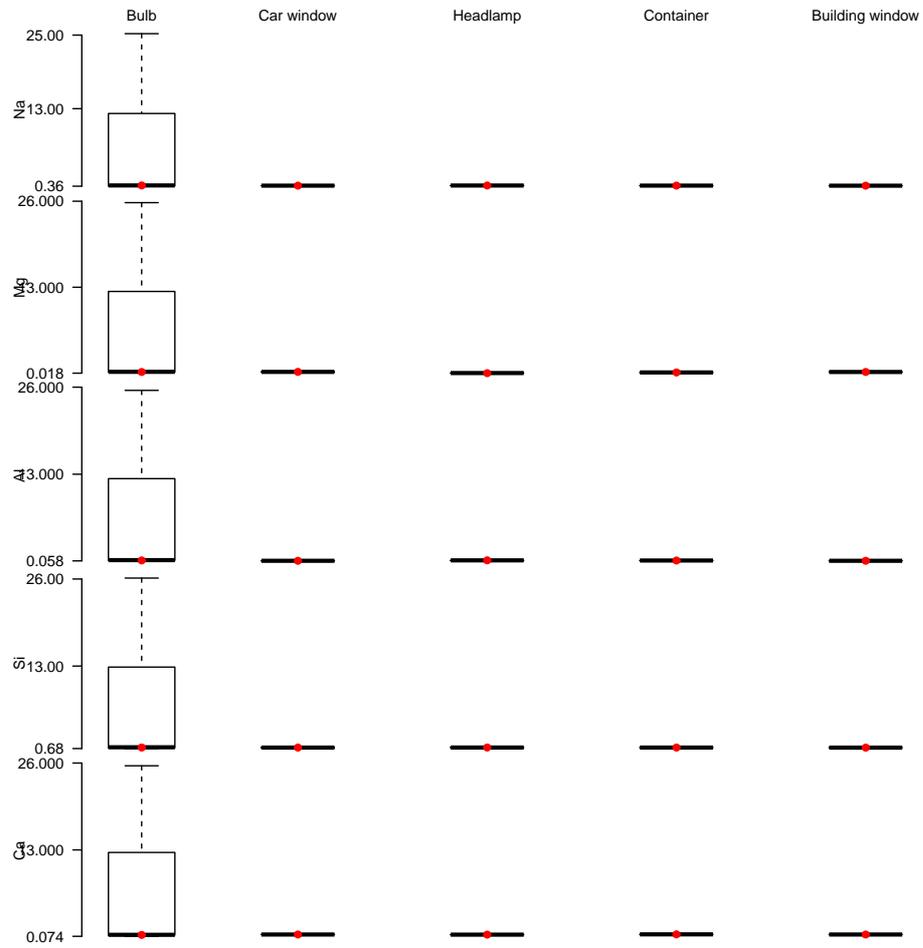


Figure E.17: Boxplots containing the mean posterior draws of θ_t from 20 simulated datasets for configuration $m = 4$ ($\overline{\text{Fe}}, \overline{\text{K}}$) for the square root ratios. The red dots indicate the true values of θ_t used to simulate the datasets.

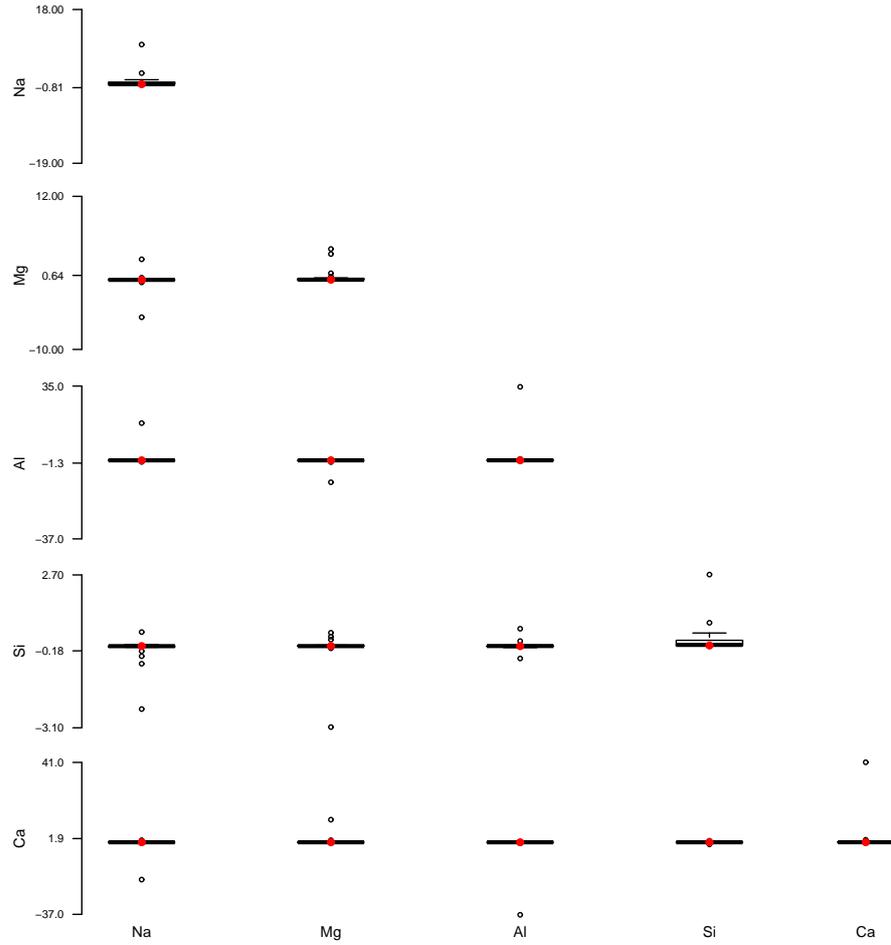


Figure E.18: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 1$ (bulb) from 20 simulated datasets for configuration $m = 4$ ($\overline{\text{Fe}}, \overline{\text{K}}$) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

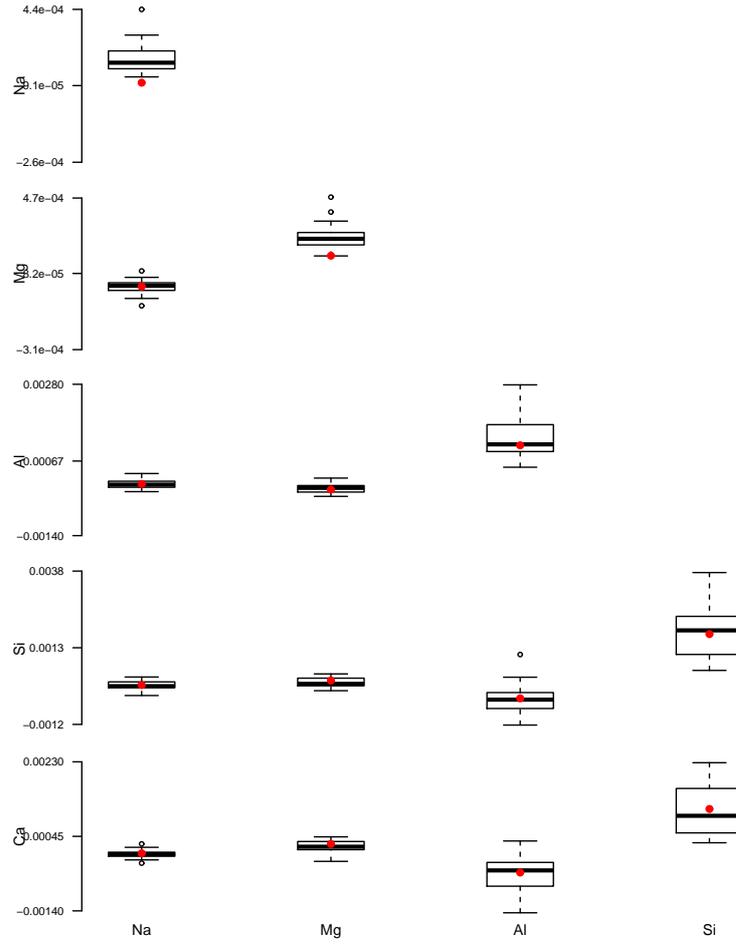


Figure E.19: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 2$ (car window) from 20 simulated datasets for configuration $m = 4$ ($\overline{\text{Fe}}, \overline{\text{K}}$) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

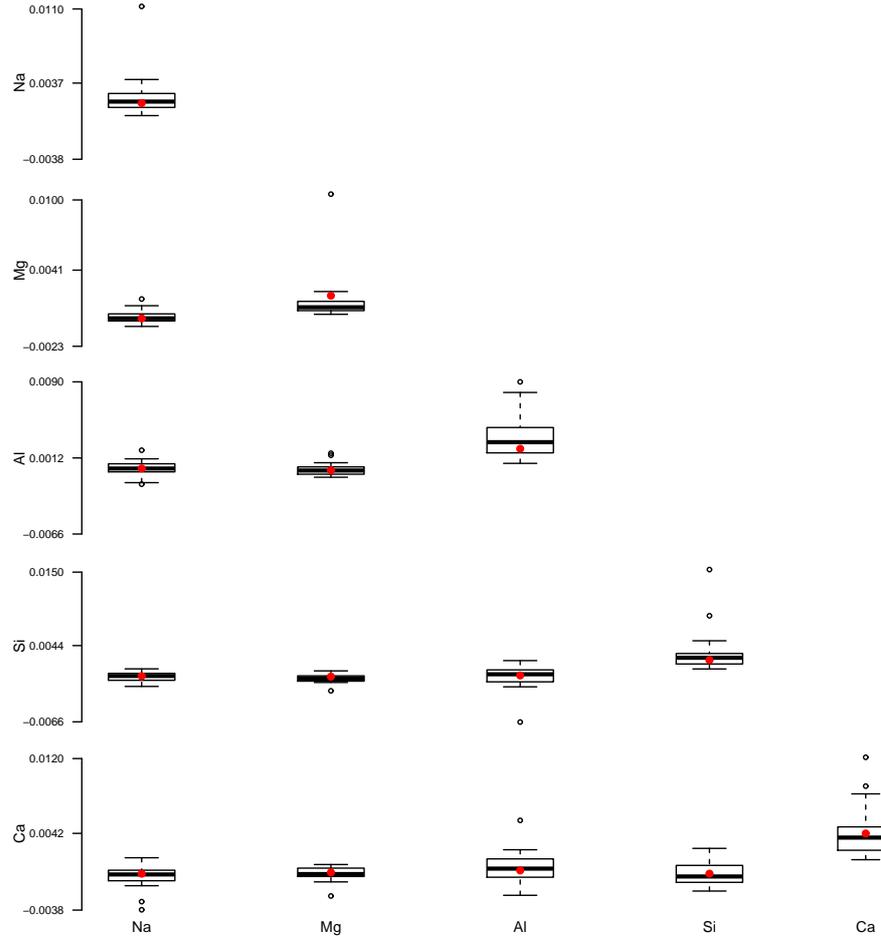


Figure E.20: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 3$ (headlamp) from 20 simulated datasets for configuration $m = 4$ ($\overline{\text{Fe}}, \overline{\text{K}}$) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

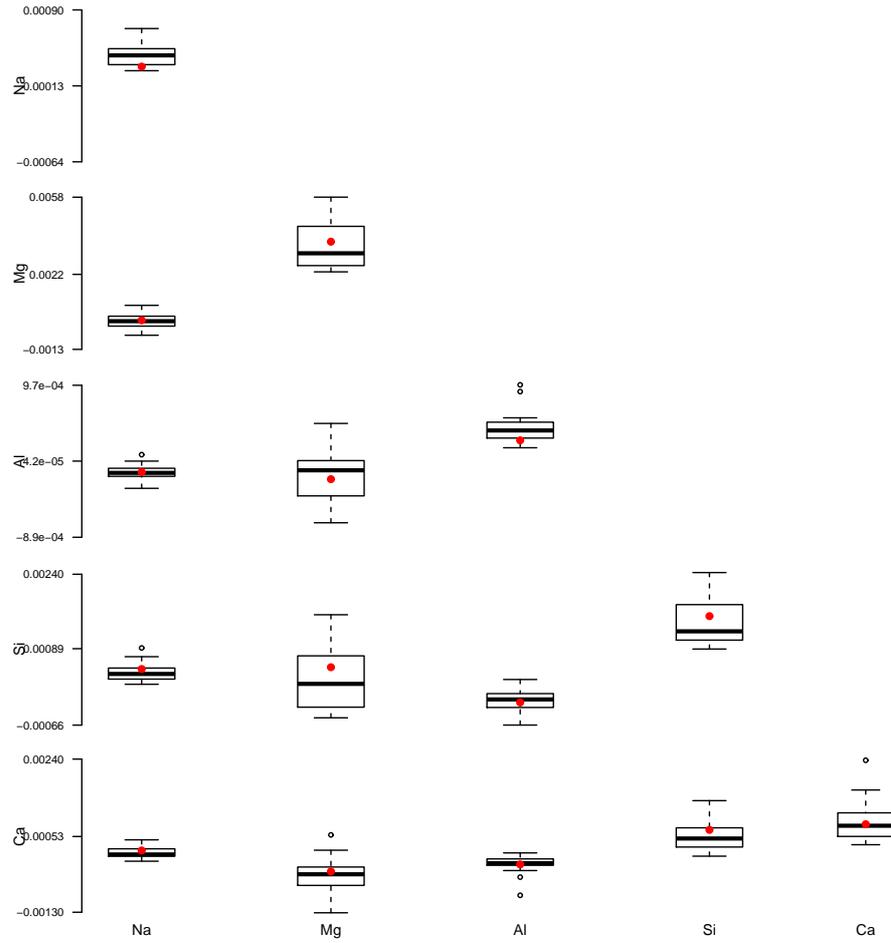


Figure E.21: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 4$ (container) from 20 simulated datasets for configuration $m = 4$ ($\overline{\text{Fe}}, \overline{\text{K}}$) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

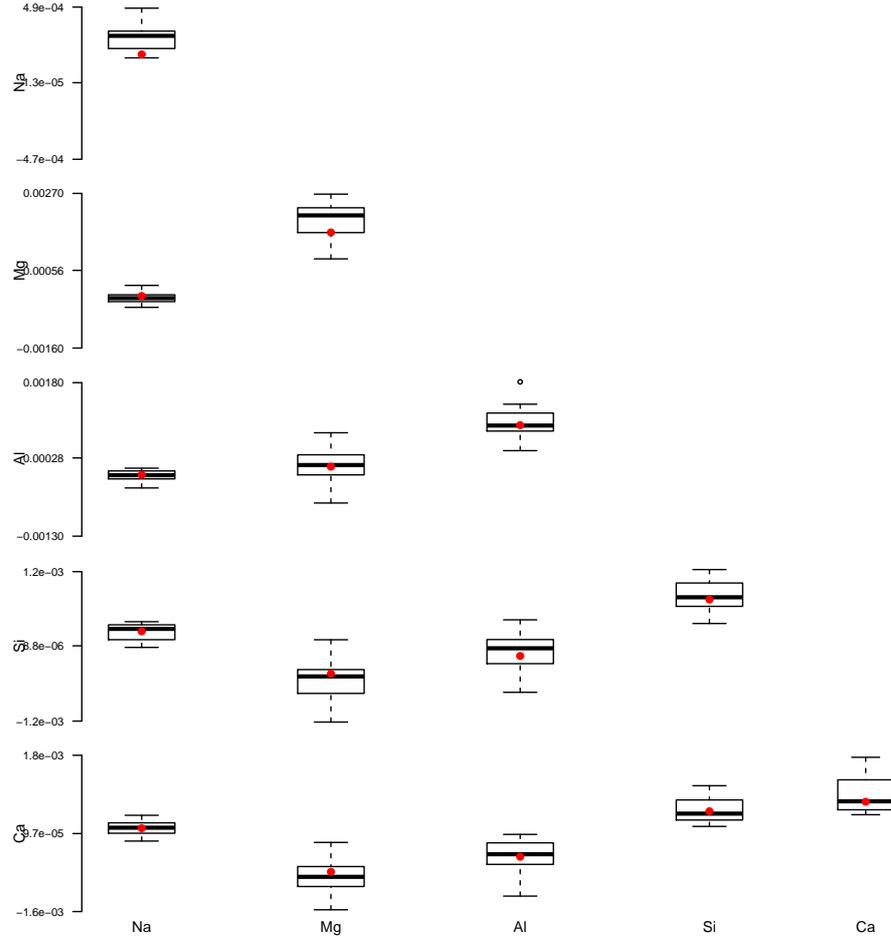


Figure E.22: Boxplots containing the mean posterior draws of Ω_t^{-1} for $t = 5$ (building window) from 20 simulated datasets for configuration $m = 4$ ($\bar{\text{Fe}}, \bar{\text{K}}$) for the square root ratios. The red dots indicate the true values of Ω_t^{-1} used to simulate the datasets.

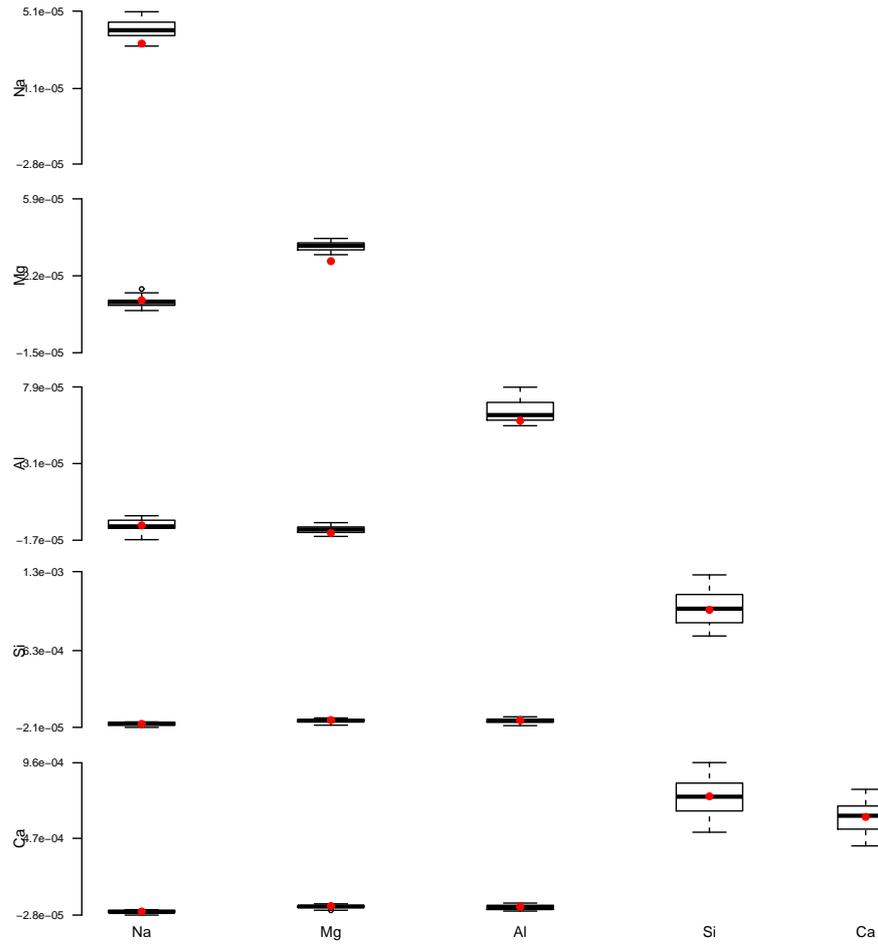


Figure E.23: Boxplots containing the mean posterior draws of Ψ^{-1} from 20 simulated datasets for configuration $m = 4$ ($\bar{\text{Fe}}, \bar{\text{K}}$) for the square root ratios. The red dots indicate the true values of Ψ^{-1} used to simulate the datasets.

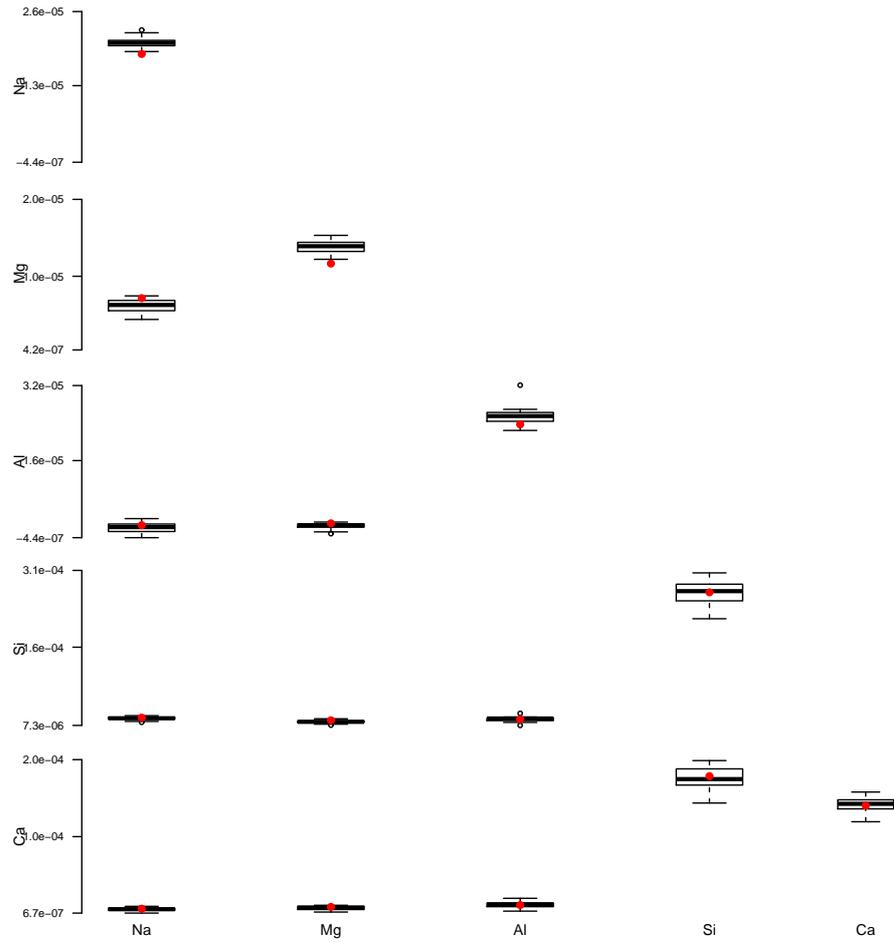


Figure E.24: Boxplots containing the mean posterior draws of Λ^{-1} from 20 simulated datasets for configuration $m = 4$ ($\bar{\text{Fe}}, \bar{\text{K}}$) for the square root ratios. The red dots indicate the true values of Λ^{-1} used to simulate the datasets.

Appendix F

Computing the evidence V for the data augmentation approach

If the elemental configurations are not considered for the control fragments \mathbf{x} and the recovered fragments \mathbf{y} then the value of the evidence V given in (5.2) is written as

$$\begin{aligned} V &= \frac{p(\mathbf{x}, \mathbf{y} | \mathcal{T}_{\mathbf{x}} = t, D, H_p)}{p(\mathbf{x}, \mathbf{y} | \mathcal{T}_{\mathbf{x}} = t, D, H_d)} \\ &= \frac{\int p(\mathbf{x}, \mathbf{y} | \mathcal{T}_{\mathbf{x}} = t, \xi, D, H_p) p(\xi | \mathcal{T}_{\mathbf{x}} = t, D, H_p) d\xi}{\int p(\mathbf{x}, \mathbf{y} | \mathcal{T}_{\mathbf{x}} = t, \xi, D, H_d) p(\xi | \mathcal{T}_{\mathbf{x}} = t, D, H_d) d\xi}. \end{aligned} \tag{F.1}$$

The first term of the integrand in the numerator of (F.1) is then

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

which gives the numerator of (F.1) as

$$\int p(\mathbf{x}, \mathbf{y} | \mathcal{T}_{(\mathbf{x}, \mathbf{y})} = t, \xi) p(\xi | D) d\xi = E_{\xi | D} [p(\mathbf{x}, \mathbf{y} | \mathcal{T}_{(\mathbf{x}, \mathbf{y})} = t, \xi)]. \quad (\text{F.2})$$

For the denominator \mathbf{x} and \mathbf{y} are assumed independent under H_d so that

$$\begin{aligned} p(\mathbf{x}, \mathbf{y} | \mathcal{T}_{\mathbf{x}} = t, \xi, D, H_d) &= p(\mathbf{x} | \mathcal{T}_{\mathbf{x}} = t, \xi, D, H_d) p(\mathbf{y} | \mathcal{T}_{\mathbf{x}} = t, \xi, D, H_d) \\ &= p(\mathbf{x} | \mathcal{T}_{\mathbf{x}} = t, \xi) p(\mathbf{y} | \xi). \end{aligned} \quad (\text{F.3})$$

As the use type $\mathcal{T}_{\mathbf{y}}$ of \mathbf{y} is uncertain the second term on the right-hand side of (F.3) can be written as

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

where $p(\mathcal{T}_{\mathbf{y}} = s)$ reflects prior knowledge of the likely use types for forensic samples. The denominator of (F.1) is then

$$\begin{aligned} \int p(\mathbf{x} | \mathcal{T}_{\mathbf{x}} = t, \xi) \left[\sum_{s=1}^T p(\mathbf{y} | \mathcal{T}_{\mathbf{y}} = s, \xi) p(\mathcal{T}_{\mathbf{y}} = s) \right] p(\xi | D) d\xi \\ = \sum_{s=1}^T p(\mathcal{T}_{\mathbf{y}} = s) E_{\xi | D} [p(\mathbf{x} | \mathcal{T}_{\mathbf{x}} = t, \xi) p(\mathbf{y} | \mathcal{T}_{\mathbf{y}} = s, \xi)]. \end{aligned} \quad (\text{F.4})$$

Substituting (F.2) and (F.4) into (F.1) gives the value V of the evidence in (5.7).

Bibliography

- Agresti, A. (1990). *Categorical data analysis*. John Wiley and Sons.
- Aitchison, J. (1982). The statistical analysis of compositional data (with discussion). *Journal of the Royal Statistical Society, Series B*, **44**(2), 139–177.
- Aitchison, J. (1986). *The statistical analysis of compositional data*. Chapman & Hall.
- Aitchison, J. (1992). On criteria for measures of compositional difference. *Mathematical Geology*, **24**(4), 365–379.
- Aitchison, J. and Egozcue, J. J. (2005). Compositional data analysis: Where are we and where should we be heading? *Mathematical Geology*, **37**(7), 829–850.
- Aitchison, J. and Kay, J. W. (2003). Possible solutions of some essential zero problems in compositional data analysis. In S. Thió-Henestrosa and J. A. Martín-Fernández, editors, *Proceedings of CODA-WORK'03, The First Compositional Data Analysis Workshop, October 15-17, University of Girona (Spain)*.

- Aitken, C. G. G. and Lucy, D. (2004). Evaluation of trace evidence in the form of multivariate data. *Journal of the Royal Statistical Society, Series C, Applied Statistics*, **53**(1), 109–122.
- Aitken, C. G. G. and Taroni, F. (2004). *Statistics and the evaluation of evidence for forensic scientists*. John Wiley and Sons.
- Bacon-Shone, J. (2003). Modelling structural zeros in compositional data. In S. Thió-Henestrosa and J. A. Martín-Fernández, editors, *Compositional data analysis workshop, Girona, Spain*. <http://ima.udg.es/Activitats/CoDaWork03/>.
- Baxter, M. J., Beardah, C. C., Cool, H. E. M., and Jackson, C. M. (2005). Compositional data analysis of some alkaline glasses. *Mathematical Geology*, **37**(2), 183–196.
- Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, **26**(2), 211–252.
- Brier, G. W. (1950). Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, **78**(1), 1–3.
- Browne, W. J. (2004). An illustration of the use of reparameterisation methods for improving MCMC efficiency in crossed random effects models. *Multilevel Modelling Newsletter, CMM*, **16**(1), 13–25.
- Brümmer, N. (2010). *Measuring, refining and calibrating speaker and language information extracted from speech*. Ph.D. thesis, School of Electrical Engineering, University of Stellenbosch, South Africa.

- Brümmer, N. and du Preez, J. (2006). Application independent evaluation of speaker detection. *Computer Speech & Language*, **20**(2-3), 230–275.
- Butler, A. and Glasbey, C. (2008). A latent Gaussian model for compositional data with zeros. *Journal of the Royal Statistical Society, Series C, Applied Statistics*, **57**(5), 505–520.
- Campbell, G. P., Curran, J. M., Miskelly, G. M., Coulson, S., Yaxley, G. M., Grunsky, E. C., and Cox, S. C. (2009). Compositional data analysis for elemental data in forensic science. *Forensic Science International*, **188**(1–3), 81–90.
- Casella, G. and George, E. I. (1992). Explaining the Gibbs Sampler. *The American Statistician*, **46**(3), 167–174.
- Chib, S. and Greenberg, E. (1995). Understanding the Metropolis-Hastings Algorithm. *The American Statistician*, **49**(4), 327–335.
- Cortes, C. and Vapnik, V. N. (1995). Support-vector networks. *Machine Learning*, **20**(3), 273–297.
- Daunis-i-Estadella, J., Martín-Fernández, J. A., and Palarea-Albaladejo, J. (2008). Bayesian tools for count zeros in compositional data sets. In *Proceedings of the 3rd international workshop on compositional data, Co-DaWork'08*.
- DeGroot, M. H. (1970). *Optimal statistical decisions*. McGraw-Hill Book Company.
- Egozcue, J. J., Pawlowsky-Glahn, V., Mateu-Figueras, G., and Barceló-

- Vidal, C. (2003). Isometric logratio transformations for compositional data analysis. *Mathematical Geology*, **35**(3), 279–300.
- Evetts, I. W., Jackson, G., Lambert, J. A., and McCrossan, S. (2000). The impact of the principles of evidence interpretation and the structure and content of statements. *Science & Justice*, **40**(4), 233–239.
- Fry, J. M., Fry, T. R. L., and McLaren, K. R. (2000). Compositional data analysis and zeros in micro data. *Applied Economics*, **32**(8), 953–959.
- Gelfand, A. E., Sahu, S. K., and Carlin, B. P. (1995). Efficient parametrizations for normal linear mixed models. *Biometrika*, **82**(3), 479–488.
- Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B. (2004). *Bayesian data analysis*. Chapman and Hall.
- Gneiting, T. and Raftery, A. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, **102**(477), 359–378.
- Green, P. J. (2003). Trans-dimensional Markov chain Monte Carlo. In P. J. Green, N. L. Hjort, and S. Richardson, editors, *Highly structured stochastic systems*, pages 179–196. Oxford: Oxford University Press.
- Hand, D. J. (2012). Assessing the performance of classification methods. *International Statistical Review*, **80**(3), 400–414.
- Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their application. *Biometrika*, **57**(1), 97–109.
- Hicks, T. C., Sermier, F. M., Goldmann, T., Brunelle, A., Champod, C., and Margot, P. (2003). The classification and discrimination of glass fragments

- using non destructive energy dispersive X-ray microfluorescence. *Forensic Science International*, **137**(2-3), 107–118.
- Kass, R. E. and Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association*, **90**(430), 773–795.
- Leininger, T. L., Gelfand, A. E., Allen, J. M., and Silander Jr, J. A. (2013). Spatial regression modeling for compositional data with many zeros. *Journal of Agricultural, Biological, and Environmental Statistics*, **18**(3), 314–334.
- Lindley, D. V. (1977). A problem in forensic science. *Biometrika*, **64**(2), 207–213.
- Link, W. A. and Eaton, M. J. (2012). On thinning of chains in MCMC. *Methods in Ecology and Evolution*, **3**(1), 112–115.
- Liu, C., Rubin, D. B., and Wu, Y. N. (1998). Parameter expansion to accelerate EM: The PX-EM algorithm. *Biometrika*, **85**(4), 755–770.
- Liu, J. S. and Wu, Y. N. (1999). Parameter expansion for data augmentation. *Journal of The American Statistical Association*, **94**, 1264–1274.
- Lucy, D. (2005). *Introduction to statistics for forensic scientists*. John Wiley and Sons.
- Martín-Fernández, J. A. and Thió-Henestrosa, S. (2006). Rounded zeros: some practical aspects for compositional data. In A. Buccianti, G. Mateu-Figueras, and V. Pawlowsky-Glahn, editors, *Compositional data analysis: from theory to practice, vol 264*. The Geological Society, London, pages 191–201.

- Martín-Fernández, J. A., Barceló-Vidal, C., and Pawlowsky-Glahn, V. (2000). Zero replacement in compositional data sets. In H. Kiers, J. Rasson, P. Groenen, and M. Shader, editors, *Proceedings of the 7th Conference of the International Federation of Classification Societies, (University of Namur (Belgium))*, pages 155–160. Springer-Verlag, Berlin.
- Martín-Fernández, J. A., Barceló-Vidal, C., and Pawlowsky-Glahn, V. (2003). Dealing with zeros and missing values in compositional data sets using nonparametric imputation. *Mathematical Geology*, **35**(3), 253–277.
- McCulloch, C. E. and Neuhaus, J. M. (2011). Misspecifying the shape of a random effects distribution: why getting it wrong may not matter. *Statistical Science*, **26**(3), 388–402.
- Metropolis, N., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machines. *The Journal of Chemical Physics*, **21**(6), 1087–1092.
- Neocleous, T., Aitken, C. G. G., and Zadora, G. (2011). Transformations for compositional data with zeros with an application to forensic evidence evaluation. *Chemometrics and Intelligent Laboratory Systems*, **109**(1), 77–85.
- Nordgaard, A., Ansell, R., Drotz, W., and Jaeger, L. (2012). Scale of conclusions for the value of evidence. *Law, Probability and Risk*, **11**(1), 1–24.
- Palarea-Albaladejo, J., Martín-Fernández, J. A., and Gómez-García, J. (2007). A parametric approach for dealing with compositional rounded zeros. *Mathematical Geology*, **39**(7), 625–645.

- Pinheiro, J. C. and Bates, D. M. (2000). *Mixed effects models in S and S-plus*. Springer.
- R Development Core Team (2011). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Ramos, D. and Gonzalez-Rodriguez, J. (2013). Reliable support: Measuring calibration of likelihood ratios. *Forensic Science International*, **230**(1-3), 156–169.
- Rayens, W. S. and Srinivasan, C. (1991a). Box-Cox transformations in the analysis of compositional data. *Journal of Chemometrics*, **5**(3), 227–239.
- Rayens, W. S. and Srinivasan, C. (1991b). Estimation in compositional data analysis. *Journal of Chemometrics*, **5**(4), 361–374.
- Royall, R. (2000). On the probability of observing misleading statistical evidence. *Journal of the American Statistical Association*, **95**(451), 760–768.
- Sanford, R. F., Pierson, C. T., and Crovelli, R. A. (1993). An objective replacement method for censored geochemical data. *Mathematical Geology*, **25**(1), 59–80.
- Scealy, J. L. and Welsh, A. H. (2011). Regression for compositional data by using distributions defined on the hypersphere. *Journal of the Royal Statistical Society, Series B*, **73**(3), 351–375.
- Scealy, J. L. and Welsh, A. H. (2014). Fitting Kent models to compositional data with small concentration. *Statistics and Computing*, **24**(2), 165–179.

- Stewart, C. (2013). Zero-inflated beta distribution for modeling the proportions in quantitative fatty acid signature analysis. *Journal of Applied Statistics*, **40**(5), 985–992.
- Stewart, C. and Field, C. (2011). Managing the essential zeros in quantitative fatty acid signature analysis. *Journal of Agricultural, Biological, and Environmental Statistics*, **16**(1), 45–69.
- Trejos, T. and Almirall, J. R. (2005). 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*, **67**(2), 396–401.
- Wang, H., Liu, Q., Mok, H. M. K., Fu, L., and Tse, W. M. (2007). A hyperspherical transformation forecasting model for compositional data. *European Journal of Operational Research*, **179**(2), 459–468.
- Wang, H. Y., Yang, Q., Qin, H., and Zha, H. (2008). Dirichlet component analysis: Feature extraction for compositional data. In *Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland*.
- Zadora, G. (2009). Classification of glass fragments based on elemental composition and refractive index. *Journal of Forensic Sciences*, **54**(1), 49–59.
- Zadora, G., Neocleous, T., and Aitken, C. (2010). A two-level model for evidence evaluation in the presence of zeros. *Journal of Forensic Sciences*, **55**(2), 371–384.
- Zadora, G., Martyna, A., Ramos, D., and Aitken, C. (2014). *Statistical anal-*

ysis in forensic science: evidential value of multivariate physicochemical data. John Wiley and Sons.