An online application for the classification and evidence evaluation of forensic glass fragments

Gary Napier[ab] Agostino Nobile[c] Tereza Neocleous[b]

June 19, 2015

Abstract

We present an easy-to-use and freely accessible online application for the analysis of forensic glass fragments. The application is browser based and takes as input `.csv` or `.txt` files containing measurements from glass fragments obtained using a scanning electron microscope with an energy-dispersive X-ray (SEM-EDX) spectrometer. The application was developed to (i) classify glass fragments into use-type categories (classification), and (ii) compute the evidential strength of two sets of fragments under competing propositions (evidence evaluation). Detailed examples of how to use the application for both tasks are given, which highlight its user-friendly interface. The suitability of the statistical methods used by the application was checked using simulation studies, and improvements upon previous methods were found in both tasks.

Keywords: Bayes factor; chemical composition; SEM-EDX analysis

1 Introduction

Glass fragments are one of many sources of forensic evidence. Fragments from a broken item can be recovered from a crime scene and sent to a forensic laboratory, where various measurements are recorded for analysis.

[a]Correspondence to: Gary Napier, School of Mathematics and Statistics, University of Glasgow, Glasgow, UK. E-mail: gary.napier@glasgow.ac.uk

[b]School of Mathematics and Statistics, University Gardens, University of Glasgow, Glasgow, G12 8QW, UK

[c]Department of Mathematics, University of York, Heslington, York, YO10 5DD, UK

Analysis of glass fragments is mainly focused on evidence evaluation, which involves computing the evidential strength of two sets of fragments (from the crime scene and from a suspect) under two competing propositions (the prosecution and defence propositions). Measurements obtained from glass fragments can also be used to help determine their use-type, thus providing additional information about the type of the glass item from which the fragments obtained from a suspect may have originated. As most glass fragments analysed are very small, their use-type cannot always be determined by their thickness or colour [1], and so measurements of physicochemical features are obtained. Here, focus is placed upon chemical composition measurements acquired from using a scanning electron microscope with an energy-dispersive X-ray (SEM-EDX) spectrometer [1]. The elemental composition data consist of the percentage weights (wt%) of the main elements comprising a glass fragment.

In this paper, we present an easy-to-use online application for the purposes of (i) classifying glass fragments into use-type categories (classification), and (ii) computing the evidential strength of two sets of fragments under complementary propositions (evidence evaluation). The application is easily accessible and straightforward to use for both tasks. It is available at `http://gnapier.shinyapps.io/GlassClassificationAndEvaluation/` and was developed using the `shiny` package, which is part of the statistical programming language R [2].

The paper is organised as follows: Section 2 describes the database used in the development of the application. Section 3 summarises the statistical model, and the classification and evidence evaluation methods developed in [3] and used by the application. Section 4 provides examples of how to use the application in the classification and evidence evaluation tasks. Section 5 discusses how the evidence evaluation results are reported. Concluding remarks are provided in Section 6.

2 Training data

The database used in the development of the application was provided by the Institute of Forensic Research, Krakow, and it consists of measurements obtained in an experimental setting using a scanning electron microscope with an energy-dispersive X-ray (SEM-EDX) spectrometer [1]. SEM-EDX analysis produces measurements, in the form of percentage weights (wt%), on the main chemical elements that comprise the composition of a glass fragment. These are oxygen (O), sodium (Na), magnesium (Mg), aluminium (Al), silicon (Si), potassium (K), calcium (Ca) and iron (Fe). The database consists of glass fragments from 320 glass items across five use-types (26 bulbs, 94 car windows, 16 headlamps, 79 containers and 105 building windows). The chemical compositions of four glass fragments from each item were measured three times. Thus, the database has a hierarchical structure: three replicate measurements on four fragments from each of 320 glass items of five possible use-types.

As the measurements obtained from SEM-EDX are compositional, there is, as frequently happens with compositional data, a large number of zero measurements. Prior to model building, the percentage weights were transformed, by taking square roots of the ratios between each element weight and the weight of oxygen. The square root transformation was employed because it turned out to be more effective at stabilizing the variability of the ratios; see [3] for further details on the choice of this transformation. The statistical model used by the application, and the methods used for classification and evidence evaluation, will be described in Section 3.

3 Methods

This section only provides a summary of the model and methods developed in [3], to which the reader is referred for full details.

3.1 Statistical model

A Bayesian mixed-effects model was used to account for the hierarchical structure of the database. The model incorporates a fixed effect for the mean of each use-type and three random effects: at item level, fragment level, and replicate measurement level. Denote the square root ratios from the k-th replicate measurement on the j-th fragment of the i-th glass item of use-type t by the p-dimensional vector $\mathbf{z}_{tijk}$. It is then assumed that

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

The fixed effect for the mean of use-type t is denoted by $\boldsymbol{\theta}_t$; the item level random effect by $\boldsymbol{b}_{ti}$; the fragment within-item random effect by $\boldsymbol{c}_{tij}$; and the error at measurement level by $\boldsymbol{\epsilon}_{tijk}$. The random effects are assumed to have multivariate normal distributions, with unknown precision (i.e. inverse covariance) matrices Ω_t, Ψ and Λ. Then, for a glass item $\mathbf{z}$ of use type $\mathcal{T}_{\mathbf{z}} = t$ with JK measurements, the distribution of $\mathbf{z}$ is

$$\mathbf{z}|\mathcal{T}_{\mathbf{z}} = t, \xi \sim N_{JKp}(\mathbf{1}_{JK} \otimes \boldsymbol{\theta}_t,\, \Sigma_t), \tag{2}$$

where $\xi = \{\boldsymbol{\theta}, \Omega, \Psi, \Lambda\}$ collectively denotes the model parameters and $\mathbf{1}_d$ is a column vector of d 1's. The covariance matrix Σ_t is given by

$$\Sigma_t = (\mathbf{1}_{JK}\mathbf{1}'_{JK}) \otimes \Omega_t^{-1} + \left[\mathbb{I}_J \otimes (\mathbf{1}_K\mathbf{1}'_K)\right] \otimes \Psi^{-1} + \mathbb{I}_{JK} \otimes \Lambda^{-1}, \qquad (3)$$

where $\mathbb{I}_d$ is the $d \times d$ identity matrix.

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

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

The covariance matrix Φ^{-1} is fixed. The precision matrices for the random effects have conjugate Wishart priors placed upon them:

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

For more details on the prior and the Markov Chain Monte Carlo (MCMC) methods used see [3]. It is worth highlighting that the application does not need to run any MCMC as it uses the posterior draws obtained from modelling the database, thus making the application quick to use. The flip side of this point is that the application is not designed to re-estimate the model using a different background database, possibly available to the potential user.

As briefly mentioned in Section 2, the database contains a large proportion of zeros. To handle these zeros the background database was partitioned into subsets based on elemental configurations. The elemental configurations denote whether an element is present (above detection limit) or absent (below detection limit) from the composition of a glass item. The background database consists of glass items with ten different elemental configurations, as shown in Figure 1. However, as the elements iron and potassium are responsible for the majority of the zeros, focus is placed on the presence or absence of these two elements only, thus reducing the number of configurations from ten to four. A Bayesian hierarchical model like (1) is then estimated for each subset of the background database for the four elemental configurations. For details on how the Bayesian hierarchical models for the four elemental configurations are brought together to form a composite model see [3].

3.2 Classification

Being able to predict the use-type of a glass fragment can help at the investigation stage of a legal case. To classify fragments, the application uses the

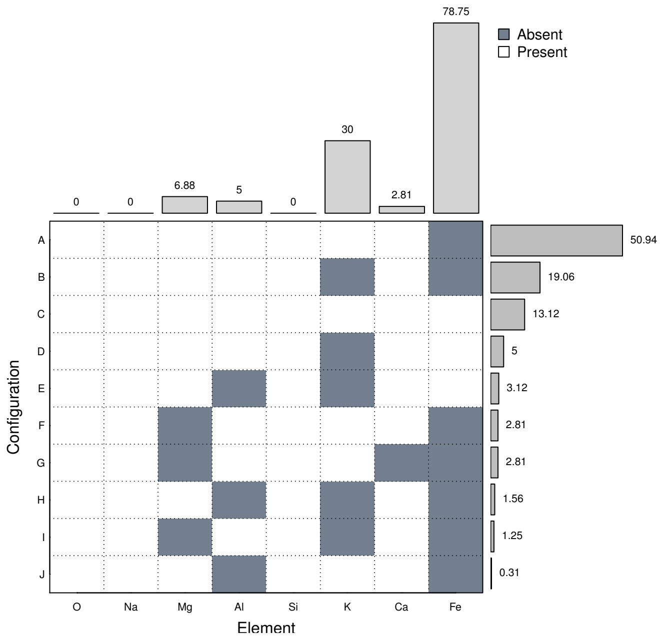


Figure 1: Plot displaying the ten elemental configurations at item level observed in the glass database. The percentage of compositional zeros by element is also shown in the barplot at the top; the percentage of data associated with each configuration is shown in the barplot on the right. The configurations used in modelling are coarser and only consider presence and absence of Fe and K. The map between these and the ones in the plot is as follows: $(\mathrm{Fe}, \mathrm{K}) = \{C\}$, $(\overline{\mathrm{Fe}}, \mathrm{K}) = \{A, F, G, J\}$, $(\mathrm{Fe}, \overline{\mathrm{K}}) = \{D, E\}$, $(\overline{\mathrm{Fe}}, \overline{\mathrm{K}}) = \{B, H, I\}$, where absence is denoted with a bar over the chemical element.

posterior predictive distribution of the use type $\mathcal{T}_{\mathbf{y}}$ of a newly observed glass item's measurement vector $\mathbf{y}$ to be classified, conditional on the background database D described in Section 2, and the new item $\mathbf{y}$. Let $\mathbf{y}$ be a vector consisting of $\tilde{K}$ replicate measurements on each of $\tilde{J}$ fragments from the same glass item. The use-type probability of $\mathbf{y}$ is given by

$$p(\mathcal{T}_{\mathbf{y}} = t|\mathbf{y}, D) \propto p(\mathcal{T}_{\mathbf{y}} = t) \frac{\alpha_{tm} + N_{tm}}{\sum_{r=1}^{M}(\alpha_{tr} + N_{tr})} E_{\xi_m|D_m}[p(\mathbf{y}|\mathcal{T}_{\mathbf{y}} = t, \mathcal{C}_{\mathbf{y}} = m, \xi_m)]. \tag{4}$$

The first two expressions on the right-hand side of (4) derive from modelling the counts N_{tm} of items in D that are of use-type t and configuration m. The expressions give the use-type probabilities for a newly observed glass item of use-type $\mathcal{T}_{\mathbf{y}}$ given it has elemental configuration $\mathcal{C}_{\mathbf{y}} = m$, without conditioning on the actual measurements $\mathbf{y}$. These use-type probabilities are reported in Table 1 for the case where one assumes that $p(\mathcal{T}_{\mathbf{y}} = t) = 1/5$ and $\alpha_{tm} = 0.1$ for all t and m, see [3] for further details on how they were obtained.

Table 1: Use-type probabilities for a new item with measurements $\mathbf{y}$ of use-type $\mathcal{T}_\mathbf{y}$ given it has elemental configuration $\mathcal{C}_\mathbf{y} = m$. The presence of iron and potassium is denoted by Fe and K, while their absence is denoted by $\overline{\mathrm{Fe}}$ and $\overline{\mathrm{K}}$.

Glass type	$\mathcal{C}_\mathbf{y} = m$			
	1: Fe, K	2: $\overline{\mathrm{Fe}}$, K	3: Fe, $\overline{\mathrm{K}}$	4: $\overline{\mathrm{Fe}}$, $\overline{\mathrm{K}}$
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 third term on the right-hand side of (4) is the predictive density of $\mathbf{y}$, according to the Bayesian hierarchical model (1) for the subset D_m of the background database with elemental configuration m. This predictive density is given as the expectation of the density $p(\mathbf{y}|\mathcal{T}_\mathbf{y} = t, \mathcal{C}_\mathbf{y} = m, \xi_m)$ with respect to the posterior distribution of the model parameters ξ_m given the data D_m. In practice, the expectation is estimated by averaging the densities $p(\mathbf{y}|\mathcal{T}_\mathbf{y} = t, \mathcal{C}_\mathbf{y} = m, \xi_m)$ over a posterior sample of ξ_m's, obtained by means of MCMC methods.

The application classifies glass fragments into one of the five use-types contained in the database: bulb, car window, headlamp, container and building window. The classification results consist of the five posterior probabilities of the use-types computed according to (4), with the glass fragments classified to the use-type with the largest posterior probability. Results of the performance of the model in the classification task can be found in [3].

3.3 Evidence evaluation

Measurements from glass fragments obtained from a suspect can be used as evidence in support (or against) the proposition that the suspect was involved in the case under investigation. The statistical approach used to evaluate the strength V of such evidence stems from Lindley [4]. A recent overview is provided in [5, Chapter 10], whose terminology we adopt. The strength of the evidence is given by the Bayes factor

$$V = \frac{p(E|H_p, I)}{p(E|H_d, I)}, \tag{5}$$

where $E = (\mathbf{x}, \mathbf{y})$ is the evidence, H_p and H_d the prosecution and defence propositions, and I additional background information related to the case.

From the evidence E, $\mathbf{x}$ denotes the measurements obtained from a sample of glass fragments found at the crime scene, and $\mathbf{y}$ denotes the measurements from glass fragments found on the suspect. The prosecution proposition, H_p, would be that the measurements $\mathbf{y}$ are from the same item as the measurements $\mathbf{x}$. The defence proposition, H_d, would be that $\mathbf{y}$ is not from the same item as $\mathbf{x}$. Using the Bayesian hierarchical model (1), with the details found in [3], V can be rewritten as

$$V = \frac{E_{\xi_m|D_m}\left[p(\mathbf{x}, \mathbf{y}|\mathcal{T}_{(\mathbf{x},\mathbf{y})} = t, \mathcal{C} = m, \xi_m)\right]}{\sum_{s=1}^{T} p(\mathcal{T}_{\mathbf{y}} = s|\mathcal{C} = m, D)\, E_{\xi_m|D_m}[p(\mathbf{x}|\mathcal{T}_{\mathbf{x}} = t, \mathcal{C} = m, \xi_m)\, p(\mathbf{y}|\mathcal{T}_{\mathbf{y}} = s, \mathcal{C} = m, \xi_m)]}. \tag{6}$$

Under H_p both $\mathbf{x}$ and $\mathbf{y}$ are assumed to come from the same glass item, thus their use-type is known. The application provides a drop-down menu for the user to choose the use-type of the control fragments collected from the crime scene. Under H_d the two sets of fragments are assumed to come from different sources, and are therefore independent. This means that there is uncertainty surrounding the origin, and thus use-type, of the recovered fragments from the suspect. This is why the denominator of V in (6) is a weighted average across the different use-types. As a reference, the application returns the evidential value V on two different scales. The scales used are the verbal scale of Evett [6], and the scale used by the Swedish National Laboratory of Forensic Science (SKL) [7]. Performance of the model in the evidence evaluation task can be found in [3]. The next section contains examples of how to use the application for both the classification and evidence evaluation tasks.

4 Examples

This section is a walk-through of some examples of how to use the application, starting first with examples of classifying glass fragments into use-type categories.

4.1 Classification examples

To use the application to classify glass fragments, the user selects the classification tab and then loads a `.csv` or `.txt` data file containing measurements from glass fragments obtained using SEM-EDX analysis. Figure 2 displays the classification application screen and shows where the file with the data is to be loaded. Figure 2 displays an example data file and a description of what the data files uploaded to the application are expected to contain.

Figure 3 displays an example from classifying measurements known to come from a glass item of the use-type bulb. Once data has been uploaded to the application the example data file is replaced by the uploaded data, for the user to check that the data has been loaded correctly. This is also shown in Figure 3 with the bulb measurements consisting of three replicate measurements from two fragments. In this example the measurements where correctly classified, with probability one, as being from a glass item of use-type bulb. As shown from the simulation study in [3], the application is very reliable at classifying measurement from bulbs, however, due to their similar chemical make up, there is more uncertainty surrounding the classification of car and building windows. This is shown in Figure 4 with measurements from a car window being incorrectly classified as being from a building window; whenever a window type is misclassified it is most often misclassified as the other window type.

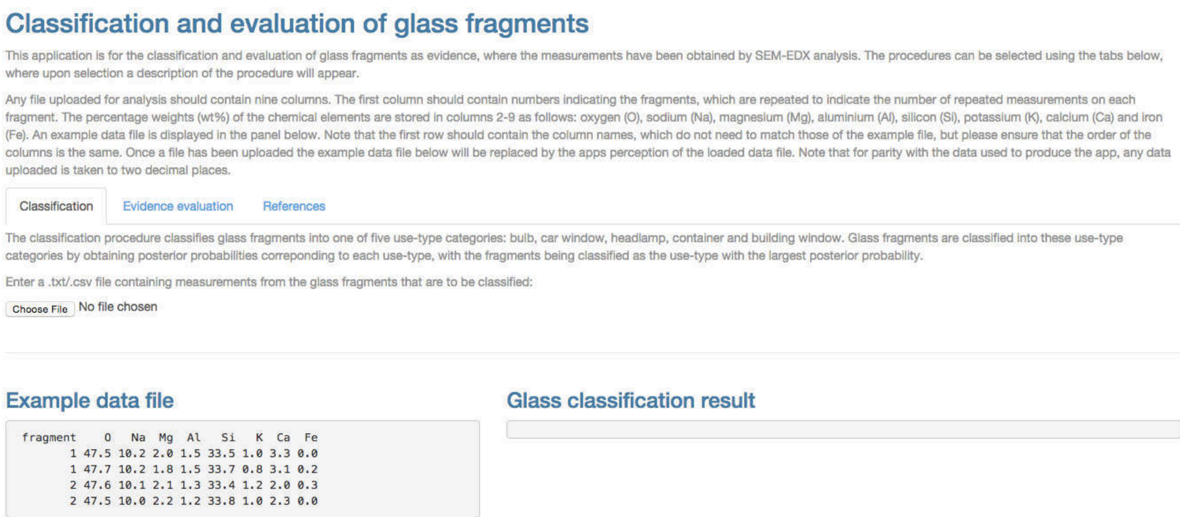


Figure 2: Screenshot of the classification application screen. Here the application informs the user on the types of data file to be used, that is .csv/.txt files containing nine columns, with importance placed on the ordering of the columns. To assist the user, an example data file of measurements from two fragments, each with two replicate measurements, is also displayed.

Loaded data file

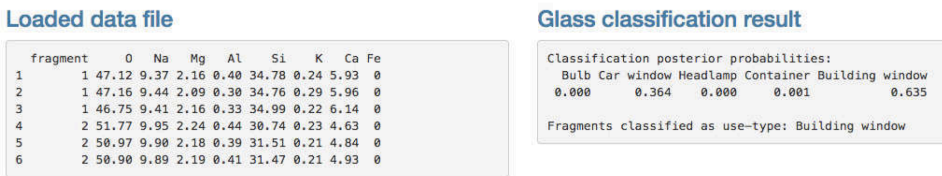
```
  fragment     O    Na   Mg   Al    Si    K   Ca Fe
1        1 52.32  5.22 0.72 1.78 34.19 4.32 1.45  0
2        1 51.91  5.15 0.72 1.91 34.37 4.49 1.45  0
3        1 52.10  5.17 0.78 1.91 34.17 4.35 1.52  0
4        2 47.35 10.39 2.06 0.88 33.67 0.63 5.02  0
5        2 48.44 10.88 2.05 0.87 32.59 0.62 4.55  0
6        2 49.12 10.81 2.07 0.88 31.97 0.56 4.59  0
```

Glass classification result

```
Classification posterior probabilities:
 Bulb Car window Headlamp Container Building window
    1          0        0         0               0

Fragments classified as use-type: Bulb
```

Figure 3: Classification example 1: classifying measurements from a bulb. The results consist of the five use-type posterior probabilities and the use-type to which the fragments have been classified to. Here the measurements from the bulb have been correctly classified with posterior probability 1.



Loaded data file

```
  fragment     O   Na   Mg   Al    Si    K   Ca Fe
1        1 47.12 9.37 2.16 0.40 34.78 0.24 5.93  0
2        1 47.16 9.44 2.09 0.30 34.76 0.29 5.96  0
3        1 46.75 9.41 2.16 0.33 34.99 0.22 6.14  0
4        2 51.77 9.95 2.24 0.44 30.74 0.23 4.63  0
5        2 50.97 9.90 2.18 0.39 31.51 0.21 4.84  0
6        2 50.90 9.89 2.19 0.41 31.47 0.21 4.93  0
```

Glass classification result

```
Classification posterior probabilities:
  Bulb Car window Headlamp Container Building window
 0.000      0.364    0.000     0.001           0.635

Fragments classified as use-type: Building window
```

Figure 4: Classification example 2: classifying measurements from a car window. The results consist of the five use-type posterior probabilities and the use-type to which the fragments have been classified to. Here the measurements from the car window have been classified with posterior probability 0.635 as coming from a building window, and with posterior probability 0.364 as coming from a car window.

4.2 Evidence evaluation examples

To use the application to evaluate the evidential strength of two sets of fragments, the user first selects the evidence evaluation tab before uploading two separate `.csv`/`.txt` files. The first set of measurements to be uploaded to the application should be from fragments obtained from the crime scene, as seen in Figure 5. As the use-type of fragments obtained from the crime scene is known, this must be selected using the displayed drop-down menu. There is no drop-down menu to select the use-type of the fragments obtained from a suspect, as there is uncertainty surrounding their origin.

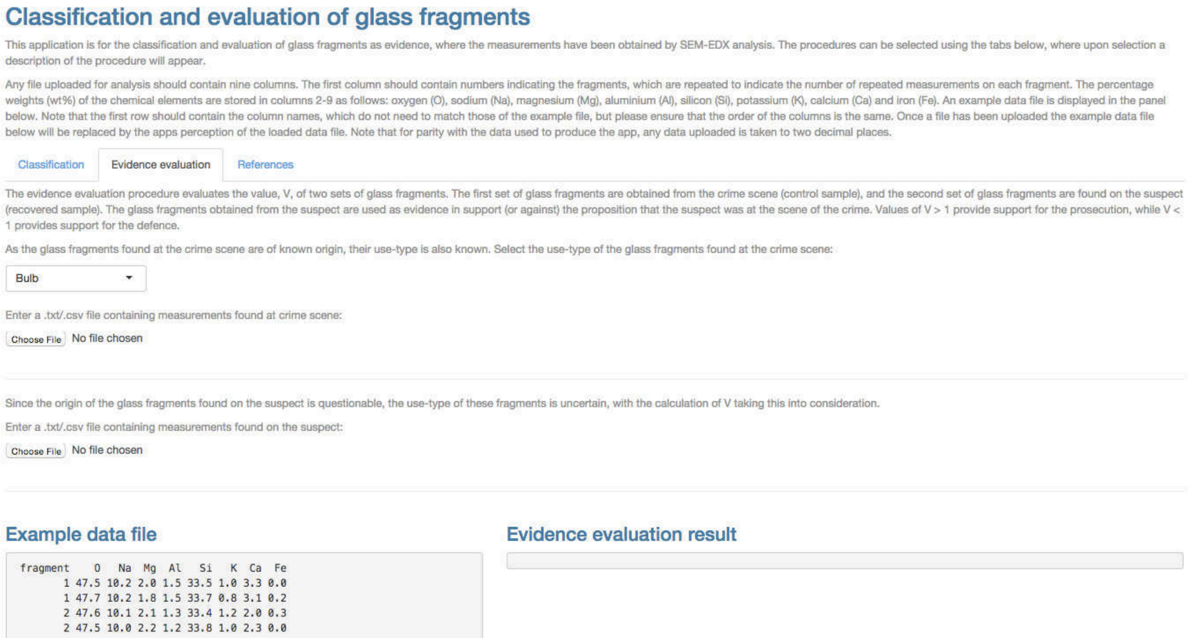


Figure 5: Screenshot of the evidence evaluation tab showing where the two files containing the two sets of measurements should be uploaded. As the glass item from which the fragments obtained from the crime scene is known, its use-type is also known, and so a drop-down menu for the user to select the use-type of these fragments is included. There is no option to select the use-type of the fragments obtained from the suspect, as their is uncertainty surrounding their origin.

Figure 6 shows the results from a same-source comparison. This example contains measurements known to be from the same glass item for the control (from crime scene) and recovered (from suspect) samples. The two separate files containing the measurements from the crime scene and suspect are also displayed in Figure 6, for the user to ensure that the files were uploaded correctly. For a same-source comparison we would expect V to have a large value. Here the result concludes moderate support in favour of H_p (same-source proposition).

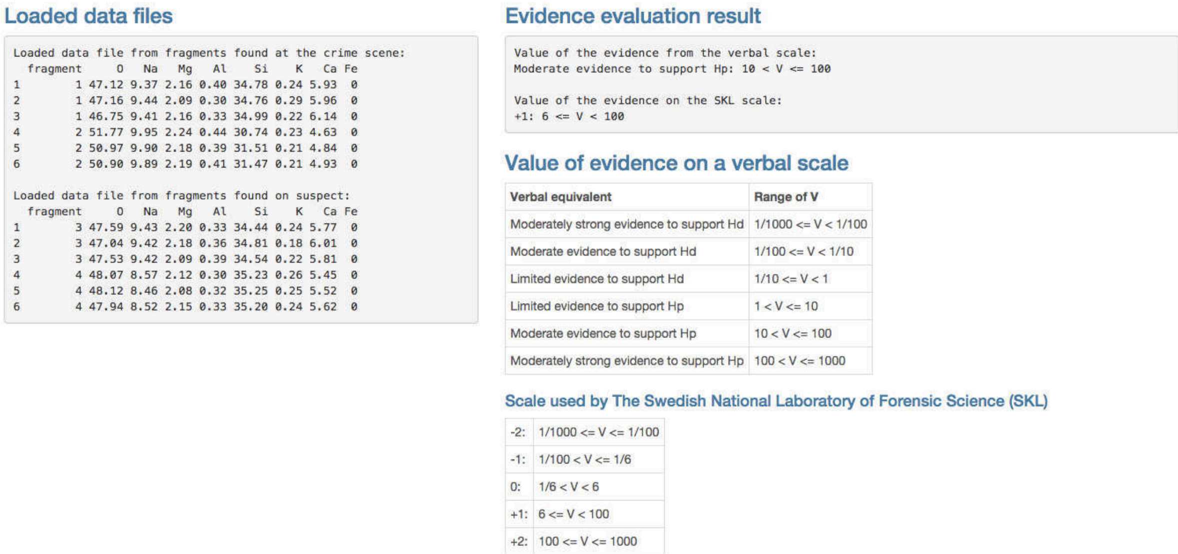


Figure 6: Evidence evaluation example 1: The strength of the evidence V obtained from two sets of measurements that are known to come from the same glass item (same-source comparison). For a same-source comparison like this we would expect a large value for V. Here the evidence results in moderate support in favour of H_p (same-source proposition).

Figure 7 shows the results from a different-source comparison. This example consists of measurements from fragments from the crime scene and subject that are known to come from different glass items. For a different-source comparison we would expect to obtain support for the defence proposition H_d (no support for H_p), that is $V < 1$. As can be seen from Figure 7 the strength of this evidence was found to provide moderately strong evidence to support H_d (different-source proposition).

Loaded data files

```
Loaded data file from fragments found at the crime scene:
  fragment     O    Na   Mg   Al    Si    K   Ca   Fe
1        1 52.21 10.51 2.24 0.42 29.92 0.17 4.24 0.29
2        1 51.78 10.50 2.28 0.36 30.17 0.23 4.38 0.30
3        1 51.30 10.26 2.17 0.42 30.75 0.22 4.58 0.30
4        2 50.20 10.25 2.35 0.35 31.52 0.23 4.79 0.31
5        2 49.29 10.05 2.22 0.36 32.26 0.21 5.17 0.44
6        2 49.89 10.09 2.27 0.37 31.92 0.12 4.99 0.35

Loaded data file from fragments found on suspect:
  fragment     O   Na   Mg   Al    Si    K   Ca   Fe
1        1 48.50 9.73 2.27 0.20 32.62 0.11 6.31 0.26
2        1 48.98 9.65 2.28 0.15 32.35 0.09 6.30 0.20
3        1 48.57 9.66 2.30 0.23 32.61 0.13 6.31 0.19
4        2 46.84 9.35 2.21 0.21 34.07 0.11 7.01 0.20
5        2 47.01 9.30 2.21 0.23 33.97 0.11 6.96 0.21
```

Evidence evaluation result

```
Value of the evidence from the verbal scale:
Moderately strong evidence to support Hd: 1/1000 <= V < 1/100

Value of the evidence on the SKL scale:
-2: 1/1000 <= V <= 1/100
```

Value of evidence on a verbal scale

Verbal equivalent	Range of V
Moderately strong evidence to support Hd	1/1000 <= V < 1/100
Moderate evidence to support Hd	1/100 <= V < 1/10
Limited evidence to support Hd	1/10 <= V < 1
Limited evidence to support Hp	1 < V <= 10
Moderate evidence to support Hp	10 < V <= 100
Moderately strong evidence to support Hp	100 < V <= 1000

Scale used by The Swedish National Laboratory of Forensic Science (SKL)

-2:	1/1000 <= V <= 1/100
-1:	1/100 < V <= 1/6
0:	1/6 < V < 6
+1:	6 <= V < 100
+2:	100 <= V <= 1000

Figure 7: Evidence evaluation example 2: The strength of the evidence V obtained from two sets of measurements that are known to come from different glass items (different-source comparison). For a different-source comparison like this we would expect to obtain $V < 1$. Here the evidence results in moderately strong support in favour of H_d (different-source proposition).

Whenever the measurements from two sets of fragments have different elemental configurations, they are reported as coming from different glass items, resulting in the strength of the evidence reported being $V = 1/1000$. This is shown in Figure 8 where the measurements from the fragments found at the crime scene have elemental configuration 1, while the measurements from the fragments found on the suspect have elemental configuration 2. This can easily be seen from looking at the columns for Fe and K from the two files uploaded in Figure 8. The value $V = 1/1000$ comes from restricting the strength of this type of evidence to ensure that it does not provide too strong support in favour of a wrong proposition. Details on why this value was chosen, and why this restriction was applied are discussed in Section 5.

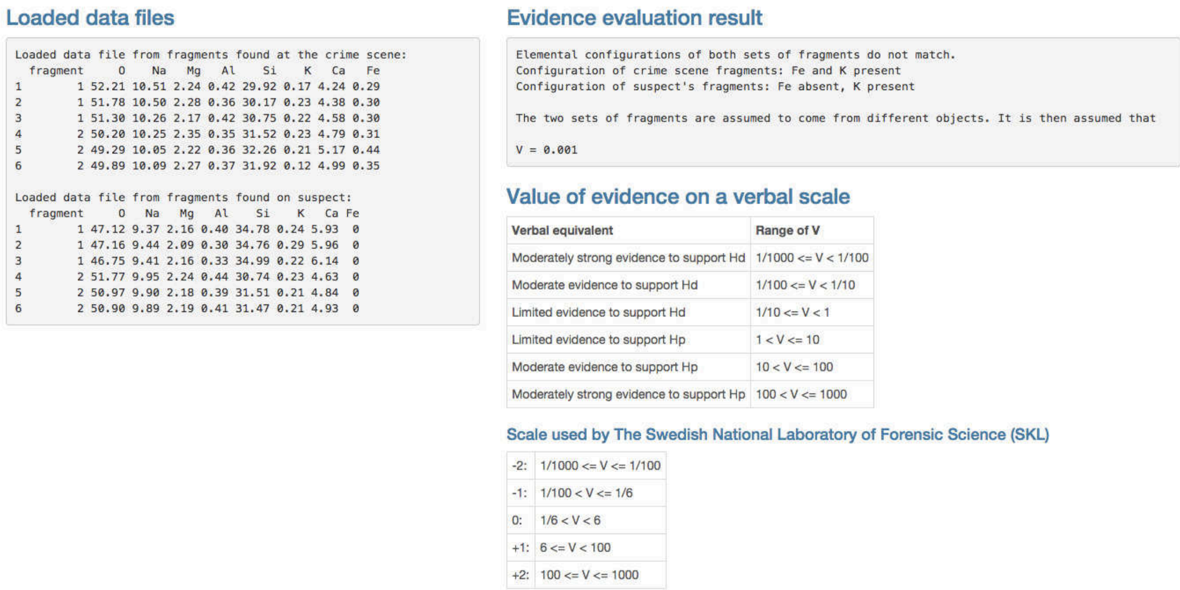


Figure 8: Evidence evaluation example 3: Results from comparing two sets of measurements with different elemental configurations. As can be seen from the loaded example files, the file containing the measurements from the fragments obtained from the crime scene contain Fe, while the measurements from the fragments found on the suspect do not. This means that the elemental configurations of the two sets of fragments differ and are therefore assumed to come from different glass items. This results in $V = 1/1000$, that is, moderately strong support in favour of H_d (different-source proposition).

5 Reported evidential value of glass fragments

Details of the simulation study carried out to assess the performance of the model in the evidence evaluation task can be found in [3]. The model was assessed in terms of the percentage of false positive (FP) and false negative (FN) answers. A FP occurs when two sets of glass fragments are from different sources, but $V > v$, for some critical value v, so that they are evaluated as coming from the same source. A FN occurs when two sets of fragments are from the same source, but $V \leq v$, so they are evaluated as coming from different sources. This provides the percentage of different-source and same-source errors, respectively. For the critical value $v = 1$ the different-source and same-source error rates obtained from the simulation study were 1.4% and 4.4%, respectively, which are improvements on previous publications using similar glass databases; see [8].

In a forensic context it is especially important to guard against the possibility of strongly misleading evidence [9, 10], that is, extremely low/high values of V that are in strong support of the wrong proposition.

Strongly misleading evidence may occur for several reasons including: (a) the models used in the computation of V are at best approximations, (b) the relatively small size of the data used to estimate them, and (c) the possible influence of prior specification on the numerator and denominator of (5). A further reason, specific to our application, has already been mentioned at the end of Section 4.2: whenever two sets of fragments have different elemental configurations, they are assumed to have originated from different sources, yielding $V = 0$ and hence extreme support for the defence proposition. This assumption, although reasonable, may not always hold in practice: in the glass dataset some objects have fragments with different configurations, although this happens in less than 1% of the cases.

A simple procedure to avoid overwhelming support in favour of the wrong proposition consists in restricting the values of V to the range $(1/v_T, v_T)$, for some threshold v_T. The verbal scale of [6] regards values of V between 1000 and 10000 as "Strong evidence to support" H_p, and values in excess of 10000 as "Very strong evidence to support" H_p. This suggests using either $v_T = 10000$ or $v_T = 1000$ as threshold, depending on whether or not we wish to allow glass evidence to count as "Strong evidence".

It may also be of interest to consider the effect of the chosen v_T on the empirical cross-entropy (ECE) measure of [9]. The ECE is a measure of the accuracy of the values of V, computed for a wide range of prior odds on H_p. Figure 9 displays ECE plots (using the R package [11]) from our simulation study, for different threshold values v_T. The solid curves represent the accuracy and the dashed curves perfect calibration, while the poorest performance is represented by the dotted neutral reference curves. The closer the solid and dashed curves are to one another the better the calibration of the V values; see [9]. From Figure 9 it can be seen that the calibration is very good for small (less than 1) prior odds on H_p but deteriorates for larger prior odds. Calibration improves for reduced thresholds v_T, with performance much better for the threshold $v_T = 1000$.

Taking this into consideration, we decided to use $v_T = 1000$ and restrict the strength of the evidence V to the range $(1/v_T, v_T)$. The scales of [6] and [7] used by the application have been truncated to reflect this range for the evidence.

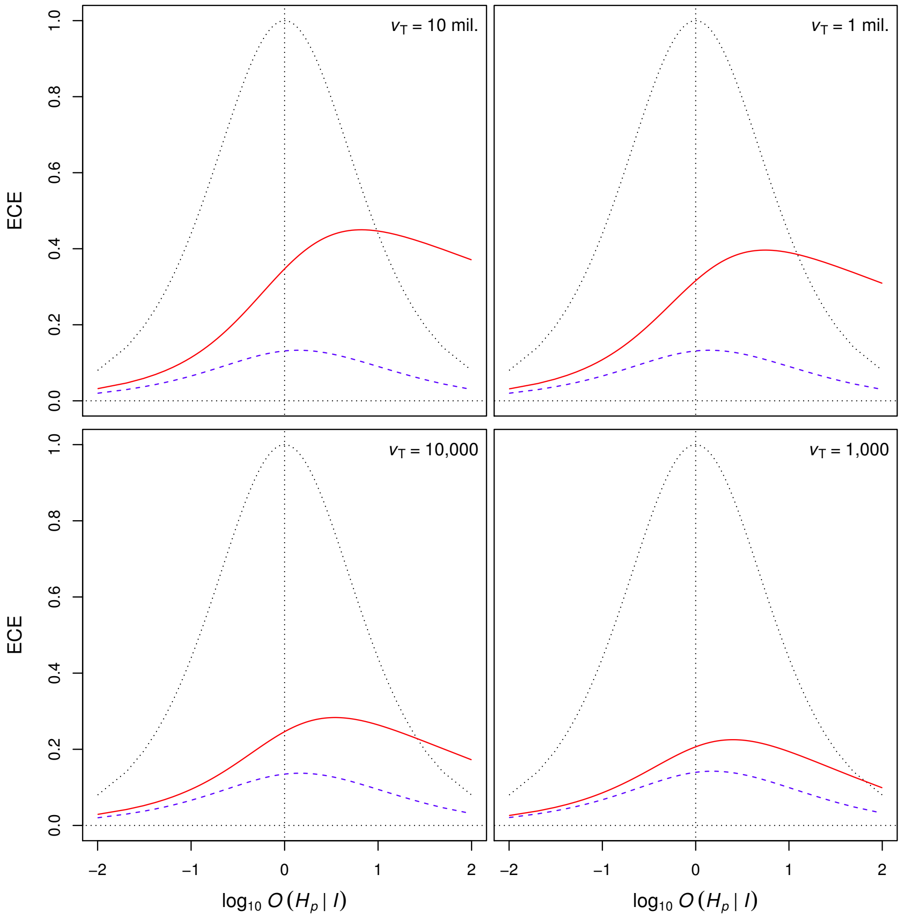


Figure 9: ECE plots from a simulation study with different thresholds v_T on V. The solid curves represent the accuracy; the dashed curves the calibrated accuracy; and the dotted curves the neutral reference. The x-axis is $\log_{10}$ of the prior odds in favour of H_p, that is $O(H_p|I) = \Pr(H_p|I)/\Pr(H_d|I)$.

6 Conclusion

We have presented a quick and easy-to-use online application for forensic scientists who analyse SEM-EDX measurements from glass fragments as forensic evidence in real-life casework. The application is a simple tool that can be loaded from a web browser and allows for quick and easy classification of glass fragments into use-type categories, and also for the evaluation of two sets of glass fragments as forensic evidence. The application can be accessed at http://gnapier.shinyapps.io/GlassClassificationAndEvaluation/. The application only requires users to upload data files, as either .csv or .txt, and outputs the required results immediately. The statistical methods used by the application were validated using simulation studies, and are improvements upon previous methods in both the classification and evidence evaluation tasks.

Independent testing

This is a nice app for classifying and calculating evidential value of SEM-EDX measurements on sets of glass particles. The GUI is well-organized and uploading measurements can be done by simply dragging and dropping the `txt/csv` files. Making use of an online application also has the immediate advantage that one does not need to go through the process of installing software before use, however a forensic expert may hesitate to send his data through the internet.

There are a number of other pros for this app. First, it uses state-of-the-art Bayesian statistical techniques in its calculations, relying on a statistical description of three levels of uncertainty commonly encountered in the analyses of glass particles, and it uses built-in probability density functions obtained by MCMC techniques. Therefore, it is also relatively fast since no additional MCMC draws are required in the calculations. For small datasets (a total of ten measurements each) they proceed in a matter of seconds.

The downsides of this app are that it gives no flexibility to include your own background data, since model training is not possible in the app. Furthermore, when a comparison is performed, it truncates LRs to be no smaller or larger than a certain value. Even though the authors motivate this choice in the article, SEM-EDX profiles may suggest more extreme values (either smaller or larger), for example when for the questioned set of particles no Fe and no K are observed and for the reference set Fe and K are observed.

Independently tested by:

Dr. P. Vergeer

Netherlands Forensic Institute

Department of Forensic science, Interdisciplinary investigations, Statistics and Knowledge management

Postbus 24044 / P.O. Box 24044 | 2490 AA | The Hague | The Netherlands

p.vergeer@nfi.minvenj.nl

Acknowledgements

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

References

[1] G. Zadora, Classification of glass fragments based on elemental composition and refractive index, J. Forensic Sci. 54(1) (2009) pp. 49–59.

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

[3] G. Napier, T. Neocleous, A. Nobile, A composite Bayesian hierarchical model of compositional data with zeros, J. Chemometr. 29(2) (2015) pp. 96–108.

[4] D. V. Lindley, A problem in forensic science, Biometrika 64(2) (1977) pp. 207–213.

[5] C. G. G. Aitken, F. Taroni, Statistics and the evaluation of evidence for forensic scientists, John Wiley and Sons, 2004.

[6] I. W. Evett, G. Jackson, J. A. Lambert, S. McCrossan, The impact of the principles of evidence interpretation and the structure and content of statements, Sci Justice 40(4) (2000) pp. 233–239.

[7] A. Nordgaard, R. Ansell, J. W. Drotz, L. Jaeger, Scale of conclusions for the value of evidence, Law, Probability and Risk 11(1) (2012) pp. 1–24.

[8] T. Neocleous, C. G. G. Aitken, G. Zadora, Transformations for compositional data with zeros with an application to forensic evidence evaluation, Chemometr. Intell. Lab 109(1) (2011) pp. 77–85.

[9] D. Ramos, J. Gonzalez-Rodriguez, Reliable support: Measuring calibration of likelihood ratios, Forensic Sci Int 230(1-3) (2013) pp. 156–169.

[10] R. Royall, On the probability of observing misleading statistical evidence, J Am Stat Assoc 95(451) (2000) pp. 760–768.

[11] D. Lucy, comparison: Multivariate likelihood ratio calculation and evaluation, R package version 1.0-4 (2013) URL http://CRAN.R-project.org/package=comparison.