Empirically exploring the effect of oxygen on the isotopic mapping of cremated and uncremated bones of a Central European Alpine passage.

Markus Mauder, Eirini Ntoutsi, Gisela Grupe, Peer Kroeger

6th International Symposium on Biomolecular Archaeology
27. - 30.8.2014, Basel, Switzerland
Outline

• A short overview of our project and the important of isotopic fingerprinting

• Employing data mining for isoscaping

• Studying the effect of oxygen on the extracted models

• Discussion and outlook
Transalpine mobility and cultural transfer project

• An interdisciplinary project of the Archaebiocenter, LMU, Munich
• Research Unit of the German Science Foundation, DFG (FOR 1670)

• **Goal:** Establishment of an isotopic fingerprint for bioarchaeological finds, especially cremations, and its application to archaeological and cultural-historical problems.

• **Reference region:** the transalpine Inn-Eisack-Etsch-Brenner passage. Specific archaeological contexts from Late Bronze Age until Roman times.

• Project www: [http://www.for1670-transalpine.uni-muenchen.de](http://www.for1670-transalpine.uni-muenchen.de)
Isotopic mapping

- Samples: animal findings
- Isotopes considered: Strondium, Lead, Oxygen
Building an isotopic fingerprint

• Isoscaping is a task of paramount importance in order to
  – describe/ “understand” an area
  – predict the most probable (spatial) origin of new samples

• Two data mining approaches towards this goal:

  1) The supervised way:

     Given the locality of the samples, can we generate a model that captures the key characteristics of the localities and is able to predict the locality of new samples?

     – Spatial coordinates of the samples are also part of the model.
     – The list of localities (problem classes) is predefined.

  2) The unsupervised way

     Can we group samples based solely on their isotopic values and check how the extracted isotopic-clusters are spatially scattered?

     – Only isotopic values of the samples are used for clustering.
     – Their coordinates are used for spatial validation/ exploration.
Our data

- Dataset consists of ~100 samples
- Each sample described in terms of:
  - Spatial coordinates (lat, long)
  - 3 isotopes (Sr, Pb, O) and
  - 7 isotope ratios
    - $^{87}\text{Sr}/^{86}\text{Sr}$
    - $^{208}\text{Pb}/^{204}\text{Pb}$
    - $^{207}\text{Pb}/^{204}\text{Pb}$
    - $^{206}\text{Pb}/^{204}\text{Pb}$
    - $^{208}\text{Pb}/^{207}\text{Pb}$
    - $^{206}\text{Pb}\ 207\text{Pb}$
    - $^{18}\text{O}\text{PO}_4$

Geographic distribution of the samples
Unsupervised learning

How do the clusters of isotopic-similar samples correlate with the actual locations of the samples?

Settings

• All 7 isotope features used for clustering
• Assumption that data are generated by Gaussian mixture models
• EM algorithm to estimate the model parameters
• Cluster population

```
<table>
<thead>
<tr>
<th>Cluster</th>
<th>Count</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>14</td>
<td>15%</td>
</tr>
<tr>
<td>1</td>
<td>29</td>
<td>30%</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3%</td>
</tr>
<tr>
<td>3</td>
<td>16</td>
<td>17%</td>
</tr>
<tr>
<td>4</td>
<td>27</td>
<td>28%</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
<td>7%</td>
</tr>
</tbody>
</table>
```

Detected clusters versus locations of the samples
Are region-specific models good predictors for the origin of new samples?

**Settings**
- The data were categorized into classes “north”, “middle”, “south” Alps based on sample coordinates.
- 10-fold cross validation (9 folds for training, 1 for testing)
- A kNN classifier is built upon the training set
- The model is evaluated upon the test set

<table>
<thead>
<tr>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>ROC Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.833</td>
<td>0.115</td>
<td>0.837</td>
<td>0.833</td>
<td>0.832</td>
<td>0.868</td>
</tr>
</tbody>
</table>
The effect of oxygen

- The problem: Oxygen is sensitive to cremation, in contrast to strontium and lead.

- Question: Is oxygen necessary for our analysis?
  - Quality might get lower of course but how worse?

- Why are interested in this?
  - A practical issue: we have a small uncremated sample set (~100 instances), it would be great if we can increase it by including uncremated samples.
  - A research question: how important is oxygen for fingerprinting?
  - A broader research question (for Data Mining): stability of data mining models under reduced feature spaces.

- Methodology:
  - Repeat the experiments by omitting oxygen
  - Find out how and where the with and without oxygen results “differ”.
  - the "differ" term depends on the Data Mining task per se.
Unsupervised learning

Detected clusters versus locations of the samples

Migration table

<table>
<thead>
<tr>
<th></th>
<th>cluster 0</th>
<th>cluster 1</th>
<th>cluster 2</th>
<th>cluster 3</th>
<th>cluster 4</th>
<th>cluster 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxygen clustering</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>cluster 0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.0</td>
</tr>
<tr>
<td>cluster 1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.97</td>
<td>0.63</td>
</tr>
<tr>
<td>cluster 2</td>
<td>0</td>
<td>0</td>
<td>1.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>cluster 3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>cluster 4</td>
<td>0</td>
<td>0.56</td>
<td>0</td>
<td>0</td>
<td>0.44</td>
<td>0</td>
</tr>
<tr>
<td>cluster 5</td>
<td>1.0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

No-oxygen clustering
Isotope distribution per cluster (Oxygen case)
Supervised learning

Are region-specific models good predictors for the origin of new samples?

Settings
- The data were categorized into classes “north”, “middle”, “south” Alps based on sample coordinates.
- 10-fold cross validation (9 folds for training, 1 for testing)
- A kNN classifier is built upon the training set
- The model is evaluated upon the test set

Evaluating the oxygen effect
- Experiments with and without oxygen

<table>
<thead>
<tr>
<th></th>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>ROC Area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxygen</td>
<td>0.833</td>
<td>0.115</td>
<td>0.837</td>
<td>0.833</td>
<td>0.832</td>
<td>0.868</td>
</tr>
<tr>
<td>No oxygen</td>
<td>0.76</td>
<td>1.168</td>
<td>0.768</td>
<td>0.76</td>
<td>0.759</td>
<td>0.785</td>
</tr>
</tbody>
</table>
How oxygen is correlated to other isotopes and location

- Oxygen isotope by location

- Oxygen correlation to other attributes
Discussion on the findings and next steps

• Our sample is to small to make general statements
  – ~100 samples
  – Even less for the unsupervised case, since 10% is kept out for model testing
• Our initial analysis seems promising though
• Both supervised and unsupervised learning show that the omission of oxygen does not completely destroy the mining models, models are stable to a certain extend.
  – In the unsupervised case, most of the clusters of the oxygen case “survive” to the non-oxygen case.
  – In the supervised case, still acceptable performance scores
  – Lower scores are to be expected due to information loss incurred by oxygen omission

• A real crash test though would be the evaluation of models performance when the cremated samples are available.
• Combination of uncremated and cremated samples for model improvement.
Thank you for your attention

Questions?

More information on the technical report at the project’s website: http://www.for1670-transalpine.uni-muenchen.de