Lectures at CMA4CH 2010, Mediterraneum Meeting

Application of Multivariate Analysis and Chemometry to Cultural Heritage and Environment
3rd ed., Taormina, Sicily island, Italy, Europe, 26-29 September 2010

Use of Gas Phase Sensors and Pattern Recognition to Determine the Condition of Books in Archives

R.G. Brereton[1], G.R. Lloyd[1], B. Knight[2], V. Horie[2]
[1]Centre for Chemometrics, School of Chemistry, University of Bristol, Cantocks Close, Bristol BS8 1TS, UK
[2]The British Library, 96 Euston Road, London, NW1 2DB, UK

Abstract

The work involves monitoring VOCs (Volatile Organic Compounds) using a cost effective portable gas phase sensor to determine the condition of books in the British Library. Multivariate Pattern Recognition techniques are employed to interpret these signals, and show potential as methods for automated monitoring of collections of old books.

Introduction

The British Library (BL) is the major library in the UK and, with the Library of Congress, one of the top libraries in the world. Because its responsibilities include the survival of archival material, considerable effort and research is devoted to understanding the composition, state and deterioration of its collections. The problems of the Library are comparable to those of every other library, for which reason the Library led the recent Identical Books Project (IBP) in collaboration with the other five UK legal deposit libraries and two national archives and with the involvement of national libraries in Europe and USA.

Library collections are accessed and used as individual objects, but are managed on an industrial scale. One of the major problems for the BL is the conservation of individual books within its collection of around 150 million items, which can be done by finding a small set of parameters able to characterise the "condition" in a single book before its deterioration. Only around 3.67 million books are delivered to readers each year, so that only 2.5% of the collection is ever opened by readers annually, and most books are never visually checked for signs of decay. A rigorous test would cost £226 per book or nearly £34,000,000,000 to test all books which is not economically feasible.

The British Library (BL) has developed an experimental procedure (using FAIMS technology) for analysing the volatile organic compounds (VOCs) found on books. It is hypothesised that using the proposed BL method it is possible to distinguish between different books based upon the VOC profile obtained by pointing a cheap sensor at each book, taking seconds. The aim of the study presented here is to apply chemometric and pattern recognition methods in an exploratory manner to gain insight into the validity of this hypothesis and suggest further work that might be required in order to prove (or disprove) the hypothesis.

The project has involved the development of techniques to use sensors to "sniff" old books to determine whether they are decaying or about to decay. A small pilot study captured VOC signatures using a novel "nose", Lonestar as alternative to well know methods [1].

A pattern recognition method using the dissimilarity between sample matrices that allows exploratory chemometric analysis has been applied using several approaches including Principal Component Analysis (PCA), Principal Coordinates Analysis (PCO) and Self Organising Maps (SOM) was used.

Materials & Methods

The Field Asymmetric Ion Mobility Spectrometer (FAIMS) is a high speed, gas phase ion separation technique, incorporated in Lonestar Portable Gas Analyzer (Owlstone Nanotech Inc. USA) permit to detect positive and negative ions by industrial gases and VOCs in an air matrix [2].

The real-time chemical spectra is the output of the instrument and this is a measure of how quickly an ion moves through an electric field. The mobility relates to size and mass, and is used to specifically distinguish and identify the chemical substances of interest. A dataset consisting of 27 data ASCII data files was obtained. All chemometric analysis was carried out using Matlab R2008a and so the raw ASCII files were to be converted into a suitable format. A bespoke Matlab routine was developed by author that extracts the relevant information from the data files and converts them into the Matlab format. The following data was extracted from each file: a) a positive ion data matrix (51x512), b) a negative ion matrix (51×512), c) a vector (51x1) of Dispersion Frequencies (DF) for each of the ion matrices, d) a vector (1x512) of Cross Voltages (CV) for each of the ion matrices.

Results

The 27 data files can be divided into two groups. The first (A), consisting of 22 data files, were analysed at different machine settings to the remaining five files that make up the second group (B). The results will focus on group A.

PCA showed fair distinction between different books, particularly those with different materials e.g. leather binding which largely dominate the analysis.

11PCO also indicates that the age or quality of the books might be determined using later components, however the factor is confounded with the book label and as such further data would be required.

SOMs showed excellent distinction between different books (better than PCO) including those with leather bindings whilst preventing them dominating the analysis.

A novel method using the dissimilarity between sample matrices was developed that allows exploratory chemometric analysis, see Fig. 1.

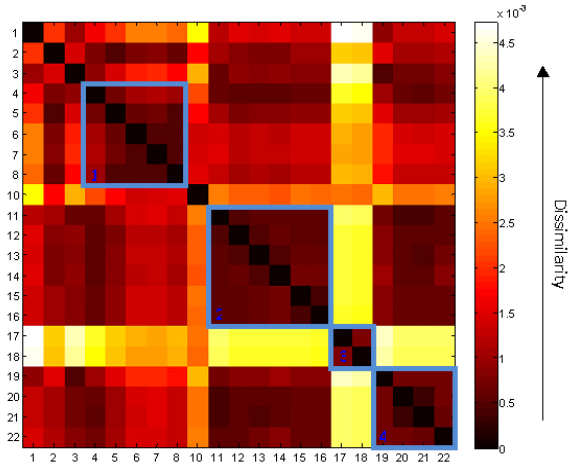

Fig. 1 Dissimilarity matrix for samples from group A. The intensity of the colours indicates the level of dissimilarity.

Conclusions

Taking into account the preliminary nature of this work several conclusions can be reached. Using VOCs together with pattern recognition shows great potential. Because of the unbalanced numbers of samples for each book some of the groupings observed e.g. using SOMs may be due to chance rather than 'real' groupings. However, since the results for both PCO and SOM are similar it seems less likely that the groupings are due to chance only.

In order to achieve the best possible outcome from this type of analysis the experimental design has to be considered very carefully, particularly as there are a large number of potentially interesting factors to investigate. Careful attention must also be paid to factors that can confound the analysis, such as the order in which samples are analysed, the machine they are analysed on, the person that does the analysis and ensure that the influence of these confounding factors from can be distringuished from those of interest.

References

1) M. Strlič, I. K. Cigić, J. Kolar, G. de Bruin, B. Pihlar, Non-Destructive Evaluation of Historical Paper Based on pH Estimation from VOC Emissions, *Sensors*, 7(12), (2007) 3136-3145

2) R. Guevremont, High-Field Asymmetric Waveform Ion Mobility Spectrometry (FAIMS), *Canadian Journal of Analytical Sciences and Spectroscopy*, 49(3), (2004) 105-113