COST D42 workshop

"*NIR/Chemometrics for Cultural Heritage*"
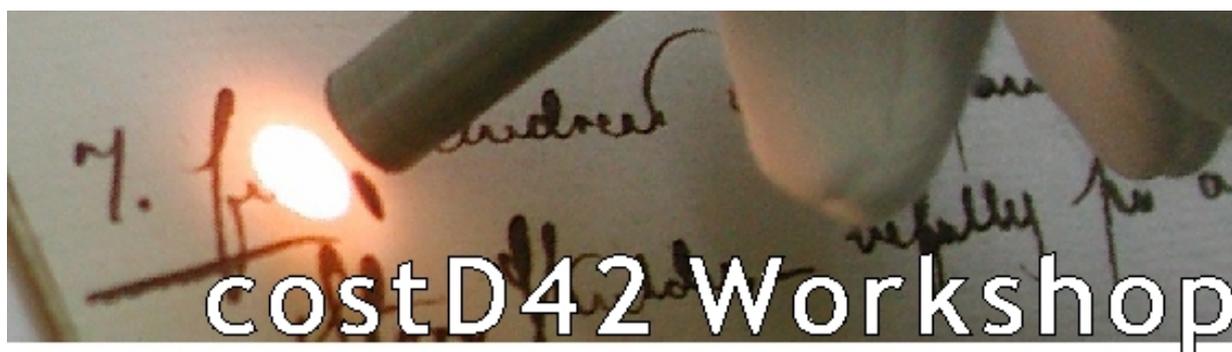
Programme

Date: 14 November 2008

Place: Hotel Slon, Slovenska cesta 34, Ljubljana, Slovenia

Coordinators: **Matija Strlič** – Centre for Sustainable Heritage University College, London, U.K.
Jana Kolar – Morana RTD d.o.o., Ivančna Gorica, Slovenia
Irena Kralj Cigić – University of Ljubljana, Faculty of Chemistry and Chemical Technology, Ljubljana, Slovenia

Session 1: 09:00-10:55
Annemie Adriaens (15 min): *COST D42 - Chemical Interactions between Cultural Artefacts and Indoor Environment*
Matija Strlič (15 min): *NIR spectrometry in CH studies: an introduction*
Marjana Novič (30 min): *Basic chemometrics in NIR spectroscopy I*
Matevž Pompe (30 min): *Basic chemometrics in NIR spectroscopy II*
Dirk Lichtblau (25 min): *The SurveNIR instrument and software*

Coffee break (30 min)

Session 2: 11:25-13:05
Jana Kolar (25 min): *The Morana/NIR instrument and software*
Mark Ormsby (keynote lecture, 25 min): *NIR/Chemometrics for rag paper diagnostics*
Michal Čeppan (25 min): *Chemometric identification of iron gall inks based on FORS NIR spectra*
Linda Cséfalvayová (25 min): *NIR/Chemometrics and genetic algorithms for gelatine determination*

Lunch break (2 h)

Session 3: 15:05-16:45
Manfred Schreiner (25 min): *Chemometric interpretation of IR and XRF for pigment identification*
Paul Garside (25 min): *NIR/Chemometrics for silk*
Emma Richardson (25 min): *NIR/Chemometrics for polyamides*
Tanja Trafela (25 min): *Classification of paper and pulps based on NIR/chemometrics*

COST Action D42 – ENVIART: A European Network in Conservation Research

A. Adriaens[1], J. Havemans[2]

[1] Department of Analytical Chemistry, Ghent University, Belgium

[2] Department Indoor Air Quality and Health, TNO, The Netherlands

COST stands for cooperation in science and technology and is the longest running network funding organization in Europe. Within the COST networks, Action D42 is dedicated to the conservation of our cultural heritage. More specifically the main objective of the Action is to explore chemical interactions between cultural artefacts and typical indoor environmental conditions through field studies and laboratory experiments and to translate the results into preventive conservation practice.

This paper gives an introduction to COST and highlights the networking activities of COST Action D42.

Near Infrared Spectroscopy/Chemometrics for Cultural Heritage

Matija Strlič

Centre for Sustainable Heritage, The Bartlett School of Graduate Studies, University College London, London, U.K.

m.strlic@ucl.ac.uk

Among the instrumental approaches to heritage material characterisation, near infrared spectroscopy (NIR) is certainly gaining in momentum. The spectroscopic technique itself has lately witnessed widespread use and acceptance in the food and pharmaceutical industries, especially in quality control, and is also becoming important in clinical analysis. This is easy to explain from the analytical point of view: the spectra are incredibly information-rich as they are composed of overtones and combination vibrations, which means that vibrations of many bonds appear at several positions in the NIR spectrum. However, due to the increased complexity of spectra an analytical approach to their interpretation is rarely possible.

Only the development of chemometrics for spectral interpretation enabled us to exploit the NIR spectral region to the full. Using a variety of chemometric tools, it is possible to extract qualitative and quantitative information from the spectra. To do so, analyses of known samples need to be performed first – an obstacle, which is difficult to overcome in the case of heritage material analysis because of the unavailability of well characterised sample sets.

The recent availability of small, portable, versatile and even hand-held NIR spectrometers with an appropriate wavelength range has made this technique even more attractive.

The number of research groups developing interesting and extremely useful applications based on NIR/chemometrics is increasing. These developments are supported in international projects within the EU Framework Programme, i.e. the SurveNIR (2005-2008) and the just initiated POPART (2008-2012) projects, but also Eureka! 3483 Lascan (2004-2007). Additional projects are running in the USA.

This very first workshop on NIR/Chemometrics for cultural heritage summarizes the efforts to date, and there is a good variety of applications, from paper to textiles and historic plastics. However, it is particularly encouraging that two companies are launching their products and services based on NIR/Chemometrics, which should increase the acceptance of this novel technique.

Principal Component Analysis - PCA

Marjana Novič

National Institute of Chemistry, Hajdrihova 19, POB 660, 1001 LJUBLJANA, Slovenia

marjana.novic@ki.si

Principal Component Analysis (PCA) is an analysis of multidimensional data regarding their variance. It was invented in 1901 by Karl Pearson. Now it is mostly used as a tool in exploratory data analysis and for making predictive models.

The variance is the driving force for the PCA of a given set of data. The labels n and p will be used for n objects and p variables. First let us define the variance of the j-th variable as $_n s^2$:

$$_n s_j^2 = \frac{\sum_{i=1}^{n}(x_{i,j} - \bar{x}_j)^2}{n}$$

The distinction between $_{n-1}S$ and $_n S$ is a common source of confusion, and extreme care should be exercised when consulting the literature to determine which convention is in use, especially since the uninformative notation is commonly used for both. The bias-uncorrected sample variance $_n S$ is implemented as "Variance" in PCA. How do we examine variance in plotted data? We determine the mean value and distances to the mean in each dimension, see Figure below.
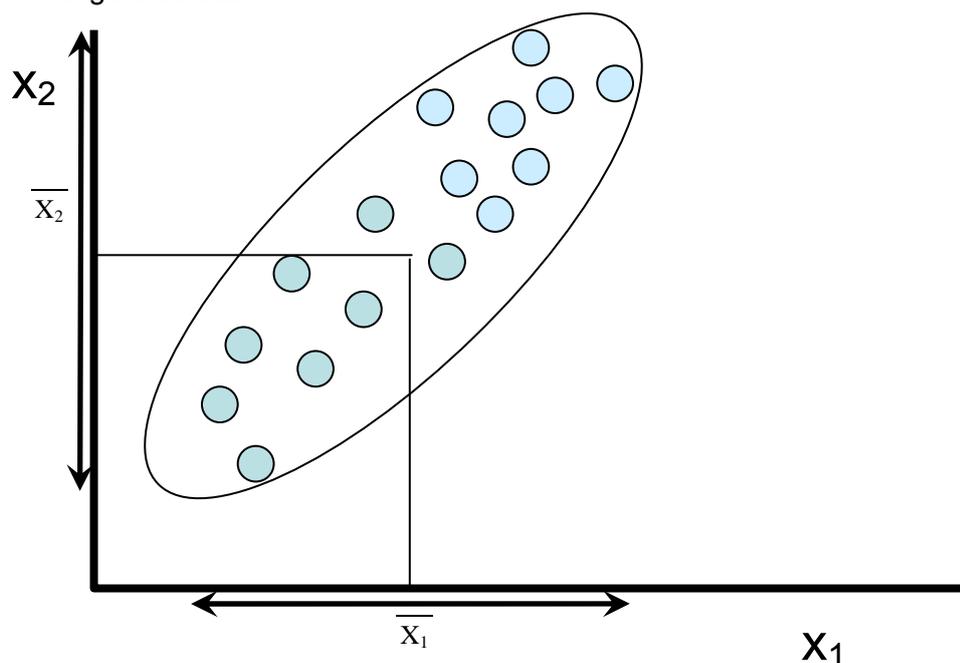

Fig. 1. Two-dimensional data in the original coordinate system (X_1, X_2).

From the above figure we can see that the maximal variance in the two-dimensional data is not parallel with original axes. In the Principal Component Analysis we search for the directions (i.e., new axes), in which the variance is maximal. In a stepwise procedure we determine the axes with maximal variances: the first one with the largest variance, the second one (orthogonal to the first one!) with the largest **remaining** variance, and so on, till the number of new axes is equal to the number of original coordinates. To summarize, PCA is a way to describe multivariate data by maximizing variances.
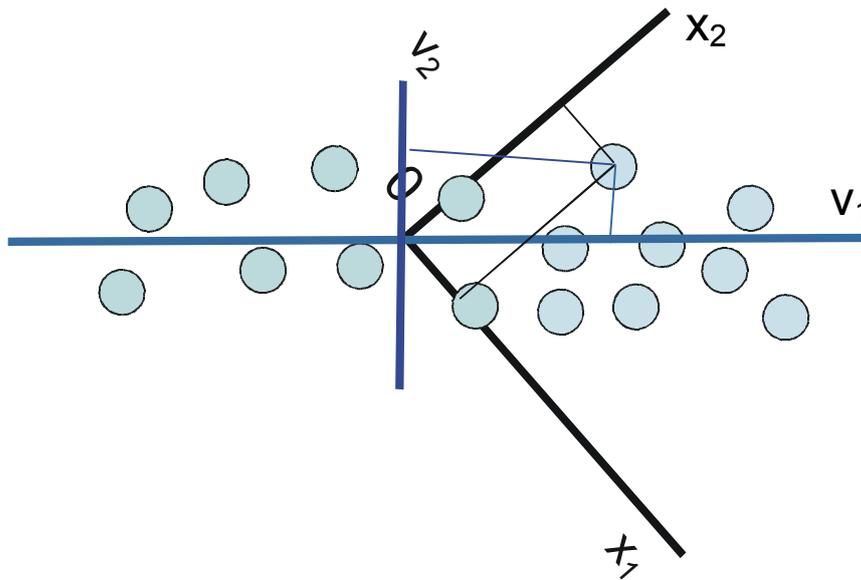

Fig. 2. Two-dimensional data in the original coordinate system (X_1, X_2) with the new, rotated coordinates (V_1, V_2).

The objects (observations) are not affected by the PCA; only the coordinate system is rotated and represents a new, so called latent space. The objects are projected onto the new coordinates in the latent space.

PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on. PCA is theoretically the optimum transform for a given data in least square terms.

Properties of PCA - convenient for exploitation

— Quality of the latent variables: One obtains maximal variance of scores in first few axes

— PCA plot preserves best the Euclidean distances in the p-dimensional feature space

— PCA scores are uncorrelated new variables

Main applications of PCA

— Projection of multivariate data onto a plane (exploratory data analysis, visualization, reduction of the information space – reduction of variables)

— Transformation of X into a set of uncorrelated variables

— Regression models - PCR (Principal Component Regression)

Mathematics behind PCA

In **mathematical language**, the determination of the new axes is formalized by matrix algebra. For the sake of completeness, the equations are given here, but as long as you understand the principles explained above, you do not need to go into mathematical details, there are many computer programs available to obtain the transformation matrix (**loadings**) and new coordinates (**scores**) of your data.

Let **X** be the data matrix of *n* objects or observations (rows) and *p* variables (columns).

We have to decompose the data matrix **X** as follows:

$$X = V^T * U$$
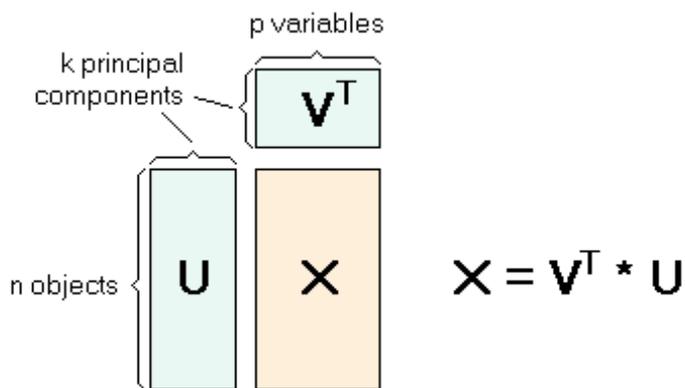
The two matrices **V** and **U** are orthogonal. The matrix **V** is usually called the loadings matrix, and the matrix **U** is called the scores matrix. The loadings can be understood as the weights for each original variable when calculating the principal component. The matrix **U** contains the original data in a rotated coordinate system.

There are a few common plots which are always used in connection with PCA:

— the scores/scores plot (left part of the figure below),

— the corresponding loading/loading plot (right part of the figure below)

7
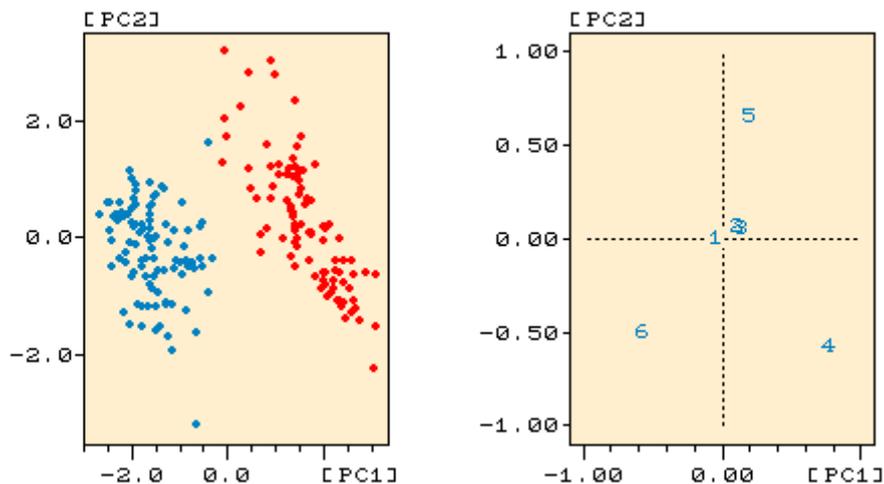
Mathematical way to obtain **V** (and consequently also **U)** is described as follows:

First we calculate so called scattered matrix

Z = **X**^{T}**X**.

We obtain a square matrix **Z** of dimension p by p, which can have at most p eigenvectors. These are the diagonal elements in the so called diagonalized matrix:

This can be denoted in matrix notation as

ZV = **V**diag(λ_1, ... λ_p)

or

Zv$_i$ = $\lambda_1 v_i$

The eigen vectors v_i are rows of the loading matrix **V**.

The eigenvectors are orthogonal to each other, and the product **V**^{T}**V** is the identity matrix **I** (**V** is an orthonormal matrix). The matrix **Z** can be expressed by its eigenvectors

Z = **V**diag(λ_1, ... λ_p)**V**^{T}

Note that the inverse matrix of **Z** can be written simply as

Z^{-1} = **V**diag(1/ λ_1, ... 1/ λ_p)**V**^{T}

Conclusions

Principal Components Analysis (PCA) is a way to describe multivariate data by maximizing variances. It is also a useful tool for data compression and information extraction. It is often exploited to find combinations of variables (factors) that describe major trends in the data. In mathematical sense PCA is an eigenvalue decomposition process of the covariance (correlation) matrix of the variables – columns in the data matrix. PCA is only applied to multivariate data that cannot be displayed on 2-dimensional paper or computer screens. For more than two dimensions, we have to project the data onto a plane. This projection changes with its direction; or, in other words, the projected image changes if the data points are rotated in the p-dimensional space. PCA defines the way of rotation of the data (or of the axes - which is quite the same) which displays a maximum of information in the projected image.

The principal components may be calculated by eigenanalysis of one of three different matrices:

— the **scatter matrix**: this is simply the product of $\mathbf{X}^T\mathbf{X}$, and does not imply any scaling

— the **variance-covariance matrix**: which is equal to the scatter matrix after mean-centring the data

— the **correlation matrix**: which is equal to the scatter matrix after standardizing the data

The choice of normalization is problem dependent. With a suitable computer program, one can play easily by experimenting with all three approaches. Generally speaking, the matrix to be used is determined by the importance of either the absolute numbers in the data (scatter matrix), or the relationships between the variables (correlation matrix). If a fixed offset in the variables causes problems, one may use the covariance matrix.

Literature

1. K. Pearson, "On Lines and Planes of Closest Fit to Systems of Points in Space". Philosophical Magazine, 1901, 2, 559–572.

2. I.T. Jolliffe, "Principal Component Analysis", Series: Springer Series in Statistics, 2nd ed., Springer, NY, 2002.

3. D.L. Massart, B.G.M. Vandeginste, S.N. Deming, Y. Michotte, and L. Kaufman, "Chemometrics: A Textbook, Amsterdam", Elsevier,1988.

4. L. Eriksson, E. Johansson, N. Kettaneh-Wold, and S. Wold, "Multi- and Megavariate Data Analysis- Principles and Applications", UMETRICS, 2001.

5. Teach/Me, SDL - Software Development Lohninger; Teach/Me DataLab 2.002 © 1999 Springer, Berlin, Developed by H. Lohninger and the Teach/Me people.

6. "Matlab" online Manual, The MathWorks, Inc.

Multivariate modeling of spectroscopic data

Matevž Pompe

University of Ljubljana, Faculty of Chemistry and Chemical Techology, Aškerčeva 5, 1000 Ljubljana

matevz.pompe@fkkt.uni-lj.si

There are only rare cases when simple univariate calibration can be used in order to obtain correlation between independent variable, that is, spectroscopic intensity at a single wavelength and chemical composition of the sample. In order to reduce matrix effect or to obtain simultaneous information about several chemical species present in the sample, a multivariate calibration is more frequently used. The simplest form of multivariate calibrations is multivariate linear regression, which represents an extension of simple linear regression model into the multi-dimensional description space. However this technique faces several problems especially in cases when we are using collinear variables in order to describe our modelling space. Such cases are quite frequent when we are modelling spectroscopic data, since there is usually significant dependence between absorbances obtained at different wavelengths. In order to reduce these problems partial least square regression (PLS) is more frequently used for the modelling of the spectroscopic data. The main advantages of the mentioned techniques will be presented together with some minor disadvantages.

The Non-destructive Survey of Paper Collections with SurveNIR

Dirk Lichtblau

ZFB Zentrum für Bucherhaltung GmbH, Mommsenstraße 7, D – 04329 Leipzig, Germany

lichtblau@zfb.com

Accessibility of a collection as well as conservation needs take into account some general criteria or strategic guidelines. Such decisions are based on the condition of the collection in question. The condition can be defined by assessing the mechanical and chemical state of the paper material. Whereas mechanical damages can be visually at least estimated, chemical damages are usually only detectable with analytical tools. Since surveys until now are mostly based on visual assessments, the survey of originals was so far quite subjective.

Contributing to a well informed collection management which will make strategic decisions on that base is there a significant risk of misinterpretation of the actual situation.

To improve the situation, an analytical tool was developed within the SurveNIR project. It allows to analyse qualitative as well as semi-quantitative the degradation state of the paper. Since it is based on NIR-spectroscopic (near infrared) the techniques is fast, efficient and completely non-destructive. User-friendly software enables the user to plan as well as to accomplish different types of surveys like statistical evaluation. It provides also the possibility to combine the analysed chemical-physical parameters to user-defined indices and categories regarding the different interest of each user. SurveNIR thus enables us to evaluate large collections within a reasonable timeframe and on a scientific basis. It enables collection manager to plan efficient preservation strategies with an individual expertise.

Literature

J. Taylor, S. Stevenson, Investigating subjectivity within collection condition surveys, Mus. Manag. Curat., 1999, 18, 19-42.

T. Trafela, M. Strlič, J. Kolar, D. A. Lichtblau, M. Anders, D. Pucko Mencigar, B. Pihlar, Nondestructive analysis and dating of historical paper based on IR spectroscopy and chemometric data evaluation, Anal. Chem., 2007, 79, 6319-6363.

D. Lichtblau, M. Strlič, T. Trafela, J. Kolar, M. Anders, Determination of mechanical properties of historical paper based on NIR spectroscopy and chemometrics - a new instrument, Appl. Phys. A 92, 2008, 191-195.

SurveNIR web page, http://www.science4heritage.org/survenir/

The Morana/NIR instrument and software

Jana Kolar

Morana RTD d.o.o., Ivančna Gorica, Slovenia

jana.kolar@morana-rtd.com

The non-destructiveness and portability of NIR spectrometers has led to a considerable interest in the use of NIR for identification and characterization of cultural heritage materials. As a result, several recent papers cover its applications to historical papers, plastics and textiles. However, while the acquisition of spectra is a rapid and facile process, their analyses require extensive knowledge of chemometrics, thus preventing the widespread use of this promising technique by conservators. To alleviate this obstacle, Morana RTD is developing a software application for a NIR instrument. The guiding principles of the system are the following:

- High quality of data

- User-friendliness

- Flexibility in data acquisition (e.g. using optical cable)

- Identification/characterization of a variety of cultural heritage materials.

So far, promising NIR calibrations have been obtained for pH and DP values of a variety of paper substrates, including papers containing iron gall inks. The work is a collaborative effort between MORANA RTD and University of Ljubljana.

In addition to characterisation of paper, an application for plastic materials will be developed within the 7 framework project POPART, while an application for proteinaceous materials is a subject of a Ph.D. study at the University of Ljubljana.

NIR/Chemometrics for rag paper diagnostics

Mark Ormsby

National Archives and Records Administration, USA

mark.ormsby@nara.gov

Gelatine sizing was a key component of handmade paper produced in Europe and America until the mid-1800s. A number of recent studies have suggested that the gelatine may play a role in helping to preserve these documents. Quantitative data on the concentration of the gelatine in historic papers is limited, however, in part because previous research required destructive sampling from rare, valuable specimens. In this project, collaboration among the National Archives, the University of Iowa, and the Getty Conservation Institute, a non-destructive method using NIR chemometrics was developed to determine the gelatine concentration. Spectra were obtained from a set of 40 specimens produced from the 15th to the 18th century. The model was calibrated by removing samples and analyzing them by gas chromatography/mass spectroscopy (GC/MS) to determine the concentration of seven amino acids found in gelatine. If the NIR measurement predicted a gelatine concentration in the range 0 to 6 percent by weight then there is a 95% probability that a destructive amino acid measurement would match the result within -1.6 and +1.3%. At higher concentrations the NIR model is less precise in matching the amino acid measurements, in part because of limited number of calibration specimens at the higher ranges. In a study of over 500 specimens from books, loose leaves, and artworks printed from 1460-1791, the mean gelatine concentrations for all papers were quite high in the 15th century but dropped significantly over time. This ongoing survey will expand to 1,500 specimens. The instrumentation also measures in the UV and visible ranges, and preliminary colour difference measurements will also be discussed.

Chemometric identification of iron gall inks based on FORS NIR spectra

Michal Ceppan

Faculty of Chemical and Food technology, Slovak University of Technology, Bratislava, Slovakia

michal.ceppan@stuba.sk

The method of identification of iron-gall inks in historical drawings and documents based on Factor Analysis of VIS-NIR Fibre Optics Reflectance Spectra (FORS) in the range 500-1050 nm was developed. Reference database set consists of 132 spectra of samples of model iron gall inks. Chemometric detector is based on testing of correspondence of studied individual spectrum with the database. This testing is performed by the method of Target Factor Analysis. This method evaluates the deviation, call AET parameter, of tested spectrum from the projection of tested spectrum into the factor space of database spectra set. Great value of AET parameter indicates that the tested spectrum does not correspond to the database. On the other hand, low value of AET parameter indicates, that the tested spectrum corresponds to the spectra of databases. In this case we select the value of AET = 0.006 as a threshold (corresponds to the statistically evaluated reproducibility of the measurement of the database spectra).

The method was tested on the set of control samples of bistre, sepia and iron-gall inks. Examples of analysis of historical drawings and documents are presented.

NIR/Chemometrics and genetic algorithms for gelatine determination

L. Cséfalvayová[1], M. Pelikan[2], I. Kralj Cigić[1], J. Kolar[3], M. Strlič[4]*

1. University of Ljubljana, Faculty of Chemistry and Chemical Technology, Slovenia

2. University of Missouri at St. Louis, Missouri Estimation of Distribution Algorithms Lab, USA

3. Morana RTD d.o.o., Ivančna Gorica, Slovenia

4. Centre for Sustainable Heritage, University College London, United Kingdom

*linda.csefalvayova@fkkt.uni-lj.si

The complex composition of historical papers complicates the quantitative determination of individual constituents. In this work, we examine the possibilities offered by near-infrared spectroscopy in combination with appropriate data analysis. To utilise the selectivity inherent to spectroscopic multivariate measurements, a chemometric approach to the analysis of spectra is employed. Genetic algorithms are used to implement a selection of spectral data derived from information rich UV-VIS-NIR measurements for use in building multivariate calibration models based on least squares regression that relate spectra to gelatine content in papers. A selective but laborious liquid chromatographic method for the quantification of hydroxyproline has been developed to provide the reference data on gelatine content. The derivatizing agent, 9-fluorenylmethyl chloroformate (FMOC) has been applied to convert the amino acid hydroxyproline, followed by reverse-phase liquid chromatographic separation and fluorimetric detection. However, sample destruction is necessary for this. The in situ non-destructive spectroscopic analysis, described in this work, removes the need for physical separation of the analyte from its matrix and is both rapid and selective, thus meeting the highly restrictive needs to maintain the integrity of heritage objects. The sampling flexibility afforded by small-size field-portable spectroscopic equipment combined with chemometric methods, undoubtedly represent an attractive addition to existing analytical techniques for cultural heritage materials.

Genetic algorithms are optimization methods based on the principles of genetics and natural selection. A brief discussion of the GA implementation used in this research is also provided.

Chemometric Interpretation of XRF and FTIR Results in Art and Archaeology

M. Schreiner[*1], M. Melcher[1], A. Schäning[1], K. Varmuza[2]*

1. Institute of Science and Technology in Art, Academy of Fine Arts Vienna, Schillerplatz 3, A-1010 Vienna, Austria

2. Laboratory for Chemometrics, Institute of Chemical Engineering, Vienna University of Technology, Getreidemarkt 9/166, A-1060 Vienna, Austria

*m.schreiner@akbild.ac.at

Scientists working in the fields of art or archaeology are often confronted with a great number of data obtained from analyzing a variety of findings or artefacts. There, questions concerning the material composition as well as the present condition of an object are predominant and these results are increasingly used to allocate an art work to a particular historic or prehistoric context, to determine the correctness or incorrectness of the claimed provenance or to explore the technology used for the preparation. The analysis of dated or precisely datable artefacts attributed to certain regions, artistic landscapes, schools or even individual workshops enables us to set up a history and technology of materials and to set up a "terminus post quem" or a "terminus ante quem" for e.g. pigments or dyes or also for methods for materials processing or fabrication. Statistic methods are increasingly applied to carry out the evaluation and support the interpretation of such data.

Three examples will be presented as a result of interdisciplinary co-operations between archaeologists, art historians and conservators on one side and natural scientists, especially analytical chemists on the other side:

 - Ancient Roman glass (106 fragments), excavated in Ephesos/Turkey and dating from the late Hellenistic to the late Ancient/Byzantine period, has been subject of analysis using SEM/EDX (energy dispersive x-ray microanalyses in a scanning electron microscope) and μ-XRF (micro-x-ray fluorescence analysis). The results of both methods were combined to obtain the full quantitative information of the main, minor, and trace elements. The results have been treated statistically using factor analysis and interpreted on the basis of an archaeological classification yielding to 19 different groups.

 - Synthetic organic pigments of the 19th and 20th century are known as the largest group of colorants used in contemporary artists´ paints. For the identification of pure pigments Fourier transform infrared (FTIR) spectroscopy has been established. FTIR spectra of these pigments have sharp and characteristic absorptions in the fingerprint region. Due to this fact, the identification of synthetic organic pigments by similarity search of FTIR spectra is usually successful if the spectrum of the pigment is contained in the used database. If the pigment investigated is not contained, application of multivariate classification methods has been studied. The used data set contains FTIR spectra from 281 synthetic organic pigments belonging to 11 classes;

methods applied are discriminant partial least-squares regression (D-PLS) and k-nearest neighbour classification (KNN). In cross validation 97 - 100% have been correctly classified in each class.

- In addition to the identification of materials of art objects, statistic methods have been also applied for the explanation of material ageing or destruction phenomena. Within an international exposure program (UNECE – Economic Commission for Europe in the United Nations) glass samples with a chemical composition similar to medieval glass was exposed and the degree of weathering determined by using SEM/EDX. The depths of the so-called leached layers formed during the natural weathering were correlated with the environmental data using multiple linear regression analysis. The results showed a significant influence of temperature, relative humidity, SO_2 and NO_2 in the ambient atmosphere on the stability of the potash-lime-silica glasses investigated.

NIR/Chemometrics for Silk

P. Garside

Textile Conservation Centre, University of Southampton, Winchester Campus, Winchester, SO23 8DL, UK.

p.garside@soton.ac.uk

Silk is a culturally important textile, found in many artefacts of historic significance including clothing, upholstery, banners and decorations. However, it is a fragile material and is prone to deterioration via a variety of mechanisms, particularly after certain historically common methods of processing, such as bleaching and weighting. Therefore it is important to be able to accurately characterise the material in order to inform the most appropriate strategies for conservation, display and storage. NIR spectroscopy allows the non-invasive, *in situ* investigation of these textiles, and when combined with chemometric (multivariate) analysis to draw correlations with data obtained by other methods, can provide a wide range of valuable information. Using these techniques, we have demonstrated that it is possible to monitor stresses in silk fabrics, highlighting those objects (or regions within a particular object) at particular risk of mechanical damage due to self-loading or inappropriate display. Also of great value to conservation is the ability to gain information about the general state of deterioration of the silk; although this is complicated by the range of degradative reactions which silk can undergo, it is possible to draw correlations between NIR spectra and characteristics such as mechanical properties (derived from tensile testing) or degree of polymerisation (from HPLC experiments). Finally, broad information can be obtained about processing which the material has undergone, including metal salt weighting; other techniques, such as XRF, often allow a more precise investigation of these treatments, but the ability to obtain this further information from the NIR spectrum minimises unnecessary interference with an artefact, and highlights those objects which may warrant additional investigation via other methods.

On-site Characterisation of Polyamide Material by Near Infrared Spectroscopy and Principle Component Analysis

Emma Richardson[*1], *Graham Martin*[2], *Paul Wyeth*[1]

1. Textile Conservation Centre, University of Southampton, Winchester, UK

2. Science Section, Victoria and Albert Museum, London, UK

*ejcr@soton.ac.uk

Collection management strategies are best informed by a detailed knowledge of the materials comprising the objects. This can be especially true when dealing with contemporary collections, where identification of synthetic material can prove problematic. The differentiation of polyamide based textiles, namely silk and Nylon, can pose particular problems for curators of historic textiles due to the similarities of their visual nature. Our work has highlighted a number of cases where false classification has occurred within collections, which may impact on appropriate storage, conservation treatments and historical documentation. The application of near-infrared reflectance spectroscopy (NIR) coupled with multivariate analysis has proven successful for the non-destructive, *in situ* differentiation of the polyamide sub-classes, which are found extensively within textile collections as fabrics, accessories and as conservation netting.

Our classification has been carried using Soft Independent Modelling of Class Analogies (SIMCA), finding widespread application across the sciences for the determination of groups within data. SIMCA utilises principle component analysis to build separate models for each known subclass, and assigns new samples depending on their orientation in PC space with respect to the model and the distance from the model boundaries.

The second method investigated in this work is Partial Least Squares Discriminant Analysis (PLS-DA). The PLS method is a multidimensional regression method, which has typically been applied for quantitative applications. However it can be applied to classification problems through the construction of a false Y matrix, using a binary system for each class in the data set.

Our work makes use of these techniques, and we will illustrate the steps taken to determine whether it is possible to distinguish the different polyamide classes. Since near infrared spectroscopy is a diffuse reflectance technique, the spectra are affected by factors other than just the underlying chemical structure of a material. As such, these needed to be accounted for before the final calibration could be made. This was to ensure that it was only the variations of interest that were influencing the models, rather than other extraneous properties

Classification of paper and pulps based on NIR/chemometrics

Tanja Trafela[1], Matija Strlič*[2], Jana Kolar[3], Simona Peterlin[4], Tjaša Drnovšek[4]

1. University of Ljubljana, Faculty of Chemistry and Chemical Technology, Ljubljana, Slovenia

2. University College London, Centre for Sustainable Heritage, The Bartlett School of Graduate Studies, U.K.

3. Morana RTD d.o.o., Ivančna Gorica, Slovenia

4. Pulp and Paper Institute, Ljubljana, Slovenia

*m.strlic@ucl.ac.uk

We have used different chemometric approaches to classify paper and pulp samples on the basis of chemical and spectroscopic information. More than 1400 paper samples from the SurveNIR reference collection and more that 400 commercially available softwood and hardwood pulp samples have been used for this purpose. On the basis of the acquired chemical and mechanical data the samples were arranged into groups using PCA (principal component analysis), which is a display method, and by the DA (discriminant analysis), which is a classification method. The papers can be conveniently grouped in four categories: rag paper, lignin-containing paper, bleached pulp-containing paper and coated paper.

We also report on the use of near- and mid-FTIR in combination with multivariate data analysis for prediction of mechanical properties of hardwood and softwood pulps. Tensile strength, tearing resistance and bursting index were determined using standard methods and FTIR spectra of the samples were measured. A Perkin-Elmer Spectrum GX (Waltham, MA) equipped with a 76-mm Labsphere RSA-PE-200-ID (North Sutton, NH) integration sphere coated with Infragold, with DTGS detector was used for measuring the spectra. The reflectance spectra were collected in the interval 6500-500 cm^{-1}, 128 scans per sample, using four layers of paper. We used partial least squares to correlate spectral information with the measured data.