

APPLICATIONS OF STATISTICAL METHODS IN INVESTIGATING HERITAGE LEATHER ITEMS

APLICATII ALE METODELOR STATISTICE LA INVESTIGAREA OBIECTELOR DE PATRIMONIU DIN PIELE

Dana Corina DESELCNICU*

National Research & Development Institute for Textile and Leather – Division: Leather and Footwear Research Institute, 93 Ion Minulescu St., Sect 3, 031215-Bucharest, Romania, e-mail: icpi@icpi.ro, d_deselnicu@yahoo.com

APPLICATIONS OF STATISTICAL METHODS IN INVESTIGATING HERITAGE LEATHER ITEMS

ABSTRACT. Characterization of protein materials, such as leather, is still an unsolved problem, due to its non-homogeneity, being a natural product. Chemical and physical-mechanical methods currently available for leather characterization are laborious and consume large quantities of chemical reagents and material, and their application requires a lot of time. By using non-destructive methods, such as NIR spectroscopy, and by using studies of statistical-mathematical modeling of analytical data, "hidden" connections between related chemical systems or between stages of physical-chemical processes a given system undergoes can be highlighted. In this paper 130 leather samples belonging to heritage items have been analyzed by NIR spectroscopy. The obtained data have been gathered in a comprehensive, searchable database, which structures complex analysis results in a simple, easily accessible and intelligible manner in view of applying statistic techniques. The paper presents the results obtained by applying SPSS and chemometric statistical software.

KEY WORDS: statistical methods, leather, SPSS, chemometrics.

APLICATII ALE METODELOR STATISTICE LA INVESTIGAREA OBIECTELOR DE PATRIMONIU DIN PIELE

REZUMAT. Caracterizarea materialelor proteice, cum este pielea, a rămas încă o problemă ce trebuie rezolvată, datorită neomogenității ei, fiind un produs natural. Metodele chimice și metodele fizico-mecanice care sunt actualmente la dispoziție pentru caracterizarea pieilor sunt laborioase și sunt consumatoare de cantități mari de reactivi chimici și material, și necesită mult timp pentru execuție. Prin utilizarea unor metode nedestructive, cât și prin utilizarea unor studii de modelare statistică-matematică a datelor analitice, se pot pune în evidență relații „ascunse” între sisteme chimice înrudite, sau între etape ale proceselor fizico-chimice pe care le parcurge un sistem dat. În cadrul acestei lucrări au fost analizate 130 de probe de piele aparținând unor obiecte de patrimoniu prin spectroscopie NIR. Datele obținute s-au constituit într-o bază de date comprehensivă, interogabilă, care structurează de o manieră simplă, ușor accesibilă, inteligibilă, rezultatele complexe ale analizelor în vederea aplicării tehniciilor statistice. Lucrarea prezintă rezultatele obținute prin aplicarea programelor statistică SPSS și chemometrice.

CUVINTE CHEIE: metode statistice, piele, SPSS, chemometrie.

APPLICATIONS DES MÉTHODES STATISTIQUES DANS L'INVESTIGATION DES OBJETS EN CUIR DU PATRIMOINE

RÉSUMÉ. La caractérisation des matériaux protéiques, comme la peau, est encore un problème à résoudre, dû à sa non-homogénéité, étant un produit naturel. Les méthodes chimiques et les méthodes physico-mécaniques qui sont actuellement disponibles pour caractériser les peaux sont laborieuses et consommatrices de grandes quantités de réactifs chimiques et des matériaux, et leur application exige beaucoup de temps. En utilisant des méthodes non destructives et des études de modélisation statistique et mathématique des données d'analyse, on peut révéler des liens « cachés » entre les systèmes chimiques liés, ou entre les étapes des processus physiques et chimiques qu'un système donné parcourt. Dans ce papier, 130 échantillons de cuir des objets du patrimoine ont été analysés par spectroscopie NIR. Les données obtenues ont formé une base de données complète, qui permet la recherche, qui structure les résultats complexes des analyses d'une manière simple, facilement accessible, intelligible, pour en appliquer les techniques statistiques. Le papier présente les résultats obtenus en appliquant les logiciels statistiques SPSS et la chimiométrie.

MOTS CLÉS: méthodes statistiques, cuir, SPSS, chimiométrie.

INTRODUCTION

Current policies of preserving global cultural heritage, aiming at saving the cultural heritage of each nation, impose a change of the general theory regarding the fundaments of scientific investigation, of preservation, restoration and display.

It is known, that, as a result of natural ageing, both inorganic materials and especially organic ones, such as

INTRODUCERE

Politicele actuale de conservare a patrimoniului cultural mondial, care vizează pentru fiecare națiune în parte salvarea moștenirii sale culturale, impun o schimbare a teoriei generale privind fundamentele investigării științifice, prezervării, restaurării și etalării.

Se știe că, în urma îmbătrânriri naturale, atât materialele anorganice, dar îndeosebi cele organice,

* Correspondence to: Dana Corina DESELCNICU, National Research & Development Institute for Textile and Leather – Division: Leather and Footwear Research Institute, 93 Ion Minulescu St., Sect 3, 031215-Bucharest, Romania, e-mail: icpi@icpi.ro, d_deselnicu@yahoo.com

tanned leather, undergo a series of micro or macrostructural destruction processes underlying deterioration effects which alter the physical state of structural-functional elements and a series of chemical, electrochemical, biochemical and radiative (thermal and photonic radiation) alteration processes, which alter the chemical nature of the material, and which leave their mark on the alteration of chemical characteristics (for instance, altering the nature of fatty substances, etc.), physical-structural characteristics (altering the crystallinity degree of a structure, shifts or chromatic degradations etc.). Elaborating methods of determining these changes allows creating new co-assisting or corroborating systems in order to accurately establish elements involved in the alteration of materials over time.

Characterization of protein materials, such as leather and parchment, is still an unsolved problem, due to their non-homogeneity, being natural products. Their chemical and physical mechanical properties are of great interest for the user, particularly in the field of restoring historical heritage items. Besides raw material (raw hides), a special interest must be paid to production technologies, dyes used, amino acid content, tanning method, etc.

Chemical and physical-mechanical methods currently available for leather characterization are laborious and consume large quantities of chemical reagents and material, and their application requires a lot of time. In heritage items, application of these methods is not possible, because it would cause their destruction. Therefore, modern, non-destructive methods are required for fast quality control. Such methods are based on IR (infrared) or NIR (near infrared) spectroscopy. NIR spectra particularly are very rich in combined bands containing a multitude of information which is rarely examined in an analytical manner.

Studies of statistical-mathematical modeling (statistics, chemometric) of analytical data can highlight "hidden" connections between related chemical systems or between stages of physical-chemical processes a given system undergoes. To ensure correct solving of chemometrics problems, procedural and/or numerical algorithms have been elaborated for each of the critical stages.

cum sunt pieile tăbăcite, suferă o serie de procese de destrucție micro sau macrostructurală, care stau la baza efectelor de deteriorare, ce modifică starea fizică a elementelor structural-funcționale și o serie de procese de alterare chimică, electrochimică, biochimică și radiativă (radiația termică și fotonică), ce modifică natura chimică a materialului, și care își pun amprenta asupra modificării unor caracteristici chimice (de exemplu modificarea naturii unor substanțe grase, etc.), fizico-structurale (modificarea gradului de cristalinitate a unei structuri, deplasările sau degradările cromatice etc.). Elaborarea unor metode de determinare a acestor modificări permite realizarea unor sisteme noi de coasistare sau coroborare în vederea stabilirii cu precizie a unor elemente implicate în modificările în timp ale unor materiale.

Caracterizarea materialelor proteice, cum sunt pielea și pergamantul, a rămas încă o problemă ce trebuie rezolvată, datorită neomogenității lor, fiind produse naturale. Proprietățile lor chimice, cât și cele fizico-mecanice sunt de mare interes pentru utilizator, în special în domeniul restaurării obiectelor istorice de patrimoniu. În afară de materia primă (pieile crude), un interes special trebuie acordat și tehnologiilor de producție, coloranților folosiți, conținutului în aminoacizi, metodei de tăbăcire, etc.

Metodele chimice și metodele fizico-mecanice care sunt actualmente la dispoziție sunt laborioase și sunt consumatoare de cantități mari de reactivi chimici și material și necesită mult timp pentru execuție. La obiectele de patrimoniu nu este posibilă aplicarea acestor metode, deoarece ar produce distrugerea acestora. Ca urmare, sunt necesare metode moderne pentru controlul rapid al calității, nedestructive. Astfel de metode sunt bazate pe spectroscopie IR (infraroșu) sau NIR (infraroșu apropiat). În special spectrele NIR sunt foarte bogate în benzi combinate care conțin o multitudine de informații care doar rareori sunt analizate într-o manieră analitică.

Studiile de modelare statistico-matematică (statistice, chemometrice) a datelor analitice pot pune în evidență relații „ascunse” între sisteme chimice înrudite, sau între etape ale proceselor fizico-chimice pe care le parcurge un sistem dat. Pentru asigurarea soluționării corecte a problemelor cu specific chemometric, pentru fiecare dintre etapele critice s-au elaborat algoritmi procedurali și/sau numerici.

SPSS Software (Statistical Package for the Social Sciences)

For the statistical interpretation of obtained data, the SPSS Software was used, one of the most used statistical data analysis software. The first version appeared in 1968, evolving to version 16 up to now, and the applicability area extended from one version to the next, together with the operating method and offered facilities. In addition to statistical analysis, the software provides strong components for data management (selecting, reconfiguring, creating new data), and data documentation (there is a metadata dictionary, which remembers data characteristics). To this, a remarkable flexibility as far as the types of data accepted and the report building method is added.

Important statistics which can be made by means of SPSS Base:

Linear Regression explores relationships between predictors and what is desired to be predicted (for instance, sales according to price and customer type).

Factor Analysis identifies variables or factors explaining correlations between a set of noticed variables. For instance, in data reduction, factor analysis is useful to identify a smaller number of factors explaining complex phenomena in which a very large number of variables is involved.

TwoStep Cluster analysis: this algorithm allows manipulation of continuous and categorical variables or attributes; it implies only one step in the procedure and discovers the right number of clusters. It is used for large data sets. For instance, it can be applied to data describing buying behaviour, gender, age, income, then the marketing strategies will be created for each customer group, in order to increase sales and build loyalty.

K-means Cluster Analysis groups data from larger sets of data. It involves a known number of clusters. A marketing analyst may wish to group cities in homogenous groups in order to discover comparable cities to which to address differently.

Hierarchical Cluster Analysis: SPSS takes clusters from a single entry and forms groups until they are all united. It has more than 40 measures of similarity and dissimilarity, of standardizing data according to different methods, groups of case or groups of

Programul SPSS (Statistical Package for the Social Sciences)

Pentru interpretarea statistică a datelor obținute sa utilizat programul SPSS, unul dintre cele mai utilizate programe software de analiză statistică a datelor. Prima versiune a apărut în anul 1968, a evoluat până în prezent la versiunea 16, și aria de aplicabilitate s-a extins de la versiune la versiune, o dată cu modul de operare și cu facilitățile oferite. În afară de analizele statistice, programul oferă componente puternice pentru managementul datelor (selectare, reconfigurare, creare de date noi), și pentru documentarea datelor (există un dicționar metadata, care reține caracteristici ale datelor). La acesta se adaugă o deosebită flexibilitate în ceea ce privește tipurile de date acceptate și modulul de construire a rapoartelor.

Statisticile importante ce se pot realiza cu ajutorul SPSS Base:

Linear Regression – regresia liniară – explorează relațiile dintre predictori și ceea ce se dorește prezis (de exemplu, vânzările în funcție de preț și tipul de client).

Factor Analysis – analiza factorială – identifică variabilele sau factorii care explică corelațiile între un set de variabile observate. De exemplu, în reducerea datelor, analiza factorială e utilă la identificarea unui număr mai mic de factori care explică fenomene complexe în care este implicat un număr foarte mare de variabile.

TwoStep Cluster analysis: acest algoritm permite manipularea variabilelor sau atributelor continue și categoriale; presupune doar un singur pas în procedură și descoperă numărul potrivit de clustere. Se folosește pentru seturi foarte mari de date. De exemplu, poate fi aplicată datelor care descriu comportamentul de cumpărare, sexul, vârstă, venitul, după care vor fi croite strategiile de marketing pentru fiecare grup de clienți, pentru a spori vânzările și a construi loialitatea.

K-means Cluster Analysis grupează datele din seturi mari de date. Presupune un număr cunoscut de clustere. Un analist de marketing poate dori să grupeze orașele în grupuri omogene pentru a descoperi orașe comparabile, cărora să se adreseze diferit.

Hierarchical Cluster Analysis: SPSS ia clusterele dintr-o singură înregistrare și formează grupuri până când toate sunt unite. Dispune de mai mult de 40 măsuri ale similarității și disimilarității, de standardizare a datelor după diferite metode, grupuri

variables. It generates measures of distance and similarity. It displays statistics in each stage to help select the optimal solution in terms of cluster number. The procedure is recommended for smaller sets of data, such as target groups. A marketing analyst can use Hierarchical Cluster Analysis to identify television shows with similar ratings for each type of show or it can group them into homogenous clusters based on viewer characteristics, to identify segments to which to address.

Description of chemometric techniques

Chemometrics elements

According to the generic definition, accepted by The International Chemometrics Society (ICS), chemometrics is "... *the science of correlating, by statistic-mathematic techniques, of measurements conducted on a chemical system or process, its state being in given conditions or at a given time.*" Although the term of chemometrics has been introduced in 1972 by the Swede Svante Wold [1] and the American Bruce R. Kowalski, the operative definition of chemometrics, stated by practitioners, was established in 1986 and presented in the editorial from the first issue of Chemometrics and Intelligent Laboratory Systems journal, in the following sentence: „*Chemometrics is that discipline in the field of chemistry which uses methods of mathematics and statistics to formulate or select experimental plans and procedures of physical-chemical analysis, in view of extracting maximum information by algorithmic processing of data obtained as a result of studies conducted at laboratory or industrial level.*”

The pragmatic purpose of chemometrics is that of establishing quantitative relationships between related or logically derived forms of one and the same product or of a series of products in different states, as well as numeric descriptions of those states connected in physical-chemical processes clearly defined in time, resorting to chemical or instrumental analytical determinations. In order to achieve this goal, the data obtained experimentally are processed through advanced techniques of mathematical statistics, numeric analysis and information theory, employing highly optimized algorithms transposed in professional software applications.

Usually, chemometrics is resorted to in situations

de cazuri sau grupuri de variabile. Generează măsuri ale distanței și similarității. Afisează statistici la fiecare stadiu pentru a ajuta în selectarea soluției optime ca număr de clustere. Procedura e recomandată pentru seturi de date mai mici, ca grupurile-țintă. Un analist de marketing poate folosi Hierarchical cluster analysis pentru a identifica emisiunile TV care atrag audiente similare pentru fiecare tip de emisiune sau le poate grupa în clusteri omogeni în baza caracteristicilor telespectatorului, pentru a identifica segmentele cărora să li se adreseze.

Descrierea tehnicilor chemometrice

Elemente de chemometrie

Conform definiției generice, acceptată de către Societatea Internațională de Chemometrie (ICS, The International Chemometrics Society), chemometria este „... *știința corelării, prin tehnici statisticomatematice, a măsurătorilor efectuate asupra unui sistem sau a unui proces chimic, cu starea acestuia în condiții date sau la un moment dat în timp.*” Deși termenul de chemometrie a fost introdus în 1972 de către suedezul Svante Wold [1] și americanul Bruce R. Kowalski, definiția operativă a chemometriei, enunțată prin prisma practicienilor, a fost stabilită în 1986 și prezentată în editorialul din primul număr al revistei Chemometrics and Intelligent Laboratory Systems, cu următorul enunț: „*Chemometria este acea disciplină din sfera chimiei, care utilizează metode ale matematicii și statisticii pentru a formula sau selecta planuri experimentale și proceduri de analiză fizico-chimică, în vederea extragerii unui maxim de informații prin procesarea algoritmică a datelor obținute în urma studiilor conduse la nivel de laborator sau industrial.*”

Scopul pragmatic al chemometriei este acela de a stabili relații cantitative între forme înrudite sau logic derivate ale unui același produs sau ale unei serii de produse aflate în diverse stări, precum și descrieri numerice ale respectivelor stări înlăntuite în procese fizico-chimice net definite în timp, recurgând la determinări analitice de factură chimică sau instrumentală. În vederea atingerii acestui scop, datele obținute pe cale experimentală sunt prelucrate prin tehnici avansate ale statisticii matematice, analizei numerice și teoriei informației, apelând la algoritmi înalt optimizați, transpuși în aplicații software profesionale.

Uzual, la chemometrie se apelează în situațiile în

where theoretical knowledge of physical-chemical features of processes and states is not advanced enough to completely describe or correctly solve problems arisen in the study of physical-chemical systems. From this point of view, chemometric techniques can be considered to be semi-empirical, because they offer nondeterministic and sometimes stochastic models, or non-parametric mathematical relations to describe physical-chemical systems. The objective of chemometric techniques is that of highlighting complex and non-obvious relationships ("hidden") between a large number of variables associated with studied systems, resorting to statistic-mathematic models capable of describing states of physical-chemical systems, at a given time or in their dynamic. Figure 1 schematizes the typical process of chemometric techniques.

care cunoașterea teoretică a fizico-chimismului proceselor și stărilor nu este suficient de avansată pentru a descrie complet sau pentru a soluționa corect problemele apărute în studiul sistemelor fizico-chimice. Din acest punct de vedere, tehniciile chemometrice pot fi considerate ca fiind semi-empirice, deoarece oferă modele nedeterministe și uneori stocastice, sau relații matematice neparametrice, pentru a descrie sistemele fizico-chimice. Obiectivul tehniciilor chemometrice este acela de a pune în evidență relații complexe și neevidente („ascunse”) între un număr mare de variabile asociate sistemelor studiate, apelând la modele statistico-matematice apte a descrie stările sistemelor fizico-chimice, la un moment dat ori în dinamica acestora. Figura 1 schematizează demersul tipic al tehniciilor chemometrice.

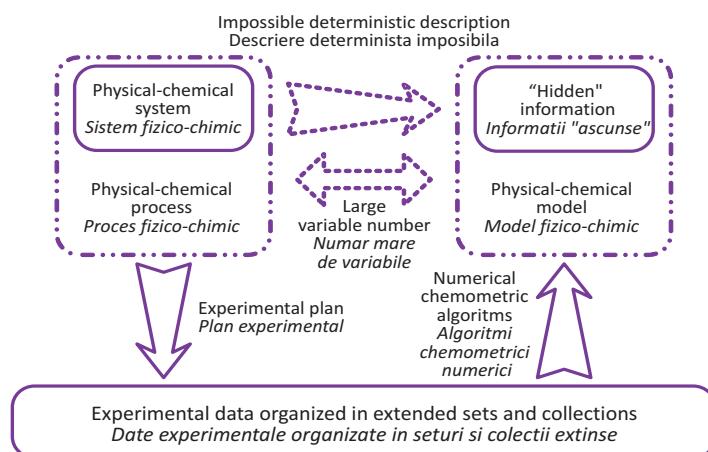


Figure 1. Process of chemometric techniques
Figura 1. Demersul tehniciilor chemometrice

Chemometric studies are conducted going through the following stages:

- defining the chemometric experiment;
- collecting, preprocessing and systematizing chemometric data;
- chemometric modeling and interpreting experimental data from a physical-chemical point of view.

The general strategy of chemometric studies consists in obtaining a large number of experimental data, organizing them in classes and categories, followed by statistical-mathematical modeling and validation of models attributed to those data. The

Studiile chemometrice se derulează parcurgând următoarele etape:

- definirea experimentului de factură chemometrică;
- colectarea, preprocesarea și sistematizarea datelor chemometrice;
- modelarea chemometrică și interpretarea în termeni fizico-chimici a datelor experimentale.

Strategia generală a studiilor chemometrice constă în obținerea unui mare număr de date experimentale, organizarea lor în clase și categorii, urmată de modelarea statistico-matematică și validarea modelelor asociate respectivelor date. Modelele astfel obținute pot fi apoi utilizate pentru interpretarea,

models thus obtained can be then used to interpret, simulate and predict introducing new experimental data in the place of variables considered when planning experiments. Data masses, organized in structured databases, can also be used in Data Mining studies, to extract new "hidden" information regarding chemical systems and processes, as the chemometric analytical data gather up.

The central concept in designing and conducting chemometric studies is that of describing, analyzing, classifying and multidimensional modeling of numeric data, generated by techniques of quantitative chemical analysis. This concept derives from the large number of variables and experimental responses considered (manipulated and measured) in chemometric experiments. The algorithms used are capable of extracting and systematically presenting obvious and "hidden" characteristics of studied systems or processes. In this respect, analytical numeric data must lend themselves to being organized in a matrix, according to at least three criteria: (i) studied systems or the sequence of stages in the studied processes, (ii) variables manipulated within the experimental plan and (iii) the nature of experimental responses quantitatively measured in the experiment. Thus, a minimal chemometric study will provide a matrix of at least three dimensions, whose numeric processing requires special algorithms. Figure 2 presents the organization of chemometric data in matrix form with three dimensions: lines (attributed to experimental variables), columns (attributed to measured responses) and layers (attributed to studied systems). Each intersection of lines and columns represents an obvious defining characteristic of the system to which it is attributed.

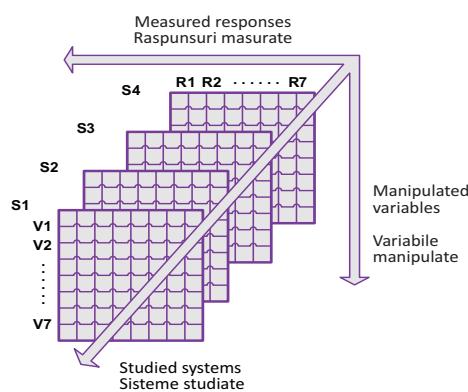
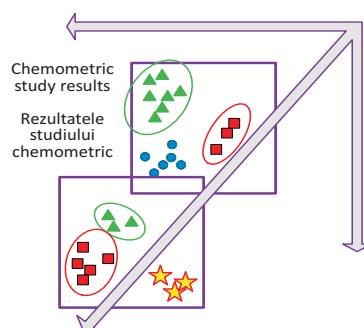


Figure 2. Multidimensional matrix organization of chemometric data and one of the ways of classifying obtained experimental data
Figura 2. Organizarea matricială multidimensională a datelor chemometrice și unul dintre modurile de clasificare a datelor experimentale obținute

simularea și predicția amplasării unor noi date experimentale în spațiul variabilelor avute în vedere în momentul planificării experimentelor. Masivele de date, organizate în baze de date structurate, pot fi, de asemenea, utilizate în studii de Data Mining, pentru extragerea de noi informații „ascunse” asupra sistemelor și a proceselor chimice, pe măsură ce datele analitice cu specific chemometric se acumulează.

Conceptul central în proiectarea și derularea studiilor chemometrice este acela al descrierii, analizei, clasificării și modelării multidimensionale a datelor numerice, date generate prin tehnici ale analizei chimice cantitative. Acest concept derivă din numărul mare de variabile și răspunsuri experimentale avute în vedere (manipulate și respectiv măsurate) în experimentele chemometrice. Algoritmii utilizați sunt capabili a extrage și a prezenta în mod sistematizat caracteristicile evidente și pe cele „ascunse” ale sistemelor ori proceselor studiate. În acest sens, datele analitice numerice trebuie să poată fi organizate matricial, în funcție de cel puțin trei criterii: (i) sistemele studiate ori secvența etapelor în procesele studiate, (ii) variabilele manipulate în cadrul planului experimental și (iii) natura răspunsurilor experimentale măsurate cantitativ în experiment. Așadar, un studiu chemometric minimal va furniza matrice cu cel puțin trei dimensiuni, a căror procesare numerică necesită algoritmi speciali. Figura 2 prezintă modul de organizare a datelor chemometrice sub formă matricială cu trei dimensiuni: linii (asociate variabilelor experimentale), coloane (asociate răspunsurilor măsurate) și straturi (asociate sistemelor studiate). Fiecare intersecție a liniilor și coloanelor reprezintă o caracteristică definitorie evidentă a sistemului căruia îi este asociată.



By exploiting the multidimensional matrix resulted at the end of the experiment, chemometric algorithms are capable of highlighting "hidden" characteristics of studied systems and group all available characteristics (either obvious or "hidden") into classes, both attributed to a given system, and common to the set of studied systems. Figure 3 presents an example in this sense.

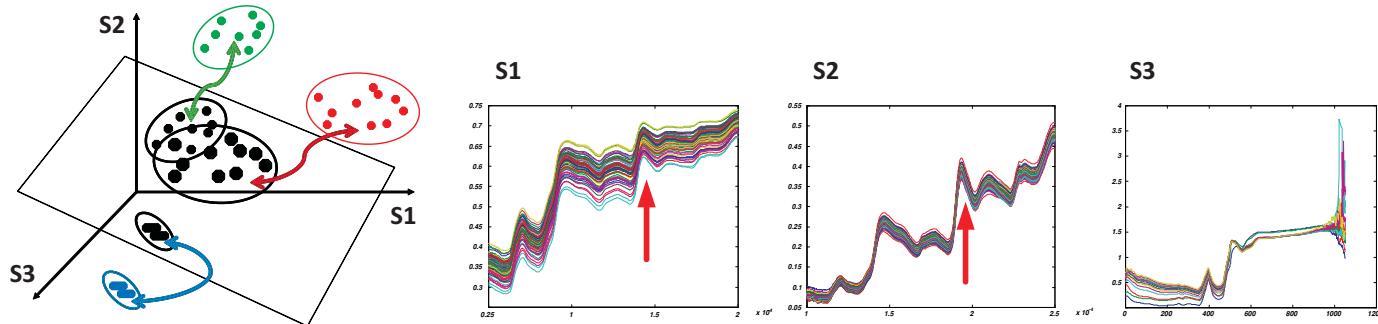


Figure 3. Example of grouping obvious characteristics of three systems studied by NIR spectroscopy, with the extraction of one common "hidden" characteristic for two of those systems

Figura 3. Exemplu de grupare a caracteristicilor evidente a trei sisteme studiate prin spectroscopie NIR, cu extragerea unei caracteristici „ascunse” comune pentru două dintre respectivele sisteme

Initiating a chemometric study is carried out going through the following steps:

- *establishing chemometric descriptors*, by identifying measurable properties of the studied system, properties which are found in the characteristics by means of which the given system will be compared to logically associated systems or with its successive states in developing a process;
- *establishing variables with univocal effect on descriptors*, through theoretic or experimental study of the correlation between the two categories;
- *selecting measurable experimental responses*, starting from the list of descriptors; descriptors with the highest values of correlation coefficients between these and the manageable variables previously established will be selected as responses;
- *postulating an adjustable model to experimental responses*, choosing from the range of those applicable to multidimensional data;
- *establishing applicable experimental plans*, in terms of type and amplitude;
- *establishing ranges of variation of manageable variables*, in correlation to the nature of the adopted experimental plan.

Algoritmii chemometrici sunt capabili ca, exploatajând matricea multidimensională rezultată la finalul experimentului, să pună în evidență caracteristici „ascunse” ale sistemelor studiate și să grupeze toate caracteristicile disponibile (fie ele evidente sau „ascunse”) în clase, atât asociate unui sistem dat, cât și comune setului de sisteme studiate. Figura 3 prezintă un exemplu în acest sens.

Inițierea unui studiu chemometric se realizează parcurgând următorii pași:

- *stabilirea descriptorilor chemometrici*, prin identificarea proprietăților măsurabile asupra sistemului studiat, proprietăți ce se vor regăsi între caracteristicile prin care sistemul dat va fi comparat cu sisteme asociate logic sau cu stări succesive ale sale în derularea unui proces;
- *stabilirea variabilelor cu efect univoc asupra descriptorilor*, prin studiu teoretic sau experimental al corelației între cele două categorii;
- *selectarea răspunsurilor experimentale măsurabile*, pornind de la lista descriptorilor; se vor selecta drept răspunsuri descriptorii cu cele mai ridicate valori ale coeficienților de corelație între ei și variabilele manipulabile anterior stabilite;
- *postularea unui model ajustabil răspunsurilor experimentale*, alegând din gama celor aplicabile datelor multidimensionale;
- *stabilirea planurilor experimentale aplicabile*, din punctul de vedere al tipului și amplorii;
- *stabilirea plajelor de variație a variabilelor manipulabile*, în corelație cu natura planului experimental adoptat.

All the above mentioned information will set the premises for conducting the selected experimental plan.

Techniques of modeling chemometric data can be classified into:

- *quantitative, regression techniques*, that provide numeric parameters of postulated models;
- *qualitative techniques, of recognizing models/patterns/structures of multicriterial grouping of experimental data*, providing images and sometimes numerical indexes of similarity/dissimilarity, on the basis of which classes and/or filiations of studied systems or of states of a system going through a given process can be identified.

The first category of techniques is the most used, given the fact that the chemometric process also aims at obtaining statistical-mathematical methods useful in classification and prediction.

In most situations, chemometric experiments provide a multitude of raw correlated data, whose algorithmic interpretation is difficult. Therefore, before applying techniques of chemometric modeling, raw data are subjected to a pre-processing in order to significantly remove or eliminate random or systematic effects of variability irrelevant within the study. Pre-processing is mainly carried out through four numerical techniques, namely [2]:

- *normalizing*, to eliminate systematic variability affecting an entire set of experimental data; it is carried out by splitting each entry from the data set at a constant value, calculated in terms of the characteristics of the data set as a whole;

- *weighting*, to eliminate systematic deviations with quasi-constant values; it is carried out by multiplying each entry in the data set with a constant value, calculated depending on the relative importance of individual entries;

- *smoothing*, to reduce "background noise" and improve the "useful signal-noise" ratio, in order to obtain data sets easily moldable by classic linear or non-linear methods; smoothing can be carried out by mediating values in a given proximity, by applying a local polynomial adjustment, through Fourier or wavelets techniques etc.;

- *baseline correction*, to eliminate inferior harmonics of variations captured in data sets by instrumental analytical techniques; it is carried out by

Toate informațiile mai sus specificate vor sta la baza conducerii planului experimental selectat.

Tehnicile de modelare a datelor chemometrice se pot clasifica în:

- *tehnici cantitative, de regresie*, ce furnizează parametri numerici ai modelelor postulate;
- *tehnici calitative, de recunoaștere a modelelor/sabioanelor/structurilor de grupare multicriterială a datelor experimentale*, ce furnizează imagini și uneori indici numerici de similaritate/disimilaritate, în virtutea cărora se pot identifica clase și/sau filiații ale sistemelor studiate, ori ale stărilor unui sistem ce parcurge un proces dat.

Prima categorie de tehnici este cel mai frecvent utilizată, dat fiind faptul că demersul chemometric vizează și obținerea de modele statistică-matematice utile în clasificare și predicție.

În majoritatea situațiilor, experimentele chemometrice furnizează o multitudine de date brute corelate, a căror interpretare algoritmică este dificilă. De aceea, înaintea aplicării tehniciilor de modelare chemometrică, datele brute sunt supuse unei preprocesări pentru a îndepărta sau a diminua semnificativ efectele variabilității irelevante în cadrul studiului, de faptură aleatoare ori sistematică. Preprocesarea se realizează, în principal, prin patru tehnici numerice, respectiv [2]:

- *normare*, pentru eliminarea variabilității sistematice care afectează un întreg set de date experimentale; se realizează prin împărțirea fiecărei înregistrări din setul de date la o valoare constantă, calculată ținând cont de caracteristicile setului de date în ansamblul său;

- *ponderare*, pentru eliminarea abaterilor sistematice cu valori cvasi-constante; se realizează prin înmulțirea fiecărei înregistrări din setul de date cu o valoare constantă, calculată în funcție de importanța relativă a înregistrărilor individuale;

- *netezire*, pentru reducerea „zgomotului de fond” și îmbunătățirea raportului „semnal util – zgomot”, în vederea obținerii unor seturi de date ușor modelabile prin metode liniare sau neliniare clasice; netezirea se poate realiza prin medierea valorilor într-o vecinătate dată, prin aplicarea unei ajustări polinomiale locale, prin tehnici Fourier ori wavelets etc.;

- *corecția liniei de bază*, pentru eliminarea armonicelor inferioare ale variațiilor capturate în seturile de date prin tehnici analitice instrumentale; se realizează prin aplicarea unor ajustări optimale,

applying optimal adjustments, using polynomial functions or the derivation technique on the whole raw database.

Following pre-processing, data sets may suffer contractions from the point of view of number of values per set, but they will be more "expressive" and easier to mold.

EXPERIMENTAL

Protein materials, such as leather and parchment, are widespread and have multiple uses. However, their characterization is still a problem to be solved, due to their non-homogeneity, being natural products. Their chemical and physical mechanical properties are of great interest for the user, particularly in the field of restoring historical heritage items. Besides raw material (raw hides), a special interest must be paid to production technologies, dyes used, amino acid content, tanning method, etc. Characterization of these materials is obtained by chemical and physical-mechanical methods which involve using a relatively large product sample. In heritage items, application of these methods is not possible, because it would cause their destruction.

Materials and methods used to characterize leathers

Chemical and physical-mechanical methods currently available for leather characterization are laborious and consume large quantities of chemical reagents and material, and their application requires a lot of time. Therefore, modern, non-destructive methods are required for fast quality control. Such methods are based on IR (infrared) or NIR (near infrared) spectroscopy. NIR spectra particularly are very rich in combined bands containing a multitude of information which is rarely examined in an analytical manner.

In this paper 130 leather and parchment samples belonging to heritage items have been analyzed by NIR spectroscopy (carried out in Slovenia, at the Ljubljana University, Faculty of Chemistry and Chemical Technology, with *Spectrometer NIR – LabSpec 5000*).

The obtained data have been gathered in a comprehensive, searchable database, which structures

utilizând funcții polinomiale ori tehnica derivării pe întreg setul de date brute.

În urma preprocesării, seturile de date pot suferi contracții din punctul de vedere al numărului de valori per set, dar vor fi mai „expresive” și mai ușor de modelat.

PARTEA EXPERIMENTALĂ

Materialele proteice, cum sunt pielea și pergamentul sunt foarte răspândite și au multiple utilizări. Cu toate acestea, caracterizarea lor a rămas încă o problemă ce trebuie rezolvată, datorită neomogenității lor, fiind produse naturale. Proprietățile lor chimice, cât și cele fizico-mecanice sunt de mare interes pentru utilizator, în special în domeniul restaurării obiectelor istorice de patrimoniu. În afară de materia primă (piele crude), un interes special trebuie acordat și tehnologilor de producție, coloranților folosiți, conținutului în aminoacizi, metodei de tăbăcire, etc. Caracterizarea acestor materiale se obține prin metode chimice și metode fizico-mecanice, care presupun utilizarea unei mostre relativ mari de produs. La obiectele de patrimoniu nu este posibilă aplicarea acestor metode, deoarece ar produce distrugerea acestora.

Materiale și metode utilizate pentru caracterizarea pieilor

Metodele care sunt actualmente la dispoziție sunt laborioase și sunt consumatoare de cantități mari de reactivi chimici și material și necesită mult timp pentru execuție. Ca urmare, sunt necesare metode moderne pentru controlul rapid al calității, nedestructive. Astfel de metode sunt bazate pe spectroscopie IR (infraroșu) sau NIR (infraroșu apropiat). În special spectrele în NIR sunt foarte bogate în benzi combinate care conțin o multitudine de informații care doar rareori sunt analizate într-o manieră analitică.

În cadrul acestei lucrări au fost analizate 130 de probe de piele și pergament aparținând unor obiecte de patrimoniu prin spectroscopie NIR (realizate în Slovenia, la Universitatea din Ljubljana, Facultatea de Chimie și Tehnologie Chimică cu *Spectrometer NIR – LabSpec 5000*).

Datele obținute s-au constituit într-o bază de date comprehensivă, interogabilă, care structurează de o manieră simplă, ușor accesibilă, inteligibilă, rezultatele complexe ale analizelor.

complex analysis results in a simple, easily accessible and intelligible manner.

Information found in the database for each analyzed sample comprises:

- name of the leather or parchment heritage item from which the sample has been taken;
- dating it according to year;
- area from which the sample was taken (recto or verso of the item);
- absorbance according to wavelength, investigating wavelengths between 350 nm and 2500 nm;

RESULTS AND DISCUSSIONS

Statistical descriptive analysis of data with SPSS software

For the statistical processing, another database was built with the help of SPSS statistical software. To facilitate processing, data has been rearranged (Transpose command) so that the cases (samples) would be listed in rows and variables (year, sample area, wavelengths) would be listed in columns.

Next, the variable type (number or string), the number of characters for their visualization, measuring scale (nominal, ordinal, scalar), missing values for variables etc. have been defined.

According to specific of the used measuring device (spectrometer with 3 detectors), elimination of 3 wavelength ranges, in which the three detectors overlap, altering results, namely 350 – 450 nm, 990-1100 nm, 1790-1810 nm, was necessary for all analyzed samples.

Database distribution according to age of heritage items

In the database, average age of analyzed samples is 169 years (samples dating from 1840). The oldest ones date from 1560, and the newest from 2006-2009 (Table 1):

Table 1: Statistic indicators for years of origin of heritage items
Tabelul 1: Indicatori statistici pentru anii de proveniență ai obiectelor de patrimoniu

N Valid Valabil	N	N Arithmetic mean Medie aritmetică	N Skewness Aplecare	N Standard kurtosis error Eroare Std. de boltire
109	109	1840,04	,167	,459
7	7	1831,00	,231	1560
		2009	,388	2009
		Mode Mod	Kurtosis Boltire	Maximum Maxim

Informațiile ce se regăsesc în baza de date pentru fiecare probă analizată cuprind:

- denumirea obiectului de patrimoniu din piele sau pergament din care a fost prelevată proba;
- datarea acestuia după an;
- zona de prelevare a probei (față sau verso-ul obiectului);
- absorbanță în funcție de lungimea de undă, fiind investigate lungimi de undă între 350 nm - 2500 nm;

REZULTATE ȘI DISCUȚII

Analiza statistică descriptivă a datelor cu programul SPSS

Pentru prelucrarea statistică a fost construită o altă bază de date cu ajutorul programului statistic SPSS. Pentru ușurință prelucrării, datele au fost rearanjate (comanda Transpose) astfel încât cazurile (probele) să fie listate pe rânduri, iar variabilele (anul, zona de prelevare a probei, lungimile de undă) să fie listate pe coloane.

În continuare, s-au definit: tipul de variabile (de tip număr sau string), numărul de caractere pentru vizualizarea lor, scala de măsurare (nominală, ordinală, scalară), valorile lipsă pentru variabile etc.

În conformitate cu specificitatele aparatului de măsurare utilizat (spectrometru cu 3 detectori), a fost necesară eliminarea pentru toate probele analizate, a 3 intervale de lungimi de undă în care cei trei detectori se suprapun, alterând rezultatele, și anume: 350-450 nm, 990-1100 nm, 1790-1810 nm.

Distribuția bazei de date în funcție de vechimea obiectelor de patrimoniu

În cadrul bazei de date, vechimea medie a probelor analizate este de 169 ani (probe datând din 1840). Cele mai vechi probe datează din 1560, iar cele mai noi din anii 2006-2009 (Tabelul 1):

Distribution of analyzed samples in terms of the year of origin is presented in Figure 4:

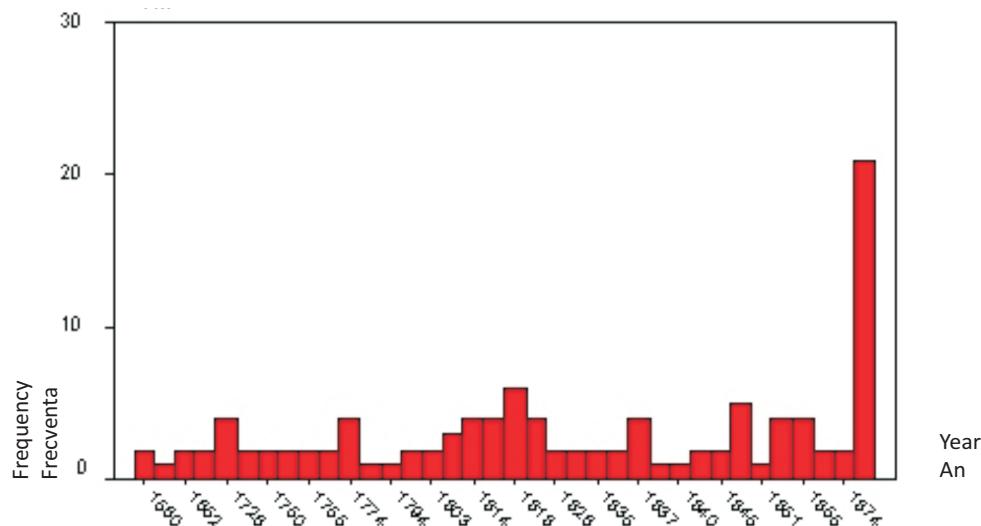


Table 2: Statistic indicators for century of origin of heritage items
 Tabelul 2: Indicatori statistici pentru secolele de proveniență ale obiectelor de patrimoniu

	Century Secol	Frequency Frecvență	Percent Procentaj	Valid percent Procentaj valabil	Cumulative percent Procentaj cumulativ
Valid Valabil	16	2	1.7	1.8	1.8
	17	3	2.6	2.8	4.6
	18	22	19.0	20.2	24.8
	19	61	52.6	56.0	80.7
	21	21	18.1	19.3	100.0
Total		109	94.0	100.0	
Missing Lipsă		7	6.0		
Total		116	100.0		

For analyzed heritage leathers, contraction temperatures have ranged between 12 and 70°C, according to the method used for tanning and the degradation stage they exhibit.

Inferential statistical analysis of data with the SPSS software

As far as the inferential statistic procedures, these have been very difficult, if not impossible to carry out in the case of the complete database, given its very large size, the extremely large number of variables, impossible to process with the available statistical software, SPSS. The data analysis was made using the Unscrambler, MVA (Multivariate Analyis) Software.

Statistical (chemometric) analysis of data obtained with the Unscrambler, MVA (Multivariate Analysis) Software

NIR spectra contain a lot of information, but the spectral features are few and not very well resolved. Multivariate analysis is therefore necessary to extract the information and correlate it with the observed or measured chemical or physical properties. Principal Component Analysis (PCA) and Partial Least Squares (PLS) are commonly used methods used in multivariate analyses.

Principal component analysis (PCA) is a standard tool in data analysis, in diverse fields from neuroscience to computer graphics, because it is a simple, non-parametric method for extracting relevant information from large data sets [3].

Pentru pieile de patrimoniu analizate temperaturile de contractie s-au situat intre 12-70°C, in functie de modul in care au fost tabacite si de stadiul de degradare pe care-l prezinta.

Analiza statistică inferențială a datelor cu programul SPSS

În ceea ce privește procedurile de statistică inferențială, acestea au fost foarte dificil, dacă nu chiar imposibil de realizat în cazul bazei de date complete, dată fiind dimensiunea foarte mare a acesteia, numărul foarte mare de variabile, imposibil de prelucrat cu programul statistic disponibil, SPSS. S-a realizat analiza statistică a datelor cu programul Unscrambler, MVA (Multivariate Analyis) Software.

Analiza statistică (chemometrică) a datelor obținute cu programul Unscrambler, MVA (Multivariate Analyis) Software.

Spectrele NIR conțin o mulțime de informații, dar caracteristicile spectrale sunt puține și nu sunt foarte bine rezolvate. Prin urmare, analiza multivariată este necesară pentru a extrage informația și pentru a corela cu proprietățile fizice sau chimice observate sau măsurate. Analiza Componentelor Principale (PCA) și Partial Least Squares (PLS) sunt metode utilizate frecvent utilizate în analiza multivariată.

Analiza componentelor principale (PCA) este un instrument standard în analiza datelor, în diverse domenii, de la neuroștiință la grafică pe calculator, deoarece este o metodă simplă, non-parametrică pentru extragerea informațiilor relevante din seturi mari de date [3].

PCA is a mathematical procedure that transforms a number of (possibly) correlated variables into a (smaller) number of uncorrelated variables called principal components. The objective of principal component analysis is to reduce the dimensionality (number of variables) of the dataset but retain most of the original variability in the data. The first principal component accounts for as much of the variability in the data as possible, and each succeeding component accounts for as much of the remaining variability as possible. A principal component analysis is concerned with explaining the variance/covariance structure of a high dimensional random vector through a few linear combinations of the original component variables [4].

Partial least-squares (PLS) modeling is a powerful multivariate statistical tool that has been successfully applied to quantitative analyses of ultraviolet, near-infrared and electrochemical data [5]. The general idea of PLS is to try to extract latent factors, accounting for as much of the manifest factor variation as possible while modeling the responses [6]. Using PLS, more than one variable (e.g. content, concentration) can be used simultaneously avoiding the problem of co-linearity. PLS is a method for constructing predictive models when the factors are many and highly collinear. The emphasis is on predicting the responses and not necessarily on trying to understand the underlying relationship between the variables. For example, PLS is usually not appropriate for screening out factors that have a negligible effect on response. However, when prediction is the goal and there is no practical need to limit the number of measured factors, PLS can be a useful tool [7].

PCA este o procedură matematică care transformă un număr de variabile (posibil) corelate într-un număr (mai mic) de variabile necorelate, numite componente principale. Obiectivul analizei componentelor principale este de a reduce dimensiunile (numărul de variabile) setului de date, dar păstrând o mare parte din variabilitatea originală a datelor. Prima componentă principală justifică o parte cât mai mare din variabilitatea datelor, și fiecare componentă ulterioară justifică o parte cât mai mare din variabilitatea rămasă. O analiză a componentelor principale se ocupă cu explicarea structurii de varianță/covarianță a unui vector aleatoriu de mari dimensiuni prin câteva combinații liniare ale variabilelor componente originale [4].

Modelarea Partial least-squares (PLS) este un instrument puternic de statistică multivariată care a fost aplicat cu succes în analize cantitative ale datelor în ultraviolet, infraroșu apropiat și date electrochimice [5]. Ideea generală a PLS este de a încerca să extragă factori latenți, justificând cât mai mult din variația evidentă a factorilor, în timp ce modelează răspunsurile [6]. Folosind PLS, pot fi folosite simultan mai multe variabile (de exemplu, conținut, concentrație), evitând problema coliniarității. PLS este o metodă de construire a modelelor de predicție, atunci când există mai mulți factori și sunt foarte coliniari. Accentul se pune pe anticiparea răspunsurilor și nu neapărat pe încercarea de a înțelege relația dintre variabile. De exemplu, de obicei, PLS nu este adecvat pentru sortarea factorilor care au un efect neglijabil asupra răspunsului. Cu toate acestea, atunci când predicția este scopul și nu există o nevoie practică de a limita numărul de factori măsuраti, PLS poate fi un instrument util [7].

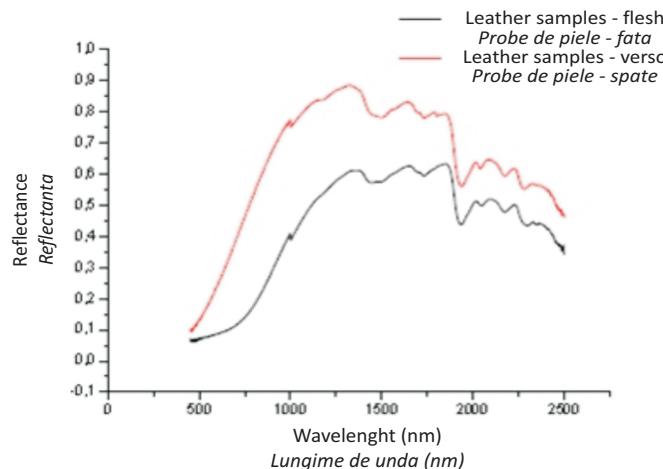


Figure 6. NIR spectrum - Difference between spectra of the recto side and the verso side of leather

Figura 6. Spectrul NIR – Diferența dintre spectre pe față și pe spatele pielii

An abrupt drop at 1000 and 1700 nm in both curves on Figure 6 can be detected. This is the result of using three independent detectors in the NIR instrument; however, it can be easily corrected. The first detector is used in the region between 350 and 1000 nm, the second one between 1000 and 1700 nm, and the third one between 1700 and 2500 nm.

O scădere bruscă la 1000 și la 1700 nm la ambele curbe se poate observa în Figura 6. Acesta este rezultatul utilizării unui număr de trei detectori independenți la instrumentul NIR; cu toate acestea, se poate corecta ușor. Primul detector se utilizează în regiunea dintre 350 și 1000 nm, al doilea între 1000 și 1700 nm, iar al treilea între 1700 și 2500 nm.

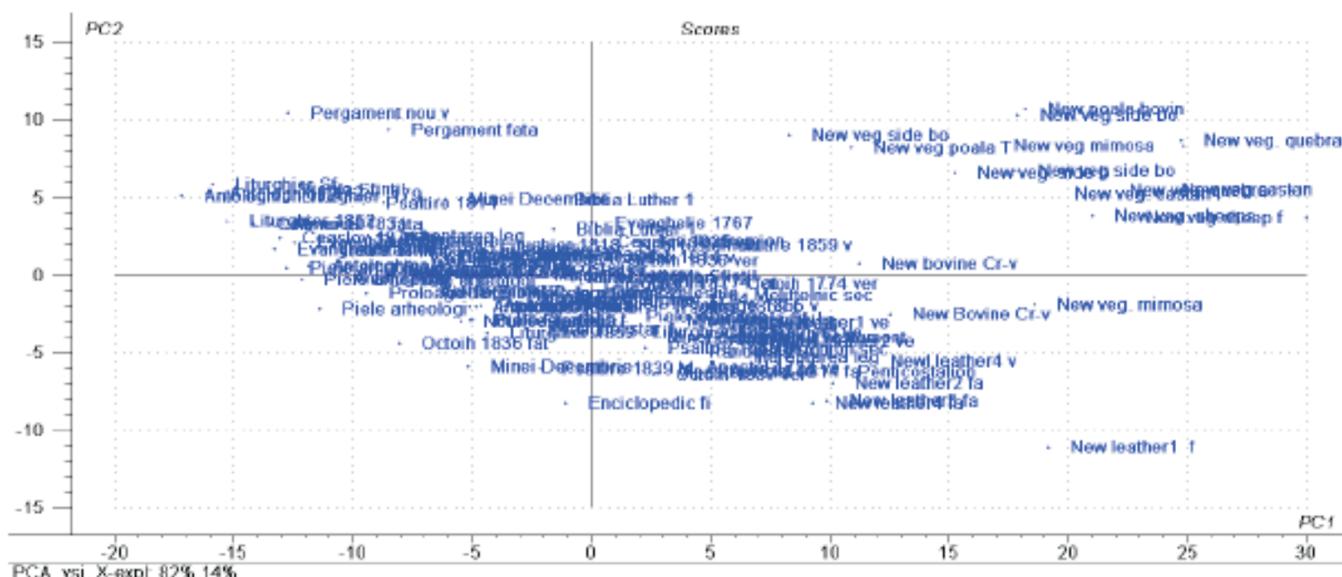


Figure 7. PCA analysis: all samples
 Figura 7. Analiza PCA: toate probele

With PCA analysis performed on all samples (Figure 7), two separate groups are immediately evident. The first group is represented by all historical samples from between 1560 and 1850, and the second group is represented by new leathers, made in 2007, 2008 and 2009. These samples are separated out in the right top quadrant. This result indicates that there is a systematic difference between the two groups of spectra, which is why in further PLS analysis only the historical samples are used.

With further PCA analysis we can determine that there is no significant difference between the recto and the verso sides of leather samples. In Figure 8, the recto (A) and the verso (B) sides are mixed.

După analiza PCA efectuată asupra tuturor probelor (Figura 7), ies imediat în evidență două grupuri separate. Primul grup constă din toate probele istorice din perioada 1560-1850, iar cel de-al doilea grup constă din pieile noi, realizate în 2007, 2008 și 2009. Aceste probe sunt separate în cadrul din dreapta-sus. Acest rezultat indică faptul că există o diferență sistematică între cele două grupe de spectre, de aceea, în analiza PLS ulterioară se folosesc doar probele istorice.

Cu ajutorul analizei PLS ulterioare putem determina faptul că nu există o diferență semnificativă între partea de pe față și cea de pe spate ale probelor de piele. În Figura 8, partea de pe față (A) și cea de pe spate (B) sunt amestecate.

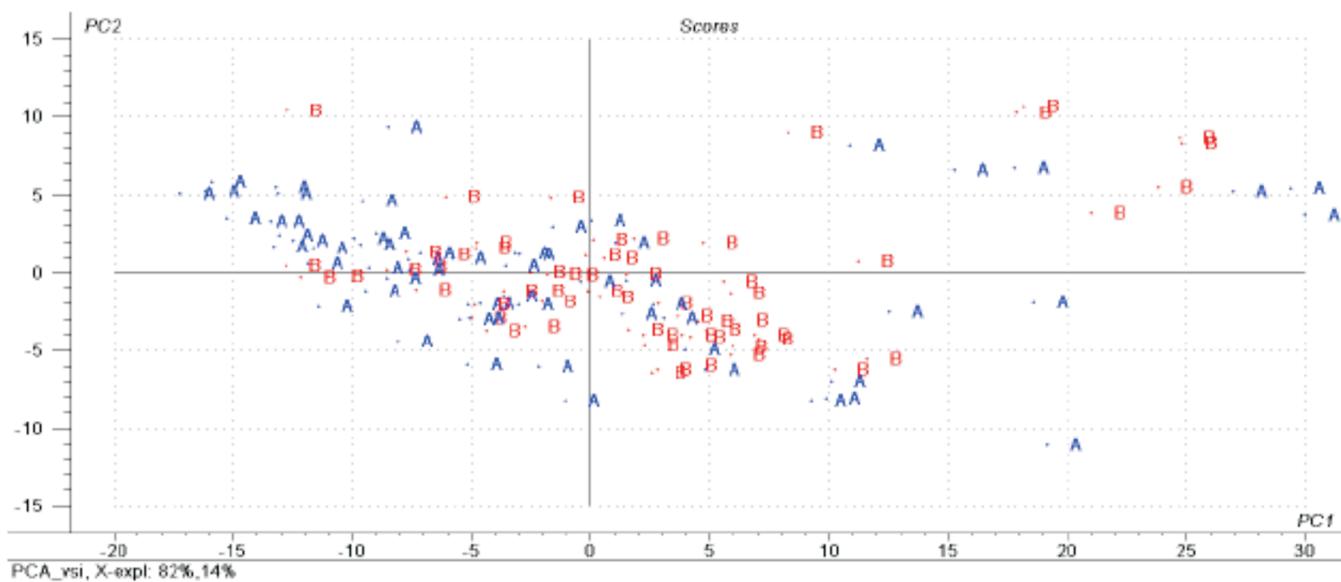


Figure 8: PCA analysis: all samples – (A) marks recto side of leather and (B) marks verso side of leather

Figura 8: Analiza PCA: toate probele – (A) marchează partea de pe față pielii; (B) marchează partea de pe spatele pielii

The PLS analysis was thus performed with 98 historical samples of leather (Figure 9 and Figure 10). In the calibration and the cross validation sets there are 62 samples and in the validation set 36 samples were used. All samples of leather had a known year of production, i.e. between 1550 and 1850.

Analiza PLS a fost efectuată cu 98 de probe istorice de piele (Figura 9 și Figura 10). În seturile de calibrare și contra-validare există 62 de probe, iar în setul de validare s-au folosit 36 de probe. Pentru toate probele de piele s-a cunoscut anul producției, și anume, între 1550 și 1850.

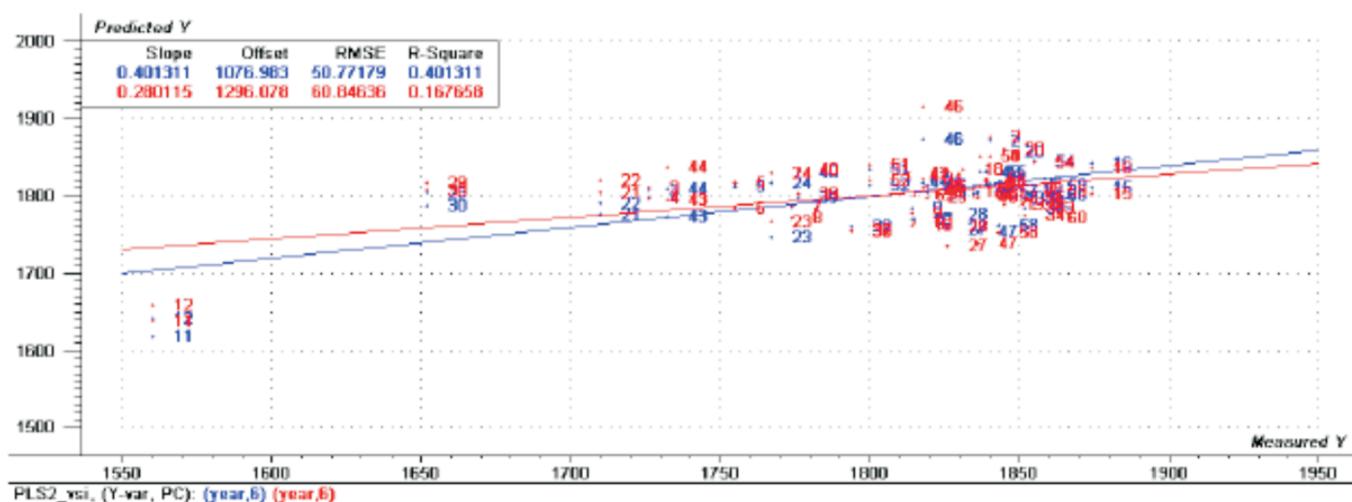


Figure 9. PLS analysis: historical samples between 1550 and 1850. The spectra for face side and verso side are combined together - they are not averaged

Figura 9. Analiza PLS: probe istorice din perioada 1550-1850. Spectrele pentru partea de pe față și cele de pe spate sunt combinate – nu s-a calculat media

Table 3: Parameters of PLS model
Tabelul 3: Parametrii modelului PLS

	Skewness <i>Înclinație</i>	Offset <i>Decalaj</i>	RMSE	R^2
Calibration <i>Calibrare</i>	0.40	1077	50.8	0.40
Cross validation <i>Contra-validare</i>	0.28	1296	60.8	0.17

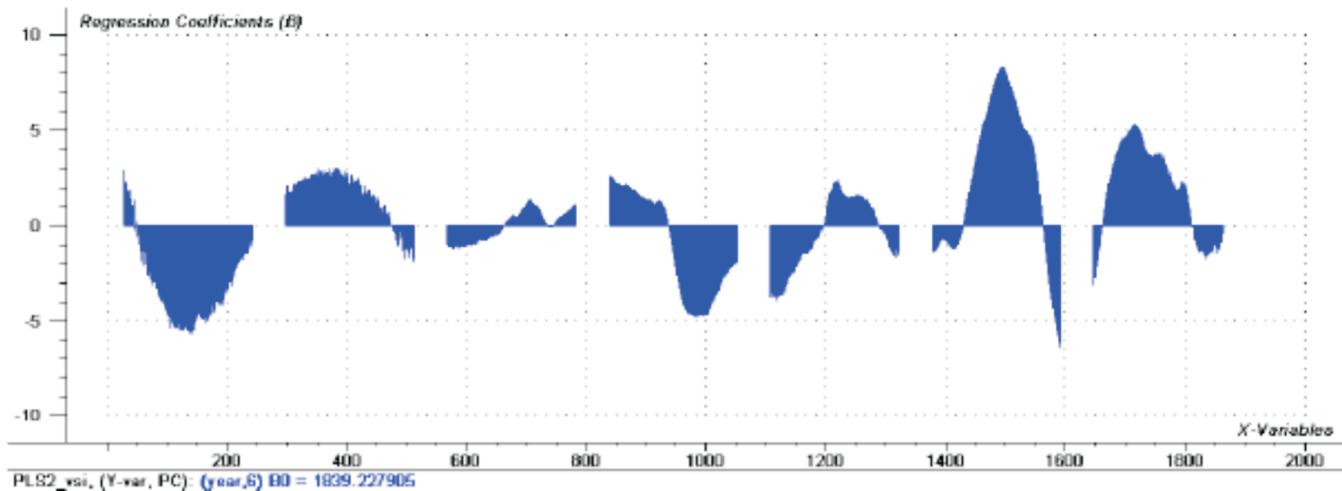


Figure 10. PLS loadings
Figura 10. Încărcăturile PLS

Even if different spectral pretreatments are used, the results are not satisfactory and no significant correlation between NIR spectra and the year of production of the historical leather (Figure 9, Figure 10). The slope of the calibration line is 0.4, which indicates a strong averaging effect of PLS, i.e. the predictions seem to be unrealistically close to each other, despite the fact that the prediction error is calculated to be 60 years.

There could be several reasons:

- Narrow time span: most samples are from 1700-1900.
- Too much variation between the samples (different kinds of leather) and too few samples, thus not covering the full space of variability.
- Remains of other organic material on leather (e.g. glues, cleaning materials, dressings).

Chiar dacă se utilizează pre-tratamente spectrale, rezultatele nu sunt satisfăcătoare și nu există nicio corelație semnificativă între spectrele NIR și anul de producție al pieilor istorice (Figura 9, Figura 10). Înclinația liniei de calibrare este de 0,4, ceea ce indică un efect puternic al PLS de a calcula media, adică predicțiile par să fie prea apropiate una de alta, în ciuda faptului că eroarea de predicție este de 60 de ani.

Ar putea exista mai multe motive:

- Perioada scurtă de timp: majoritatea probelor sunt din 1700-1900.
- Prea multă variație între probe (diferite tipuri de piele) și prea puține probe, neacoperind astfel spațiul complet al variabilității.
- Urme ale altor materiale organice pe piele (de ex., cleiuri, materiale de curățare, soluții de tratare).

CONCLUSIONS

Applying statistical-mathematic studies of modeling analytical data can highlight “hidden” connections between related chemical systems or between stages of physical-chemical processes a given system undergoes.

The applications of SPSS software in investigating leather heritage items are limited, due to the fact that a large data volume cannot be processed.

Applications of quantitative chemometric regression techniques are the most used, given the fact that the chemometric process aims at obtaining statistical-mathematical methods useful in classification and prediction. Chemometric experiments provide a multitude of correlated raw data, whose algorithmic interpretation is difficult. Therefore, before applying techniques of chemometric modeling, raw data are subjected to a pre-processing in order to significantly remove or eliminate random or systematic effects of variability irrelevant within the study. In this study, investigations will be continued in order to find a new data pre-processing formula.

Acknowledgements

This paper has been elaborated within the Romania – Slovenia bilateral collaboration programme, entitled: “Development of spectroscopic methods for quality control of proteinaceous materials”, contract 60CB/2008, financed by ANCS, in the period 2008-2009, and within the Nucleu Project entitled: “Application of spectroscopic methods for the qualitative control of proteinaceous materials - leather and parchment”, contract PN 09 10 03 01, financed by ANCS, in the period 2009-2011.

REFERENCES

1. Wold, S., *Spline-funktioner-ett nytt verktyg i data-analysen*, Svensk Kemisk Tidskrift, **1972**, 3, p. 34-37.
2. Beebe, K.R., Pell, R.J., Seasholtz, M.B., *Chemometrics – A Practical Guide*, **1998**, John Wiley & Sons Inc., New York.
3. Shlens, J., *A Tutorial on Principal Component*, **2009**, Version 3.01.
4. http://www.resample.com/xlminer/help/PCA/pca_intro.htm (Dated: 3.5.2009).

CONCLUZII

Aplicarea studiilor de modelare statistico-matematică a datelor analitice pot pune în evidență relații „ascunse” între sisteme chimice înrudite sau între etape ale proceselor fizico-chimice pe care le parcurge un sistem dat.

Aplicațiile programului SPSS la investigarea obiectelor de patrimoniu din piele sunt limitate, datorită faptului că nu poate fi procesat un volum mare de date.

Aplicațiile tehniciilor chemometrice cantitative de regresie sunt cel mai frecvent utilizate, dat fiind faptul că demersul chemometric vizează obținerea de modele statistico-matematice utile în clasificare și predicție. Experimentele chemometrice furnizează o multitudine de date brute corelate, a căror interpretare algoritmică este dificilă. De aceea, înaintea aplicării tehniciilor de modelare chemometrică, datele brute sunt supuse unei pre-procesări pentru a îndepărta sau a diminua semnificativ efectele variabilității irelevante în cadrul studiului, de natură aleatoare ori sistematică. În cazul prezentului studiu se vor continua investigațiile pentru găsirea unei formule noi de pre-procesare a datelor.

Mulțumiri

Această lucrare a fost elaborată în cadrul proiectului de colaborare bilaterală România – Slovenia, cu titlul: “Development of spectroscopic methods for quality control of proteinaceous materials”, contract 60CB/2008, finanțat de ANCS, în perioada 2008-2009, și în cadrul proiectului Nucleu cu titlul “Aplicații ale metodelor spectroscopice pentru controlul calitativ al materialelor proteice - piei și pergamente”, contract PN 09 10 03 01, finanțat de ANCS, în perioada 2009-2011.

5. Haaland, D.M., Thomas, E.V., *Partial Least-Squares Methods for Spectral Analyses. 1. Relation to Other Quantitative Calibration Methods and the Extraction of Qualitative Information*, *Anal. Chem.*, 60, **1988**, 1193-1202.
6. Tobias, R.D., *An Introduction to Partial Least Squares Regression*, SAS Institute Inc., Cary, NC.
7. Trafela, T., Strlič, M., Kolar, J., Lichtblau, D.A., Anders, M., Pucko Mencigar, D., Pihlar, B., *Non-destructive analysis and dating of historical paper based on IR spectroscopy and chemometric data evaluation*, *Anal. Chem.*, 79, **2007**, 6319-6323.