FOLIA FORESTALIA POLONICA
Seria B, zeszyt 24, 1993

CHEMOMETRICS IN CHEMOTAXONOMY AND FOREST DECLINE MONITORING

I. CLASSIFICATION OF SOME GENERA OF GYMNOSPEARES ACCORDING TO THE COMPOSITION AND CONTENT OF TERPENES

Peter Jurášek[1], *František Berta*[2], *Štefan Vodný*[1] *and* *Martin Košík*[1]

Slovak Technical University, Faculty of Chemical Technology, Bratislava, Slovakia[1]
Slovak Academy of Science, Arboretum Mlyňany, Slovakia [2]

Dedicated of the memory of Prof. Dr. Martin Košík

Components of essential oils from foliage (leaf and needles) were analyzed by gas chromatography. Determined mono- and sesquiterpenes and terpenes served for the estimation of 4 genera (*Juniperus, Pinus, Chamaecyparis, Abies*) of wood plants. Differences in the chromatographic results related to occurancy of significant compounds are considered as a diagnostic parameter. Formation of separated clusters and relevant terpenes is presented by the principal component analysis. Pattern recognition techniques were applied for classification of the species to 4 genera with a total classification ability of 89%.

INTRODUCTION

The utilization of terpenes analysis for phytogenetic and taxonomic purposes is well-known. Essential oils obtained by hydrodistillation of foliage or turpentine from oleoresin had been first used in the study of pine genus [1]. The most reliable way of analyzing monoterpenes is the gas chromatography – mass spectrometry technique (GC/MS). During the past 20 years, the development of the headspace and capillary column has permitted a rapid and high resolution separation. Quantification is typically achieved by internal standards [6]. The identification is based on mass spectral and GC retention time comparison with model substances. However, model compounds are not always available and, in many cases, structural isomers give almost identical mass spectra. Therefore IR and ^{1}H NMR spectroscopies can be useful to distinguish those isomers and ring compounds [3].

Extensive results of the composition analysis of plant extracts of various species have been published in plenty of papers from the point of view of

chemotaxonomy. In most of them the results are presented in large tables e.g. the comparison of epicuticular waxes of seed coats for Cistaceae [4], GC percentage compositions of the leaf oils of Eucalyptus [5] or in simple graphs e.g. turpentine composition of *Pinus contorta* and *Pinus banksiana* hybrids [6]. The other form of the presentation and comparison of chemical component concentrations is the use of the basic statistical terms (means, standard errors, normality tests, correlations) e.g. in terpenoid chemisystematic study of *Abies grandis* [7], and for determining seasonal changes during needle maturation [8] or the use of overlapped concentration profiles e.g. as in the study of the decomposition and chemical changes in *Pinus silvestris* needle litter [9].

A promising alternative in the taxonomic studies is the chemometric approach in the treating of large data sets by using selected methods of multivariate statistical analysis: exploratory data analysis (EDA) and pattern recognition (PR) [10, 11, 12, 13, 14, 15].

The exploratory data analysis (especially principal component analysis) can be used to show in a graphical form the information contained in the data; to find similarities, anomalies or errors and in many cases to get the idea of the strategy for classification or correlation. This technique was used for confirmation of distinctiveness in the bud exudate composition of a number of poplar species [16], for chemosystematic differentation of *Quercus ilex* and *Quercus rotundifolia* trees based on acorn steroids [17], for differentiating among Malaysian *Hospitalitermes* species [18], for separation of the *Combertum* and *Albizia* species gums from each other and from *Acacia* species gums [19].

Pattern recognition methods can assign an uknown sample to one of the known groups and evaluates relevancy or redundancy of the variables in order to differentiate between groups. A computer database of essential oils has been developed [4] in combination with the pattern recognition method (based on K-nearest neighbor classification rule) for the rapid identification of components in a complex mixture, the comparison between the complete GC profiles of different samples, and for the recognition of a group of compound as a single entity [21].

In this contribution we have introduced the chemometric approach in data treatement. On the basis of terpene composition from mature needles sampled according to conventional rules the formation of separated clusters was presented by the principal component analysis and the classification of several conifers was performed by several pattern recognition methods.

EXPERIMENTAL

The essential oil samples of four genera of wood species were investigated, namely 5 species of *Thuja*, 12 species of *Juniperus*, 17 species of *Pinus* and 9 species of *Chamaecyparis*.

The quantitative content of compounds in the individual species is different and variable, which depends not only on climatic conditions but also on the locality, time and place of sampling and on the increment sectors (one-, two- or three-year old needles). These factors were standardized by taking the samples in the day-time period, at the same height and, as far as it is possible, on the sunny side of the crown. The influence of increments sectors was eliminated by taking the needles from the last increment sector but one. A choice was made of the undamaged and uncut trees of about the same age. The material was sampled during the vegetation in activity at about 9.00 a.m. and immediately caried into the laboratory for its treatment. 100 g of fresh needles were picked and essential oils were obtained by hydrodistillation performed in an apparatus for the detemination of volatile oils. The other part served for the determination of volatile oils. The other part served for the determination of solids. Water residues found in oil were removed with addition of anhydrous potassium sulphate.

The analysis itself was done on a gas chromatograph CHROM 5 using a silicon capillary column (internal diameter 0.2 mm and 25 m lenght) impregnated with Carbowax 20 M under following conditions: temperature gradient 50°C (2 min), increase 4°C/min, final temperature 180°C (14 min); sensitivity 64; dosing 0.5 μl; split ratio 1:250; initial pressure of the carrier gas (N_2) 50 kPa; hydrogen flow 25 ml/min; air flow 500 ml/min; FID detector (flame-ionization).

The identification of each compound was done by comparing the retention times of individual standards and investigated compound as well as by applying internal standarts. For determination of the retention times and the percentual proportion of individual terpenes a CI 100 integrator was used.

An evaluation of the results attained was carried out in two steps. First, the occurence of all compounds established was observed for each species of the individual genera to find out characteristic compounds for particular genus. The the Exploratory data analysis was employed together with the manual processing of measured data for individual genera [22]. In the next step, the data associated only with the most characteristic 16 substances were processed. To discern the genera of wood species the projection of analyzed samples was used by means of the principal components analysis [14]. The classification of a training set was made by pattern recognition methods such as KNN, the linear and quadratic discriminant analysis [11].

RESULTS AND DISCUSSION

Among plenty of volatile compounds contained in the tree bitumen only one group of compounds was chosen which is represented by terpenes constituting an essential part of volatile oils contained, to a large extent, just in softwoods.

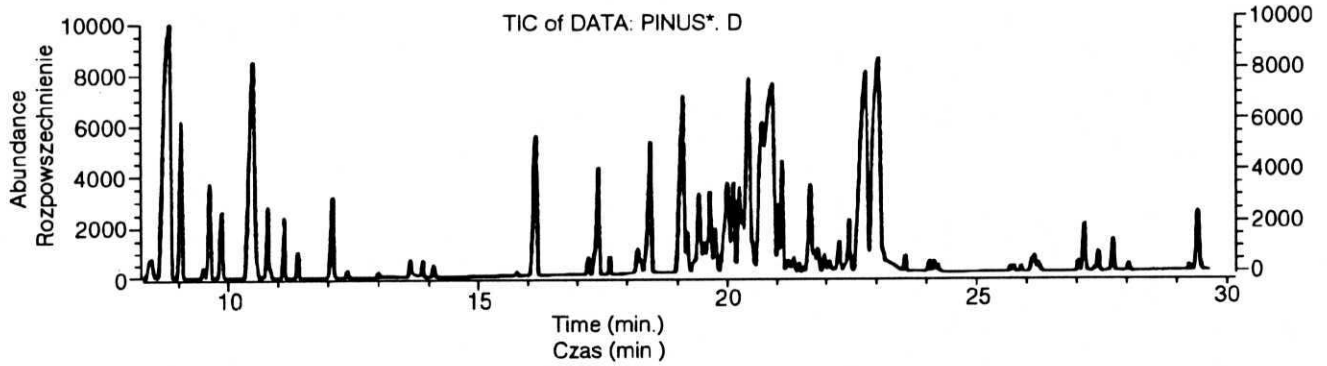

Fig. 1. Gas chromatogram profile of the hydrocarbon fraction of the essential oil of *Pinus* species
Rys. 1. Chromatogram gazowy frakcji węglowodorowej olejków eterycznych rodzaju *Pinus*

Fig. 1 illustrates chromatographic profile of terpene compounds for the genus *Pinus*. The evaluation of several samples of such rich mixtures and the processing of an acquired information are rather demanding. It is simplification to consider only a group of the most important components in mixtures. Some of them are specific for a particular genus or species, as for example β-Pinen for genus *Abies*, p-Cymol for *Chamaecyparis*, Thujon for *Thuja* and other ones. On determining these characteristic compounds the possibility of a simple and immediate demonstration of the measured data by computer contributed a great deal [22]. By analyzing all 43 samples via the Exploratory data analysis (profiles), and also manually, we chose from the total amount of 38 identified terpenes 16, that is those which were the most specific for individual genera (Table 1): α-Pinen (1), Kamfen (2), Sabinen (3), β-Pinen (4), α-Terpinen (5), 3-Karen (6), 1,8-Cineol (7), *p*-Cymol (8), α-Tujon (9), Linalool (10), Pulegon (11), Borneol (12), Dihydrokarveol (13), β-Citronellol (14), Nerol (15) and Farnezol (16).

Table 1

The most characteristic terpenes for 4 genus of conifers
Najbardziej charakterystyczne terpeny dla czterech rodzajów drzew iglastych

Genus – Rodzaj			
Thuja	*Juniperus*	*Pinus*	*Chamaecyparis*
Terpenes – Terpeny			
α-Tujon Farnezol β-Pinen	Sabinen Linalool 1,8-Cineol 3-Karen	α-Pinen β-Pinen Kamfen β-Citronellol Nerol Dihydrokarveol Pulegon	α-Cymol Borneol α-Terpinen β-Pinen Pulegon

Thus the choice of the characteristic features applied later in the recognition (classification) of individual genera of the species was made. The quantitative determination of features was expressed as the percentual proportion of terpenes (also some other modes could be used such as: the ratio of some chromatographic peaks, the number of peaks, the amount of non-identified compounds, the presence of some characteristic terpenes and others). The data were normalized and the data matrix for the next chemometric processing was created.

The principal component analysis of distribution 43 species was used in comparison to genus considered. The data points scattered in the 16-multidimensional space were projected onto a plane which was spanned by the two principal components with the highest amount of information, thus visualizing grouping a similar species. The PCA plot shows evident separation of all 4 genera (Fig. 2).

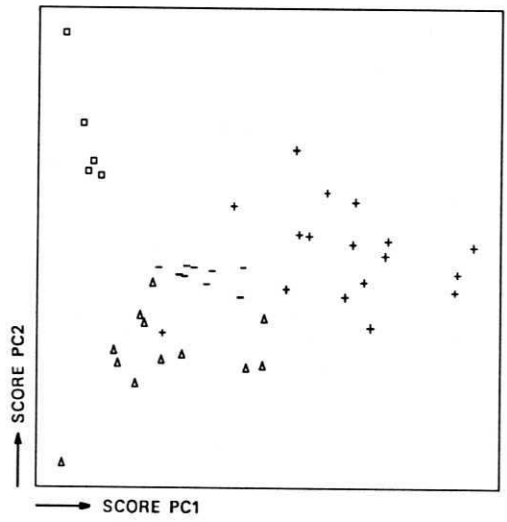

Fig. 2. Principal component analysis of the gas chromatographic data of 43 samples showing the separation of the genus species. Features: relative concentrations of the 16 terpenoid compounds (see Tabl. 1). Axes: the first and third principal component, 16% and 11% of total variance, respectively. Symbols: □ – *Thuja*, △ – *Juniperus*, + – *Pinus*, – – *Chemaecyparis*

Rys. 2. Analizy głównych składników według wyników chromatografii gazowej 43 próbek ukazująca rozróżnienie rodzajów drzew. Cechy: względna zawartość 16 związków terpenowych (patrz Tabl. 1). Osie: pierwszy i trzeci składnik główny, 16% i 11% całkowitej wariacji, odpowiednio. Symbole: □ – *Thuja*, △ – *Juniperus*, + – *Pinus*, – – *Chamaecyparis*

The most relevant features for data structure presented follows from loading-loading plot of the first two principal components (Fig. 3). The position of the feature on plot Fig. 3 reflects the influence of the feature to position of the objects on Fig. 2. This relationship is useful to find out which spectral features are responsible for a separation of classes of species. The most distinctive terpenes were: α-Tujon (9), Farnezol (16), Kamfen (2), α-Pinen (1), Nerol (15), β-Citronellol (14), 3-Karen (6), Sabinen (3), Linalool (10), 1,8-Cineol (7), etc. The final results considering those obtained by the manual processing are shown in Table. 1.

As it is evident from Fig. 2, 16 selected terpene compounds contain the sufficient amount of information for creating the model for individual genus, which enables one to recognize the particular species by classification methods. During classification the data matrix created a training set of data. In view of the smaller number of investigated samples of the genera of *Thuja* and *Chamaecyparis* a reliability of the applied technique was tested on the training set (the recognition ability). The 3 pattern recognition methods were used: KNN (K-nearest neighbours), LDArep (repeating linear discriminant analysis:

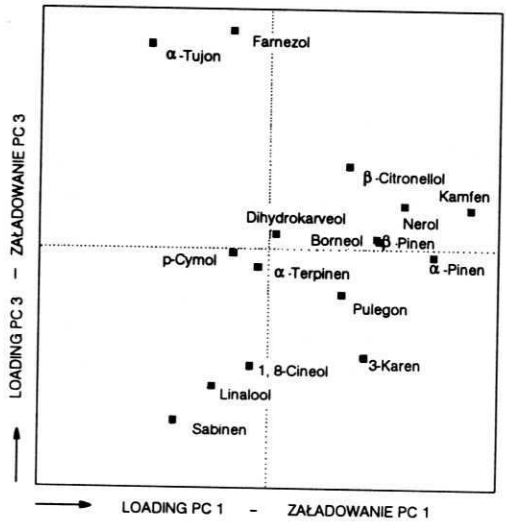

Fig. 3. Loading-loading plot of the first and third principal components of the genus species (16 selected terpenes, Tabl. 1)
Rys. 3. Diagram rozproszenia głównych składników (pierwszego i trzeciego) dla czterech rodzajów drzew (16 wybranych terpenów, Tabl. 1)

samples are randomly divided into several number of groups – each group is successively used as unknown samples) and BAYES (quadratic discriminant analysis based on Bayesian analysis and disjoint multivariate normal distribution for each category – first 4 principal components were used for calculation due to the required data dimension reduction). The average success of methods was following: the KNN method – 84%, the LDArep method – 86% and the BAYES method – 97%. The results for individual genera and methods are indicated in Table 2. The less favourable results were achieved for *Juniperus* and *Pinus*.

Table 2
Classification ability of 3 pattern recognition methods (KNN, LDArep, BAYES) obtained for training set (together 43 species)
Ocena przydatności 3 wzorcowych metod do klasyfikacji surowca drzewnego (KNN, LDArep, BAYES), na podstawie analizy zbioru 43 próbek

Genus Rodzaj	Number of samples Liczba próbek	Classification ability of PR methods Wyniki oceny		
		KNN	LDArep	BAYES
Thuja	5	5	5	5
Juniperus	12	10	9	12
Pinus	17	14	14	16
Chamaecyparis	9	7	9	9
Total – Ogółem	43	36 (84%)	37 (86%)	42 (97%)

The results obtained in the sphere of the study of volatile compounds under investigation contributed to a closer knowledge of differences and the composition of ethereal oils in softwood species and have shown an approach in comparing the differences among individual genera and wood species as well as in their classification. The employment of the computer data processing (mainly by the library searching, target analysis and by the multivariate statistical analysis methods) enables the examined species to be relatively quickly characterized. It can be stated that the more profound analysis of the terpene compounds content by chemometric approach may serve, similarly as various taxonomic data in the IDENT system [23, 24], in the diagnostics of wood genera. The next work will lead to the recognition among species as well as to the determination of the damage degree of wood species.

Received in May 1993

REFERENCES

1. Mirov N.T.: Chemical analysis of the oleoresins as a means of distinguishing Jeffrey pine and western yellow pine. J. Forest. 1929, 27 176-187.
2. Corkill J.A.: Analysis of Conifer Needles by Headspace Cappilary Gas Chromatography. J. of High Res. Chromat. & Chromat. Commun. 1988, Vol. 11, 211-214.
3. Veijanen A., Kolehmainen E., Kauppinen R., Lahtiperä M., Paasivirta J.: Methods for the identification of tainting terpenoids and other compounds from Algae. Wat. Sci. Tech. 1992, Vol. 25, No. 2, 165-170.
4. Krollmann P., Eich C., Gülz P.G.: Epicuticular Waxes of Seed Coats from Species of the Genus *Cistus* L. (*Cistaceae*). Z. Naturforsch 1984, 39c, 521-524.
5. Lassa E.V., Southwell I.A.: The steam volatile leaf oils of some species of Eucalyptus subseries Strictinae. Phytochemistry 1982, Vol. 21, No. 9, 2257-2261.
6. Zavarin E., Critchfield W.B., Snajberk K.: Turpentine Composition of *Pinus contorta* and *Pinus bansiana* hybrids and hybrid derivatives. Can. J. of Botany 1969, 47, 1443-1453.
7. Zavarin E., Snajberk K., Critchfield W.B.: Terpenoid Chemosystematic Studies of Abies grandis. Biochem. Systemat. Ecology 1977, Vol. 8, 81-93.
8. Wagner M.R., Clancy K.M., Tinus R.W.: Maturational variation in needle essential oils from *Pseudotsuga menziesii, Abies concolor* and *Picea engelmanii*. Phytochemistry 1989, Vol. 28, No. 3, 765-770.
9. Berg B., Hannus K., Popoff T., Theander O.: Changes in organic chemical components of needle during decomposition. Long-term decomposition in a Scots pine forest. I Can. J. of Botany 1982, Vol. 60, No. 8, 1310-1319.
10. Brereton R.G.: Chemometrics, Application of mathematics and statistics to laboratory system, 1990, Ellis Horwood, New York and London.
11. Forina M., Leardi R., Armanino C., Lanteri S., Conti P., Princi S.: PARVUS: An extendable packadge of programs for data exploration, classification and correlation. 1988, Elsevier Science Publishers, Amsterdam.
12. Massart D.L., Vandeginste B.G.M., Deming S.N., Michotte Y., Kaufman L.: Chemometrics: a textbook. Data Handling in Science and Technology, 2. 1988, Elsevier Science Publishers, Amsterdam.
13. Varmuza K., Lohninger H.: Principal component analysis of chemical data, in Zupan J. (ed.): PCs for chemist, 1990, p. 43, Elsevier, Amsterdam.
14. Varmuza K.: EXCERPT: Chemometric software for exploratory data analysis of spectra, Technical University Vienna, 1993.

15. Wold S.: Multivariate data analysis: Converting chemical data tables to plots, in Brant J., Ugi I.K. (eds.): Computer application in chemical research and education, 1989, p. 101, Hüthig-Verlag, Heidelberg.
16. Greenway W., English S., May J., Whatley F.R.: Chemotaxonomy of section Leuce poplars by GC/MS of bud exudate. Biochem. Syst. Ecol. 1991, Vol. 19, No. 6, 507-518.
17. Rafii Z.A., Zavarin E., Pelleau Y.: Chemisystematic differentiation of *Quercus ilex* and *Quercus rotundifolia* based on acorn steroids. Biochem. Syst. Ecol. 1991, Vol. 19, No. 3, 249-252.
18. Chuah C.H., Goh S.H., Blunt J.W.: Intra- and interspecific variations in the defence secretions of the Malaysian termite Hospitalitermes (*Isoptera: Nasutitermitinae*). Biochem. System. Ecol., Vol. 19, No. 1, 35-46.
19. Jurášek P., Košík M., Phillips G.O.: A chemometric study of the *Acacia* (gum arabic) and related natural gums. Food Hydrocolloids 1993, Vol. 7, No. 1, 73-85.
20. Chien M.J.: A computer database of essential oils. A world perspective, Proceedings of the 10th International Congress of Essential Oils, Fragrance and Flavors, Washington, DC, U.S.A., 16-20 November 1986, Elsevier Science Publishers 1988, Amsterdam.
21. Chien M.: Analysis of Complex Mixtures by Gas Chromatography/Mass Sectrometry Using a Pattern Recognition Method. Anal. Chem. 1985, Vol. 57, 348-352.
22. Statistical Graphics System, Statgraphics ver. 2.6, STSC, 1987.
23. Miller R.B.: Wood Identification via computer. IAWA Bulletin n.s. 1980, Vol. 1, No. 2, 154-160.
24. Miller R.B., Pearson R.G., Wheeler E.A.: Creation of a large database with IAWA standard list characters. IAWA Bulletin n.s. 1987, Vol. 8, No. 3, 219-232.

CHEMOMETRIA W CHEMOTAKSONOMII I MONITOROWANIU ZAMIERANIA LASÓW

I. KLASYFIKACJA NIEKTÓRYCH RODZAJÓW NAGONASIENNYCH WEDŁUG SKŁADU I ZAWARTOŚCI TERPENÓW

Streszczenie

Składniki olejków eterycznych listowia (liście i igły) poddano analizie za pomocą chromatografii gazowej. Oznaczone mono-, sequiterpeny i terpeny posłużyły do szacunku czterech rodzajów roślin drzewiastych (*Juniperus* sp., *Pinus* sp., *Chemaecyparis* sp. i *Abies* sp.). Różnice w uzyskanych wynikach badań chromatograficznych związane z występowaniem znaczących związków uznano za parametr diagnostyczny. Powstawanie oddzielnych grup oraz odnośnych terpenów przedstawiono na podstawie analizy związków podstawowych. W badaniach zastosowano technikę rozpoznawania wzorców do oznaczania gatunków dla czterech rodzajów, z całkowitą zdolnością oznaczania rzędu 89%.

Authors addresses:
Ing. Peter Jurášek, CSc.,
Ing. Štefan Vodný, CSc.,
Doc. Ing. Martin Košík, DrSc.
Katedra dreva, celulózy a papiera
Chemickotechnologická fakulta
Slovenská Technická Univerzita
Radlinského
812 37 Bratislava – Slovakia
Ing. František Berta
Ústav dendrobiológie „Arborétum Mlyňany"
CBEV SAV
951 52 Vieska nad Žitavou – Slovakia