Proc. Estonian Acad. Sci. Chem., 1993, 42, 4, 172—178
https://doi.org/10.3176/chem.1993.4.02

ESTIMATION OF STEROLS CONTENT AND SOME OTHER PROPERTIES OF TALL OIL USING NIR SPECTROSCOPY AND MULTIVARIATE CALIBRATION

Arkadi EBBER

Eesti Teaduste Akadeemia Keemia Instituut (Institute of Chemistry, Estonian Academy of Sciences), Akadeemia tee 15, EE-0026 Tallinn, Eesti (Estonia)

Presented by K. Lääts

Received April 19, 1993; accepted June 11, 1993

Abstract. The sterols content in 16 tall oil samples was determined by gas chromatography and near-infrared (NIR) spectra of all the samples were measured. Applying methods of multivariate calibration (PCR, PLS), mathematical models for estimating the sterols content and some other properties of tall oil from NIR spectra were developed.

Key words: tall oil, analysis of sterols, NIR spectra, multivariate calibration.

Tall oil is a major by-product of the kraft or sulfate pulping process [1]. Crude tall oil (CTO) is formed as a product of acidulation of tall oil soap. It consists of three main groups of substances: fatty acids, rosin acids, and neutral materials (Fig. 1). The latter contain hydrocarbons, sterols and high molecular weight alcohols. Most of CTO is converted through separation by distillation to two valuable products, the fatty acid and the rosin acid fraction. The neutral substances, including sterols, remain in the undistilled residue, tall oil pitch (TOP). Depending on how exhaustively the distillation has been carried out, pitch may contain 2—5% free sterols (Fig. 2), but the bulk of the free sterols is esterified during distillation. After the pitch has been hydrolyzed (Fig. 3), the peaks of free acids and sterols appear again in the chromatogram and the total content of sterols may be estimated (usually 10—15%). The sterols recovered from TOP can be used in the manufacture of pharmaceutical products. So, TOP deserves attention as a source of phytosterols. But the potential of sterols in CTO should be evaluated before distillation because the profitability of any method to be designed for the isolation of sterols depends on the amount of sterols in the initial raw material.

The purpose of this article is to show how near-infrared (NIR) spectroscopy combined with the multivariate calibration approach can be applied to estimating the sterols content and some other properties of tall oil.

Experimental. 16 CTO samples from various kraft mills (12 in Sweden, 3 in Finland and 1 in Poland) were taken as a calibration set and in all of them the sterols content was measured. The following analytical procedure was carried out: a 2 g sample of CTO with approximately 2% of an added internal standard, cholesterol (Merck), was dissolved in 20 ml of hot ethanol and partitioned between two immiscible liquid layers of water (60 ml) and dichloroethane (20 ml). Extraction from the water phase with 20 ml dichloroethane was repeated twice. After separation the joint organic phase was washed with 20 ml of water and dichloroethane evaporated under reduced pressure. The residue was dissolved in ethyl acetate

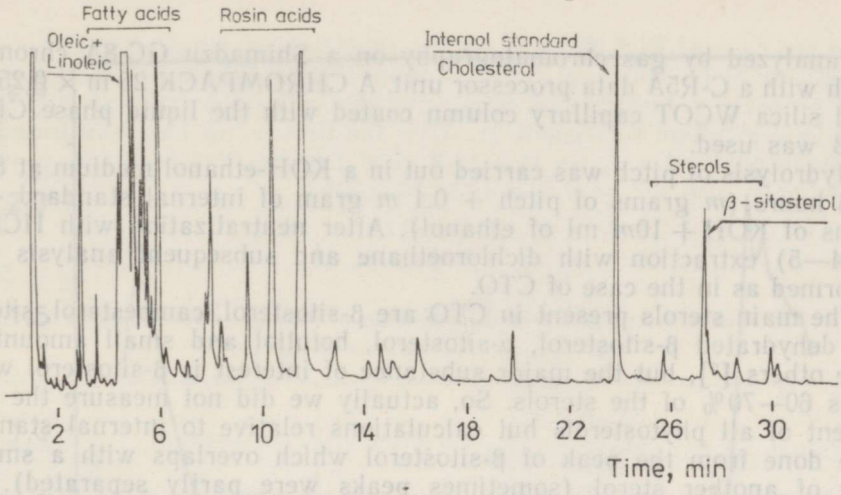

Fig. 1. Chromatogram of CTO (dichloroethane extract) sample from Karlsborg Kraft Mill. Column temperature 195 °C for 9 min, then raised to 260 °C at 4° C /min.

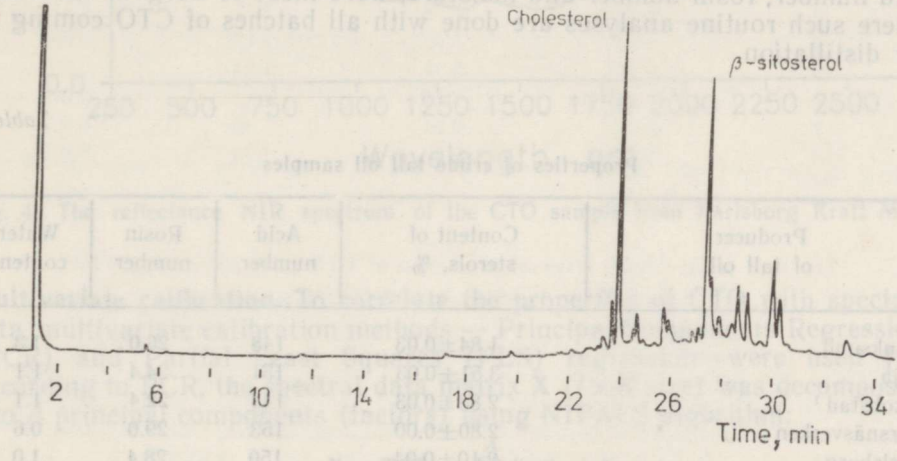

Fig. 2. Chromatogram of TOP (1.7% of free sterols). The acids have been distilled out. Prior to chromatographic analysis the pitch was dissolved in hot ethyl acetate and the solution permeated the column filled with aluminium oxide.

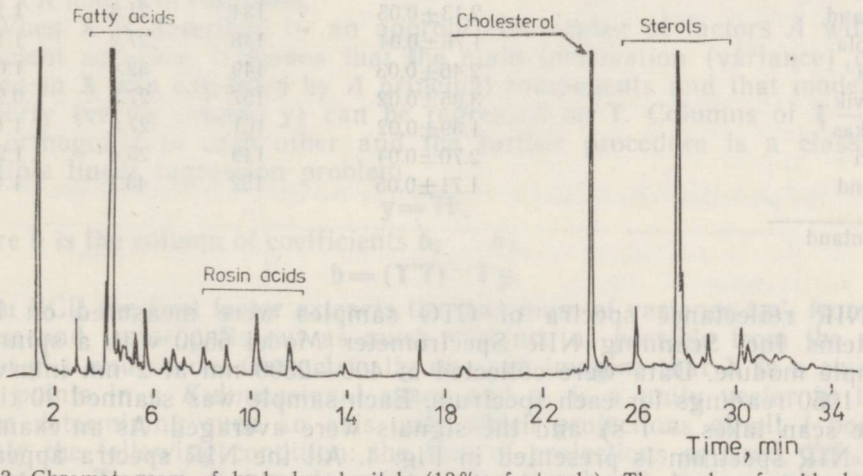

Fig. 3. Chromatogram of hydrolyzed pitch (13% of sterols). The peaks of acids bound to esters have appeared again.

and analyzed by gas chromatography on a Shimadzu GC-8A chromatograph with a C-R5A data processor unit. A CHROMPACK 25 m × 0.25 mm fused silica WCOT capillary column coated with the liquid phase CP-Sil 5 CB was used.

Hydrolysis of pitch was carried out in a KOH-ethanol medium at 80 °C for 3 hours (*m* grams of pitch + 0.1 *m* gram of internal standard + *m* grams of KOH + 10*m* ml of ethanol). After neutralization with HCl (to pH 4—5) extraction with dichloroethane and subsequent analysis were performed as in the case of CTO.

The main sterols present in CTO are β-sitosterol, campesterol, sitostanol, dehydrated β-sitosterol, α-sitosterol, betuliol and small amounts of some others [2], but the major substance of interest is β-sitosterol which forms 60—70% of the sterols. So, actually we did not measure the total content of all phytosterols but calculations relative to internal standard were done from the peak of β-sitosterol which overlaps with a smaller peak of another sterol (sometimes peaks were partly separated). The sensitivity coefficient 1.1 for the cholesterol—β-sitosterol pair was used. The results are presented in Table 1. The other properties involved, viz. acid number, rosin number and moisture, were those of Bergvik Kemi AB, where such routine analyses are done with all batches of CTO coming in for distillation.

Table 1

Properties of crude tall oil samples

Producer of tall oil	Content of sterols, %	Acid number	Rosin number	Water content
Munksund	1.84±0.03	148	26.0	1.3
Väja	3.51±0.03	151	24.4	1.1
Jakobstad [1]	2.88±0.03	143	22.4	1.1
Korsnäsverken	2.80±0.00	153	29.0	0.6
Karlsborg	2.10±0.04	150	28.4	1.0
Skutskär	3.00±0.05	148	25.4	1.0
Norrsundet	2.06±0.02	156	34.5	1.2
Lövholmen	2.65±0.02	155	30.3	0.8
Skoghall	2.87±0.03	159	28.9	1.3
Östrand	2.13±0.05	134	21.4	1.1
Obbola	1.76±0.04	148	27.2	1.1
Kemi [1]	2.46±0.03	149	32.4	1.0
Vallvik	3.05±0.02	157	27.3	0.9
Kaukas [1]	1.39±0.02	161	27.8	1.1
Frövi	2.70±0.04	149	25.4	1.2
Poland	1.71±0.05	152	43.1	1.4

[1] Finland

NIR reflectance spectra of CTO samples were measured on NIR Systems Inc. Scanning NIR Spectrometer Model 6500 with a spinning sample module. Data were collected at 400—2500 nm at 2 nm intervals, i. e. 1050 readings for each spectrum. Each sample was scanned 20 times (one scan takes ~1 s) and the signals were averaged. As an example, one NIR spectrum is presented in Fig. 4. All the NIR spectra appeared to be rather similar but sophisticated chemometrical methods enable to extract subtle changes in spectra.

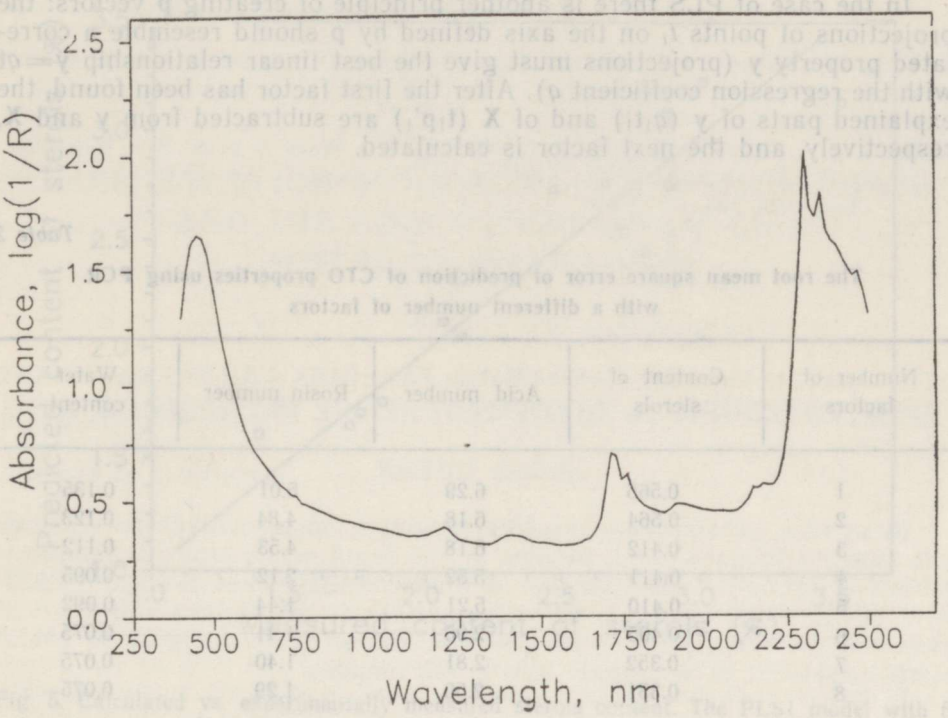

Fig. 4. The reflectance NIR spectrum of the CTO sample from Karlsborg Kraft Mill.

Multivariate calibration. To correlate the properties of CTO with spectral data, multivariate calibration methods — Principal Components Regression (PCR) and Partial Least Squares (PLS) regression—were used [3]. According to PCR, the spectral data matrix **X** ($I \times K$ size) was decomposed into A principal components (factors) using NIPALS algorithm:

$$\mathbf{X} = \mathbf{TP}' = \sum_{a=1}^{A} \mathbf{t}_a \mathbf{p}'_a + \mathbf{E},$$

where **T** is the matrix of factor scores and **P** is the loadings matrix; **t** and **p** are column vectors of **T** and **P**, respectively (**p**' is the transpose of **p**); **E** is the $I \times K$ matrix of residuals.

When **X** is described by an appropriate number of factors A with a sufficient accuracy, it means that the main information (variance) contained in **X** was extracted by A principal components and that modelled property (vector column **y**) can be regressed on **T**. Columns of **T** — (**t**) are orthogonal to each other and the further procedure is a classical multiple linear regression problem

$$\mathbf{y} = \mathbf{Tb},$$

where **b** is the column of coefficients $b_1 \ldots b_A$,

$$\mathbf{b} = (\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'\mathbf{y}.$$

In PCR the first factor extracts the maximum of variance $\mathbf{t}_1\mathbf{p}'_1$ from **X**, the second factor pulls out as much variance as possible from the rest ($\mathbf{X} - \mathbf{t}_1\mathbf{p}'_1$) and so on. Geometrically one can imagine that **X** is a swarm of I points in a K-dimensional space and $\mathbf{p}_1$ is a unity vector in that space determining such an axis onto which projections of all I points satisfy the following condition: the sum of projections squared, Σt^2_{ia}, is maximum. The set of **p** vectors (an orthonormal set) is a new basis in which data are represented.

In the case of PLS there is another principle of creating **p** vectors: the projections of points t_i on the axis defined by **p** should resemble a correlated property **y** (projections must give the best linear relationship $\mathbf{y}=q\mathbf{t}$ with the regression coefficient q). After the first factor has been found, the explained parts of **y** ($q_1\mathbf{t}_1$) and of **X** ($\mathbf{t}_1\mathbf{p}'_1$) are subtracted from **y** and **X**, respectively, and the next factor is calculated.

Table 2

The root mean square error of prediction of CTO properties using PCR with a different number of factors

Number of factors	Content of sterols	Acid number	Rosin number	Water content
1	0.566	6.29	5.01	0.135
2	0.564	6.18	4.84	0.123
3	0.412	6.18	4.53	0.112
4	0.411	5.52	2.12	0.095
5	0.410	5.21	1.44	0.092
6	0.405	2.95	1.41	0.075
7	0.352	2.81	1.40	0.075
8	0.351	2.50	1.29	0.075

Table 3

The root mean square error of prediction of CTO properties using PLS1 with a different number of factors

Number of factors	Content of sterols	Acid number	Rosin number	Water content
1	0.562	6.26	4.82	0.133
2	0.449	6.11	4.57	0.119
3	0.408	5.08	2.95	0.101
4	0.301	4.60	1.86	0.091
5	0.244	2.58	1.22	0.072
6	0.195	1.84	0.59	0.066
7	0.154	0.96	0.47	0.050
8	0.129	0.80	0.39	0.031

Table 4

The prediction error of models evaluated by the cross validation procedure

Modelled property	PCR		PLS	
	Prediction error	Number of factors	Prediction error	Number of factors
Content of sterols	0.48	3	0.42	8
Acid number	5.2	6	3.3	8
Rosin number	3.7	4	1.9	8
Water content	0.19	6	0.15	6

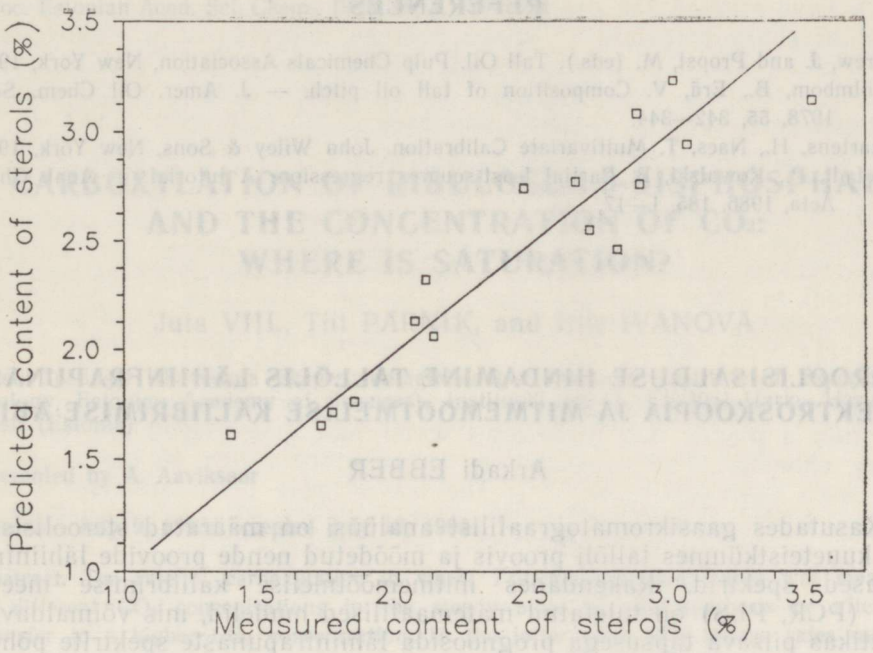

Fig. 5. Calculated vs. experimentally measured sterols content. The PLS1 model with 6 factors was used.

For the calculations, **y** and **X** data were mean-centered (from each column the mean value of the column was deducted) and the principal component regression and PLS1 with orthogonal **t**-vectors in **T** (proposed by S. Wold) were performed. For all the four properties of CTO, models with 1...8 factors were calculated (Tables 2 and 3). As could be expected, the PLS method is better than PCR and models with approximately 6 factors satisfactorily describe the properties of CTO which were used for model building (Fig. 5).

In fact, the ability of models to predict properties of new samples not used in calibration is not so good because of over-fitting or because our set of 16 samples does not properly represent the actual population. The cross validation procedure, in which objects were left out one by one, showed that in some cases the well defined minimum of PRESS (prediction residual sum of squares) [4] does not exist and that the accuracy of prediction is not very high. Table 4 represents the best results of all runs. All the programs used for calculations were written in MATLAB 386 (The Math Works, Inc.).

So, the conclusion can be made that several properties of tall oil can be quickly estimated with an industrially sufficient accuracy from the reflectance NIR spectrum using PLS regression.

ACKNOWLEDGEMENTS

The author wishes to thank the Swedish Institute for granting a scholarship and M. Karlsson, R. Olsson, J. Eriksson and Prof. K. Sjöberg for encouragement and fruitful discussions.

REFERENCES

1. Drew, J. and Propst, M. (eds.). Tall Oil. Pulp Chemicals Association, New York, 1981.
2. Holmbom, B., Erä, V. Composition of tall oil pitch. — J. Amer. Oil Chem. Soc., 1978, 55, 342—344.
3. Martens, H., Naes, T. Multivariate Calibration. John Wiley & Sons, New York, 1991.
4. Geladi, P., Kowalski, B. Partial least-squares regression: A tutorial. — Anal. Chim. Acta, 1986, 185, 1—17.

STEROOLISISALDUSE HINDAMINE TALLÕLIS LÄHIINFRAPUNASE SPEKTROSKOOPIA JA MITMEMÕÕTMELISE KALIIBRIMISE ABIL

Arkadi EBBER

Kasutades gaasikromatograafilist analüüsi on määratud steroolisisaldus kuueteistkümnes tallõli proovis ja mõõdetud nende proovide lähiinfrapunased spektrid. Rakendades mitmemõõtmelise kaliibrimise meetodeid (PCR, PLS) on tuletatud matemaatilised mudelid, mis võimaldavad praktikas piisava täpsusega prognoosida lähiinfrapunaste spektrite põhjal steroolisisaldust ja mõningaid teisi tallõli omadusi.

ОЦЕНКА СОДЕРЖАНИЯ СТЕРОЛОВ В ТАЛЛОВОМ МАСЛЕ С ПОМОЩЬЮ БЛИЖНЕЙ ИК-СПЕКТРОСКОПИИ И МНОГОМЕРНОГО КАЛИБРОВАНИЯ

Аркадий ЭББЕР

Хроматографически определено содержание стеролов в 16 образцах таллового масла и измерены ближние ИК-спектры образцов. Методами многомерного калибрования (PCR, PLS) получены математические модели, позволяющие на основе ближних ИК-спектров прогнозировать содержание стеролов и другие свойства таллового масла с практически приемлемой точностью.