

УДК 543.5

Яак АРРО\*, Эда ГРЮНЕР\*, Михкель КОЭЛЬ\*

## ВОЗМОЖНОСТИ ЯМР<sup>1</sup>H-СПЕКТРОМЕТРИИ В СОЧЕТАНИИ С ФАКТОРНЫМ АНАЛИЗОМ ПРИ ИЗУЧЕНИИ СОСТАВА ТРУДНОРАЗДЕЛИМЫХ СМЕСЕЙ

Относительная простота получения ЯМР<sup>1</sup>H-спектров и их высокая информативность сделали этот метод незаменимым при выяснении структур органических соединений и их идентификации. Этот вид спектроскопии нашел широкое применение и при изучении состава сложных смесей, таких как нефти, сланцевые смолы и др.

Однако расшифровка ЯМР<sup>1</sup>H-спектров смесей — задача не простая, так как здесь принципиально не применимы многие методы, которыми пользуются при истолковании спектров индивидуальных соединений. Поэтому для расшифровки спектров смесей разработан целый ряд методов (см., напр., [1–6]). Согласно им, спектры смеси рассматривают как сумму спектров разных видов водородов, спектральные области которых не накладываются друг на друга или это наложение может быть элиминировано эмпирическими методами. Такие методы позволяют, с определенной долей субъективности, выяснить структурно-групповой состав смеси, но не спектры индивидуальных или групповых ее составляющих. Последнюю задачу можно решить только при помощи методов, которые рассматривают спектр смесей как аддитивную сумму спектров ее составляющих. В этом направлении наиболее перспективными следует считать версии факторного анализа, которые предложены в [7, 8]. Тем более что попытка применения факторного анализа спектральной информации при изучении состава труднолетучих смол уже дала обнадеживающие результаты [9]. Данная работа является продолжением этих работ применительно к ЯМР<sup>1</sup>H-спектроскопии.

### Теоретические основы

Если спектр смеси является аддитивным, тогда для вектора спектра смеси справедливо следующее выражение

$$\vec{d}_n = \vec{r}_{n1}c_1 + \vec{r}_{n2}c_2 + \dots + \vec{r}_{nl}c_l = \sum_{i=1}^l \vec{r}_{ni}c_i, \quad (1)$$

где  $\vec{r}_{ni}$  — вектор-столбец спектра  $i$ -го соединения длиной  $n$  экспериментальных точек;  $c_i$  — концентрация  $i$ -го соединения в смеси;  $l$  — число компонентов смеси.

Как правило, спектры всех  $r_i$  можно определить лишь после полного разделения смеси. В случае высококипящих (высокомолекулярных) соединений эта трудоемкая работа часто даже не поддается осуществлению. Значительно легче получить  $m \gg l$  различных по составу смесей (фракций) этих соединений, для которых верно

\* Eesfi Teaduste Akadeemia Keemia Instituut (Институт химии Академии наук Эстонии). 200108 Tallinn, Akadeemia tee 15, Estonia.



$$\begin{aligned}
 \vec{d}_{n1} &= r_{n1}c_{11} + r_{n2}c_{21} + \dots + r_{nl}c_{l1}, \\
 \vec{d}_{n2} &= r_{n1}c_{12} + r_{n2}c_{22} + \dots + r_{nl}c_{l2}, \\
 &\vdots \\
 \vec{d}_{nm} &= r_{n1}c_{1m} + r_{n2}c_{2m} + \dots + r_{nl}c_{lm},
 \end{aligned}
 \tag{2}$$

где концентрации  $c_{ij}$  образуют векторы-строки, и систему уравнений (2) можно записать в виде матрицы

$$\mathbf{D}_{nm} = r_{n1}c_{1m} + r_{n2}c_{2m} + \dots + r_{nl}c_{lm}
 \tag{3}$$

или

$$\mathbf{D}_{nm} = \mathbf{R}_{nl}\mathbf{C}_{lm}.
 \tag{4}$$

Уравнение (4) является основной формулой факторного анализа. Другими словами, если матрица  $\mathbf{D}$  задана, тогда появляется возможность определить (выделить) спектры компонентов смеси (матрица  $\mathbf{R}$ ) и концентрации последних (матрица  $\mathbf{C}$ ), а также число компонентов в смеси ( $l$ ) [7, 8, 10].

В случае сложных смесей можно ожидать, что одному выделенному спектру (статистически надежному) будет соответствовать множество соединений, во-первых, потому, что при фракционировании смеси некоторые ее компоненты образуют псевдоочищенное вещество (к примеру, азеотропные смеси), и, во-вторых, потому, что некоторые компоненты смеси имеют практически совпадающие спектры или их области.

В математическом смысле это значит, что в уравнении (3) возможны следующие замены:

$$\vec{r}_{ni} = \sum_{j=1}^n g_{nj}c_j.
 \tag{5}$$

Легко убедиться, что существует множество величин  $\vec{g}_{nj}$  и  $c_j$ , при которых уравнение (3) остается математически строгим (независимым от величин  $c_{lm}$ ). Это означает, что выделенный спектр является в своем роде групповым спектром, а  $l$  соответствует числу групповых компонентов в смеси.

Ясно, что групповые спектры тем надежнее (статистически), чем больше они отличаются друг от друга и чем меньше они различаются для внутрigrupповых соединений. Таким образом через факторный анализ спектров фракций сложной смеси можно выявить в ней группы соединений со своим конкретным спектром. Для проверки этой версии нами изучались ЯМР<sup>1</sup>H-спектры модельных смесей, компоненты которых выбраны с целью имитировать состав труднолетучей смолы полукоксования сланца-кукерсита.

исходные  
образы — 1, 2

### Экспериментальная часть

Из совокупности 30 соединений (исходная смесь, табл. 1), каждое по 15 весовых единиц  $q = (50 \pm 2,5 \text{ мг})$ , составили 15 случайных смесей. ЯМР<sup>1</sup>H-спектры растворов одинаковой концентрации снимали на приборе ВС 487 (80 МГц) фирмы «Tesla» (ЧСФР) в четыреххлористом углероде. Для подготовки математической обработки спектров измеряли интенсивность сигнала в области от 0,7 до 8,65 м. д. через интервал 0,05 м. д. ( $n=160$ ). Вычисления проводили на ЭВМ ЕС 1052 при помощи пакета программ, которые использовались уже в [11]. К ним относятся програм-



ма выделения спектров (ВС), включающая как составную часть определение уникальных точек [8], и абстрактный факторный анализ (метод главных компонент). В первую очередь при помощи абстрактного факторного анализа определяли для исходной матрицы данных ( $n$  значений интенсивностей и  $m$  фракций) число существенных абстрактных факторов  $l$ . Для определения  $l$  использовали индикаторную функцию, предложенную в [7]. Потом при помощи ВС получали уже  $l$  выделенных спектров.

Таблица 1

Состав модельных смесей

Соединение	Доля водорода в исходной смеси, %	Средне-квadraticное отклонение $q$	Соединение	Доля водорода в исходной смеси, %	Средне-квadraticное отклонение $q$
Тридекан	4,82	1,07	4-Метилциклогексанон	3,70	0,76
Декалин	4,25	1,13	Камфора	3,50	1,00
Дициклопентадиен	3,13	0,76	Пропилфенилкетон (А)	2,66	0,93
Изоамилбензол	3,58	0,85	1,3-Дифенил-2-пропанон (Б)	2,25	1,69
1-Метилнафталин	2,34	0,93	<i>п</i> -Оксибензофенон (В)	1,67	0,85
Аценафтен	2,01	0,85	Ксантон (Г)	1,36	1,41
1-Децен	4,70	0,93	4-Пропилфенол	2,87	0,93
<i>цис</i> -4-Нонен	4,73	0,93	2,6-ди- <i>трет</i> -Бутил-1,4-крезол	3,60	1,07
Циклогексен	3,97	0,93	4-Гексилрезорцин	3,05	1,00
Пинен	3,86	1,07	2-Аллилфенол	2,33	1,13
Лимонен	3,96	0,76	5-Оксииндан	2,56	0,93
2-Октанон	4,05	0,76	Диизоамиловый эфир	4,51	0,85
5-Нонанон	3,87	1,06	Бутилфениловый эфир	3,85	0,70
5-Гексен-2-он	3,32	0,85	Анетол	2,83	1,13
Циклопентанон	3,15	1,13	Фарнезиллацетат	3,52	0,76

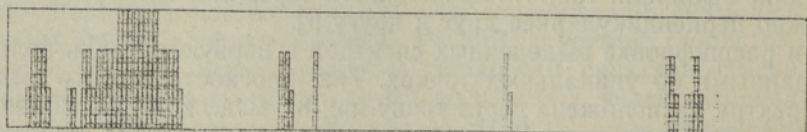
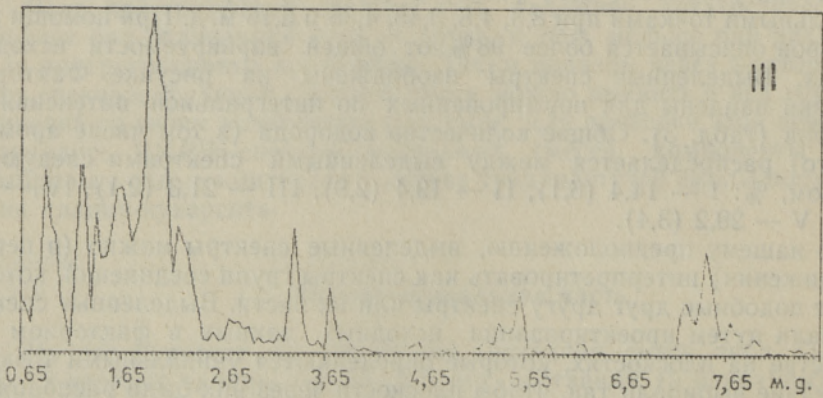
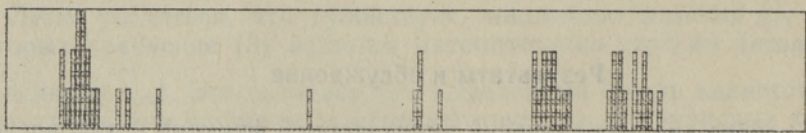
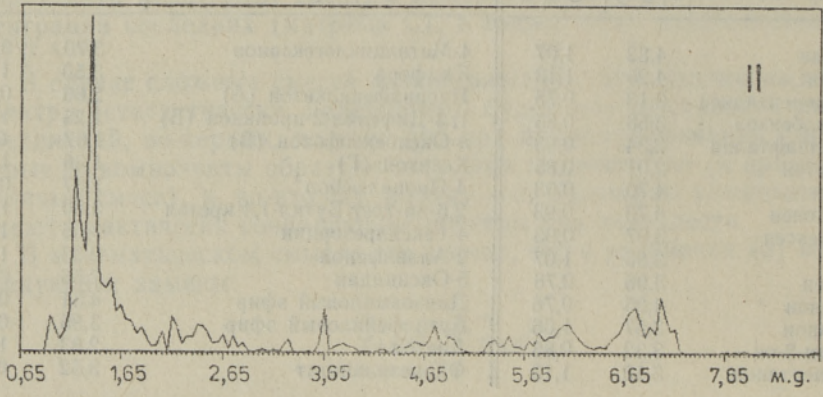
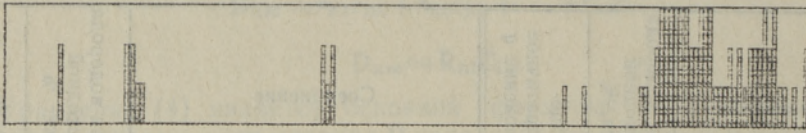
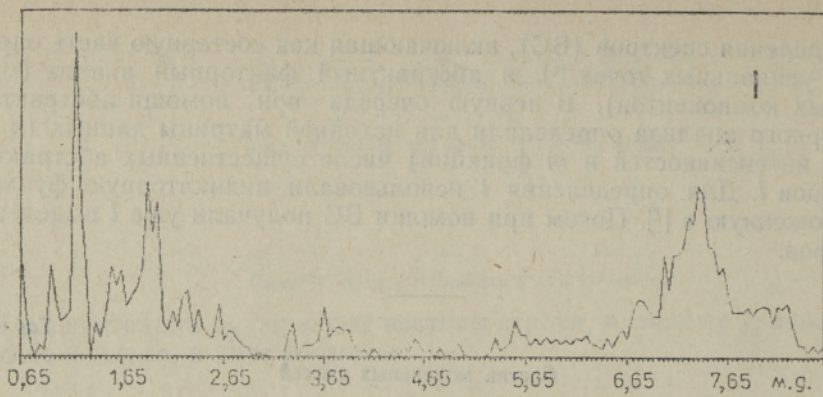
### Результаты и обсуждение

На основе анализа значений индикаторной функции установлено, что достаточно учитывать пять факторов (табл. 2). При этом пользовались уникальными точками при 8,3, 4,8, 3,45, 4,35 и 5,15 м. д. При помощи этих факторов описывается более 98% от общей варьированности исходных данных. Выделенные спектры изображены на рисунке. Факторные нагрузки найдены для нормированных по интегральной интенсивности спектров (табл. 3). Общее количество водорода (в том числе ароматического) распределяется между выделенными спектрами следующим образом, %: I — 14,4 (6,1), II — 19,4 (2,9), III — 21,3 (2,1), IV — 15,7 (1,4), V — 29,2 (3,4).

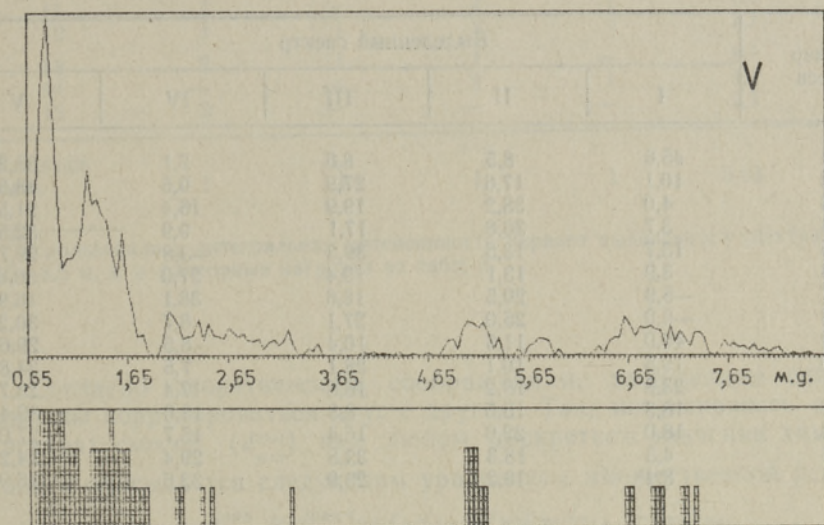
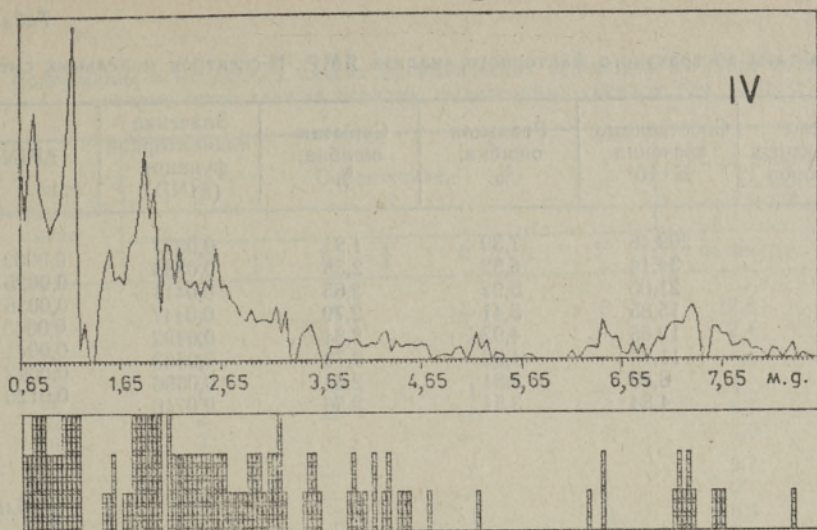
По нашему предположению, выделенные спектры можно (в первом приближении) интерпретировать как спектры групп соединений, которые имеют подобные друг другу спектры или их части. Выделенные спектры получали путем проектирования исходных данных в факторном пространстве на плоскостях, которые определяются уникальными точками. Последние выбирали так, чтобы плоскости через них были расположены возможно перпендикулярнее друг к другу [8].

При расшифровке выделенных спектров в первую очередь пользовались данными об уникальных точках. Так, плоскость первого выделенного спектра расположена через точку при 8,3 м. д., которую можно приписать спектру ксантона. Действительно, спектр ксантона существенно









Выделенные ЯМР  $^1\text{H}$ -спектры на базе модельных смесей, соответствующие пяти наиболее значимым факторам. Внизу показаны уровни значимости частной корреляции (95, 99 и 99,5%) для каждой дискретной точки спектра.

коррелируется (в области химических сдвигов 7,0—8,5 м.д.) с первым выделенным спектром ( $r=0,780$ ), для которого спектр ксантона можно считать базисным. Спектры других соединений учитываются постольку, поскольку их спектры подобны спектру базисного соединения. Так, первым выделенным спектром учитываются ксантон, *n*-оксибензофенон, 1,3-дифенил-2-пропанон и пропиленфенилкетон. Суммарное содержание ароматических водородов в этих соединениях (согласно составу модельных смесей) существенно коррелируется ( $r=0,915$ ) с факторными нагрузками первого выделенного спектра (табл. 4). Спектры этих соединений (приняты один к одному) составляют 55% от интегральной интенсивности первого выделенного спектра и на их счет приходится 7,99% от суммарного водорода в исходной смеси. В выделенных спектрах наиболее интенсивной является та часть спектра, которая совпадает со спектром базисного соединения и характеризует данную группу соединений.



Результаты абстрактного факторного анализа ЯМР  $^1\text{H}$ -спектров модельных смесей

Число абстрактных факторов	Собственные значения $E \cdot 10^4$	Реальная ошибка, %	Скрытая ошибка, %	Значение индикаторной функции (FIND)	$\Delta\text{FIND}$
1	209,46	7,39	1,91	0,0377	0,0009
2	34,14	6,52	2,38	0,0386	0,0025
3	21,05	5,92	2,65	0,0411	0,0036
4	15,85	5,41	2,79	0,0447	0,0045
5	12,65	4,92	2,84	0,0492	0,0041
6	11,93	4,32	2,73	0,0533	0,0063
7	8,23	3,81	2,61	0,0596	0,0120
8	4,84	3,51	2,56	0,0716	

Таблица 3

## Факторные нагрузки выделенных спектров, %

Номер смеси	Выделенный спектр				
	I	II	III	IV	V
1	45,6	8,5	8,6	2,1	35,2
2	10,1	17,6	27,9	0,6	43,8
3	4,0	38,2	19,9	16,4	21,5
4	3,7	36,8	17,1	9,9	32,5
5	13,7	13,5	39,3	-3,8	29,7
6	3,9	13,1	19,4	27,0	36,6
7	-5,9	20,5	18,6	38,1	16,9
8	-9,0	25,0	27,1	8,7	30,2
9	40,0	11,5	10,4	8,5	29,6
10	19,5	10,1	28,1	7,5	34,8
11	23,8	18,2	16,8	17,4	23,7
12	18,3	15,5	8,8	18,0	39,4
13	18,0	22,9	16,4	15,7	27,0
14	4,3	18,3	33,8	29,4	14,2
15	8,4	18,2	23,9	32,6	16,9

Аналогично можно показать, что спектры алкилфенолов (2,6-ди-*трет*-бутил-1,4-крезола, 4-гексилрезорцина и 4-пропилфенола) свойственны второму выделенному спектру. Количество водорода в этих соединениях пропорционально факторным нагрузкам второго выделенного спектра ( $r=0,854$ ) и охватывает 49% от интегральной интенсивности этого спектра.

Установлено, что третьему выделенному спектру принадлежат сигналы циклических (4-метилциклогексанона и циклопектанона), четвертому — сигналы алифатических кетонов (2-октанона, 5-нонанона, а также тридекана) и пятому — сигналы опять же циклических соединений (декалина, камфоры, пинена и др.).

В выделенные спектры могут попасть и «чужие» сигналы, которые не принадлежат чистым спектрам соединений данной группы. Эти примеры явно подтверждают вышеизложенные теоретические рассуждения, а также подводят к выводу, что не все части в выделенных спектрах одинаково надежны (достоверны), тем более что доля шума в каждом следующем выделенном спектре нарастает.



Содержание в модельных смесях ароматических водородов в группе соединений, которые принадлежат первому выделенному спектру (см. табл. 1)

Номер модельной смеси	Содержание, $q$				Содержание аром. водородов в А+В+Б+Г, %	
	А	Б	В	Г	истинное	рассч. *
1	—	1	—	5	17,8	19,4
2	1	—	1	—	2,4	4,3
3	1	—	—	—	1,1	1,7
4	1	1	1	—	4,1	1,6
5	2	—	2	1	6,7	5,8
6	2	1	—	1	5,0	1,7
7	—	—	2	—	2,7	0
8	—	—	—	—	0	0
9	1	5	1	3	15,4	17,0
10	—	5	1	2	12,3	8,3
11	1	—	2	1	5,9	10,1
12	1	1	2	1	7,7	7,8
13	3	1	1	—	6,1	7,7
14	—	—	2	1	4,4	1,8
15	2	—	—	—	2,1	3,6

Исходная смесь                    1                    1                    1                    1                    5,58                    6,06

\* Использовались интегральная интенсивность первого выделенного спектра в области 6,5—8,5 м. д. и факторные нагрузки из табл. 3.

Согласно теоретическим соображениям, выделенные спектры не должны коррелироваться друг с другом. Так, интенсивность сигнала в спектрах смесей ( $d_{nm}^{\text{экс}}$ ) при любом дискретном значении химического сдвига описывается следующим уравнением множественной регрессии:

$$d_{nm}^{\text{экс}} \approx d_{nm}^{\text{расч}} = r_{n1}c_{1m} + r_{n2}c_{2m} + \dots + r_{nl}c_{lm}, \quad (6)$$

где  $r_{nl}$  — интенсивность выделенного спектра  $l$  при химическом сдвиге  $n$ ;  $c_{lm}$  — факторная нагрузка спектра  $l$  для смеси  $m$ .

Если выделенные спектры не коррелируются между собой, тогда для каждой конкретной спектральной области при каждом химическом сдвиге величина  $d_{nl}^{\text{экс}}$  является линейной функцией только одного аргумента  $c_l$ . В общем случае величина  $d_{nm}^{\text{экс}}$  может коррелироваться одновременно с несколькими аргументами, и для каждого аргумента можно определить только коэффициенты частной корреляции. Так, можно охарактеризовать достоверность каждого дискретного значения в выделенных спектрах при любом значении химического сдвига (см. рисунок). Как и следовало ожидать, выделенный спектр наиболее достоверен в областях, которые являются общими для спектров данной группы соединений, и менее достоверен в других выделенных спектрах. Таким образом, знание достоверной вероятности для всей спектральной области выделенного спектра позволяет элиминировать шум в выделенном спектре и тем самым упростить его расшифровку. Так, в первом выделенном спектре выделяется область ароматических водородов, а во втором — водородов в алифатических заместителях ароматического ядра.



Отсюда следует, что выделенные (групповые) спектры можно соотнести только для тех соединений, которые по своим спектрам наиболее близки друг к другу и в то же время явно отличаются от других групп соединений. Спектры остальных соединений как бы распределяются между выделенными спектрами, обеспечивая наиболее точную репродуцируемость (согласно методу наименьших квадратов) спектров как исходной смеси, так и ее фракций. В принципе можно получить также такие выделенные спектры, среди которых один учитывает какое-то конкретное соединение, а другие — остальные соединения. Например, при добавлении к матрице исходных данных спектра ксантона (16-м), получается выделенный спектр (первый), который в области 6,5—8,65 м. д. практически совпадает со спектром ксантона ( $r=0,977$ ). Факторные нагрузки этого выделенного спектра позволяют с приемлемой погрешностью определить содержание водорода ксантона как в исходной смеси, так и в ее фракциях (табл. 5).

Таблица 5

Содержание водорода ксантона в модельных смесях, %

Номер модельной смеси	Истинное	Рассч.
Исходная	1,4	3,7
1	7,3	12,6
2	0	1,3
3	0	1,2
4	0	1,4
5	1,5	4,8
6	1,3	0,1
7	0	0
8	0	0
9	4,4	9,8
10	2,7	7,1
11	1,5	6,1
12	1,6	3,5
13	0	5,3
14	1,4	1,7
15	0	0

Из вышеизложенного следует, что выделенный спектр можно рассматривать как спектр смеси родственных соединений конкретного состава. Это значит, что по степени спектрального подобия различных смесей можно судить о степени подобия молекулярных составов этих смесей. Действительно, сопоставление выделенных спектров данной работы и полученных для реальной кукурситовой смолы [9] показывает их явное сходство. Так, первый выделенный спектр в обоих случаях учитывает водород в замещенных ароматических ядрах, а второй — в алифатических ядрах. Только в случае реальной смолы водород принадлежит чисто алифатическим соединениям, а в случае модельных смесей — алкилароматическим.

Таким образом, факторный анализ как набор статистических методов позволяет при расшифровке ЯМР<sup>1</sup>H-спектров сложных смесей отойти от грубого структурно-группового уровня их изучения и приблизиться к молекулярному уровню. Это, несомненно, расширяет возможности ЯМР<sup>1</sup>H-спектрометрии при исследовании составов различных трудноразделимых смесей, в том числе труднолетучих смол и тяжелых нефтяных остатков. При этом все математические процедуры со спектрами можно максимально автоматизировать, включая введение спектров в ЭВМ.



## ЛИТЕРАТУРА

1. *Brown, J. K., Landner, W. R.* A study of the hydrogen distribution in coal-like materials by high-resolution nuclear magnetic resonance spectroscopy. II. A comparison with infrared measurement and the conversion to carbon structure. — *Fuel*, 1960, **39**, N 1, 87—96.
2. *Chamberlain, N. F.* The Practice of NMR Spectroscopy with Spectra-structure Correlation for Hydrogen-1. Plenum Press, New York—London, 1974.
3. *Bodzek, D., Krzyzanowski, T., Marzec, A.* Heterocompounds present in asphaltenes from various products of coal hydrogenation. — *Fuel*, 1979, **58**, N 3, 196—202.
4. *Netzel, D., Miknis, F.* N. m. r. study of US Eastern and Western shale oils produced by pyrolysis and hydrolysis. — *Fuel*, 1982, **61**, N 11, 1101—1109.
5. *Shu-An Quan, Cheng-Feng Li, Peng-Zhou Shaing.* Study of structural parameters on some petroleum aromatic fractions by H n. m. r./i. r. and C H n. m. r. spectroscopy. — *Fuel*, 1984, **63**, N 2, 268—283.
6. *Огородников В. Д.* ЯМР-спектроскопия как метод исследования химического состава нефтей. — В кн.: Инструментальные методы исследования нефтей. Новосибирск, 1987, 49—67.
7. *Malinowski, E. R., Howery, D. G.* Factor Analysis in Chemistry. Wiley, New-York—Chichester, 1980, 50—99.
8. *Malinowski, E. R.* Obtaining the key of typical vectors by factor analysis and subsequent isolation of component spectra. — *Anal. Chim. Acta*, 1982, **134**, 129—137.
9. *Арро Я. В., Грюнер Э. Г.* Определение характеристических групп соединений труднотлетучих смол при помощи масс-спектрометрии и ЯМР <sup>1</sup>H-спектров. — *Хим. тв. топ.*, 1988, № 2, 53—59.
10. *Туров Ю. П.* Факторный анализ и его применение при исследовании смесей органических соединений. — В кн.: Инструментальные методы исследования нефти. Новосибирск, 1987, 67—82.
11. *Арро Я. В., Грюнер Э. Г., Халдна Ю. Л., Юга Р. Я., Метсала А. А.* Состав нейтральной тяжелой смолы полукоксования сланца-кукерсита. 3. Определение ключевых фракций при помощи факторного анализа. — *Горючие сланцы*, 1986, 3, № 4, 426—431.

Представил Ю. Канн

Поступила в редакцию  
26/IV 1990

*Jaak ARRO, Eda GRÜNER, Mihkel KOEL*

### <sup>1</sup>H TMR-SPEKTROMEETRIA JA FAKTORANALÜÜSI KOOSKASUTUSE VÕIMALUSED RASKESTILAHUTATAVATE SEGUDE KOOSTISE UURIMISEL

30 ühendist (alkaanid, alkeenid, fenoolid, alifaatsed ja aromaatsed ketoonid jt.) koosneva mudelsegu abil on näidatud, et selle segu fraktsioonide <sup>1</sup>H TMR-spektritist saab faktoranalüüsi abil ekstraheerida spektrid, mis vastavad kindla koostisega ühendite gruppidele. See võimaldab keerukate segude <sup>1</sup>H TMR-spektrite interpreteerimisel üle minna struktuurigruppide tasandilt sisuliselt täpsemale molekulaarsele tasandile.

*Jaak ARRO, Eda GRÜNER and Mihkel KOEL*

### ENHANCING POSSIBILITIES OF <sup>1</sup>H NMR SPECTROSCOPY BY FACTOR ANALYSIS FOR THE IDENTIFICATION OF COMPLEX MIXTURES

Fifteen artificial complex mixtures were made choosing randomly from among 30 substances of different classes to model shale oil fractions. Factor analysis of <sup>1</sup>H NMR spectrums of the mixtures gave five factor spectrums describing 98% of the variance in this data set. It was shown that, on the basis of known substances, these factor spectrums can be identified in greater detail and compared with spectrums from real complex mixtures. This might serve as a base for identifying the characteristic substances in real mixtures hard to separate by other methods.