Analysis of complex mixtures using self-modeling decomposition of different spectral data

Yu.B. Monakhova, S.P. Mushtakova, S.S. Kolesnikova

Department of Chemistry, Saratov State University, Astrakhanskaya ul. 83, Saratov, 410026 Russia

E-mail: yul-monakhova@mail.ru; mushtakovasp@info.sgu.ru

Advantages of spectroscopic methods:

- application to different type of spectra (IR, UV, EPR, scattering)
- non-destructive measurements
- an opportunity to separate measurements and data processing
- straightforward sample preparation and measurements
- the simplicity and low cost of equipment
- reliable quantitative analysis of known systems/analytes
- sensitivity to low or even trace concentrations

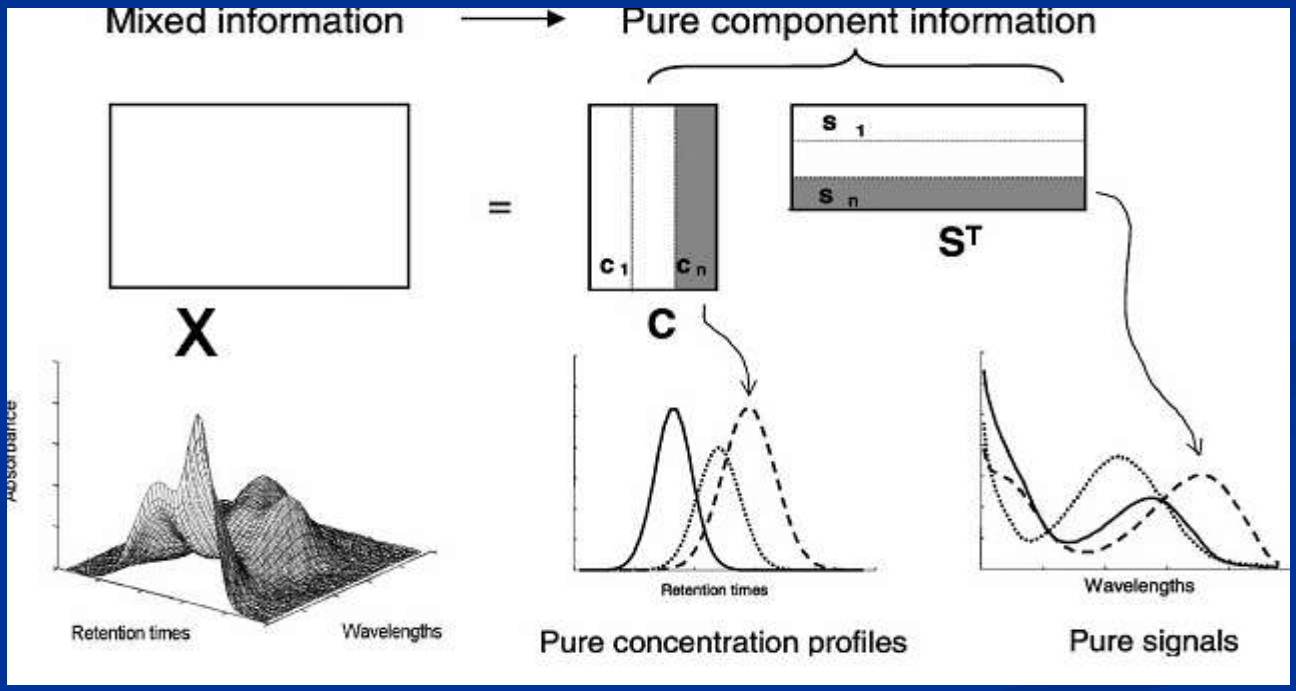
X = C S

X - a M×N matrix of M measured mixture spectra

S - a K×N matrix of K unknown spectra of pure components

C - a M×K mixture matrix (unknown concentrations)

N – the number of counts over wavelength



Advantages of ICA algorithms:

- standardless and self-modelling;
- can be applied to different spectral data (UV-VIS, IR, Raman, EPR and other);
- quickness;
- sufficient accuracy

Assumptions:

- linearity of the mixture;
- stationarity of individual signals;
- statistical independence

MILCA - **Mutual Information Least Dependent Component Analysis**

SNICA - **Stochastic Non-Negative Independent Component Analysis**

Mutual Information for random variable (X_1, X_2,… X_m):

$$I(X_1, X_2, ..., X_M) = \sum_{i=1}^{M} H(X_i) - H(X_1, X_2, ..., X_M),$$

where

$$H(X_i) = -\int \mu_i \ln \mu_i dx_i$$

- differential entropies

$$H(X_1, X_2, ..., X_M) = \int \mu \ln \mu \, dx_1 dx_2 ... dx_M$$

and

μi(xi), μ(x1, x2, ..., xM) – marginal and joint densities

Amari index:

$$P_{err} = \frac{1}{2N}\sum_{i,j=1}^{N}\left(\frac{|p_{ij}|}{\max_k |p_{ik}|} + \frac{|p_{ij}|}{\max_k |p_{kj}|}\right) - 1$$

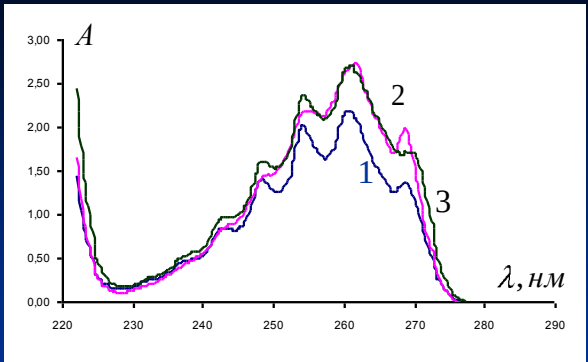
where $p_{ij} = (\hat{A}^{-1}A)_{ij}$

P < 0.05- good decomposition quality

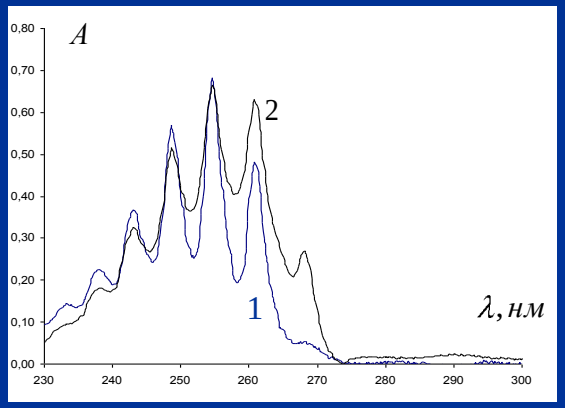
P > 0.2 - unacceptably poor performance

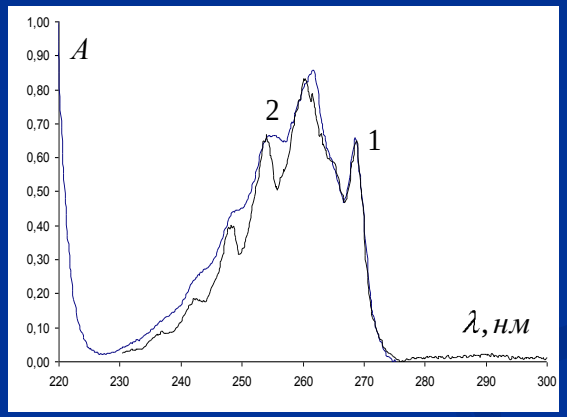
Algorithms taken for comparison:

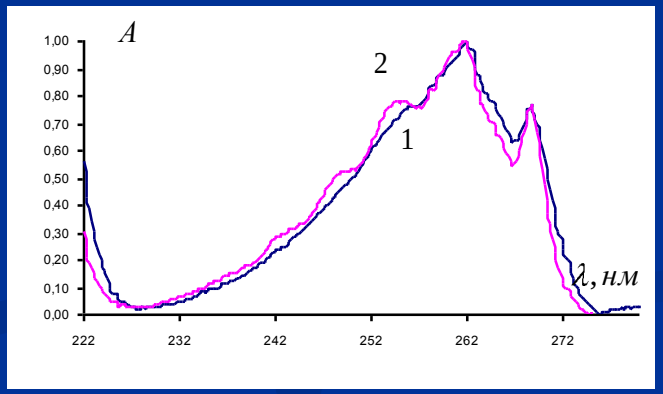
MCR-ALS – multivariate curve resolution-alternating least squares;
SIMPLISMA – simple-to-use interactive self-modeling mixture analysis



Experimental spectra of the benzene-toluene-o-xylene system:
1 -C(C6H6) = 5.0.10-2 M, C(C6H5CH3)= 5.0.10-2 M, C(C6H5(CH3)2)= 5.0.10-2 M;
2 - C(C6H6) = 5.0.10-2 M, C(C6H5CH3)= 10.10-2 M, C(C6H5(CH3)2)= 1.3.10-2 M;
3 - C(C6H6) = 10.10-2 M, C(C6H5CH3)= 2.5.10-2 M, C(C6H5(CH3)2)= 5.0.10-2 M
(l=0.1 cm)







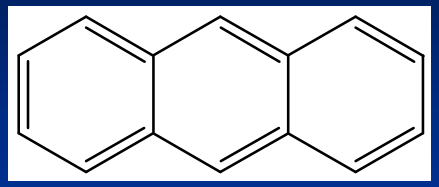
Benzene (R=0.95)

Toluene (R=0.98)

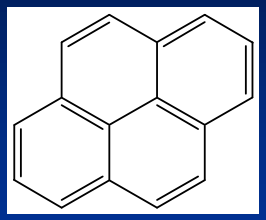
O-xylene (R=0.91)

Recovered (1) spectra of benzene, o-xylene and toluene in comparison with experimental spectra (2)

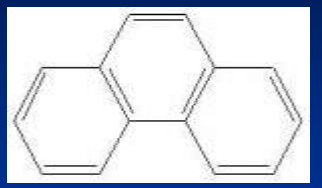
Polyaromatic hydrocarborns (PAHs)



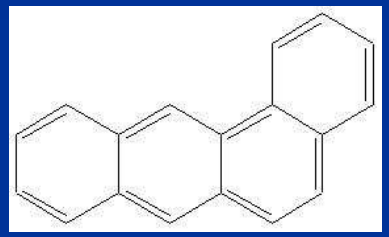
anthracene (1)



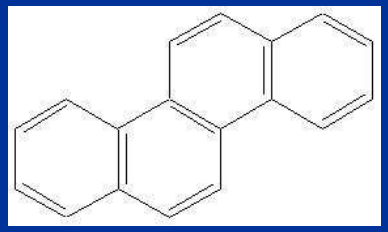
pyrene (2)



phenantrene (3)

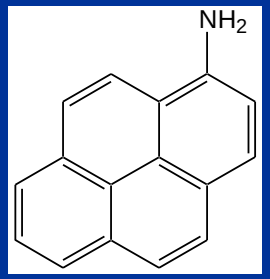


benz[a]antracene (4)

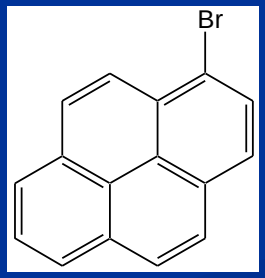


benz[a]phenantrene (5)

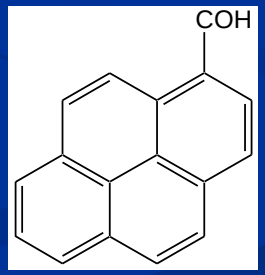
fluorantene (6)



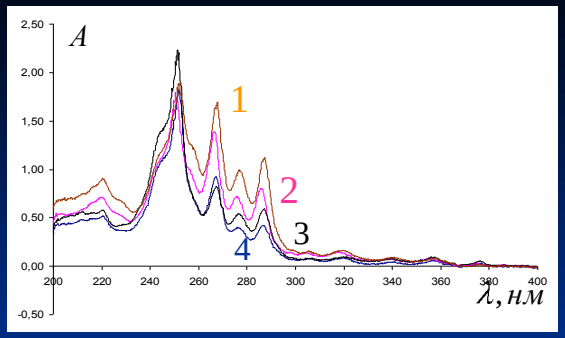
1-aminopyrene (7)



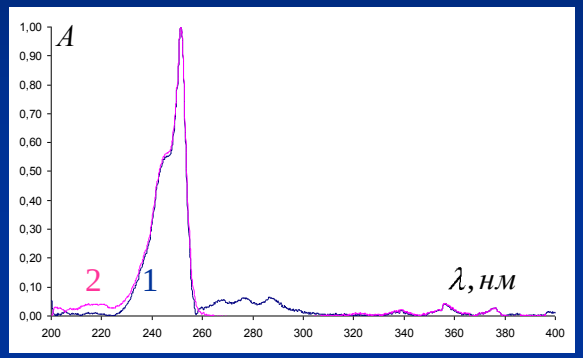
1-bromopyrene (8)



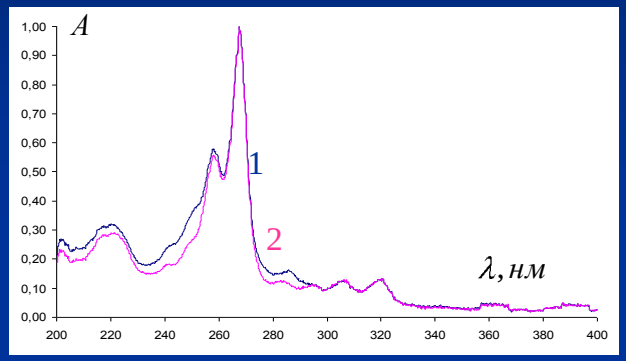
1-pyrenecarboxyaldehyde (9)



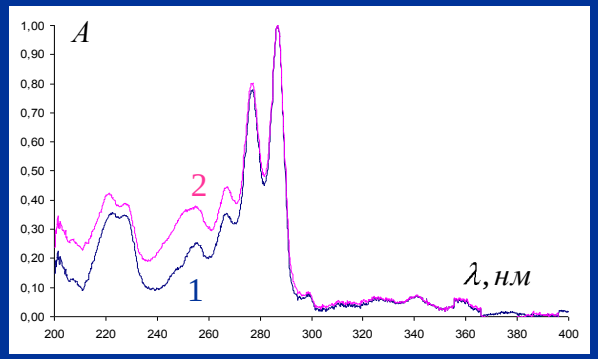
Experimental spectra of the four-component system 1 – 3 – 4 – 5:
1-C(1)=3.0.10-6 M, C(3)=6.0.10-6 M, C(4)=1.0.10-5 M, C(5)=1.0.10-5 M;
2-C(1)=4.0.10-6 M, C(3)=4.0.10-6 M, C(4)=7.0.10-6 M, C(5) =7.0.10-6 M;
3-C(1)=5.0.10-6M, C(3)=5.0.10-6M; C(4)=5.0.10-6M, C(5)=5.0.10-6 M;
4-C(1)=6.0.10-6M, C(3)=6.0.10-6M; C(4)=5.0.10-6M, C(5)=3.0.10-6 M;



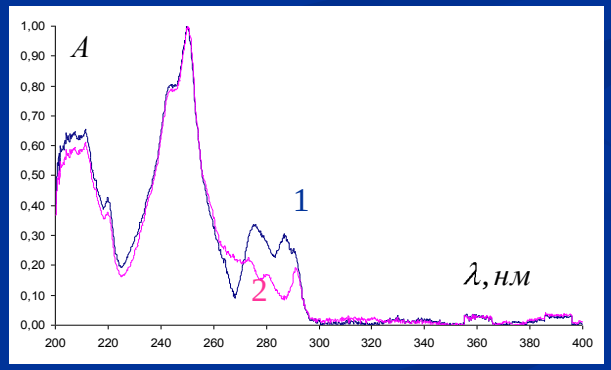
anthracene (R=0.95)



benz[a]phenantrene (R=0.99)



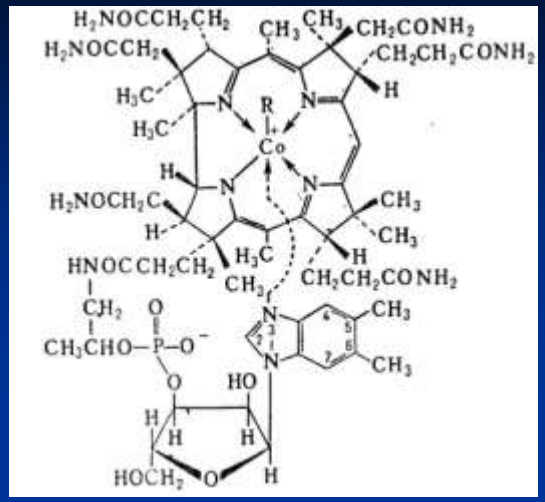
benz[a]antracene (R=0.96)



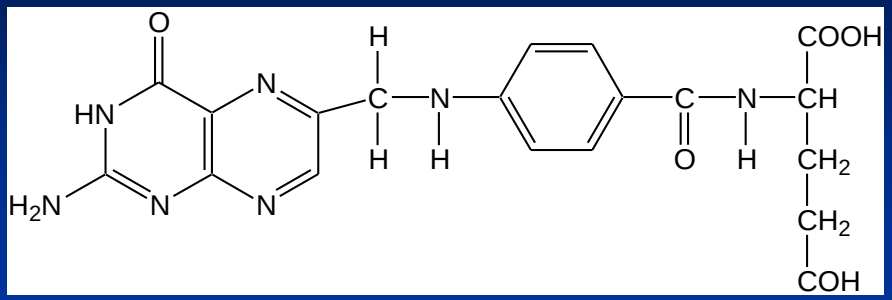
phenantrene (R=0.98)

Experimental (1) spectra of compounds in comparison with recovered spectra (2) (method MILCA)

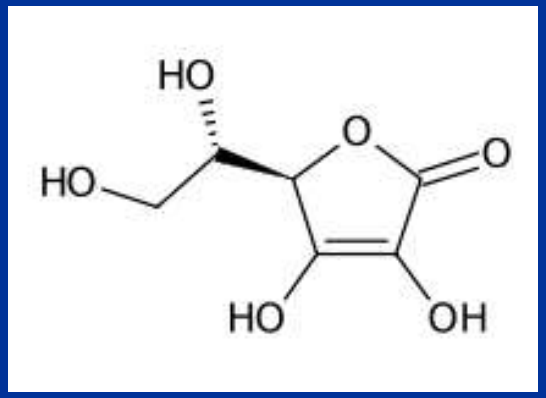
VITAMINS



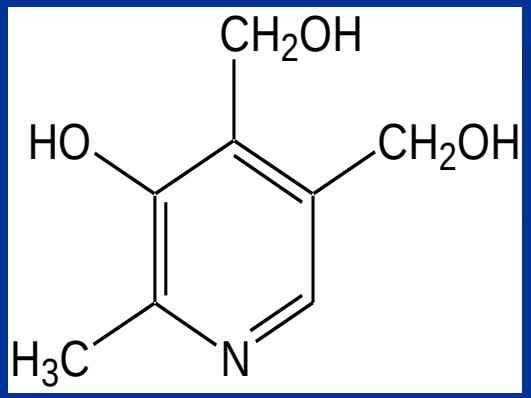
B12



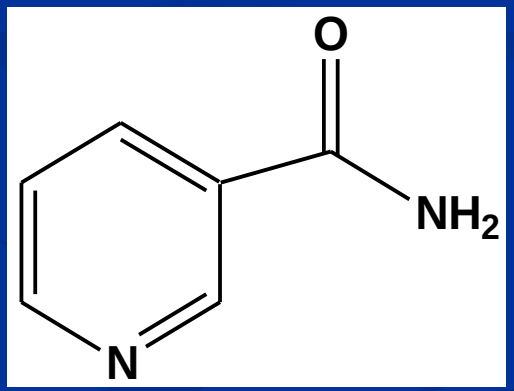
B9 (folic acid)



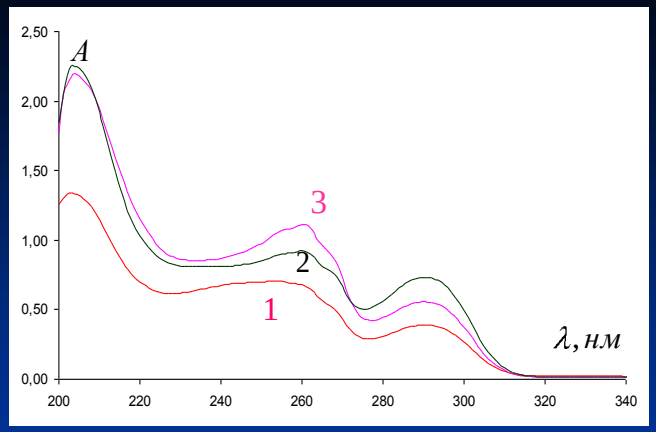
C



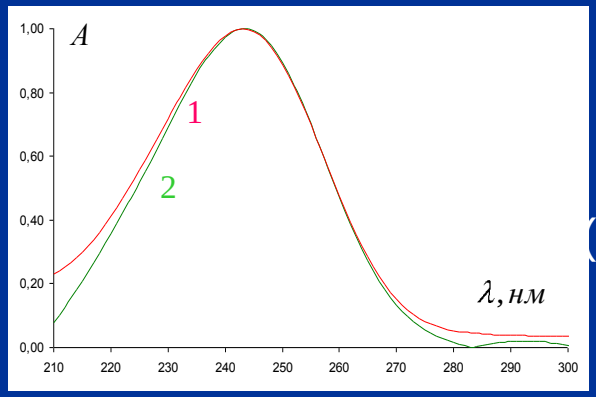
B6



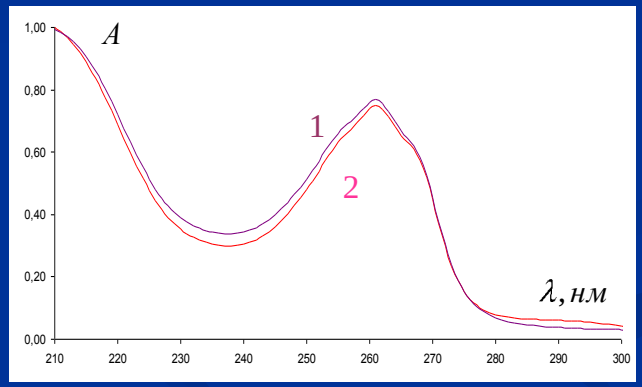
PP



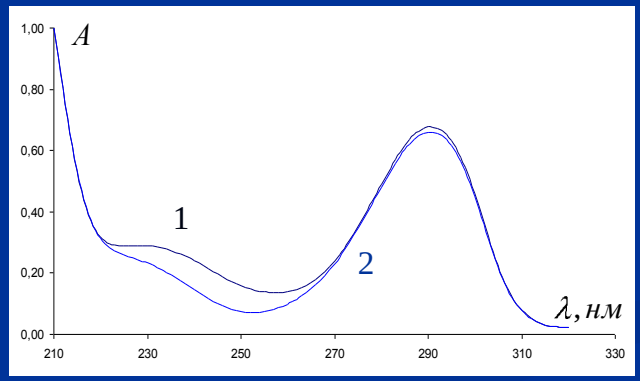
Experimental absorption spectra of the ternary system of vitamins B6, C and PP (pH=1.0):
C **B6, ·10^{-5}, M: *1* – 5.0, *2* – 7.5, *3* – 10;**
C **C: 5.0·10^{-5} M;**
C **PP, ·10^{-5} M: *1* – 5.0, *2* – 10, *3* – 1.0**



C (R=0.99)



PP (R=0.99)



B6 (R=0.98)

1 - experimental spectra, 2 - recovered spectra (method MILCA)

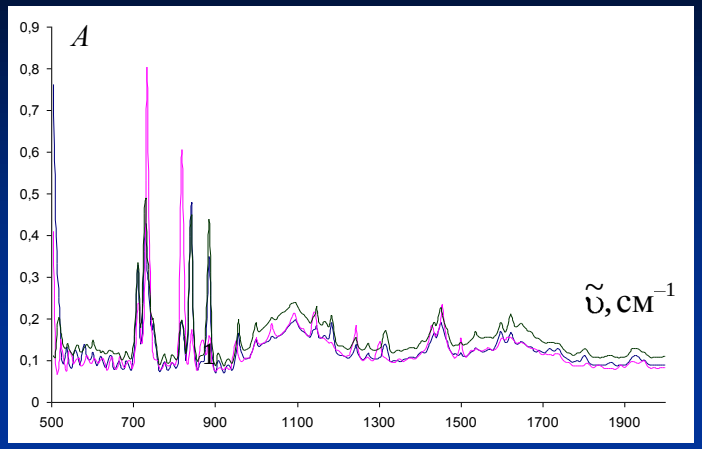
Correlation coefficients of experimental and resolved spectra (UV-VIS region)

System	Compound	Algorithm		
		MILCA	MILCA-ALS	SIMPLISMA
benzene-toluene-o-xylene	benzene	0.95	1.0	0.92
	toluene	0.98	0.99	0.84
	o-xylene	0.91	0.99	0.73
Benzene – toluene - o-xylene – m-xylene – p-xylene	benzene	0.95	0.96	0.89
	toluene	0.99	0.99	0.89
	o-xylene	0.99	1.0	0.50
	m-xylene	0.97	0.98	0.78
	p-xylene	0.89	0.92	0.65
Anthracene - pyrene - phenantrene	anthracene	1.0	0.99	1.0
	pyrene	0.98	1.0	0.96
	phenantrene	0.99	1.0	0.79
Anthracene phenantrene benz[a]antracene benz[a]phenantrene	anthracene	1.0	1.0	1.0
	phenantrene	0.89	0.95	0.65
	benz[a]antracene	0.91	0.96	0.80
	benz[a]phenantrene	1.0	0.98	0.85
Vitamins B6-B9-B12	B6	0.93	1.0	0.98
	B9	0.89	1.0	0.91
	B12	0.97	1.0	0.94

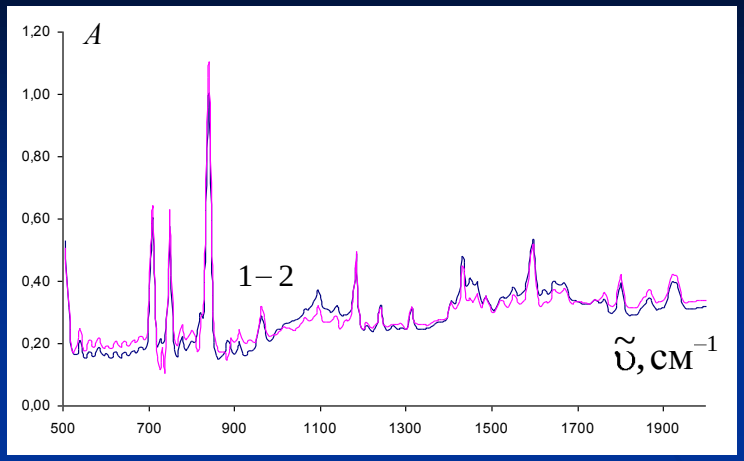
Amari indexes of similarity of experimental and calculated concentration matrixes (UV-VIS)

Algorithm / System	MILCA	MILCA+ALS	SIMPLISMA
benzene-toluene-o-xylene	0.02	0.0075	0.4
Benzene – toluene - o-xylene - m-xylene - p-xylene	0.05	0.01	0.5
Anthracene - pyrene - phenantrene	0.06	0.07	0.13
Anthracene phenantrene benz[a]antracene benz[a]phenantrene	0.08	0.01	0.07
B6-B9-B12	0.05	0.03	0.02

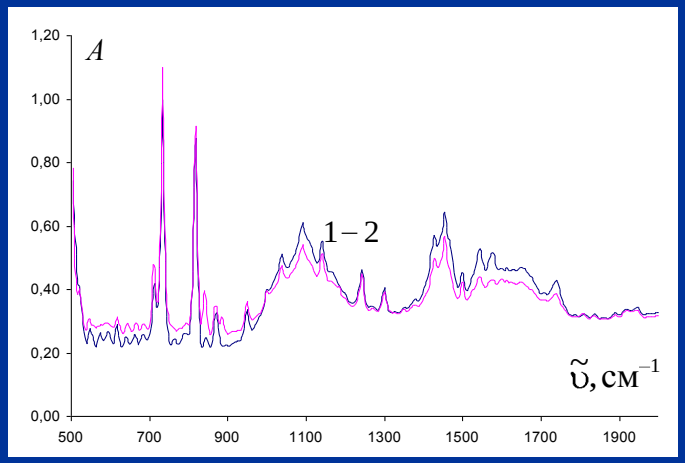
Infrared spectra



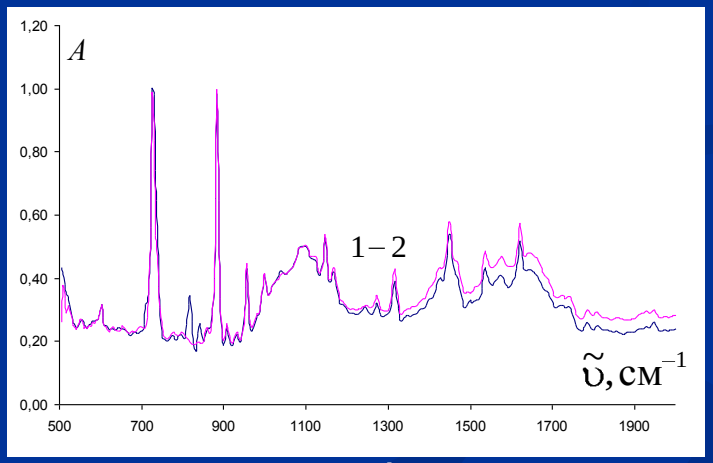
a



b



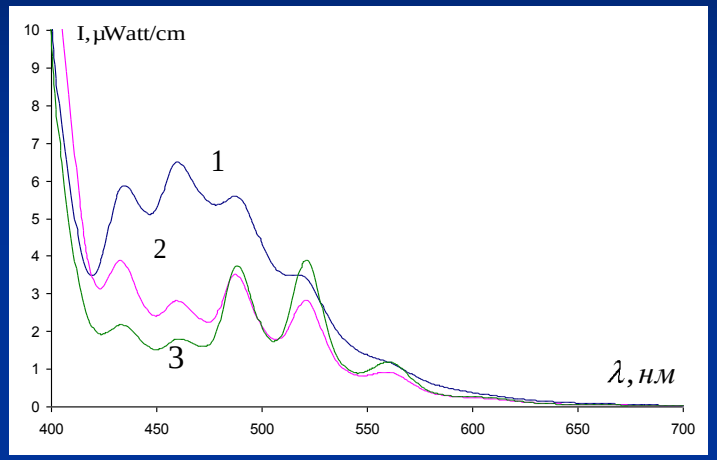
c



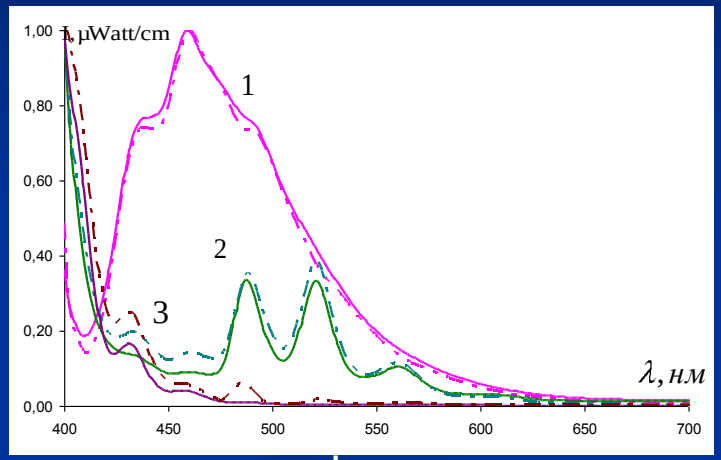
d

IR spectra of ternary mixtures (a);
b-d recovered spectra (1) of phenantrene (b), pyrene (c) and anthracene (d) in comparison with experimental spectra (2) (algorithm SNICA)

Fluorescence



a



b

Experinemtal fluorescence spectra of ternary mixtures (a):
1 - m(fluorantene):m(benz[a]phenantrene):m(phenantrene) = 1 : 1 :1;
2 - m(fluorantene):m(benz[a]phenantrene):m(phenantrene) = 4 : 2 :1;
3 - m(fluorantene):m(benz[a]phenantrene):m(phenantrene) = 1 : 4 :2;

b – experimental (solid line) and recovered (italic)
spectra of fluorantene (1), benz[a]phenantrene (2) и phenantrene (3)
accordingly (algorithm SNICA)

Correlation coeficients of experimental and resolved spectra (IR, fluorescence)

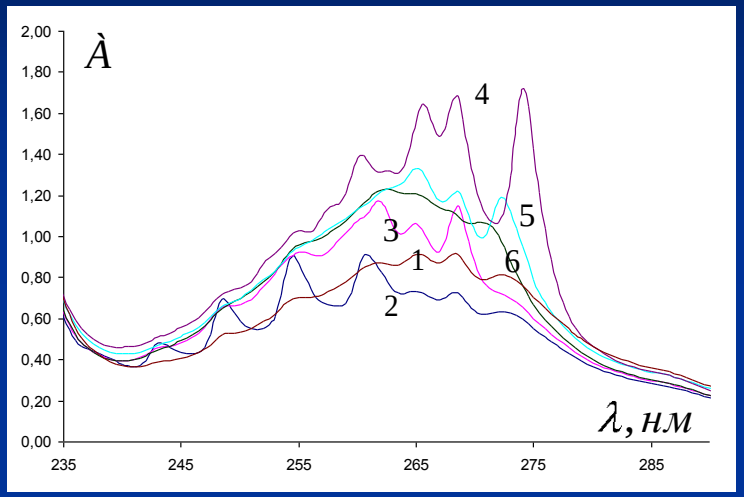
Type of spectra	System	Compound	MILCA	SNICA	MILCA+ ALS	SIMPLISMA
IR	benzene - toluene – isooctane (solvent CCl_4)	benzene	0.89	0.95	0.81	0.88
		toluene	0.95	0.98	0.88	0.89
		isooctane	0.98	0.90	1.0	0.93
IR	fluorantene – anthracene - pyrene – phenantrene (solvent CCl_4)	fluorantene	0.88	0.92	0.90	0.85
		phenantrene	0.90	0.89	0.91	0.88
		Anthracene	0.78	0.93	0.94	0.98
		pyrene	0.66	0.90	0.89	0.86
IR	Anthracene - pyrene – phenantrene (solid state)	Anthracene	0.91	0.98	0.96	0.88
		pyrene	0.90	0.95	0.93	0.91
		phenantrene	0.93	0.90	0.98	0.94
Fluorescence	fluorantene - benz[a]phenantrene - phenantrene (solid state)	fluorantene	0.80	1.0	1.0	0.87
		phenantrene	0.80	0.98	1.0	0.89
		benz[a]phenantrene	0.87	0.98	1.0	0.93

Amari indexes of similarity of experimental and calculated concentration matrixes (IR, fluorescence)

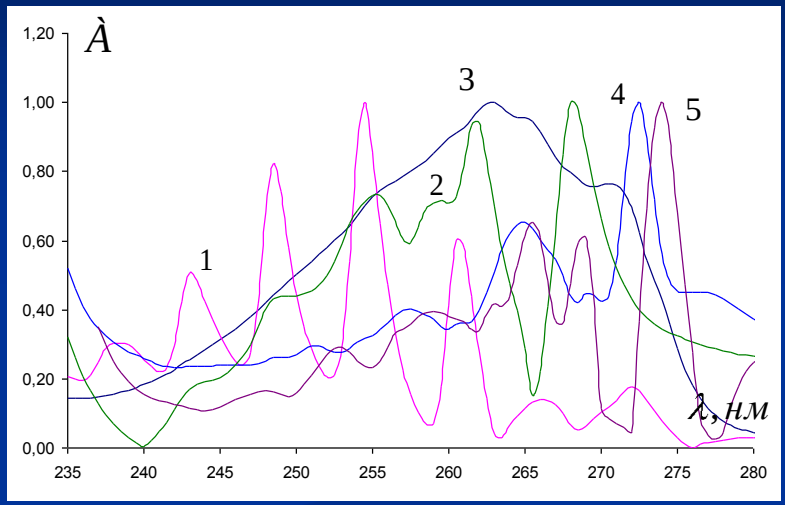
Type of spectra	System	MILCA	SNICA	MILCA+ ALS	SIMPLISMA
IR	benzene - toluene - isooctane	0.09	0.01	0.04	0.11
IR	fluorantene – anthracene - pyrene – phenantrene	0.10	0.02	0.03	0.05
Fluorescence	fluorantene - benz[a]phenantrene - phenantrene (solid state)	0.18	0.005	0.01	0.05

Analysis of real objects

gasoline



a



b

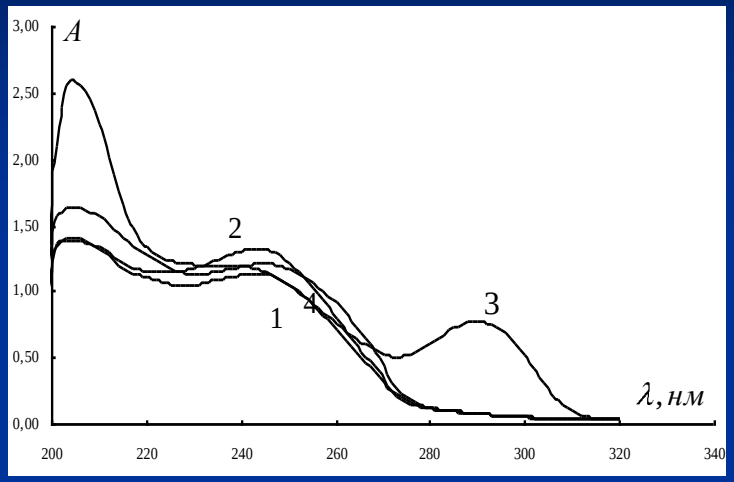
1 – gasoline (1:100 delution); 2-6 gasoline (1:133 delution) with standard addition of benzene, toluene, o-xylene, m-xylene, p-xylene ($c=2 \cdot 10^{-3}$ M) accordingly;

b – resolved spectra of individual compounds: benzene (1, r=0.95), toluene (2, r=0.91), o-xylene (3, r=0.90), p-xylene (4, r=0.93) and m-xylene (5, r=0.92) (algorithm MILCA)

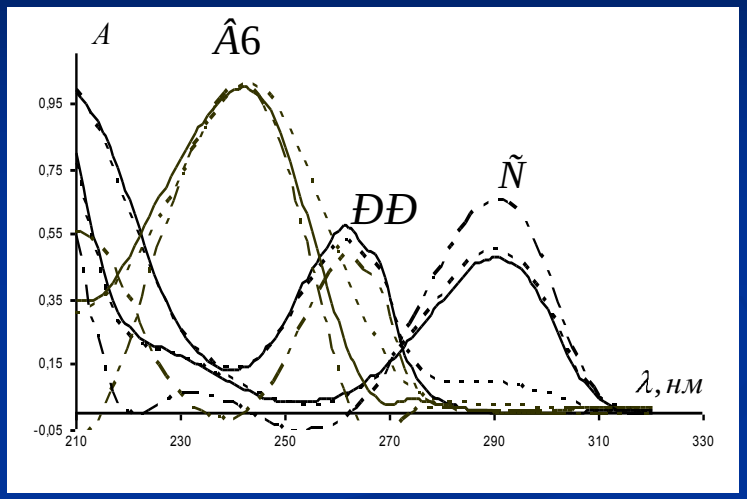
Determination of aromatic compounds in gasoline (n=3, p=0.95)

Compound / Method	Concentration, $\cdot 10^{-3}$ M				
	benzene	toluene	o-xylene	m-xylene	p-xylene
MILCA	2.2 0.21	4.5 0.38	2.2 0.18	3.5 0.26	1.1 0.08
Gas chromatography	2.5	4.6	2.0	3.8	1.1

Multivitamin drugs



a



b

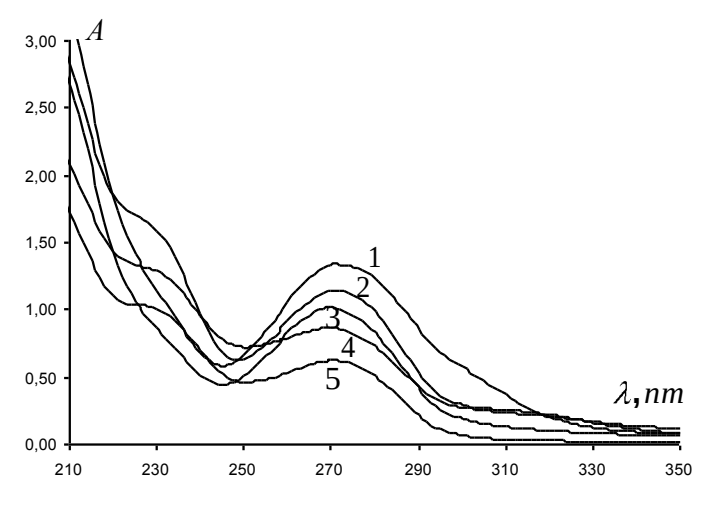
Absorption spectra of drug “Sana-Sol” without (1) and with standard addition of vitamins C (*2*), PP (*3*), B6 (*4*) (a) (pH=1.0): $c \cdot 10^{-5}$, M: 2 - 2.0, 3 – 2.5; *4* – 10;

b – resolved spectra of vitamins B6 (R=0.92-0.95) – C (R=0.91-0.98) – PP (R=0.93-0.97):
MILCA (italic), MCR-ALS (solid line), SIMPLISMA (. _ . _ . _ . _)

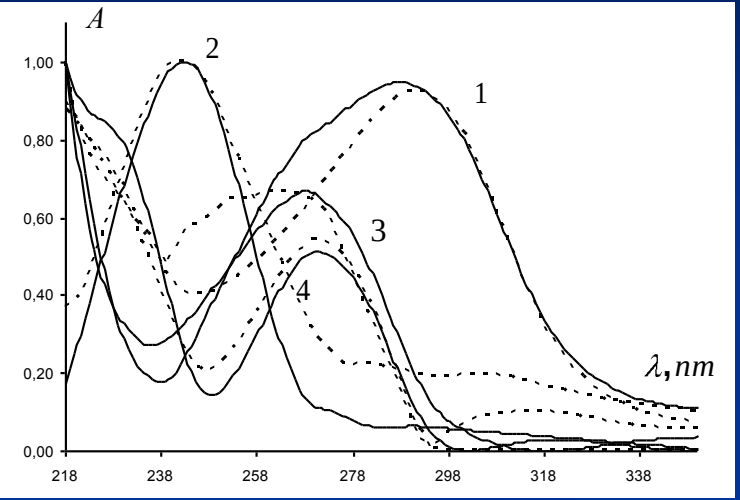
Determination of vitamins in multivitamin drugs (weight %) (n=3, p=0.95)

Drug	Vitamin	C	PP	B6	A	E
"Nitamin"	Declared	10	---	---	2.9	5.0
	HPLC	10 0.5	---	---	3.0 0.2	4.8 0.3
	MILCA	11 1	---	---	3.2 0.3	4.6 0.5
	SIMPLISMA	9.0 0.5	---	---	2.0 0.2	5.5 0.4
	MCR-ALS	10.2 0.8	---	---	3.0 0.3	5.2 0.3
"Picovit Forte"	Declared	5.2	1.8	0.18	---	---
	HPLC	4.7 0.4	2.0 0.2	0.23 0.02		
	MILCA	4.7 0.5	2.2 0.3	0.27 0.06		
	SIMPLISMA	2.9 1.2	3.3 0.9	0.8 0.3		
	MCR-ALS	5.1 0.9	2.4 0.5	0.3 0.02		
"Sana-Sol"	Declared	1.7	0.40	0.04	---	---
	HPLC	2.0 0.1	0.50 0.05	0.06 0.01		
	MILCA	1.7 0.1	0.48 0.08	0.09 0.01		
	SIMPLISMA	1.8 0.2	1.2 0.6	0.04 0.01		
	MCR-ALS	2.6 0.8	0.4 0.6	0.5 0.3		
"Complevit"	Declared	5.9	0.89	0.59	---	---
	HPLC	5.0 0.5	0.86 0.03	0.60 0.04		
	MILCA	4.5 0.5	0.86 0.04	0.51 0.06		
	SIMPLISMA	2.0 0.8	2.5 1.5	1.0 0.4		
	MCR-ALS	4.1 0.6	0.80 0.06	0.49 0.5		

Energy drinks



a



b

Experimental absorption spectra of energy drinks (a) (pH 1.0):
1- Red Bull; 2-Red Devil; 3-Bullit (1/20 dilution); 4-Strike (1/30 dilution); 5-Absenter Energy (1/20 dilution);

b – resolved spectra of vitamins B6 (1) (R=0.94-0.95), C (2) (R=0.95-0.99), PP (3) (R=0.90-0.97)
and caffeine (R=0.93-0.96) (4): MILCA (italic), MCR-ALS (solid line)

Determination of caffeine and vitamins in energy drinks (weight %) (n=3, p=0.95)

Energy drink	Vitamin	Vitamin C	Vitamin PP	Vitamin B6	Caffeine
«Bullit»	Declared	---	---	3.0	97
	MILCA			2.5 0.4	98 1
	MCR-ALS			3.5 0.3	97 2
«Red Bull»	Declared	---	---	7.1	93
	MILCA			8.4 0.5	91 3
	MCR-ALS			8.2 0.3	92 5
«Absenter Energy»	Declared	---	9.1	0.79	90
	MILCA		8.6 0.3	0.53 0.05	91 1
	MCR-ALS		9.1 0.2	0.65 0.04	90 2
«Red Devil»	Declared	---	0.52	1.9	98
	MILCA		0.58 0.06	2.5 0.3	96 1
	MCR-ALS		0.48 0.03	2.3 0.2	97 2
«TenStrike»	Declared	32	12	1.3	55
	MILCA	30 3	12 1	1.8 0.1	56 2
	MCR-ALS	31 4	10 1	1.5 0.1	51 2