

УДК 541.6

В.В.Важев, А.Ю.Кухарева, А.В.Цебрук

Костанайский государственный педагогический институт

**КОМПЬЮТЕРНОЕ ПРОГНОЗИРОВАНИЕ И ЭКСПЕРТИЗА ЗНАЧЕНИЙ
ТЕМПЕРАТУРЫ КИПЕНИЯ СПИРТОВ ПО ИК-СПЕКТРАМ**

ИК-спектрлерді дескриптор ретінде пайдалану, біратомдық спирттердің қайнау температураларына болжам жасау көзделеді. Болжаудың қорытындылары нәтижесінде эксперименталды ақпараттық мәліметтердің сараптамасы жүргізілді.

It is rotined that the use IR spectrums as descriptors allows successfully to predicting of boiling point of monoatomic alcohols. With the use of results of prognostication examination of experimental reference data is executed.

Исследование количественных соотношений структура–свойство, или QSPR — Quantitative Structure–Property Relationships, в настоящее время признается одним из приоритетных и фундаментальных направлений современной химической (и не только химической) науки. Возможно, это одна из самых важных проблем в естествознании. Помимо фундаментального аспекта, направление имеет и прикладной — предсказательный аспект, когда на основании структуры молекулы делается прогноз о свойствах вещества. Нормальная температура кипения (ТК) чаще всего фигурирует в QSPR-исследованиях как практически значимое свойство органических соединений. За последнее столетие было предложено множество аддитивных методов оценки ТК, пригодных для решения локальных прикладных задач. С другой стороны, ТК является модельным свойством, своего рода стандартом при разработке новых принципов кодирования структуры молекул и новых алгоритмов установления корреляционных соотношений структура-свойство. Очевидно, для построения эффективных моделей необходима информация о достоверных значениях ТК для веществ, входящих в тренировочную (обучающую) выборку. Известно, что любые экспериментально определенные величины, в том числе и ТК, в разных базах данных имеют часто значительный разброс, что порождает проблему выбора.

На наш взгляд, QSPR-подход, являющийся потребителем качественной информации об экспериментальных значениях свойств, может быть в то же время эффективным инструментом экспертизы для выявления надежных значений этих же свойств. Одним из вариантов работы QSPR-метода в режиме эксперта может быть исследование зависимости качества прогнозирования и моделирования свойства от использованных значений свойств веществ, полученных из разных источников.

В качестве объекта исследования выберем ТК спиртов, прогнозированию которой посвящен ряд работ, упомянем здесь наиболее удачные из них. Модель, найденная Ху и другими [1] для небольшого набора из 37 одноатомных спиртов характеризовалась коэффициентом корреляции $R = 0,978$ между экспериментальными и вычисленными значениями ТК и стандартным отклонением $s = 7,4988$. На тех же объектах в другой работе [2] было получено $R = 0,9837$ и $s = 6,35$. Для набора из 29 спиртов [3] получены результаты $R = 0,979$ и $s = 3,63$. Также небольшой набор из 54 предельных спиртов изучен в недавней работе [4], получено $R = 0,9945$ и $s = 4,26$. Рандич и Бейсак [5] для 58 спиртов построили серию моделей, лучшая из которых имеет $R = 0,9943$ и $s = 3,891$. Као и другие [6] для 106 одноатомных спиртов получена модель $R = 0,9899$ и $s = 4,5727$. Высокое качество прогнозирования $R = 0,9960$ и $s = 3,814$ достигнуто в работе Рена [7] для 108 спиртов. Одна из лучших моделей для 120 алифатических спиртов построена [8] с использованием топологических индексов и квантово-химических дескрипторов и характеризуется $R = 0,992$ и $s = 2,34$. Нами было выполнено прогнозирование ТК

спиртов с привлечением ИК-спектров в качестве дескрипторов молекулярной структуры [9]. Был использован набор из 100 спиртов, размер тренировочной выборки 60, контрольной — 40 веществ, модель в целом имела $R = 0,9972$ и $s = 2,27$, что свидетельствует в первую очередь о высокой информативности ИК-спектров как дескрипторов.

В настоящей работе исследована возможность дальнейшего повышения качества за счет экспертизы и уточнения данных о ТК, полученных из различных источников. Наиболее достоверная информация об экспериментальных значениях физико-химических свойств химических соединений, вероятно, содержится в следующих источниках: NIST Chemistry WebBook [10], справочник [11], сайт [12] справочник [13]. Используются ИК-спектры пропускания в газовой фазе 121 одноатомного алифатического спирта с числом атомов углерода от 2 до 14 из имеющихся на сервере [10]. Перед вычислениями спектры нормировались по формуле

$$d_{ij} = M_i \frac{h_{ij}}{\sum_{k=1}^{k=n} h_{ik}}, \quad (1)$$

где d_{ij} — рабочие значения дескрипторов, используемые в расчетах; n — количество точек в ИК-спектре; h_{ij} — ординаты пропускания в спектре вещества i в точке j ; M_i — относительная молекулярная масса вещества i . Использован участок спектров в интервале $550\text{--}3798\text{ см}^{-1}$. В этой области спектр линейно интерполирован с шагом 4 см^{-1} , количество точек в ИК-спектре $n = 813$. Элементы d_{ij} формируют матрицу дескрипторов \mathbf{D} , в которой выделены субматрицы дескрипторов \mathbf{A} — тренировочной и \mathbf{C} — контрольной выборок, для расчетов использовалось выражение [9]:

$$\mathbf{p} = \mathbf{C} \mathbf{V}_r \mathbf{\Gamma}_r^{-1} \mathbf{U}_r^T \mathbf{b}, \quad (2)$$

где \mathbf{p} — искомый вектор значений свойства контрольной выборки; \mathbf{b} — вектор значений свойства тренировочной выборки веществ; \mathbf{U} — матрица, состоящая из собственных векторов матрицы $\mathbf{A} \mathbf{A}^T$; \mathbf{V} — матрица, состоящая из собственных векторов матрицы $\mathbf{A}^T \mathbf{A}$; $\mathbf{\Gamma}$ — диагональная матрица из ненулевых сингулярных чисел, совпадающих для обоих матричных произведений. Нижний индекс является оценкой ранга матрицы дескрипторов $r(\mathbf{A})$ и означает, что используются r наибольших собственных чисел и соответствующих им собственных векторов. Расчеты выполнены с помощью разработанной нами [9] компьютерной программы PROGROC (PROGgram ROBustness Calculation). Тренировочная выборка состояла из 80, а контрольная — из 41 вещества.

Результаты прогнозирования ТК спиртов, полученные при оценке ранга $r(\mathbf{A}) = 78$, представлены в таблице 1, количественные показатели соответствующих корреляционных зависимостей приведены в таблице 2.

Таблица 1

Экспериментальные [10–13] и рассчитанные по ИК-спектрам значения нормальной температуры кипения спиртов (К). CAS — регистрационный номер по Chemical Abstracts

№	Соединение	CAS	Лит.	Эксп.	Расч.	Разн.
1	2	3	4	5	6	7
1	Ethanol	64–17–5	[11]	351,44	350,61	0,83
2	Isopropyl alcohol	67–63–0	[11]	355,41	355,73	–0,32
3	1-Propanol	71–23–8	[11]	370,35	370,53	–0,18
4	1-Butanol	71–36–3	[11]	390,81	390,59	0,22
5	2-Methyl-1-propanol	78–83–1	[11]	380,81	381,55	–0,74
6	2-Pentanol	6032–29–7	[11]	392,15	389,49	2,66
7	3-Hexanol	623–37–0	[11]	408,55	409,51	–0,96
8	2-Ethyl-1-butanol	97–95–0	[11]	419,65	418,65	1,00
9	3-Methyl-1-pentanol	589–35–5	[11]	425,55	427,80	–2,25
10	3-Methyl-2-pentanol	565–60–6	[11]	407,36	407,08	0,28
11	2-Methyl-1-pentanol	105–30–6	[11]	421,15	421,27	–0,12
12	3-Methyl-3-pentanol	77–74–7	[11]	395,56	394,64	0,92
13	3,3-Dimethyl-1-butanol	624–95–3	[11]	416,15	415,64	0,51

1	2	3	4	5	6	7
14	4-Methyl-1-pentanol	626-89-1	[13]	425,05	423,45	1,60
15	2,3-Dimethyl-2-butanol	594-60-5	[11]	391,75	390,86	0,89
16	5-Methyl-1-hexanol	627-98-5	[11]	445,15	444,71	0,44
17	4-Methyl-3-hexanol	615-29-2	[11]	422,15	421,69	0,46
18	2,3-Dimethyl-2-pentanol	4911-70-0	[11]	412,15	412,05	0,10
19	4-Methyl-2-hexanol	2313-61-3	[11]	424,15	425,19	-1,04
20	5-Methyl-3-hexanol	623-55-2	[11]	421,15	424,75	-3,60
21	3-Methyl-3-hexanol	597-96-6	[11]	415,95	420,89	-4,94
22	2,2-Dimethyl-3-pentanol	3970-62-5	[11]	409,15	408,75	0,40
23	3-Ethyl-2-pentanol	609-27-8	[11]	425,15	425,10	0,05
24	1-Heptanol	111-70-6	[11]	449,45	450,37	-0,92
25	2,3-Dimethyl-3-pentanol	595-41-5	[11]	413,15	413,84	-0,69
26	2-Ethyl-1-pentanol	27522-11-8	[11]	439,15	436,94	2,21
27	2-Methyl-2-hexanol	625-23-0	[11]	415,95	414,29	1,66
28	3-Methyl-2-hexanol	2313-65-7	[11]	425,15	425,15	0,00
29	5-Methyl-2-hexanol	627-59-8	[11]	424,15	426,44	-2,29
30	3-Ethyl-3-pentanol	597-49-9	[11]	415,65	415,73	-0,08
31	2,4-Dimethyl-3-pentanol	600-36-2	[11]	411,95	411,30	0,65
32	2,4-Dimethyl-2-pentanol	625-06-9	[11]	406,15	406,84	-0,69
33	3,4-Dimethyl-3-hexanol	19550-08-4	[11]	425,15	425,12	0,03
34	5-Methyl-2-heptanol	54630-50-1	[11]	445,05	444,17	0,88
35	6-Methyl-1-heptanol	1653-40-3	[11]	460,85	461,22	-0,37
36	3-Methyl-4-heptanol	1838-73-9	[11]	437,85	438,57	-0,72
37	2,2-Dimethyl-3-hexanol	4209-90-9	[11]	429,25	429,50	-0,25
38	2,3,4-Trimethyl-3-pentanol	3054-92-0	[11]	430,15	429,27	0,88
39	2,2,4-Trimethyl-3-pentanol	5162-48-1	[11]	424,15	425,51	-1,36
40	2-Ethyl-1-hexanol	104-76-7	[11]	457,75	457,50	0,25
41	3-Methyl-3-heptanol	5582-82-1	[13]	436,15	435,89	0,26
42	4-Ethyl-3-hexanol	19780-44-0	[11]	437,15	437,27	-0,12
43	3-Octanol	589-98-0	[13]	444,15	443,07	1,08
44	2-Ethyl-4-methyl-1-pentanol	106-67-2	[11]	449,65	450,73	-1,08
45	2,2,4-Trimethyl-1-pentanol	123-44-4	[11]	441,45	441,20	0,25
46	3-Ethyl-2-methyl-3-pentanol	597-05-7	[11]	433,15	433,54	-0,39
47	2,5-Dimethyl-3-hexanol	19550-07-3	[11]	432,15	428,66	3,49
48	2,4,4-Trimethyl-1-pentanol	16325-63-6	[11]	444,15	444,98	-0,83
49	2,3-Dimethyl-3-hexanol	4166-46-5	[11]	431,35	434,53	-3,18
50	3,5,5-Trimethyl-1-hexanol	3452-97-9	[11]	466,15	465,58	0,57
51	2-Methyl-4-octanol	40575-41-5	[11]	457,15	458,95	-1,80
52	3-Methyl-3-octanol	5340-36-3	[11]	462,15	460,45	1,70
53	2-Methyl-3-octanol	26533-34-6	[11]	457,15	457,97	-0,82
54	5-Nonanol	623-93-8	[11]	468,25	469,64	-1,39
55	3,5-Dimethyl-3-heptanol	19549-74-7	[11]	450,00	446,85	3,15
56	2,2-Dimethyl-3-heptanol	19549-70-3	[11]	451,00	450,37	0,63
57	1-Nonanol	143-08-8	[11]	486,25	486,12	0,13
58	2-Nonanol	628-99-9	[11]	471,65	471,05	0,60
59	4-Methyl-4-octanol	23418-37-3	[11]	454,15	453,30	0,85
60	4-Ethyl-4-heptanol	597-90-0	[11]	452,15	448,72	3,43
61	2-Methyl-2-octanol	628-44-4	[11]	451,15	451,48	-0,33
62	5-Methyl-5-nonanol	33933-78-7	[11]	475,15	477,19	-2,04
63	4,7-Dimethyl-4-octanol	19781-13-6	[11]	465,15	464,74	0,41
64	2-Decanol	1120-06-5	[13]	484,15	484,24	-0,09
65	3-Methyl-3-nonanol	21078-72-8	[11]	482,00	481,35	0,65
66	3,7-Dimethyl-3-octanol	78-69-3	[11]	469,15	470,40	-1,25
67	6-Ethyl-3-octanol	19781-27-2	[11]	472,00	472,25	-0,25
68	2,4,6-Trimethyl-4-heptanol	60836-07-9	[11]	455,15	454,60	0,55
69	2-Methyl-4-nonanol	26533-31-3	[11]	482,00	478,76	3,24
70	3,6-Dimethyl-3-octanol	151-19-9	[11]	465,15	465,67	-0,52
71	4-Propyl-4-heptanol	2198-72-3	[11]	467,15	468,96	-1,81

1	2	3	4	5	6	7
72	5-Undecanol	37493-70-2	[12]	502,15	499,95	2,20
73	2-Undecanol	1653-30-1	[11]	508,00	506,92	1,08
74	3-Undecanol	6929-08-4	[10]	502,15	502,82	-0,67
75	1-Undecanol	112-42-5	[11]	518,15	517,49	0,66
76	1-Dodecanol	112-53-8	[13]	533,15	533,51	-0,36
77	2-Dodecanol	10203-28-8	[11]	525,15	526,58	-1,43
78	2-Butyl-1-octanol	3913-02-8	[13]	519,65	520,99	-1,34
79	1-Tetradecanol	112-72-1	[11]	560,15	560,08	0,07
80	2-Tetradecanol	4706-81-4	[11]	557,15	557,46	-0,31
81	2-Butanol	78-92-2	[11]	372,70	371,03	1,67
82	2-Methyl-1-butanol	137-32-6	[11]	401,85	401,32	0,53
83	3-Pentanol	584-02-1	[13]	389,40	387,12	2,28
84	1-2-Methyl-1-butanol	1565-80-6	[10]	401,00	405,86	-4,86
85	3-Methyl-1-butanol	123-51-3	[11]	404,35	403,90	0,45
86	Amylene Hydrate	75-85-4	[11]	375,15	374,71	0,44
87	1-Hexanol	111-27-3	[11]	430,15	429,60	0,55
88	2-Hexanol	626-93-7	[11]	413,04	413,89	-0,85
89	2-Methyl-3-pentanol	565-67-3	[11]	399,66	400,38	-0,72
90	4-Methyl-2-pentanol	108-11-2	[11]	404,85	402,34	2,51
91	2-Methyl-2-pentanol	590-36-3	[11]	394,56	394,87	-0,31
92	2,3-Dimethyl-1-butanol	20281-85-0	[10]	417,70	421,68	-3,98
93	3-Heptanol	589-82-2	[11]	429,85	430,21	-0,36
94	2-Heptanol	543-49-7	[11]	432,35	434,74	-2,39
95	4-Heptanol	589-55-9	[13]	429,15	428,06	1,09
96	4-Methyl-4-heptanol	598-01-6	[11]	434,25	436,73	-2,48
97	2,5-Dimethyl-2-hexanol	3730-60-7	[11]	425,65	423,47	2,18
98	3,5-Dimethyl-3-hexanol	4209-91-0	[11]	425,15	427,98	-2,83
99	1-Octanol	111-87-5	[11]	468,35	468,34	0,01
100	6-Methyl-2-heptanol	4730-22-7	[11]	445,05	445,07	-0,02
101	2-Methyl-2-heptanol	625-25-2	[11]	429,85	426,83	3,02
102	2-Octanol	123-96-6	[11]	452,95	450,20	2,75
103	2,4-Dimethyl-3-hexanol	13432-25-2	[11]	433,15	435,16	-2,01
104	5-Methyl-1-heptanol	7212-53-5	[11]	459,75	462,64	-2,89
105	4-Methyl-1-heptanol	817-91-4	[11]	456,35	456,96	-0,61
106	3-Ethyl-3-hexanol	597-76-2	[11]	432,15	434,80	-2,65
107	4-Nonanol	5932-79-6	[11]	466,15	462,15	4,00
108	3-Nonanol	624-51-1	[11]	467,85	468,82	-0,97
109	5-Methyl-4-octanol	59734-23-5	[11]	455,00	456,55	-1,55
110	3-Ethyl-3-heptanol	19780-41-7	[11]	455,35	455,69	-0,34
111	2,6-Dimethyl-3-heptanol	19549-73-6	[11]	448,15	448,23	-0,08
112	2,3-Dimethyl-3-heptanol	19549-71-4	[11]	447,15	444,99	2,16
113	3,6-Dimethyl-3-heptanol	1573-28-0	[11]	446,15	444,97	1,18
114	2,6-Dimethyl-4-heptanol	108-82-7	[11]	451,00	452,02	-1,02
115	4-Decanol	2051-31-2	[11]	483,15	483,33	-0,18
116	3-Decanol	1565-81-7	[11]	483,00	482,63	0,37
117	2-Methyl-5-nonanol	29843-62-7	[11]	482,00	485,07	-3,07
118	5-Methyl-2-(1-methylethyl)-1-hexanol	2051-33-4	[11]	486,15	488,96	-2,81
119	1-Decanol	112-30-1	[13]	504,25	505,43	-1,18
120	1-Tridecanol	112-70-9	[11]	547,15	547,14	0,01
121	2-Tridecanol	1653-31-2	[11]	541,15	537,49	3,66

Примечание. Вещества с 1 по 80 составляют тренировочную выборку, с 81 по 121 — контрольную. Использованы названия веществ, принятые в базе NIST.

Распределение частот ошибок прогнозирования приведено на рисунке 1.

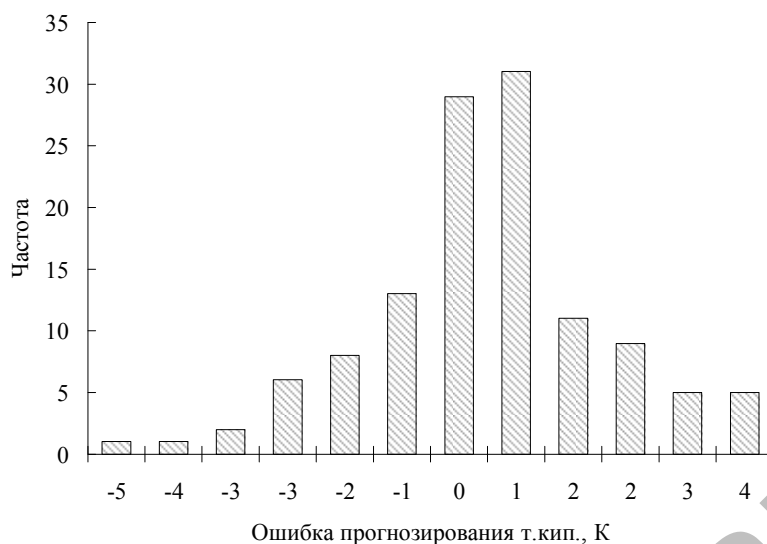


Рис. 1. Гистограмма распределения частот ошибок прогнозирования ТК спиртов

Максимальная ошибка наблюдается для 3-метил-гексанола-3 (№ 21) и достигает $-4,94$ К. Близость гистограммы распределения частот ошибок прогнозирования к нормальному подтверждает адекватность модели; 75 % отклонений расчетных значений ТК располагается в диапазоне от -2 до 2 К.

Т а б л и ц а 2

Показатели корреляции между экспериментальными и вычисленными значениями ТК спиртов

Показатели корреляции	Весь набор	Тренировочная выборка	Контрольная выборка
R	0,9991	0,9994	0,9985
S	1,71	1,47	2,12

Можно видеть, что данные таблиц 1 и 2 указывают на существенное улучшение качества прогнозирования при улучшении качества значений свойств веществ тренировочной выборки и использовании высокоинформативных дескрипторов (ИК-спектры).

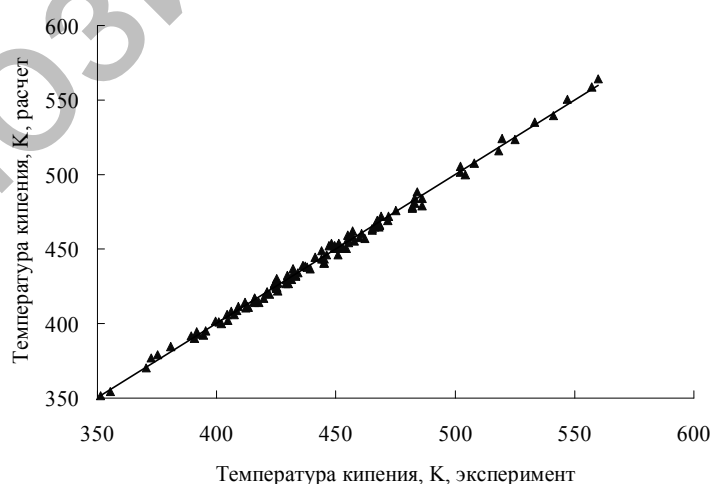


Рис. 2. Корреляционная зависимость между экспериментальными и вычисленными по ИК-спектрам значениями ТК спиртов (скользящий контроль)

Для улучшения качества прогнозирования свойств неисследованных (и даже неизвестных!) веществ нами предложен прием [14], суть которого состоит в контролировании качества прогнозирования спарринг-свойства, при его одновременном прогнозировании с целевым свойством. В качестве

спарринг-партнера удобнее всего использовать молекулярную массу, определение которой не представляет особых проблем. Тогда алгоритм решения будет заключаться в поиске условий (веществ тренировочной выборки, ранга матрицы дескрипторов), при которых достигается наиболее точное значение молекулярной массы (или находится наилучшее решение для другого спарринг-партнера). Фактически такой прием в некоторой степени подобен скользящему контролю, когда из набора поочередно извлекается каждое вещество, строится модель по оставшимся веществам, затем выполняется оценка свойства извлеченного вещества. Осуществление такого рода скользящего контроля в отношении исследуемой здесь выборки спиртов при прогнозировании их температуры кипения привело к результатам, которые компактно отражены на рисунке 2.

Этому рисунку соответствуют такие статистические параметры, как $R = 0,9978$ и $s = 2,65$. Сравнивая эти показатели с данными других авторов, отметим, что они существенно превосходят показатели любых лучших моделей.

Подводя итоги выполненного исследования, можно сделать следующие выводы:

- с использованием ИК-спектров как дескрипторов структуры молекул построена модель ТК с наиболее высокими на текущий момент показателями качества;
- впервые осуществлено прогнозирование температуры кипения спиртов в режиме скользящего контроля с качеством, существенно превосходящим качество простого моделирования;
- на основании анализа справочных данных можно рекомендовать 121 значение температуры кипения одноатомных спиртов, надежность которых подтверждена независимыми результатами прогнозирования по ИК-спектрам.
- важной особенностью предлагаемого подхода является возможность прогнозирования свойств неизвестных соединений без идентификации с использованием только их ИК-спектров.

Список литературы

1. Yao Y., Xu L., Yang Y., Yuan X. Study on structure-activity relationships of organic compounds: Three new topological indexes and their applications // *J. Chem. Inf. Comput. Sci.* — 1993. — Vol. 33. — № 4. — P. 590–594.
2. Yang Y.-O., Xu L., Hu C.-Y. Extended Adjacency Matrix Indices and Their Applications // *J. Chem. Inf. Comput. Sci.* — 1994. — Vol. 34. — № 5. — P. 1140–1145.
3. Galvez J., Garcia R., Salabert M.T., Soler R. Charge Indexes. New Topological Descriptors // *J. Chem. Inf. Comput. Sci.* — 1994. — Vol. 34. — № 3. — P. 520–525.
4. Kompany-Zareh M. A QSPR study of boiling point of saturated alcohols using genetic algorithm // *Acta Chim. Slov.* — 2003. — Vol. 50. — P. 259–273.
5. Randic M., Basak S.C. Optimal Molecular Descriptors Based on Weighted Path Numbers // *J. Chem. Inf. Comput. Sci.* — 1999. — Vol. 39. — № 2. — P. 261–266.
6. Cao C., Liu S., Li Z. On Molecular Polarizability: 2. Relationship to the Boiling Point of Alkanes and Alcohols // *J. Chem. Inf. Comput. Sci.* — 1999. — Vol. 39. — № 6. — P. 1105–1111.
7. Ren B. Novel atom-type AI indices for QSPR studies of alcohols // *Computers and Chemistry.* — 2002. — Vol. 26. — P. 223–235.
8. Penchev P., Kochev N., Vandeva V., Andreev G. Prediction Of Boiling Points Of Acyclic Aliphatic Alcohols From Their Structure // *Travaux Scientifiques d'Universite de Plovdiv.* — 2007. — Vol. 35. — № 5. — P. 53–57.
9. Важев В.В., Алдабергенов М.К. Использование инфракрасных спектров для прогнозирования температуры кипения и молекулярной массы одноатомных алифатических спиртов // *Журн. прикл. химии.* — 2005. — Т. 78. — № 3. — С. 457–460.
10. NIST Chemistry WebBook. NIST Standard Reference Database Number 69 — November 1998 Release. <http://webbook.nist.gov/chemistry/>
11. Yaws' Handbook of Thermodynamic and Physical Properties of Chemical Compounds / CARL L. YAWS, ed. — New York, Norwich, 2004.
12. UW-Madison Chemistry Library. PHYSICAL & CHEMICAL PROPERTIES: O-P. — Режим доступа: <http://esc.syrres.com/interkow/physdemo.htm>
13. CRC Handbook of Chemistry and Physics, 88th Edition / David R. Lide, ed. — CRC Press, Inc., Boca Raton, FL. — 2007. — 2640 p.
14. Важев В.В., Бектурганова Г.К. и др. Прогнозирование и экспертиза значений температуры вспышки углеводородов. I. Алкены // *Метрология.* — 2007. — № 4. — С. 13–18.