Профессор
Бекман Игорь Николаевич

МЕМБРАНЫ В МЕДИЦИНЕ

Курс лекций

Лекция 8. ПАРНЫЕ КОРРЕЛЯЦИИ КОЭФФИЦИЕНТОВ ПРОНИЦАЕМОСТИ

Предсказание величин параметров массопереноса интересующего нас газа, для которого отсутствуют данные для большинства важных для нас полимеров, но всё же есть данные по одному или нескольким полимерам, можно осуществить путём построения корреляционных зависимостей типа проницаемость одного газа в зависимости от проницаемости другого газа (двумерные корреляции), построение графиков зависимости проницаемости третьего газа в зависимости от проницаемостей первого и второго газа, и т.д., переходя к многомерным пространствам и проводя сквозь образовавшиеся совокупности («облака»), регрессионные прямые, подогнанные к экспериментальным данным методом наименьших квадратов (МНК). Конечно, многомерные графики (например, 6-ти мерные) представить визуально невозможно, но, тем не менее, расчёты по найденному уравнению регрессии позволяют установить вероятное значение коэффициента проницаемости требуемого газа в требуемом полимере. Естественно, что такие операции можно проводить и с коэффициентом диффузии и с константой растворимости.

В этой лекции мы рассмотрим интерполяцию транспортных данных (в основном – собранных в Базе Данных ИНХС) с целью прогноза величин коэффициентов проницаемости газов в полимерах, диффузионные свойства которых хоть и не очень хорошо, но всё же изучены. Проблемой экстраполяции данных на неизвестные с диффузионной точки зрения системы займёмся в следующей лекции.

1. РЕГРЕССИОННЫЙ АНАЛИЗ

В качестве базовой статистической модели обычно выбирается модель множественной линейной регрессии (МЛР), а сами расчеты проводят с использованием интегрированных систем комплексного статистического анализа и обработки данных STATISTICA 6.0 (StatSoft Inc.), Origin 6.0 (Microcal Software Inc.) и пакета Микрософт Exel.

Линейный регрессионный анализ является широко распространенным статистическим методом анализа массивов экспериментальных данных во многих областях научного познания. Этот статистический инструмент позволяет исследователю разработать математическую модель, описывающую взаимосвязь между двумя или более переменными. Модель МЛР может быть применена как для описания массивов данных, оценки параметров, так и для прогнозирования.

Регрессионный анализ – статистический метод исследования зависимости между зависимой переменной Y и одной или несколькими независимыми переменными $X_1,X_2,...,X_p$.

Цели регрессионного анализа: определение степени детерменированности вариации зависимой переменной независимыми переменными; предсказание значения зависимой переменной с помощью независимой (-ых); определение вклада отдельных независимых переменных в вариацию зависимой.

Замечание. Регрессионный анализ нельзя использовать для определения наличия связи между переменными, поскольку наличие такой связи и есть предпосылка для применения анализа.

Различают линейный и нелинейный регрессионный анализ. Нелинейность бывает двух типов: нелинейность по независимым переменным (формально этот тип легко сводится к линейной регрессии), и о нелинейность по оцениваемым параметрам (приводит к серьезным вычислительным трудностям). При нелинейности первого вида в модели появляются члены вида X_1X_2, $X_1X_2X_3$, свидетельствующее о наличии взаимодействий между признаками X_1, X_2 и т.д.

На практике линия регрессии чаще всего ищется в виде линейной функции

$$Y = b_0 + b_1X_1 + b_2X_2 + ... + b_NX_N \qquad (1)$$

(линейная регрессия), наилучшим образом приближающей искомую кривую. Делается это с помощью метода наименьших квадратов, МНК, когда минимизируется сумма квадратов отклонений реально наблюдаемых Y от их оценок $\hat{Y}$ (имеются в виду оценки с помощью прямой линии, претендующей на то, чтобы представлять искомую регрессионную зависимость):

$$\sum_{k=1}^{M}\left(Y_k-\hat{Y}_k\right)^2\to\min \qquad (2)$$

(M — объём выборки). Этот подход основан на том факте, что фигурирующая в приведенном выражении сумма принимает минимальное значение именно для того случая, когда $Y=y(x_1,x_2,...x_N)$.

Для решения задачи регрессионного анализа методом наименьших квадратов вводится понятие функции невязки:

$$\sigma\left(\bar{b}\right)=\frac{1}{2}\sum_{k=1}^{M}\left(Y_k-\hat{Y}_k\right)^2 \qquad (3)$$

Если экспериментатор рассматривает возможное влияние на свой параметр множество других параметров, то обычно он начинает с анализа парных взаимодействий, т.е. с подгонки к данным уравнения регрессии вида:

$y=b_0+b_1x$

график которого – прямая линия.

Затем переходят к уравнению вида:

$y= b_0+b_1x_1+ b_2x_2$,

график которого – некоторая плоскость

Далее рассматривается регрессия

$y= b_0+b_1x_1+ b_2x_2+b_3x_3$,

которое уже не может быть представлено графически.

И так далее…

Замечание. Возникает вопрос: увеличивается ли надёжность анализа при вовлечении в него всё большего числа переменных? Прогресс оказывается спорным: хотя коэффициент регрессии увеличивается, что свидетельствует приближение данных к линейной зависимости, это не обязательно означает улучшение качества регрессии и приводит к снижению значимости оцениваемых параметров. Рост коэффициента корреляции компенсируется «штрафом» за увеличение числа регрессоров (см. ниже).

Особенности множественной линейной регрессии проиллюстрируем на примере парных взаимодействий. Линейное уравнение связи двух переменных, учитывающее случайные отклонения, можно представить в виде:

$$y=\alpha+\beta x+\varepsilon \qquad (4)$$

где ε – отклонение от теоретически предполагаемого значения; α и β - неизвестные параметры (коэффициенты регрессии).

В этом уравнении можно выделить две части: 1) систематическую, $\hat{y}=\alpha+\beta x$, где $\hat{y}$ характеризует некоторое среднее значение y для данного значения x; 2) случайную (ε).

Коэффициенты α и β описывают вид зависимости для генеральной совокупности. Так как при выполнении подобных исследований всегда имеют дело с выборочной совокупностью, то истинные значения параметров α и β являются неизвестными, и мы можем говорить лишь об их оценках. Обозначим эти оценки, соответственно, a и b, тогда уравнение регрессии с оцененными параметрами будет иметь вид:

$$\hat{y}_i=a+bx_i,\quad i=1,n \qquad (5)$$

где n - объем выборки.

Обозначим через e_i отклонение реального значения отклика y_i от теоретически рассчитанного по уравнению $\hat{y}_i$.

Параметры a и b уравнения регрессии чаще всего оцениваются с помощью метода наименьших квадратов (МНК). Суть его состоит в том, чтобы, зная положение точек на плоскости XY, так провести линию регрессии, чтобы сумма квадратов отклонений e_i^2 этих точек от проведенной прямой вдоль оси OY была минимальной.

Математически критерий оценки параметров линейной парной регрессии записывается так:

$$Q=\sum_{i=1}^{n}e_i^2=\sum_{i=1}^{n}\left(y_i-\hat{y}\right)^2=\sum_{i=1}^{n}\left(y_i-a-bx_i\right)^2\to\min \qquad (6)$$

Расчетные формулы для нахождения коэффициентов уравнения регрессии:

$$b = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (7)$$

$$a = \bar{y} - b\bar{x}$$

Наиболее вероятные значения коэффициентов уравнения регрессии a и b можно рассчитать по формулам:

$$\bar{b} = \frac{n\sum_{i=`}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = r\frac{s_y}{s_x} \qquad (25)$$

$$\bar{a} = \bar{y} - b\bar{x} \qquad (26)$$

где $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ - средние значения x и y, соответственно, n – число экспериментальных точек.

Ошибки параметров прямой:

$$s_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}, \quad s_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2}\ . \qquad (27)$$

Величина влияния фактора на исследуемый отклик может быть оценена при помощи коэффициента линейной парной корреляции, характеризующего тесноту (силу) линейной связи между двумя переменными. Коэффициент корреляции можно определить по формуле:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} = b\sqrt{\frac{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}{n\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2}}\ .$$

Коэффициент парной корреляции обеспечивает проверку гипотезы – чем ближе значение r к единице, тем лучше прямая линия описывает экспериментальные данные и тем справедливее становится выбор классического механизма для описания изучаемого процесса диффузии. Он обладает следующими свойствами: 1) не имеет размерности, следовательно, сопоставим для величин различных порядков; 2) изменяется в диапазоне от –1 до +1. Положительное значение свидетельствует о прямой линейной связи, отрицательное – об обратной. Чем ближе абсолютное значение коэффициента к единице, тем теснее связь. Считается, что связь достаточно сильная, если коэффициент по абсолютной величине превышает 0,7, и слабая, если он менее 0,3. Величина r^2 называется коэффициентом детерминации. Он определяет долю вариации одной из переменных, которая объясняется вариацией другой переменной. Величина влияния фактора на исследуемый отклик может быть оценена при помощи коэффициента линейной парной корреляции, характеризующего тесноту (силу) линейной связи между двумя переменными. Для оценки значимости коэффициента корреляции применяется t - критерий Стьюдента. При этом фактическое значение этого критерия *($t_{эксп}$)*

$$t_{эксп} = \sqrt{\frac{r^2(n-2)}{1-r^2}} \qquad ()$$

сравнивается с критическим значением *($t_{теор}$)* которое берется из таблицы значений критерия Стьюдента $t(p,f)$ с учетом заданного уровня значимости *($p = 0.05$)* и числа степеней свободы *($n - 2$)*. Если $t_{эксп} > t_{теор}$, то полученное значение коэффициента парной корреляции признается значимым.

Перейдём теперь к оценке доверительных интервалов для параметров a и b.

$$a = \bar{a} \pm \Delta a_1 \ \text{и}\ b = \bar{b} \pm \Delta b\,, \qquad (30)$$

где

$$\Delta a = \frac{t(p,f)}{s_x}\sqrt{\frac{q(b)\cdot\sum_{i=1}^{n} x_i^2}{(n-2)(n-1)n}} \qquad (31a)$$

$$\Delta b = \frac{t(p,f)}{s_x}\sqrt{\frac{q(b)}{(n-2)(n-1)}} \qquad (31б)$$

где

$q(b) = \sum_{i=1}^{n}(y_i - \bar{y})^2 - \bar{b}^2\sum_{i=1}^{n}(x_i - \bar{x})^2 = (n-1)(s_y^2 - b^2 s_x^2)$, t(p=0,95, f=n-2) – критерий Стьюдента.

Важный этап анализа - верификация модели регрессии, т.е. процесс, в ходе которого подвергается анализу качество полученной модели. Допустим, имеется уравнение. Значения определяемые уравнением $\hat{y}_i$, тогда фактические значения можно представить как:

$$y_i = \hat{y}_i + e_i \qquad (9)$$

где e_i - случайная (остаточная) компонента.

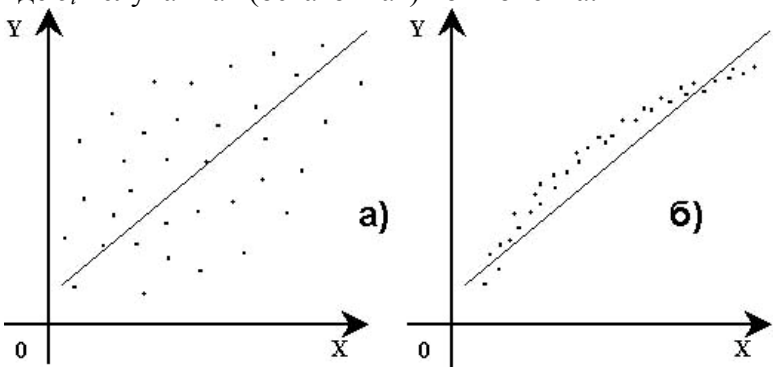


Рис. 1. Пример модели регрессии: а – модель адекватна, но не точна; б – модель точна, но не адекватна.

Анализ остаточной компоненты (остаточного ряда) позволяет оценить качество полученного уравнения регрессии. Качество характеризуется выполнением определенных статистических свойств и точностью, т.е. степенью близости к фактическим данным. Модель считается хорошей со статистической точки зрения, если она адекватна и достаточно точна. Смысл используемых терминов характеризуют **Рис. 1.** Оценить адекватность модели позволяет анализ случайной компоненты e_i. Модель считается адекватной исследуемому процессу, если: 1) математическое ожидание значений остаточного ряда близко или равно нулю; 2) значения остаточного ряда случайны; 3) независимы; 4) подчинены нормальному закону распределения.

Поиск уравнения регрессии удобно вести в рамках системы EXCEL. Остановимся на этом несколько подробнее.

Аппроксимация данных прямолинейной зависимостью состоит из следующих этапов:

1. Расчёт параметров уравнений линейной регрессии
2. Оценка тесноты связи с помощью показателей корреляции и детерминации
3. Оценка статической надежности регрессионного моделирования с помощью F- критерия Фишера и t-критерия Стьюдента.
4. Расчёт прогнозного значения результата, если прогнозное значение фактора увеличится на 10% от его среднего уровня.
5. Определение доверительного интервала прогноза для уровня значимости α = 0,05

Для парной линейной регрессии: $y_x = r_{xy}\frac{x-\bar{x}}{\sigma_x}\sigma_y + \bar{y}$

Среднее значение $\bar{x} = \frac{\sum x_i}{n}$; $\bar{y} = \frac{\sum y_i}{n}$; $\overline{xy} = \frac{\sum x_i y_i}{n}$

Среднеквадратическое отклонение $\sigma_x = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2}$; $\sigma_y = \sqrt{\frac{\sum y^2}{n} - \bar{y}^2}$

Коэффициент корреляции можно найти по формуле: $r_{xy} = a\frac{\sigma_x}{\sigma_y}$, с помощью статистической функции КОРРЕЛ (Excel), по формуле:

$$r_{xy} = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\sigma_x \sigma_y} \qquad r_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n \sigma_x \sigma_y}$$ или

по формуле: $$r_{xy} = \frac{K_{xy}}{\sigma_x \sigma_y}$$

K_{xy} - корреляционный момент (коэффициент ковариации) $$K_{xy} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n}$$

$$r = \frac{\sigma_x^2 + \sigma_y^2 - \sigma_{x-y}^2}{2 \sigma_x \cdot \sigma_y}$$

Средняя ошибка аппроксимации (рассчитываем столбцы y_x, $y_i - y_x$, A_i)

$$A_i = \left| \frac{y - y_x}{y} \right| \cdot 100$$

$$\bar{A} = \frac{\sum A_i}{n}$$

Это означает, что в среднем, расчетные значения зависимого признака отклоняются от фактического значения на A%.

Коэффициент детерминации

а) с помощью функции ЛИНЕЙН (Excel)

б) $R^2 = r^2_{xy}$; R^2 = 0,32, т.е. в 31,57% случаев изменения *x* приводят к изменению *y*. Другими словами - точность подбора уравнения регрессии 31,57% - низкая.

в) по формуле:

$$R^2 = 1 - \frac{\sum (y_i - y_{xi})^2}{\sum (y_i - \bar{y})^2}$$

Оценка статистической значимости

Использование критерия Фишера:

1. Выдвигаем нулевую гипотезу о статистической незначимости параметров регрессии и показателя корреляции $a = b = r_{xy}$

2. Фактическое значение критерия получено из функции ЛИНЕЙН (Excel)

3.Для определения табличного значения критерия рассчитываем коэффициенты $k_1 = m = 1$ и $k_2 = n - m - 1$

4. Сравниваем фактическое и табличное, значения критерия $F_{факт} > F_{табл}$ нулевую гипотезу отклоняем и делаем вывод о статистической значимости и надежности полученной модели.

$$F = \frac{R^2}{1 - R^2}(n - 2)$$

Использование критерия Стьюдента

1. Выдвигаем нулевую гипотезу о статистически незначимом отличии показателей от нуля; $a = b = r_{xy} = 0$;

2. Табличное значение *t*-критерия зависит от числа степеней свободы и заданного уровня значимости a.

Уровень значимости - это вероятность отвергнуть правильную гипотезу при условии, что она верна.

3. Фактические значения *t*-критерия рассчитываются отдельно для каждого параметра модели. С этой целью, сначала определяются случайные ошибки параметров m_a, m_b, m_r

$$m_a = S_{ocm} \frac{\sqrt{\sum x^2}}{n \sigma_x} \qquad m_a = S_{ocm} \sqrt{\frac{\sum x_i^2}{n \sum (x_i - \bar{x})^2}}$$ или

$$m_b = \frac{S_{ocm}}{\sigma_x \sqrt{n}} \qquad m_b = \frac{S_{ocm}}{\sqrt{\sum (x_i - \bar{x})^2}}$$ или

$$m_r = \sqrt{\frac{1 - r_{xy}^2}{n - m - 1}}$$

где $$S_{ocm} = \sqrt{\frac{\sum (y - y_x)^2}{n - m - 1}}$$

S^2 - необъясненная дисперсия является несмещенной оценкой дисперсии случайных отклонений
n - число наблюдений, m - число независимых переменных.

Рассчитаем фактические значения t-критерия. $t_{фа} = \frac{a}{m_a}$; $t_{фb} = \frac{b}{m_b}$; $t_{фr} = \frac{r_{xy}}{m_r}$;

Сравниваем фактические значения t-критерия с табличными значением: $t_{фа} > t_{набл;}$ $t_{фb} > t_{набл;}$ $t_{фс} > t_{набл}$
нулевую гипотезу отклоняем, параметры a = b = r_{xy} - не случайно отличаются от нуля и являются статистически значимыми и надежными.

Рассчитываем доверительный интервал для параметров регрессии *a,b* для чего находим предельную ошибку параметров: $\Delta a = t_{табл} * m_{a:}$ $\Delta b = t_{табл} * m_b$

Доверительные интервалы: $a \pm \Delta a$; $a - \Delta a \le a \le a + \Delta a$; $b \pm \Delta b$; $b - \Delta b \le b \le b + \Delta b$

Анализ верхней и нижней границ доверительного интервалов показывает, что с вероятностью $p = 1 - \alpha = 0,95$ параметры *a* и *b* не принимают нулевые значения, т.е. являются статистически значимыми и надежными. Если одна из границ доверительного интервала - меньше нуля или равна нулю - делается вывод о статистической незначимости соответствующего параметра.

Полученные оценки уравнения регрессии позволяют использовать его для прогноза. Прогнозное значение y_p определятся путем подстановки в уравнение регрессии соответствующего прогнозного значения x_p. Если прогнозное значение составит $x_p = \bar{x} \cdot 1,1$, то прогнозное значение *y*

$y_p = b + a * x_p$

Расчёт случайной ошибки прогноза:

$$m_{yp} = S_{ост}\sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x - \bar{x})^2}}$$

Предельная ошибка прогноза: $\Delta y = t_{табл} * m_{yp}$

Доверительный интервал прогноза: $\gamma_{yp} = y_p \pm \Delta y_{p;}$ $y_p - \Delta y_p \le y_p \le y_p + \Delta y_p$

С надежностью 0,95 прогнозное значение *y* заключено в данном доверительном интервале. Поскольку границы не принимают нулевых значений можно сделать вывод о статистической надежности прогноза.

Оценка качества регрессионной модели характеризуется рядом показателей:

- проверка статистической значимости коэффициентов;
- определение интервальные оценки коэффициентов уравнения регрессии;
- определение доверительных интервалов для зависимой переменной;
- проверка общего качества уравнения регрессии (коэффициент детерминации и его статистическая значимость).

Пример (без параллельных в точке).

x	y	x^2	y^2	x · y	y(x)	$(y-y_{ср})^2$	$(y-y(x))^2$	$(x-x_{ср})^2$
1	107	1	11449	107	103.19	333.06	14.5	30.25
2	109	4	11881	218	107.2	264.06	3.23	20.25
3	110	9	12100	330	111.21	232.56	1.47	12.25
4	113	16	12769	452	115.22	150.06	4.95	6.25
5	120	25	14400	600	119.23	27.56	0.59	2.25
6	122	36	14884	732	123.24	10.56	1.55	0.25
7	123	49	15129	861	127.26	5.06	18.11	0.25
8	128	64	16384	1024	131.27	7.56	10.67	2.25
9	136	81	18496	1224	135.28	115.56	0.52	6.25
10	140	100	19600	1400	139.29	217.56	0.51	12.25
11	145	121	21025	1595	143.3	390.06	2.9	20.25
12	150	144	22500	1800	147.31	612.56	7.25	30.25
78	**1503**	**650**	**190617**	**10343**	**1503**	**2366.25**	**66.23**	**143**

Уравнение прямой имеет вид y = ax + b = 4,01x+99,18.

a=4,01; b=99,18

Средние значения

$$\bar{x} = \frac{\sum x_i}{n} = \frac{78}{12} = 6.5$$

$$\bar{y} = \frac{\sum y_i}{n} = \frac{1503}{12} = 125.25$$

$$\overline{xy} = \frac{\sum x_i y_i}{n} = \frac{10343}{12} = 861.92$$

Дисперсия

$$D(x) = \frac{\sum x^2_i}{n} - \bar{x}^2 = \frac{630}{12} - 6.5^2 = 11.92$$

$$D(y) = \frac{\sum y^2_i}{n} - \bar{y}^2 = \frac{190617}{12} - 125.25^2 = 197.19$$

Среднеквадратическое отклонение

$$\sigma(x) = \sqrt{D(x)} = \sqrt{11.92} = 3.45$$

$$\sigma(y) = \sqrt{D(y)} = \sqrt{197.19} = 14.04$$

Коэффициент корреляции

$$r_{xy} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sigma(x) \cdot \sigma(y)} = \frac{861.92 - 6.5 \cdot 125.25}{3.45 \cdot 14.04} = 0.9859$$

Связь между признаком Y фактором X сильная и прямая

Уравнение регрессии

$$y_x = r_{xy} \frac{x - \bar{x}}{\sigma(x)} \sigma(y) + \bar{y} = 0.9859 \frac{x - 6.5}{3.45} 14.04 + 125.25 = 4.01x + 99.18$$

Коэффициент регрессии (коэффициент эластичности)

k = a = 4.01

Индекс корреляции (эмпирическое корреляционное отношение)

$$i = \sqrt{\frac{\sigma_{y_x}}{\sigma_{y_0}}} = \sqrt{\frac{2366.25}{2432.48}} = 0.9728$$

где $\sigma_{y0} = 2366.25 + 66.23 = 2432.48$

Фактор x существенно влияет на y.

Коэффициент детерминации $R^2 = 0.99^2 = 0.97$, т.е. в 97.2009 % случаев изменения x приводят к изменению y. Точность подбора уравнения регрессии - высокая

Значимость коэффициента корреляции определяется t-критерием Стьюдента:

$$t_{набл} = r_{xy} \frac{\sqrt{n-2}}{\sqrt{1-r^2_{xy}}} = 0{,}9859 \sqrt{\frac{10}{1-0{,}972}} = 18{,}63$$

По таблице Стьюдента находим $t_{табл}$

Tнабл(n-m-1,a)=(10, 0,05)=1,812

Поскольку $t_{набл} < t_{табл}$, то отклоняем гипотезу о равенстве 0 коэффициента корреляции.

Интервальная оценка для коэффициента корреляции (доверительный интервал)

$$\left(r - \varepsilon \frac{1-r^2}{\sqrt{n}} ; r + \varepsilon \frac{1-r^2}{\sqrt{n}} \right)$$

Доверительный интервал для коэффициента корреляции: r(0.9713; 1.0005).

Анализ точности определения оценок коэффициентов регрессии

$$S_{ост} = \sqrt{\frac{\sum (y_i - y_x)^2}{n - m - 1}}$$

$$S_{ост} = \frac{66.23}{10} = 2.57$$

$$S_b = S_{ост} \frac{\sqrt{\sum x^2}}{n \sigma_x}$$

$$S_b = 2.5736 \frac{\sqrt{630}}{12 \cdot 3.45} = 1.5839$$

$$S_a = \frac{S_{ост}}{\sqrt{n} \sigma_x}$$

$$S_a = 0.2152$$

Доверительные интервалы для зависимой переменной

$$(b + ax - t S_{ост} \sqrt{1 + \frac{1}{n} + \frac{(\bar{x} - x_p)^2}{(y_i - \bar{y})^2}})$$

Границы интервала, в котором будет сосредоточено 95% возможных значений Y при неограниченно большом числе наблюдений и X = 7: (122.4; 132.11)

Проверка гипотез относительно коэффициентов линейного уравнения регрессии

1) t-статистика

$$t_a = \frac{a}{m_a}$$

$$t_a = \frac{4.01}{0.22} = 18.63$$

Статистическая значимость коэффициента регрессии подтверждается

$$t_b = \frac{99.18}{1.58} = 62.62$$

Статистическая значимость коэффициента регрессии подтверждается

Доверительные интервалы коэффициентов регрессии с надежностью 95%:

$(a - t_a S_a; a + t_a S_a)$

(3.6205;4.4005)

$(b - t_b S_b; b + t_b S_b)$

(96.3117;102.0519)

Критерий Фишера (F-статистика):

$$R^2 = 1 - \frac{\sum (y_i - y_n)^2}{\sum (y_i - \bar{y})^2} = 1 - \frac{66.23}{2366.25} = 0.972$$

$$F = \frac{R^2}{1 - R^2} \frac{(n - m - 1)}{m} = 347.2547$$

$F_{табл} = 4.96$ (для α=0,05).

Поскольку $F > F_{табл}$, то коэффициент детерминации статистически значим.

В данном курсе лекций линейный регрессионный анализ мы будем применять как для построения парных корреляций так и для вывода законов, связывающих транспортные параметры газов в полимерах со свойствами тих газов.

2. ЛИНЕЙНАЯ РЕГРЕССИЯ ТРАНСПОРТНЫХ ПАРАМЕТРОВ ГАЗОВ В ПОЛИМЕРАХ

Множественная линейная регрессия может быть применена к любому транспортному параметру: проницаемости, диффузии или растворимости. Мы основное внимание уделим корреляции между коэффициентами проницаемости разных газов в полимерах.

Для нахождения коэффициента проницаемости интересующего нас газа использовался так называемый метод парных корреляций, суть которого заключается в предварительном построении двухмерных зависимостей значений коэффициента проницаемости одного газа от проницаемости другого. Подобные зависимости можно строить как в простой, так и в логарифмической форме. После построения двухмерной зависимости для пары газов с использованием всех доступных экспериментальных данных полученные значения аппроксимируют той или иной зависимостью, которая позволила бы предсказывать коэффициент проницаемости интересующего нас газа по известному значению коэффициента проницаемости другого газа (например, коэффициент проницаемости кислорода по значению коэффициента проницаемости азота, или коэффициент проницаемости радона по коэффициентам проницаемости азота и кислорода, одновременно вовлечённым в рассмотрение).

Применительно к коэффициенту газопроницаемости, основное уравнение множественной линейной регрессии (МЛР) выражается в следующей форме (в десятичных логарифмах):

$$\lg P_i = b_0 + \sum_{n=1} b_{nj} \lg P_{nj}\,, \qquad \text{или} \qquad \lg P^{\,расч} = b_0 + \sum_{n=1} b_n \lg P_n^{эксп} \tag{10}$$

где n – количество используемых в расчете газов, b_0 – константа (свободный член), b_j – коэффициенты регрессии, P_j – экспериментальные значения коэффициентов газопроницаемости полимерного материала, P_i – расчетные.

Более наглядно основное уравнение МЛР выражается в следующей форме (в десятичных логарифмах):

$$\lg P_i = b_0 + b_1 \lg P_1 + b_2 \lg P_2 + \ldots + b_n \lg P_n \tag{10a},$$

где n - количество используемых в расчете газов, b_0 - константа, b_n - коэффициенты регрессии, P_n - экспериментальные значения коэффициентов проницаемости полимерного материала.

Частными случаями уравнения МЛР (1) являются двух- и трехпараметрические уравнения, которые можно представить графически – в виде прямой и плоскости соответственно:

$$\lg P_i = b_0 + b_1 \lg P_1 \tag{11a},$$

$$\lg P_i = b_0 + b_1 \lg P_1 + b_2 \lg P_2 \qquad (11б).$$

где параметры b_0, b_1 и b_2 зависят от природы газа и от природы полимера.

Замечание. В рассмотрение можно вовлечь селективность, α, (как отношение проницаемостей). Например, в случае проницаемости какого-то газа (например, ксенона) относительно проницаемостей кислорода, азота и метана:

$$\lg P_i = \ln P_{O2} + b_1 \lg(\alpha_{O2/N2}) + b_2(\alpha_{O2/CH4}) \qquad (12)$$

но оказалось, что это не даёт заметных преимуществ по сравнению с использованием коэффициентов газопроницаемости.

Линейная регрессия 1-го порядка (2D модель). Уравнение МЛР (1) при $n = 1$ записывается в виде:

$$\lg P_i = b_0 + b_1 \lg P_1, \qquad (13)$$

где b_0, b_1 – коэффициенты регрессии, P_1 – экспериментальные значения коэффициентов проницаемости полимерного материала, P_i – расчетные.

Проверку гипотезы линейности осуществляли коэффициентом парной корреляции – чем ближе значение r к единице, тем лучше прямая линия описывает экспериментальные данные и тем справедливее становится выбор классического механизма для описания изучаемого процесса диффузии.

Двухпараметрическая модель является наиболее простой с точки зрения расчётов и визуализации результата – решение представляет собой прямую (**Рис. 2).**

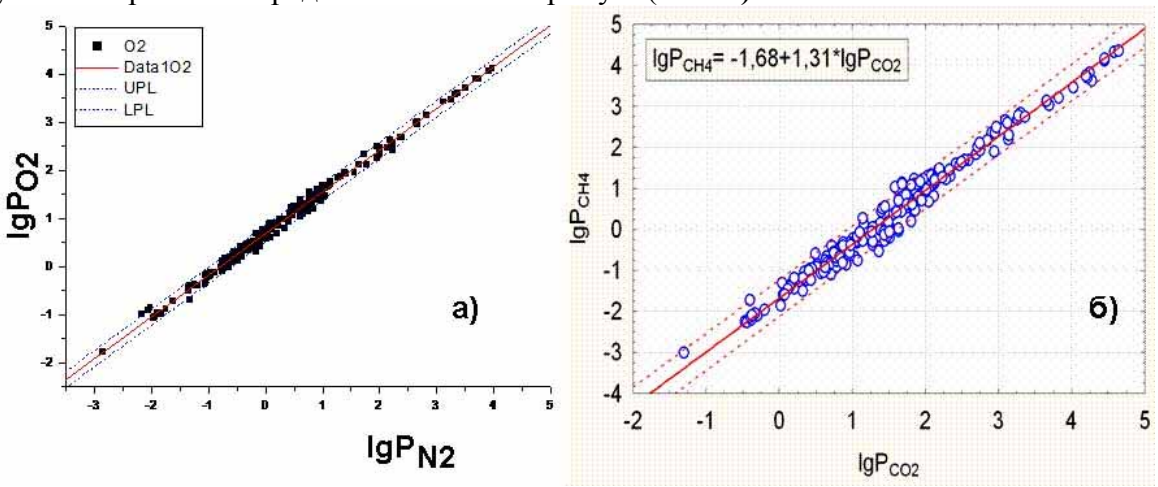


Рис. 2. Корреляционные зависимости (логарифмический масштаб): а - зависимость коэффициента проницаемости кислорода от проницаемости азота; б - зависимость коэффициента проницаемости метана от проницаемости углекислого газа. Точки – экспериментальные данные, прямая – линия регрессии, пунктир – 95% уровень значимости.

Рассмотрим пример расчета проницаемости по формуле (13) на примере пары газов O_2/N_2. На **Рис. 2** представлена зависимость коэффициента проницаемости O_2 в зависимости от проницаемости N_2. Видно, что экспериментальные данные в двойной логарифмической шкале хорошо описываются линией регрессии, уравнение для данной пары газов с рассчитанными коэффициентами регрессии (**Табл. 1**) можно записать в виде:

$$\lg P_{O2} = 0.679 + 0.867 \cdot \lg P_{N2}. \qquad (14)$$

Табл. 1. Результат регрессионного анализа зависимости $\lg P_{O2}$ от $\lg P_{N2}$.

Пара газов	b_0	b_1	SD	R^2	N	$\bar{\varepsilon}$,%
O_2/N_2	0.679	0.867	0.0848	0.993	235	22%

Статистические критерии, приведенные в **Табл. 1**:

$SD = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})^2}$ - стандартное отклонение для выборки;

$R^2 = \left[\frac{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{SxSy}\right]^2$ - квадрат коэффициента корреляции между x и y ($\lg P_{эксп}$ и $\lg P_{расч}$), диапазон изменения $0<R^2<1$, чем ближе к 1, тем выше степень связи между x и y;

$\varepsilon = \left|\frac{\lg P_{расч} - \lg P_{эксп}}{\lg P_{расч}}\right| \cdot 100\%$ - относительная ошибка расчетов.

На **Рис. 3** представлена гистограмма распределения ошибок расчета для пары O_2/N_2. Средняя ошибка расчета (отклонение предсказанных данных от экспериментальных) составляет 22%, нижний и верхний квартили – 7% и 31% соответственно.

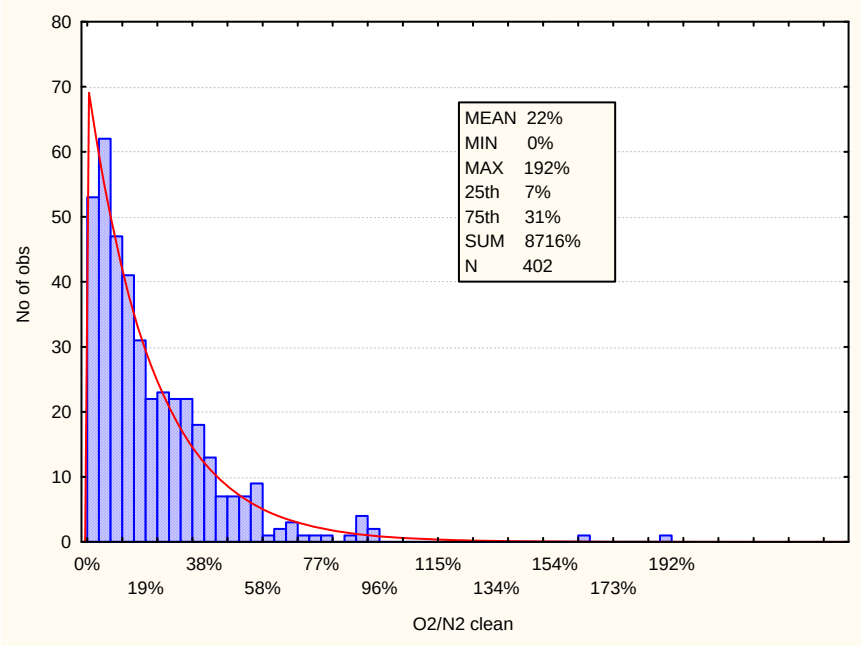


Рис. 3. Распределение относительных ошибок расчета P_{O2} от P_{N2} по формуле (13).

В **Табл. 2** приведены результаты расчетов коэффициентов проницаемости O_2 в различных полимерах с использованием рассчитанных коэффициентов регрессии и экспериментальных значений проницаемости N_2.

Табл. 2. Результаты расчета P_{O2} с использованием экспериментальных значений P_{N2} (в Баррерах) для некоторых полимеров Базы ИНХС: сопоставление рассчитанных значений с экспериментальными.

Акроним	Полимер	P_{O2} (эксп.)	P_{O2} (расч.)	ε, %
1,5-NPSF	1,5-dihydroxynaphthalene polysulfone	0.42	0.4	5%
3,4'-PSF	3,4'-Polysulfone	0.39	0.45	14%
12F PEK	12 Fluorine poly(ether ketone)	3.7	3.76	2%
12H PEK	12 Hydrogen poly(ether ketone)	1.1	1.23	11%
6FDA-3,3'-ODA	Polyimide of 2,2-bis(3,4-dicarboxyphenyl) hexafluoro-propane dianhydride and 3,3'-oxydianiline	0.68	0.65	5%
6FDA-4BDAF	Polyimide of 2,2-bis(3,4-dicarboxyphenyl) hexafluoro-propane dianhydride and 4,4'-bis(aminophenoxyphenyl) hexafluoropropane	5.4	4.7	15%
BPDA-6FpDA	Polyimide of 3,3',4,4'-biphenyltetracarboxylic dianhydride and 4,4'-(hexafluoro isopropylidene)-dianiline	6.65	5.6	18%
BPDA-BAPHF	Polyimide of 3,3',4,4'-biphenyltetracarboxylic dianhydride and 2,2'-bis(4-(4-aminophenoxy) phenyl) hexafluoropropane	1.54	1.4	9%
BPDA-BATPHF	Polyimide of 3,3',4,4'-biphenyltetracarboxylic dianhydride and 2,2'-bis(4-(4-amino-2-trifluoromethylphenoxy) phenyl) hexafluoropropane	3.11	2.9	7%
DMPSF	Dimethyl bisphenol-A polysulfone	0.64	0.6	7%
DMPSF-Z	Dimethyl bisphenol-Z polysulfone	0.41	0.4	3%
FBPC	Fluorene bisphenol polycarbonate	3.18	3.0	5%
FBPSF	Fluorene bisphenol polysulfone	2.76	2.5	9%
NBPC	Bisphenol of norbornane polycarbonate	2.4	2.5	3%
PTMSP	Poly(1-trimethylsilyl-1-propyne)	10000	9977	1%
PVTMS	Polyvinyltrimethylsilane	44	38.2	15%

PVTMS/ PAN	(77.8/22.2)	13	15.9	18%
PVTMS/ PAN	(60/40)	5	8.7	43%
PVTMS/ PDMS	(40/60)	220	259	15%
PVTMS/ PBD	(80/20)	30	29	4%
PBD-trans	Polybutadiene	6.2	9.3	34%
PBD(29%)	Hydropol	11.3	15.9	29%
CA	Cellulose acetate	0.68	1.35	50%
Cellulose	Cellophane	0.0021	0.033	94%
EC	Ethyl cellulose (Ethocel 610)	11.2	13.4	17%
MC	Methyl cellulose	0.5	1.05	52%
NR	Nitrosorubber	82	90.6	9%
SR	Silicon rubber (Silastic 500-1)	577	624	8%
PDMS	Polydimethylsiloxane	920	954	4%
PDMS (10%SiO2	Polydimethylsiloxane (10% SiO2)	490	553	11%
SILAR (38/62)	Арилат-силоксан	360	410	12%
PE Alathon14	Polyethylene, LDPE	3	4.78	37%
PE Grex	Polyethylene, HDPE	0.4	0.87	54%
PI (0.969)	Polyisoprene (natural rubber)	24	31	24%
Teflon FEP	Poly(tetrafluoroethylene-hexafluoropropylene)	5	7.2	34%

Аналогичные расчёты можно провести и для других данных, представив их в зависимости от коэффициента проницаемости азота (логарифмический масштаб).

Уравнения линейной регрессии для зависимостей можно записать в виде:

$$\lg P_i = b_0 + b_1 \lg P_{N2}, \qquad (15)$$

где i - O_2, CO_2, CH_4, H_2, He; b_0 и b_1 – регрессионные коэффициенты; P_{N2} – экспериментальные значения проницаемости N_2.

В **Табл. 3** приведены результаты расчета коэффициентов регрессии и статистические критерии расчета.

Табл. 3. Коэффициенты регрессии (b_0, b_1), стандартные отклонения (SD), степень корреляции (R^2), размер выборки (N) и средняя относительная ошибка ($\bar{\varepsilon}$) для расчета коэффициентов проницаемости газа (кислород, углекислый газ, метан, водород и гелий) в зависимости от проницаемости газа азота.

Газ	b_0	b_1	SD	R^2	N	$\bar{\varepsilon}$,%
O_2	0.624	0.910	0.180	0.977	403	27%
CO_2	1.240	0.930	0.239	0.963	369	35%
CH_4	0.025	1.121	0.239	0.967	302	48%
H_2	1.401	0.693	0.373	0.881	244	64%
He	1.412	0.550	0.421	0.754	294	74%

Отличия теоретически рассчитанных значений от определённых экспериментально достаточно велики. Использование программы отбраковки грубых сбоев позволяет лучше подогнать теорию к эксперименту (**Табл. 4**).

Табл. 4. Те же параметры, что и в предыдущей таблице, но рассчитанные по массивам из которых удалены грубые выбросы).

Газы(i/j)	b_0	b_1	SD	R^2	N	$\bar{\varepsilon}$,%
O_2/N_2	0.679	0.867	0.085	0.993	235	22%
CO_2/N_2	1.281	0.875	0.102	0.991	233	31%
CH_4/N_2	0.0147	1.145	0.203	0.979	212	49%
H_2/N_2	1.550	0.616	0.208	0.936	166	45%
He/N_2	1.517	0.510	0.250	0.835	188	58%

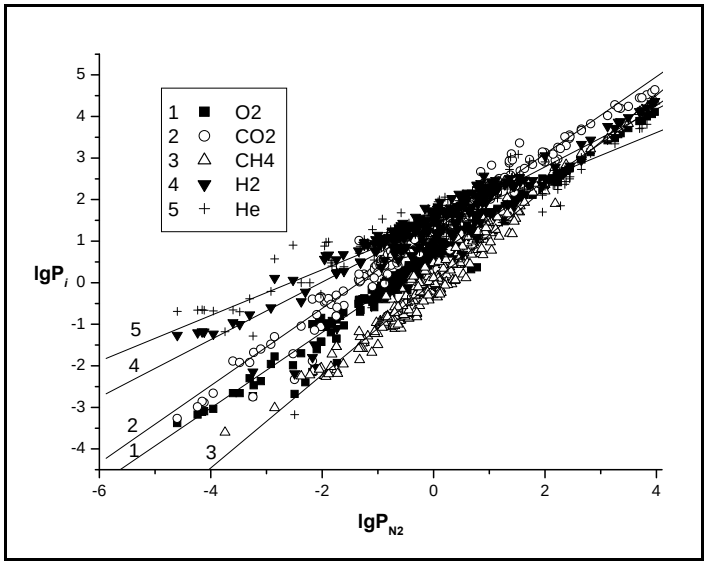


Рис. 4. Зависимость коэффициентов проницаемости: *1* - O_2, *2* - CO_2, *3* - CH_4, *4* - H_2, *5* - He от проницаемости N_2. Символы – экспериментальные данные, прямые – линии регрессии.

Табл. 5. Проницаемость некоторых газов по отношению к метану, *R* - степень корреляции.

Газ	b_0	b_1	R
O_2	-0.899	1.329	0.984
N_2	0.0147	1.145	0.989
CO_2	-1.679	1.314	0.987
H_2	-2.746	1.755	0.951
He	-2.964	1.992	0.896

Корреляционная зависимость lgP_{O2} от lgP_{N2} во всех полимерах, собранных в базе данных ИНХС, представлена на **Рис.** . Видно, что результаты распадаются по крайней мере на два участка – прямолинейный для умеренно- и высокопроницаемых полимеров и не очень прямолинейный для низкопроницаемых полимеров. Это указывает на то, что парные корреляции для высоко- и низкопроницаемых полимеров следует рассматривать отдельно.

Далее мы будем анализировать только данные по высокопроницаемым полимером, поскольку именно такие материалы используются в мембранной технологии. На **Рис.** представлена зависимость lgP_{O2} от lgP_{N2} и уравнение корреляционной прямой для неё.

Несмотря на некоторые отклонения, данные достаточно хорошо апроксимируются прямой: коэффициент корреляции близок к единице. Угловой коэффициент прямой (1,1) больше единицы.

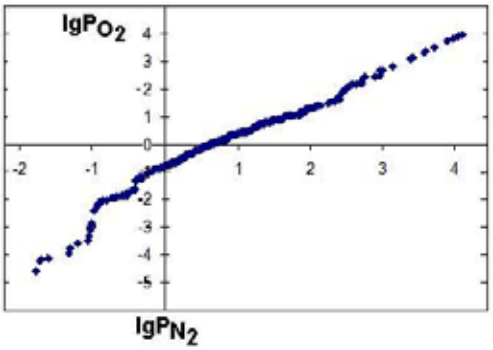
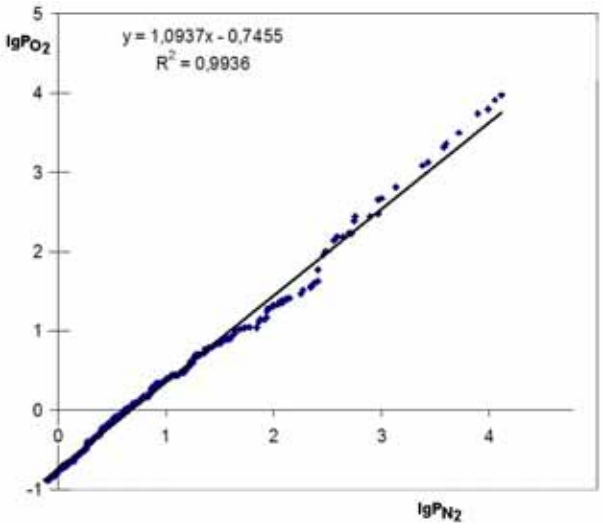


Рис. 5. Взаимосвязь коэффициентов проницаемости кислорода и азота (двойной логарифмический масштаб), собранных по всем полимерам Базы Данных ИНХС.



Аналогичные парные корреляции имеют место и для других газов. На **Рис. 5** приведена парная корреляция между проницаемостями ксенона и азота и дано уравнение апроксимирующей прямой. Данные хорошо описываются прямой, хотя коэффициент корреляции ниже, чем для пары кислород-азот (возможно, из-за малого числа данных). Угловой коэффициент (1,48) довольно высок.

Рис. 6. Корреляционная зависимость коэффициента проницаемости кислорода от коэффициента проницаемости азота в полимерах (двойной логарифмический масштаб) для умеренных и высокопроницаемых полимеров (по базе данных ИНХС).

Связь между проницаемостями кислорода и азота подтверждает вид взаимокорреляционной функции (**Рис.**), которая демонстрирует сильную корреляцию, без какого-либо лага.

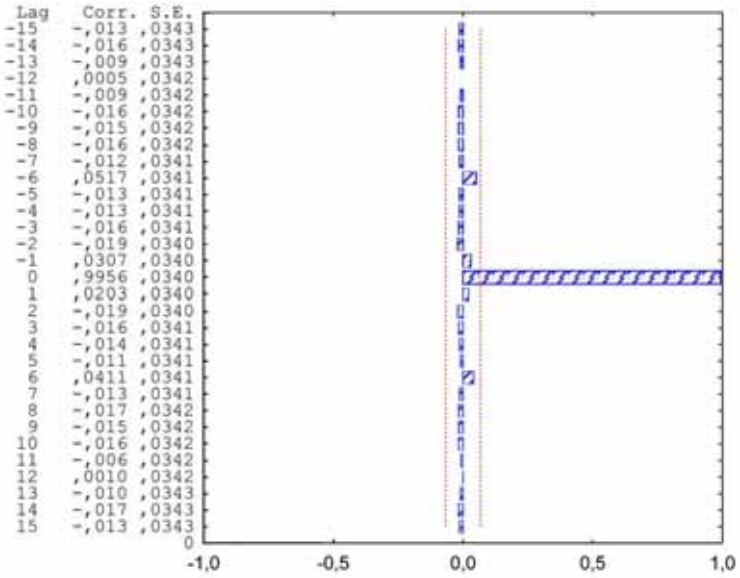


Рис. 7. Взаимокорреляционная функция коэффициентов проницаемости кислорода и азота.

В случае зависимостей lgP_{CO2}, lg_{O2} или lgP_{N2} от lgP_{Ar} (**Рис.** 9), аппроксимирующие прямые практически параллельны друг другу. Аналогичные зависимости коэффициентов проницаемости инертных газов от проницаемости азота (в двойном логарифмическом масштабе) также прямолинейны (**Рис.** 10), но они пересекаются при высоких проницаемостях (разделение невозможно). Аналогичный эффект наблюдается и для зависимостей проницаемостей CH_4, H_2, C_2H_6, и C_3H_8 от проницаемости азота (в двойном логарифмическом масштабе) (**Рис.**).

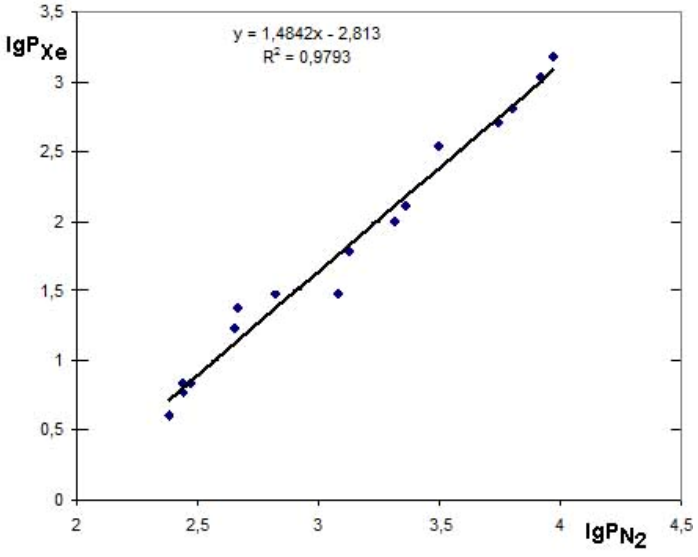


Рис. 8. Зависимость коэффициентов проницаемости ксенона от коэффициента проницаемости азота в различных полимерах (по базе данных ИНХС).

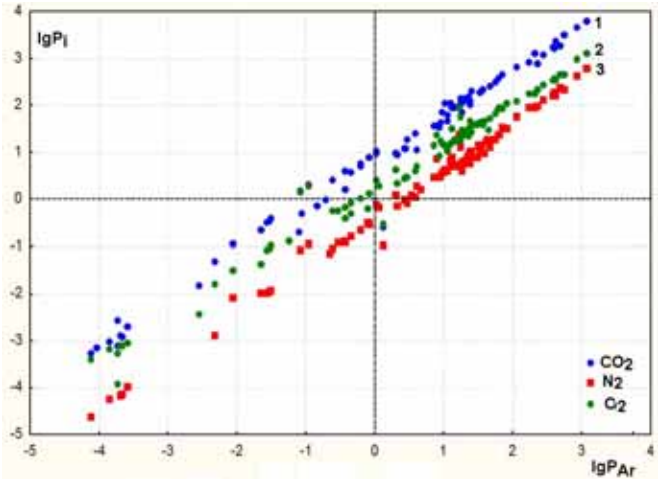


Рис. 9. Зависимость проницаемости атмосферных газов от коэффициента проницаемости аргона: 1- CO_2, 2 – O_2, 3 - N_2. (Двойной логарифмический масштаб).

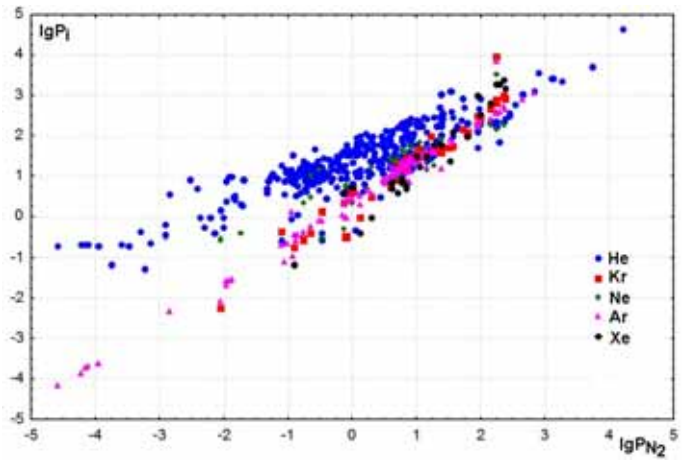


Рис. 10. Зависимости коэффициентов проницаемости инертных газов от проницаемости азота (двойной логарифмический масштаб).

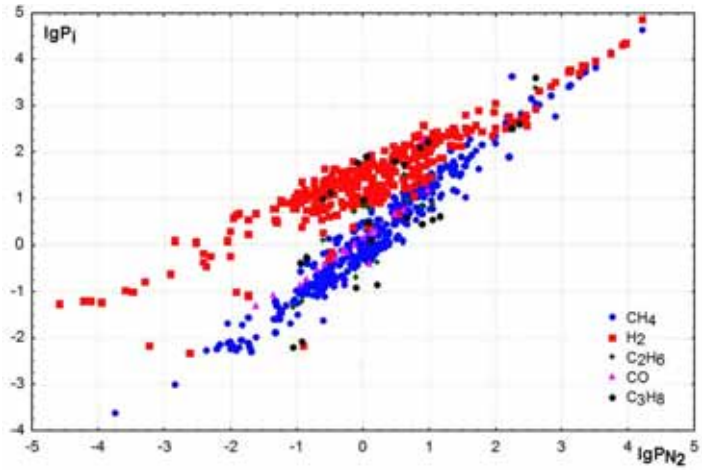


Рис. 11. Зависимость проницаемости некоторых газов от коэффициента проницаемости азота: 1- CH_4, 2 – H_2, 3 – C_2H_6, 4 – CO, 5 – C_3H_8. (Двойной логарифмический масштаб).

Линейная регрессия 2-го порядка ($ЛР_2$) (3D модель)

Увеличение количества используемых в расчете газов в правой части уравнения МЛР (1) с 1 до 2 газов приводит к следующему уравнению:

$$\lg P_i = b_0 + b_1 \lg P_1 + b_2 \lg P_2 . \qquad (16)$$

Решение этого уравнения представляет собой плоскость в трехмерном пространстве (**Рис. 5**). Из графиков видно, что экспериментальные точки образуют достаточно плотную область значений (некое подобие галактики).

Увеличение количества переменных ведет к повышению точности расчета с одной стороны (например, средняя ошибка 2D-модели 20% может уменьшиться до 15%, а при дальнейшем увеличении параметров в правой части уравнения и до 13%), и требует большего количества экспериментальных данных с другой.

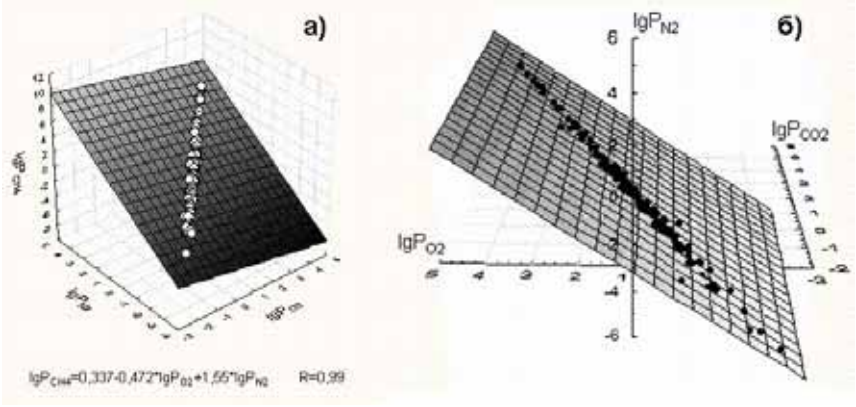
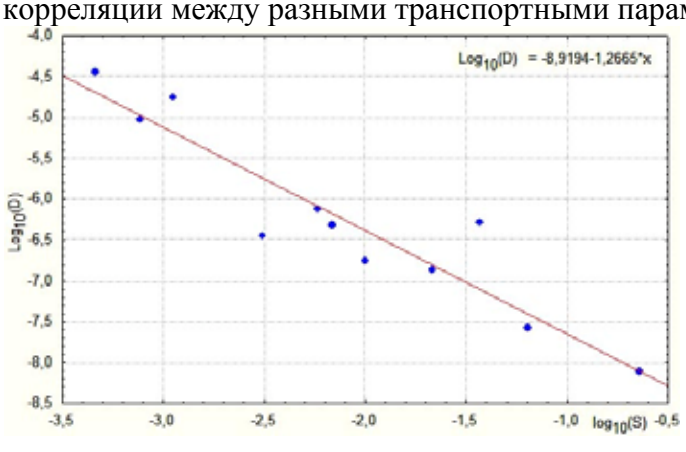


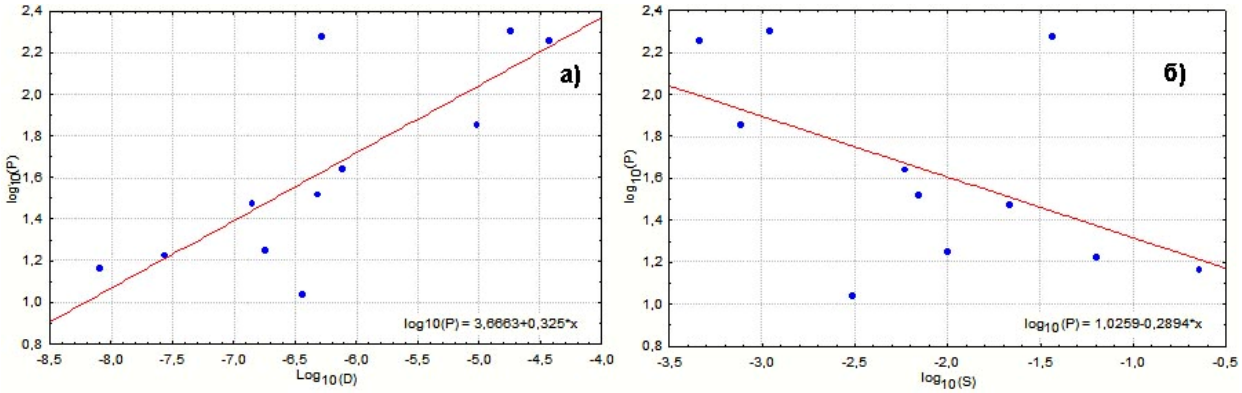
Рис. 12. Трёхмерные корреляционные зависимости: а – зависимость P_{CH4} от P_{O2} и P_{N2}; б - зависимости P_{N2} от P_{O2} и P_{CO2}.

Уравнение регрессии:

$$\lg P_{N2} = -0.8221 + 0.1349 \lg P_{CO2} + 0.9847 \lg P_{O2} \qquad (17)$$

Парные корреляции между разными транспортными параметрами





3. ПРЕДСКАЗАНИЕ ТРАНСПОРТНЫХ ПАРАМЕТРОВ В ДОСТАТОЧНО ХОРОШО ИЗУЧЕННЫХ СИСТЕМАХ ГАЗ-ПОЛИМЕР

Расчёт коэффициентов газопроницаемости плохо изученных полимерах методом множественной линейной регрессии (МЛР), требует привлечения как можно более обширного массива надёжно установленных данных для других полимеров. В Базе данных ИНХС наиболее изученными оказались процессы проницаемости в полимерах таких газов, как CO_2, N_2, O_2, Ar, He, CH_4. В качестве эталонного газа для всех расчетов удобно выбрать H_2, проницаемость которого изучена довольно хорошо.

Вычислительный процесс состоит из нескольких этапов:

Выбор из общей базы таких значений коэффициентов проницаемости полимеров, для которых имеется хоть одно значение коэффициента проницаемости для вышеуказанных газов. На основании этой

выборки строится базовая матрица, которая изначально содержит довольно большое количество незаполненных мест.

Выбор полимеров из Базы данных (БД) с данными по коэффициентам проницаемости для всех пар газов H_2-X, где X - выбранный газ, при этом осуществляется выбор только таких полимеров, для которых имеются оба значения коэффициента проницаемости, как для H_2, так и для X.

Построение графических зависимостей коэффициента проницаемости H_2, от коэффициента проницаемости газа X, на основании имеющихся данных и нахождение уравнения связи коэффициентов проницаемостей.

Связывание полученных уравнений по зависимостям коэффициентов проницаемости между собой, таким образом, что бы стало возможно нахождение значений коэффициентов проницаемостей всех газов, если известно значение хоть одного.

Задача состоит в построении системы уравнений или матрицы из коэффициентов уравнений, которую в дальнейшем можно использовать для решения.

В качестве примера на **Рис. 13** приведена зависимость коэффициента проницаемости углекислого газа от коэффициента проницаемости водорода.

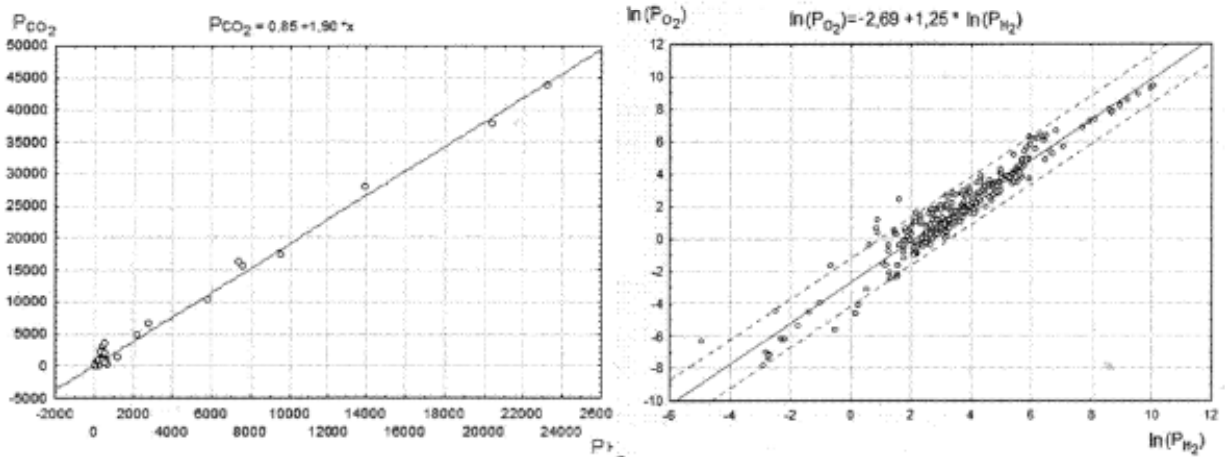


Рис. 13. Парные корреляции: зависимости коэффициентов проницаемости углекислого газа и кислорода от коэффициента проницаемости водорода для различных полимеров.

В обоих случаях наблюдались линейные зависимости, с некоторым разбросом значений около наилучшей прямой проведенной с использованием МНК. В качестве примера уравнения приведем зависимость коэффициента проницаемости N_2, от H_2:

$$P_{N2} = -32.423 + 0.384\ P_{H2}. \qquad (18)$$

Для шести газов были рассчитаны парные корреляции относительно водорода. Набор параметров уравнений регрессий (a и b) для lg$P$ в полимерах всех изученных газов относительно водорода приведены в **Табл. 6.**

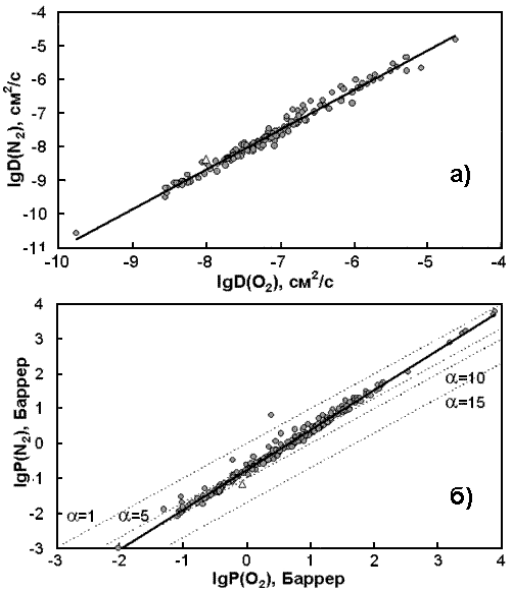


Рис. 14. Парные корреляции транспортных параметров азота и кислорода в ПВТМС: а – коэффициент диффузии, б – коэффициент проницаемости.

Табл. 6. Значения коэффициентов парных корреляций для логарифмов коэффициентов проницаемости некоторых газов относительно логарифма проницаемости водорода.

Газ	*a*	*b*
He	0,558	13,665
O_2	0.552	-8,849
N_2	0,384	-2,423
CO_2	1,900	0,850
CH_4	0,693	25,150
Ar	0,441	-8,434

Зная коэффициенты *a* и *b* можно рассчитать коэффициент диффузии любого из рассматриваемого здесь газа в конкретном полимере, если известно хотя бы значение коэффициента проницаемости какого-либо газа. Точность предсказания, однако, не высока. Лучшей сходимости теории и эксперимента можно добиться, если данные для низко- и высокопроницаемых полимеров рассматривать раздельно. Другой подход – использовать более сложные (нелинейные) функции для поиска характера зависимости проницаемости одного газа от другого.

Более точным методом анализа и предсказания коэффициента проницаемости газа является метод *N*-мерных корреляций, в котором коэффициент проницаемости искомого газа представляется в виде сложной функции от коэффициентов проницаемости или неких функций от них для *N* компонентов, где $N \geq 2$.

В общем виде такие зависимости можно представить в виде уравнения:

$$P_N = F(P_1,P_2,...,P_{N-1}). \qquad (19)$$

Метод *N* - мерных корреляций является непосредственным развитием метода парных корреляций, так как при $N=2$, получается метод парных корреляций. Если коэффициент проницаемости одного из компонентов функционально зависит от коэффициента проницаемости любого другого компонента, что подтверждает метод парных корреляций, то можно утверждать, что при наличии более трех компонентов коэффициент проницаемости любого из компонентов может быть представлен в виде некоторой функциональной зависимости от остальных компонентов.

Уравнение (19) может быть преобразовано с учетом перехода от величин коэффициентов проницаемостей к селективностям, если учитывать, что селективность по паре компонентов равна отношению их коэффициентов проницаемостей. Тогда уравнение (19) можно записать в виде:

$$P_N=F(P_1,P_2,\ldots,P_M,\alpha_{M+1/j},\alpha_{M+1/j},\ldots,\alpha_{N-1,j}) \qquad (20)$$

Величина M при этом может варьироваться от 0 до *N*-1, а компонент j - это газ относительно которого считаются все остальные селективности.

В достаточно простом виде уравнение (32) может быть записано как:

$$P_N=a_1P_1+a_2P_2+\ldots+a_{N-1}P_{N-1}+b \qquad (21)$$

Здесь коэффициент проницаемости *N*-го компонента зависит от коэффициентов проницаемостей *N*-1 выбранных компонентов с учетом неких коэффициентов, при этом, вообще говоря, наличие свободного члена *b* в уравнении (21) не является обязательным и возможен анализ следующих уравнений:

$$P_N=a_1P_1+a_2P_2+\ldots+a_{N-1}P_{N-1} \qquad (22)$$

Анализ зависимостей с *N* равным 2 или 3 возможен в графическом виде (см., например, **Рис. 8**). Логарифмическое представление в данном случае является более удобным, т.к. обладает большей наглядностью по распределению данных.

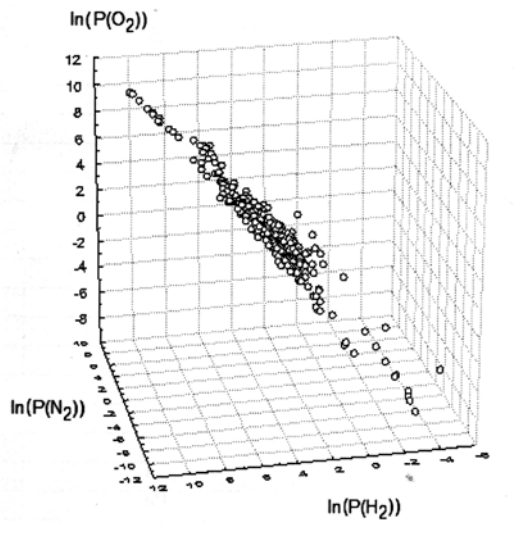


Рис. 15. Трёхмерная зависимость коэффициентов проницаемостей газов через полимерные мембраны в логарифмической форме.

База данных позволяет осуществлять такие *N*-мерные корреляции, где *N* варьировалось от 2 до 7. Отобранные для таких корреляций данные удобно представлять в виде матрицы размером *M*x*N*, где *M* - количество выбранных записей по коэффициентам проницаемости полимеров, а *N* - количество газов по которым производилась выборка.

Очевидно, что для решения системы содержащей в себе уравнения вида (21) достаточно *N* уравнений, что бы определить все коэффициенты. А для систем с уравнениями вида (22) - *N*-1 уравнений. Но в Базе имеется

большое количество экспериментальных данных, поэтому фактически количество уравнений для $N < 7$, с учетом выбора наиболее изученных газов всегда превышает число уравнений достаточное для решения системы и нахождения коэффициентов. Необходимым условием выбора полимеров является наличие данных по коэффициентам проницаемости для всех указанных газов, но для разных газов такое количество существенно разное. Поскольку приходится ориентироваться на минимальное число измерений, для современной информации в Базе, для анализа нельзя использовать более 7 газов.

Для преобразования матрицы размера *M*x*N* к матрице размером *N*x*N* используется метод МНК с последующим решением одним из матричных способов (например, методом Гаусса).

Данный алгоритм включён в БД и для его использования достаточно выбрать значения по коэффициентам проницаемости для интересующих нас газов. Далее возможно преобразовать выбранные коэффициенты проницаемости с учетом использования некоторых математических функций или перейти к селективностям. После проведения нужных преобразований необходимо передать файл с данными на обработку и на выходе получить искомые коэффициенты и среднюю погрешность расчетных данных. Важно понимать, что с увеличением количества переменных происходит уменьшение погрешности расчетных данных – факт положительный, однако при этом заметно уменьшается количество полимеров, для которых известны коэффициенты проницаемости всех компонентов. Чем больше данных по газопроницаемости различных компонентов выбрано и использовано в обработке (*N*), тем выше точность предсказания коэффициента проницаемости для интересующего нас компонента. На точность прогноза существенное влияние сильных выбросов в данных. Такие данные лучше исключить из рассмотрения. Обычно исключают данные, относительное отклонение которых от расчётных значений превышает 90%.

Для предсказания коэффициента проницаемости интересующего нас газа достаточно универсальным оказался метод основанный на уравнениях (21) и (22), при этом уравнение (21) применяют для низкопроницаемых полимеров, а (22) для высокопроницаемых. При таком подходе погрешность предсказания коэффициента проницаемости компонентов в полимере не превышает 50% для $N \geq 4$ по всем записям и на всех диапазонах значений или $N = 3$ и полимеров с коэффициентом проницаемости превышающим 200 Баррер по водороду.

В **Табл. 7** приведены коэффициенты рассчитанные исходя из уравнения (34) для разных значений *N*, при этом искомым компонентом являлся кислород. Коэффициенты a_i следуют в порядке уменьшения количества экспериментальных данных для данного газа (1 – N_2, 2 – He, 3 – CH_4, 4 – H_2, 5 – CO_2, 6 – Ar). Искомым коэффициентом проницаемости при этом является коэффициент проницаемости кислорода.

Табл. 7. Коэффициенты $a_{1..6}$ для разных значений числа N.

N	a_1	a_2	a_3	a_4	a_5	a_6
2	1.463					
3	0.735	0.805				
4	1.035	0.277	0.083			
5	0.959	0.015	0.041	0.251		
6	0.985	0.150	0.042	0.126	0.029	
7	0.294	0.062	0.123	0.149	0.021	0.768

На сегодняшний день, включённые в Базу программы оперируют с *N*=4. Способ оптимален с точки зрения среднего отклонения, и включает газы, для которых имеется достаточно большое количество экспериментальных значений.

Если для некоторого полимера измерено достаточно большое количество коэффициентов проницаемостей газов и полимер может быть условно отнесен к высокопроницаемым, то предсказание коэффициента проницаемости недостаточно хорошо изученного газа, входящего тем не менее в первую десятку наиболее изученных, является достаточно точным. Точность эта будет возрастать по мере пополнения БД. Для неизученных систем газ-полимер описанный здесь статистический подход в принципе невозможен. Предсказание становится возможным лишь при использовании корреляций между транспортными параметрами газа и физико-химическими свойствами газа и/или полимера.

5. КОРРЕЛЯЦИИ ПРОНИЦАЕМОСТЬ-СЕЛЕКТИВНОСТЬ

Уже на ранней стадии изучения процессов газоразделения была обнаружена связь между логарифмом селективности разделения пары газов от логарифма проницаемости одного из этих газов. График на **Рис. 9** показывает, что переход стеклования (от каучука к стеклу) не оказывает какого-либо влияния на корреляционную зависимость.

На данный момент O_2/N_2 является наиболее изученной парой и для нее имеется больше всего экспериментальных данных по проницаемостям через различные полимеры. На **Рис. 10** представлены данные

по полимерам, имеющиеся на данный момент. Видно, сто основная часть полимеров лежит в области, которая ограничена сверху неким пределом. Эта граница по мере появления новых данных медленно перемещается.

При исследовании зависимостей селективности от проницаемости, было обнаружено существование верхнего предела. Указанный предел наблюдался для таких пар газов как O_2/N_2, CO_2/CH_4, H_2/N_2, He/N_2, H_2/CH_4, He/CH_4, He/H_2, H_2/CO_2 и He/CO_2. При этом уравнение для верхнего предела может быть представлено в форме

$$P_i = k\alpha_{ij}^{n}, \qquad (23)$$

где P_i - коэффициент проницаемости более проницаемого газа, а α - селективность для выбранной пары газов и n - угол наклона для прямой, после представления уравнения в логарифмической форме (отрицательная величина).

Величина -1/n связана с диаметром молекулы газа соотношением:

$$-\frac{1}{n} = \left(\frac{d_j}{d_i}\right)^2 - 1 = \left(d_j - d_i\right)\frac{d_j - d_i}{d_i^2} \qquad (24)$$

В настоящее время большие усилия прилагаются для получения материалов обладающих проницаемостями и селективностями, которые лежат выше верхнего предела для обычных полимеров.

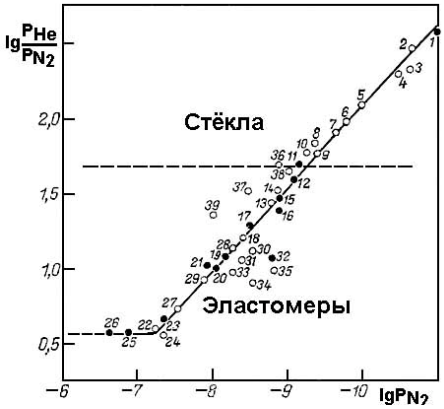


Рис. 16. Зависимость $\lg(P_{He}/P_{Ne})$ от $\lg P_{N2}$: 1 – поливинилиденхлорид; 2, 4– поли-этилентерефталат; 3 – поливинилфторид; 5 – поливинилхлорид; 6 – полиамид; 7 – поливинилиденхлорид (пластифицированный); 8 – нитрат целлюлозы; 9 – полипропилен; 10 – фторопласт 26; 11 – сополимер изопрена (74%) с акрилонитрилом (26%); 12, 18, 20 – сополимеры бутадиена с акрилонитрилом разного состава; 13 – полиакрилонитрат Д-4П; 14 – поликарбонат; 15 – полиизобутилен; 16 – бутилкаучук; 17 – сополимер винилхлорида с винилацетатом; 19, 37 – ацетобутират целлюлозы; 21 – сополимер этилена с винилацетатом; 22 – полибутадиен; 23 – СКИ-3; 24 – натуральный каучук; 25 – нитрил-силиконовый каучук; 26 – диметил-силоксановый каучук; 27 – МКС-30; 28 – СКМС-50; 30, 34, 35 – ПЭВД, СД, НД соответственно; 31 – ПЭ+5% технического углерода; 32 – сополимер этилена (90%) с пропиленом (10%); 33 – полимер этилена (95,7%) с винилацетатом (3,5%); 36 – триацетат целлюлозы; 38 – ацетат целлюлозы; 39 – полистирол.

Следующей по популярности изучения и наличия литературных данных является пара CO_2/CH_4. Для нее были исследованы аналогичные зависимости, но помимо обычных полимеров на **Рис. 17б** так же присутствуют полимеры, обладающие повышенными разделительными характеристиками (отмеченные синим цветом). Это в основном полимеры, содержащие бензоксозол-фениленовые или бензоциазол-фениленовые структуры в основной цепи. С точки зрения диффузии их можно отнести к классу молекулярных сит. Естественно верхний предел для них будет отличаться от верхнего предела для обычных полимеров, что наглядно видно на **Рис. 17б**. Можно найти предельную границу и для разделения пары кислород-азот (**Рис.18**). **Рис. 19** демонстрирует, какие селективности требуются для разделения в одном цикле смеси CH_4-H_2 состава 1:1 для получения водорода требуемой чистоты.

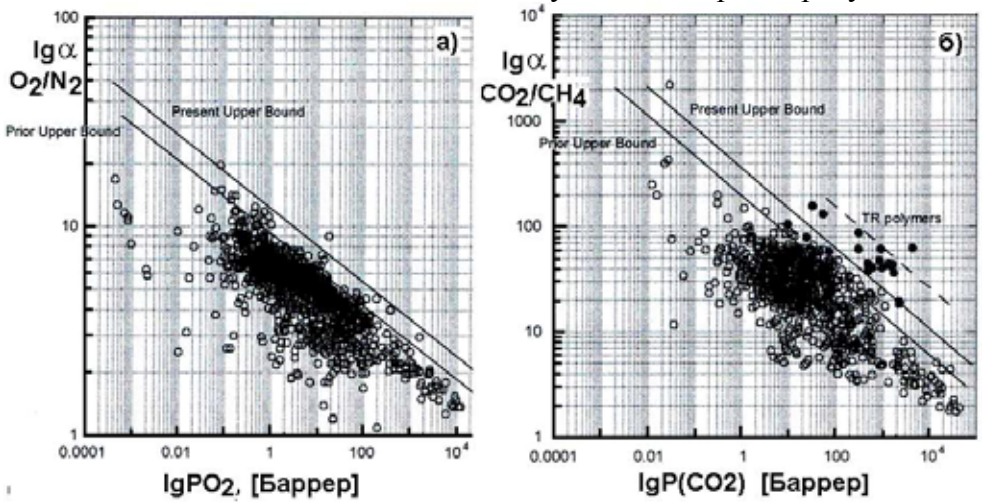


Рис. 17. Связь селективности с проницаемостью (двойной логарифмический масштаб): а – пара кислород-азот; б – пара углекислый газ-метан. Указан старый (1991) и новый (2002) предельные уровни. На **Рис. 10б** показан также уровень для полимеров – молекулярных сит (TR polymers).

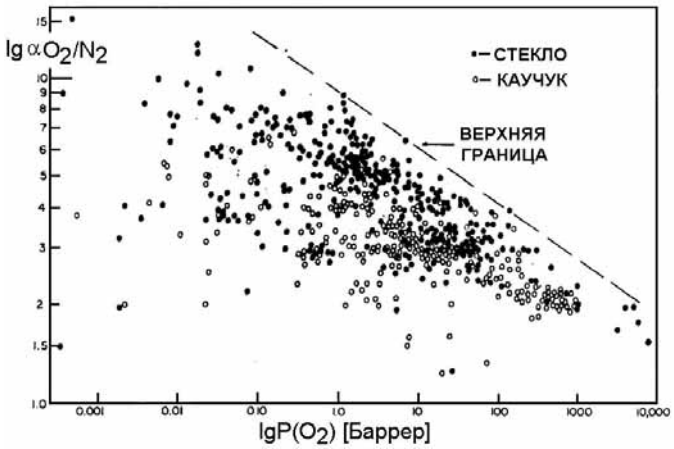


Рис. 18. Верхняя граница для пары O_2/N_2 (показан разброс данных для каучукоподобных и стеклообразных полимеров.

На **Рис 19.** представлены факторы селективности кислород/азот для различных полимеров (в обычном и логарифмическом масштабе). Из **Рис. 19а** следует, что фактор селективности в среднем равен 4,5 и имеют место выбросы в сторону высоких значений факторов селективности, а из **Рис. 19б** – выбросы в сторону низких значений факторов селективности, причём имеются полимеры которые вообще не делят эту пару газов (α=1). Кроме того в некоторых случаях имеет место инверсия разделения: если большинство полимеров пропускает преимущественно кислород, то есть полимеры (их правда мало), которые пропускают преимущественно азот.

Гистограммы распределения факторов селективности $\alpha(P_{CO2}/P_{N2})$ и $\alpha(P_{O2}/P_{N2})$ представлены на **Рис.** . Распределение $\alpha(PC_{O2}/P_{N2})$ довольно симметрично, тогда как $\alpha(P_{O2}/P_{N2})$ имеет заметную левостороннюю асимметрию. Если для $\alpha(P_{CO2}/P_{N2})$ имеются многочисленные выбросы с аномальновысокими факторами селективности, то для $\alpha(P_{O2}/P_{N2})$, наоборот, выбросы связаны с низкими значениями фактора селективности.

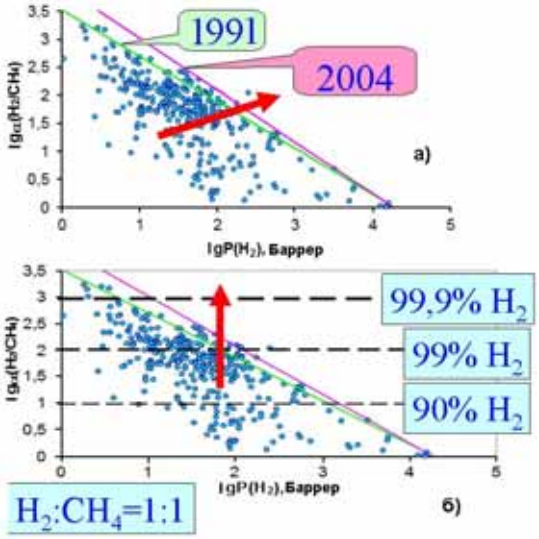


Рис. 19. Зависимость селективности разделения пары метан-водород от проницаемости целевого газа (водорода): с а - диаграмма Робсона для H_2/CH_4 ; б - одностадийное обогащение смеси H_2/CH_4.

Как можно было бы ожидать, фактор селективности разделения смеси кислород/азот линейно зависит от коэффициента проницаемости азота (в двойном логарифмическом масштабе) (**Рис.**). Имеют место выбросы, связанные с полимерами, не делящими эти газа, или, с полимерами, в которых имеет место инверсия разделения.

На **Рис. .** факторы разделения смесей водорода, гелия, аргона, метана, аргона или углекислого газа относительно кислорода построены в зависимости от величины фактора разделения пары кислород/азот (в двойном логарифмическом масштабе). Здесь также имеют место выбросы из корреляций, связанные с низкоселективными полимерами. Сами корреляции не очень чёткие, но в некоторых случаях определённо прослеживаются. Для уточнения ситуации следует использовать кластерный анализ.

Подход, основанный на анализе верхней границы проницаемости и селективности полимеров, позволяет качественно оценивать характеристики полимеров, указывая на максимально достижимые характеристики полимера по селективности для заданной пары газов, если известна проницаемость для одного из газов. Так же появляется возможность разделения полимеров по их эффективности, с

возможностью дальнейшего отнесения к той или иной группе (пример с перфторированными полимерами). Анализ позиции полимера по отношении к верхней границе позволяет судить об эффективности его разделительных свойств, т.е. чем ближе полимер к границе, тем он более эффективен для выбранного процесса разделения. В качестве основного недостатка данного метода можно отметить его качественный, а не количественный подход для массива данных по проницаемостям полимеров. Количественные же оценки селективности и проницаемости могут быть применены только для полимеров лежащих в непосредственной близи верхней границе. Основным преимуществом метода является возможность анализа полимера по сравнению с другими полимерами для конкретной пары газов и выбора дальнейшего направления в исследовании и получении новых полимерных материалов, с максимально приближенными свойствами к верхней границе или даже превышающей ее, если используется совершенно новый материал с уникальными свойствами.

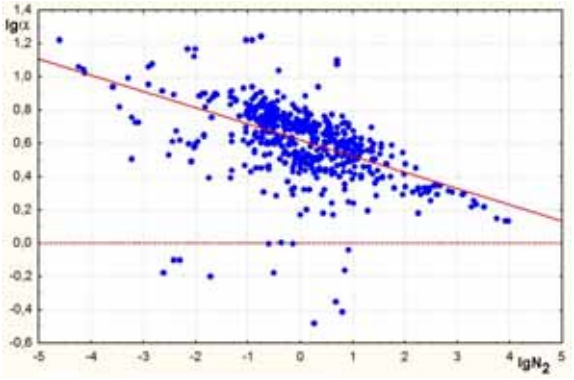


Рис. 20. Зависимость селективности разделения смеси кислород/азот от проницаемости азота (двойной логарифмический масштаб). Пунктир – отсутствие разделения.

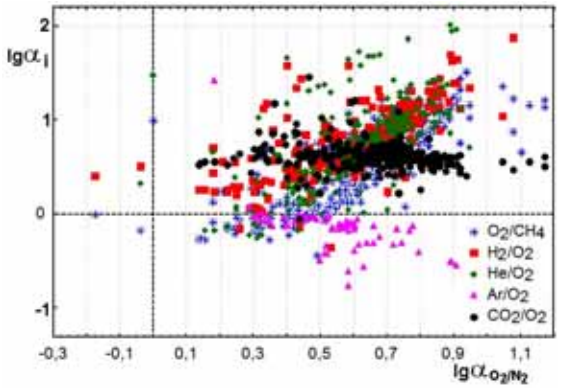


Рис. 21. Зависимость селективностей некоторых пар газов от селективности разделения пары кислород/азот.

----*

Полученные в данной работе парные корреляции можно использовать для уточнения параметров массопереноса газов для недостаточно хорошо изученных системах газ-полимер. В данной лекции мы использовали кибернетический подход (метод «чёрного ящика»), не требующий каких-либо знаний о свойствах газа, структуре полимера или о механизмах растворения и диффузии. Построение корреляций измеренных параметров для всех газов от измеренного параметра какого-либо одного газа (или нескольких газов), например, построение зависимости проницаемостей газов от проницаемости азота (или, для уточнения – от проницаемостей азота, кислорода, водорода и каких-либо других тестовых газов) и нахождение параметров уравнения регрессии для них оказывается вполне достаточным для уточнения значения транспортного параметра (например, коэффициента проницаемости) какого-либо газа в каком-либо полимере. При этом достаточно иметь значение нужного транспортного параметра для интересующего нас газа хотя бы в одном полимере. Точность предсказания (в среднем – 15%) оказывается вполне достаточной для корректировки значений эффективных диаметров молекул диффузанта в полимере (в основном – для молекул сложной формы или газов, для которых неприменимы традиционные методы измерения радиусов атомов, например, радон) и прикладных целей газоразделения.

Более точные предсказания для малоизученных систем и оценки значений для совершенно неизученных систем требуют построения корреляционных зависимостей между транспортными параметрами конкретного полимера и физико-химическими характеристиками тестовых газов. Такие корреляции будут рассмотрены в следующей лекции.